TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2015

44

Miguel Ángel Minaya Santa Cruz

A Multigenomic Approach to the Phylogeny, Evolution and Biogeography of the Grass Subfamily *Pooideae* with an Emphasis on the Subtribe *Loliinae*

Departamento
Ciencias Agrarias y del Medio Natural

Director/es
Catalán Rodríguez, Pilar

ISSN 2254-7606

Prensas de la Universidad Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

A MULTIGENOMIC APPROACH TO THE PHYLOGENY, EVOLUTION AND BIOGEOGRAPHY OF THE GRASS SUBFAMILY *POOIDEAE* WITH AN EMPHASIS ON THE SUBTRIBE *LOLIINAE*

Autor

Miguel Ángel Minaya Santa Cruz

Director/es

Catalán Rodríguez, Pilar

UNIVERSIDAD DE ZARAGOZA

Ciencias Agrarias y del Medio Natural

2015

TESIS DOCTORAL
con Mención Internacional

A Multigenomic Approach to the Phylogeny, Evolution and Biogeography of the Grass Subfamily Pooideae with an Emphasis on the Subtribe Loliinae

Autor: Miguel Ángel Minaya Santa Cruz

Directora: Dra. Pilar Catalán Rodríguez

Huesca, Abril de 2015

Escuela Politécnica Superior de Huesca

Departamento de Ciencias Agrarias y del Medio Natural

Universidad Zaragoza
1542

Escuela Politécnica Superior de Huesca
Departamento de Ciencias Agrarias y del Medio Natural
Universidad de Zaragoza

A Multigenomic Approach to the Phylogeny, Evolution and Biogeography of the Grass Subfamily Pooideae with an Emphasis on the Subtribe Loliinae

Memoria presentada por D. Miguel Ángel Minaya Santa Cruz para optar al grado de Doctor por la Universidad de Zaragoza

Directora: Dra. Pilar Catalán Rodríguez, Catedrática de Botánica del Departamento de Ciencias Agrarias y del Medio Natural de la Universidad de Zaragoza.

La Dra. Pilar Catalán Rodríguez hace constar que el trabajo recogido en la presente memoria de tesis doctoral ha sido desarrollado bajo su dirección y autoriza su presentación y defensa.

Huesca, Abril de 2015

Esta tesis doctoral ha sido desarrollada en el laboratorio del grupo de investigación Bioflora de la Escuela Politécnica Superior de Huesca (Departamento de Ciencias Agrarias y del Medio Natural, Universidad de Zaragoza), bajo la dirección de la Profesora Pilar Catalán Rodríguez. Parte de los resultados fueron obtenidos en el laboratorio del Profesor Jeffrey Palmer, Department of Biology, Indiana University (USA) y en el el laboratorio de la Profesora Roberta Mason-Gamer, Dept. Biological Sciences, University of Illinois at Chicago (USA).

La realización de esta Tesis se enmarca en los proyectos "Evolución multigenómica de las gramineas templadas (Pooideae, Poaceae). Biogeografia y filogeografia de especies modelo de pooideas " (proyecto CGL2009-12955-C02-01) y "Genómica comparada, biogeografía y evolución floral y adaptativa de gramíneas modelo" (proyecto CGL2012-39953-C02-01), financiados por el Ministerio de Ciencia e Innovación, y fue posible gracias a una beca-contrato predoctoral de formación de personal investigador (FPI) asociada al primer proyecto. Adicionalmente, se obtuvieron dos ayudas de movilidad para desarrollar sendas estancias de investigación del doctorando en las Universidades de Indiana e Illinois, financiadas por el Ministerio de Ciencia e Innovación.

A todos aquellos que respetan, entienden, y valoran
la ilusión y el trabajo que hay detrás
de una tesis doctoral

A mis padres

A Emily, Violeta e Ibai

Table of Contents

Chapter III: Horizontal Gene Transfer and Pseudogeneization of the Mitochondrial *rps3* Gene in Poaceae

Chapter IV: Diversification Rates and Chromosome Evolution in the Temperate Grasses (Pooideae) are Associated with Major Environmental Changes in the Oligocene-Miocene

Chapter V: Phylogeny and Biogeography of the Temperate Loliinae Grasses in the Southern Hemisphere

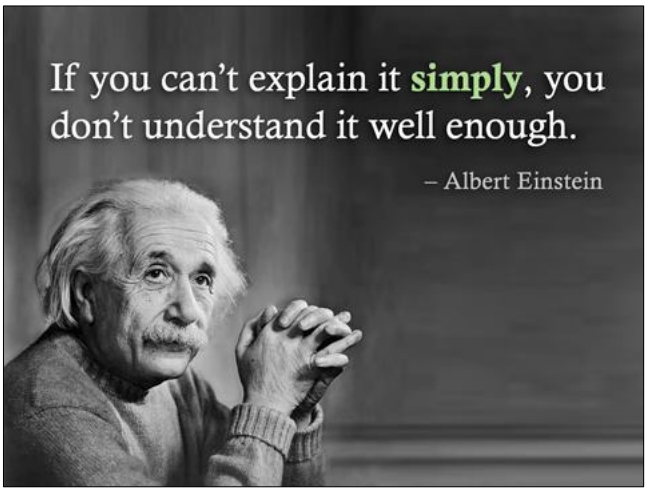
AGRADECIMIENTOS

PARA DECIR GRACIAS SÓLO TENGO QUE DECIRLO

[… Don Quijote y Sancho fueron invitados a comer con unos labradores, y al término con gran reposo levantó la voz don Quijote y dijo: *"Entre los pecados mayores que los hombres cometen, aunque algunos dicen que es la soberbia, yo digo que es el desagradecimiento, ateniéndome a lo que suele decirse: que de los desagradecidos está lleno el infierno. Este pecado, en cuanto me ha sido posible, he procurado yo huir desde el instante que tuve uso de razón, y si no puedo pagar las buenas obras que me hacen con otras obras, pongo en su lugar los deseos de hacerlas, y cuando estos no bastan, las publico, porque quien dice y publica las buenas obras que recibe, también las recompensara con otras, si pudiera; porque por la mayor parte los que reciben son inferiores a los que dan…"]*

Ustedes me van a permitir que antes de hablar diga unas palabras a modo de epílogo prescindible. Los botánicos rara vez somos noticia, uno cree que la única noticia es alguno de esos premios que se dan para recordarnos que somos humanos y una necrológica en algún diario matinal para que nuestros amigos sepan que nos hemos ido a otra parte. No, no somos noticia casi nunca. A fin de cuentas, nuestro trabajo no es más importante que el del panadero que hace pan, el del periodista que cuenta los sucesos, el maestro que educa y desasna, o el del poeta que cura las heridas en el corazón. Muchos botánicos hemos superado hace tiempo la vanidad de pensar que nuestro trabajo quedará por encima de nosotros cuando no vivamos sobre la tierra. No somos distintos a otros profesionales: los camareros, los conserjes, los conductores de autobús… Pero algunas veces queremos contar algunas cosas particulares. No, no es ego personal. Es un sentimiento y algo de educación. Tantos años delante de una clave dicotómica, muestreando en el campo, encerrado en un laboratorio y delante de un ordenador que igual, pienso, alguno de ustedes tiene la bondad de preguntarse: ¡y este joven! ¿qué ha hecho este joven investigador durante los últimos años?

Miren, estoy en una profesión complicada, muy complicada, desconocida y también anárquica. Pero es sin duda, la mejor profesión del mundo – o eso creo yo – . No hay nada que pueda con la fuerza de ser, de querer ser, un científico. Nada que haga, que evite que nuestro trabajo sea sencillo (porque es grande), simple (porque es complicado), seguro (porque es inseguro). En ciencia hay calidad, hay trabajo y hay ganas de hacerlo bien. Somos la esencia de nuestra sociedad moderna, la desesperación del dogma, el motor del cambio de todo lo que nos rodea, y por supuesto la única forma de entender razonadamente de dónde venimos y hacia dónde vamos. Los científicos hacen suyo un dicho que oía a menudo a mis abuelos: "*la vida siempre te da dos opciones: la cómoda y la difícil. Cuando dudes elige la difícil, porque así siempre estarás seguro de que no ha sido la comodidad la que ha elegido por ti*". En definitiva, yo entiendo esta profesión como un disco de Leonard Cohen rayado que nos recuerda que siempre seremos un pájaro en el alambre. Y así me siento esta mañana de viernes.



La memoria que constituye la presente tesis doctoral es fruto del trabajo conjunto de diferentes persona. Quiero agradecer en primer lugar la encomiable labor de la directora de esta tesis, la Prof. Pilar Catalán Rodriguez. Pilar, sin tu tiempo, tu apoyo, tu dedicación, tus consejos y tu ayuda no habría sido posible comenzar, ni mucho menos completar, este ingente trabajo. También quiero agradecer el apoyo técnico y las facilidades prestadas por el equipo de investigación Bioflora y la Escuela Politécnica Superior de Huesca. No puedo olvidar a los miembros del PAS y PDI pertenecientes a la Universidad de Zaragoza que con tanto esmero y rigor me han atendido y ayudado durante estos años. Me habéis demostrado que detrás de la rigidez y frialdad de una legislación existen personas sensatas y coherentes con una gran vocación de ayudar a estudiantes desorientados, que no entienden (o no quieren entender) las tergiversaciones burocráticas.

Me querría también acordar con estas líneas de los que me han ayudado a crecer como botánico y como científico desde que sólo era un pequeño cotiledón y cursaba mi segundo año en la Licenciatura de Biología. Estos amigos y compañeros me han ayudado con una asiduidad casi indecorosa a pensar mucho, a modificar pensamientos, a refrescar reflexiones, a cambiar

opiniones, y me han insistido muchísimas veces en que debo leer y releer para evitar descubrir el Mediterráneo. Bajo el lema de: Cuando no sopla el viento, ¡rema!, Andrew Alverson, Antonio Díaz-Pérez, Consuelo Cebolla, Emily Lemonds, Eugenio Pavón, Pilar Gozálo, José Gabriel Segarra, Helios Sainz, Jeffrey Palmer, Luis Ángel Inda, Manuel Pimentel, Mª Antonia Rivas-Ponce, Pilar Pavón, Roberta Mason-Gamer, … han sido para mí un modelo, y una fuente de inspiración. En este mundo no son muchas las personas que consiguen levantar lira con sus palabras, pero he tenido la inmensa suerte de charlar con ellas, y ellas me han hecho pensar, reflexionar, imaginar. Hay cosas que sólo se tienen cuando se dan: ¡¡GRACIAS!!.

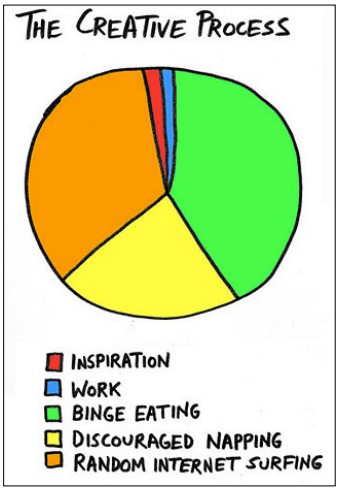
Quiero agradecer especialmente a todos aquellos amigos que han venido dándome ánimos durante estos años. Sin duda son muchos, más de los que puedo recordar estos días. Recuerdos de exámenes, prácticas de campo, jornadas pajareras, muestreos en Alameda del Valle y largos días por la facultad azul revolotean por mi cabeza cuando pienso en mis compañeros y amigos de la UAM y de la residencia Ciudad Escolar. Por ahí andará Álvaro, Ana, David, Estela, Eva, José Carlos, José Manuel, Monticola, Nagore, Pepe Botika, Piluka, Rut, Victor, etc. Gracias a todos, a vuestro lado disfruté lo indecible. Estad seguros que vosotros sois los primeros responsables de que yo acabara donde estoy. Para bien o para mal.

Curro, y Bea, a vosotros os agradezco vuestra confianza, enseñanzas y paciencia. Solo diré que echo de menos esas singulares jornadas de campo por el Pirineo o por la Sierra de Guara. Mientras contemplábamos la naturaleza hemos intentado arreglar el mundo con nuestras discusiones políticas, económicas, y filosóficas. Y Tina (DEP) amenizaba la fiesta. Ahora, en la distancia consaburense, puedo afirmar que no tengo a mi lado a nadie que reemplace todas esas vivencias. Espero que nuestras vidas se vuelvan a cruzar de nuevo en el futuro.

Y entonces marché a Estados Unidos gracias a sendas ayudas de movilidad financiadas por el Ministerio de Ciencia e Innovación de España, las cuales estaban vinculadas a los proyectos de investigación que desarrollábamos en el equipo de investigación Bioflora de la Universidad de Zaragoza. Cuando uno se marcha a un lugar que está a 12 horas en avión no sabe qué es lo que va a encontrar. Y yo me encontré a un pueblo educado, atento, trabajador, amable, servicial, y hospitalario que me

ayudó, entre otras muchas cosas, a enfocar mi carrera científica desde un punto de vista distinto. Entre las numerosas personas que conocí y que me ayudaron en Bloomington, en Chicago y en Saint Louis están Ariadne Lieber, Brandon Fonnemann, Claudio Casola, Danny W. Rice, Diane McCarthy, Francesco Catania, Jim Lemonds, Josep Ignasi Lucas, Michael Jorgensen, Mary Ann Lemonds, Mira Markova, Moody Zheng, Nathan Taylor, Zhanar Abil, Zhi Zhou. Gracias por vuestros consejos, ayuda, compañía, y paciencia, pero sobre todo gracias por estar ahí. Quisiera destacar a Emily Lemonds por tentar tantas veces tu paciencia, por tu apoyo y cariño, y por un mundo de detalles que no podría enumerar aquí. Emily tienes bien merecida en esta disertación tu posición como asistente gráfico, psicóloga y lingüista.

En resumen, quisiera agradecer a las personas me han acompañado durante este viaje, algunas solo al principio, algunas solo al final. Pero todas ellas han estado disponibles, a veces estando a mi lado, a veces estando demasiado lejos, pero con las que he podido contar siempre, de forma incondicional, dejando que les intrigue con mis pensamientos y dudas. Gracias a vuestra ayuda he salido de numerosos atolladeros y he dado menos palos de ciego, que aun así han sido innumerables. Y les quiero decir que tengo muchos ánimos, tengo salud y ganas de continuar. Muchas gracias amigos.



Durante los años que he dedicado a este doctorado he tenido el placer y el honor de visitar diversas instituciones de enseñanza e investigación, acercándome a ellas como estudiante, becario de investigación o simplemente como visitante. Entre ellos, el Centro de Investigación y Tecnología Agroalimentaria de Aragón, the Chicago Botanic Garden, Indiana University, el Instituto Pirenaico de Ecología, the Missouri Botanical Garden, el Museo Nacional de Ciencias Naturales, el Real Jardín Botánico de Madrid, la Universidad Complutense de Madrid, the University of Illinois at Chicago, Kew Gardens, y Morton Arboretum. Queden estas palabras como signo de agradecimiento y admiración por su encomiable labor. Y como Lope de Vega escribió:

"Dando voy pasos perdidos
por tierra que toda es aire,
que sigo mi pensamiento
y no es posible alcanzarle".

Llegados a este punto, queridos lectores, querría yo añadir unas líneas del poema "*The Road Not Taken*", del poeta estadounidense Robert Frost, es posible que estos versos sinteticen la esencia del escrito que ustedes han leído o van a leer.

"*Two roads diverged in a yellow wood,*
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim
Because it was grassy and wanted wear,
Though as for that the passing there
Had worn them really about the same,

And aboth that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference."

También quisiera dar las gracias a los lectores de esta tesis por leído, en un mundo, en una sociedad de la información donde todos queremos opinar, escribir, y hablar, pero muy pocas veces estamos dispuestos a atender y reflexionar. Ya decía Don Manuel Azaña: "*… si los españoles habláramos solo y exclusivamente de lo que sabemos, se produciría un gran silencio que nos permitiría pensar…*" Seguro estoy de que los lectores de esta tesis no son mis lectores, son de la ciencia, lectores que como el autor de estas líneas, aman su oficio por encima de tantas cosas, sobretodo en estos tiempos que corren de desesperación, pero también de esperanza. Cuando quien no logra morir, continúa andando.

Sigue la ciencia, amigos míos, siguen los sueños, la magia de los datos, las gráficas y los pensamientos. Sigue el caballero de la triste figura luchando contra los molinos de viento en la Mancha, él ya sabe que no podrá vencerlos… lo ha descubierto, por fin lo ha descubierto, aunque lo volverá a intentar. Y ellos, los molinos, no saben que ya no son gigantes. A estas

alturas de mi vida me siento como un Quijote del siglo XXI cuando, habiendo nacido en La Mancha, me veo rodeado de libros, con algunos títulos reconocidos, incomprendido y abordando desesperadamente a gigantes avatares. Solo falta lanzarme a la aventura mundana a lomos de un caballo gris…, pero todo se andará. Porque esta tesis se escribió con las muchas convicciones y pocas (ninguna) certezas que me han dado estos años de estudio. Como en los buenos libros, salgo de esta aventura mejor de lo que entro, soy otro, soy más y mejor. Y no es un mérito mío, es de los demás. Como diría Don José Mujica: "*… pagamos precios enormes, pero seguimos por milagro vivos, templados y aprendiendo con la adversidad. Porque al cabo de tanto trajín sabemos que la lucha que se pierde es la que se abandona. Y también sabemos que no hay ningún final, sino el camino mismo…*" Con la firme convicción de que la taxonomía no es más que una dulce introducción al caos, acabo una etapa importante de mi vida, otra empezará y en esta etapa que empiece, como en la anterior, yo solamente aportaré lo que puedo aportar; mi honestidad crítica y mi empeño en mejorar mis habilidades y defectos. Les doy las gracias a quién me ha acompañado y a quién me ha otorgado su confianza, y lo hago con una mezcla de alegría, nostalgia y tristeza porque supone que habré terminado este capítulo de mi vida. Otros vendrán, repetirán los pasos que yo he andado, y sacarán adelante

otros proyectos, de eso estoy seguro. Por anticipado les deseo suerte, salud, ánimo, y más suerte. En nombre de todas las personas que han hecho que este texto sea una realidad, les escribe con mucho respeto y siempre con mucha admiración: Miguel Ángel Minaya Santa Cruz.

RESUMEN

El principal foco de interés de la presente tesis ha sido explorar diversos procesos de duplicación, recombinación, pseudogenización, conversiones génicas y poliploidización en las gramíneas (Poaceae), poniendo un mayor acento en la subfamilia Pooideae. Para ello nos hemos valido del análisis de genes nucleares, cloroplásticos y mitocondriales. Estos estudios, junto con la construcción de una robusta filogenia desde un enfoque molecular, la estimación de los tiempos de divergencia y los patrones biogeográficos históricos, logran explicar de lo micro a lo macroevolutivo la aparición de nuevos linajes de gramíneas templadas, y sus patrones de dispersión y de distribución histórica. En su conjunto, esta tesis doctoral ofrece una importante contribución en el campo de la sistemática molecular y evolutiva, logrando avances significativos y ofreciendo nuevas herramientas para el estudio de los genes en las plantas. La presente memoria doctoral está organizada en una introducción, los objetivos, cinco capítulos y las conclusiones finales.

En el primer capítulo se evalúa el potencial filogenético del gen nuclear copia simple *ß-amylase*, así como la existencia de pseudogenización, paralogía, homeología, y recombinación dentro de una amplia representación de gramíneas (142 especies, 16 muestras de ellas clonadas, clasificadas en 88 géneros y 7 subfamilias). Para ello comparamos la filogenia basada en el gen *ß-amylase* con una robusta filogenia organísmica de las gramíneas basada en genes nucleares (ITS) y cloroplásticos (*mat*K, *ndh*F, *trn*TL and *trn*LF). Además, se usan las técnicas de detección de recombinación (RDP4) y detección de códones terminales prematuros (Premature Terminal Codons: PTCs), así como el modelo Branch-Specific Branch-Site REL (Random Effects Likelihood), que nos indica las posibles posiciones nucleotídicas bajo selección purificadora (negativa), neutral y positiva. Nuestros resultados demuestran la existencia de siete secuencias recombinantes y otras tantas bajo selección neutral y/o positiva, lo cual podría indicar la presencia de pseudogenes en el grupo de gramíneas analizado. También hemos descrito la presencia de semi-paralogía (clones de *Vulpia alopecuros*) y homeología (clones de *Festuca simensis, F. fenas* and *F. arundinacea*). Una vez analizadas y propiamente consideradas aquellas secuencias bajo una irregular dinámica evolutiva, el gen nuclear copia simple *ß-amylasa* resultó ser un buen candidato para resolver diversos conflictos filogenéticos pobremente resueltos hasta la fecha en las Poaceae.

El segundo capítulo analiza la distribución y dinámica evolutiva de un tipo muy concreto de elementos transponibles, conocidos en lengua inglesa como *Stowaway* MITEs

(Miniature Inverted repeat Transposable Elements). Dichos elementos transponibles fueron localizados en el cuarto intrón del gen nuclear copia simple *ß-amylasa* tras analizar muestras de 117 especies de gramíneas, que representaban a 88 géneros y 4 subfamilias, siendo la subfamilia Pooideae las más ampliamente muestreada. Para analizar la presencia de los "*Stowaway* MITEs" se realizó en primer lugar un árbol organísmico de Poaceae basado en genes nucleares (*ITS*, *ß-amylasa*) y cloroplásticos (*trn*TL, *trn*LF, *ndh*F y *mat*K). A continuación se comparó la presencia de dichos elementos móviles con el marco filogenético aportado por el árbol organísmico. Nuestros resultados indican la presencia de hasta tres tipos distintos de "*Stowaway* MITEs" en el clado BEP de las gramíneas, y su llamativa ausencia de las muestras que representan al clado PACCMAD. Se describen tres hipótesis alternativas para explicar la dinámica evolutiva de dichos elementos en las gramíneas templadas: (i) su adquisición temprana en el ancestro de las Pooideae, seguido de múltiples e independientes pérdidas; (ii) inserciones independientes de hasta tres tipos distintos de "*Stowaway* MITEs" en diferentes ancestros en el mismo locus del gen *ß-amylasa*; (iii) una adquisición temprana de dichos elementos móviles en el ancestro común, seguidas por múltiples pérdidas, readquisiciones y transferencias horizontales a lo largo de la historia del clado BEP. Este capítulo ha sido publicado: Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

El tercer capítulo de esta memoria doctoral aborda uno de los más recientemente descritos y trascendentales procesos evolutivos en las angiospermas: la transferencia de secuencias génicas entre especies alejadas filogenéticamente o transferencia horizontal de genes (HGT). Para llevar a cabo esta tarea se secuenció parte del gen mitocondrial *rps*3 en 141 especies de gramíneas, las cuales representaban a 97 géneros y 10 subfamilias. Al igual que en los capítulos anteriores, se comparó la filogenia basada en el gen *rps3* con una robusta filogénia organísmica de las gramíneas basada en genes nucleares (ITS, *ß-amylasa*) y cloroplásticos (*mat*K, *ndh*F, *trn*TL and *trn*LF). Se llevaron a cabo técnicas de detención de recombinación mediante inspecciones visuales y usando programas informáticos específicos como el RDP4. También se estimó la existencia de pseudogenización mediante la localización visual de códones terminales prematuros (Premature Terminal Codons: PTCs), y del modelo Branch-Specific Branch-Site REL (Random Effects Likelihood). Nuestros resultados indican la presencia de hasta 32 recombinaciones (incluyendo 12 casos de micro-conversiones génicas), y 13 secuencias con una alta proporción de sitios bajo selección positiva y/o neutral. Todo lo ello

demuestra la naturaleza quimérica del marcador mitocondrial analizado. Estas evidencias nos llevan a concluir que la transferencia desde el genoma mitocondrial al genoma nuclear y/o cloroplástico del gen mitocondrial *rps3* en gramíneas es la hipótesis más plausible para explicar la atípica evolución de dicho marcador mitocondrial en gramineas.

El contenido del cuarto capítulo, además de seguir explorando la filogenia de la subfamilia Pooideae, incide en la evolución cromosómica y el ritmo de divergencia y las tasas de diversificación en los linajes de Poaceae. Para ello se secuenciaron cinco regiones cloroplásticas, dos codificantes (*ndhF*, *matk*) y tres no codificantes (*trnH-psbA*, *trn*TL y *trn*LF) en una selección de 170 especies de gramíneas, las cuales representa a 9 subfamilias de Poaceae. La evolución cromosómica fue analizada mediante el programa chromEvol, que utiliza el método máximo-verosímil, los tiempos de divergencia fueron estimados en el programa bayesiano BEAST, y las tasas de diversificación fueron evaluados con el programa MEDUSA. Los resultados filogenéticos corroboraron otras reconstrucciones recientes de Poaceae e indicaron que las Pooideae comenzaron su diversificación en el Eoceno medio, lo cual podría estar relacionado con el descenso de temperatura global detectado a partir de esa época. Los análisis sobre la evolución del número de cromosomas mostraron que el número de cromosomas permanece estable a través de la filogenia de las Pooideas. Además, nuestros resultados indican que la evolución poliploide ha sido predominante sobre la diploide, lo cual contradice las hipótesis postuladas recientemente para diversas angiospermas.

El quinto y último capítulo de esta tesis doctoral explora la filogenia, los tiempos de divergencia calibrados con dataciones fósiles y los patrones biogeográficos históricos en Loliinae (Pooideae), con un especial enfoque en las regiones australes más pobremente estudiadas hasta la fecha (Sudamérica, África tropical, Sudáfrica, Australia y Nueva Zelanda). Este trabajo utilizó aproximaciones bayesianas (BEAST) para la estimación de los tiempos de divergencia y métodos de reconstrucción biogeográfica mediante análisis de dispersión, extinción, y cladogénesis (DEC models). Se utilizaron secuencias del gen nuclear (ITS) y de los genes cloroplásticos (*trn*TL, *trn*LF) en 178 especies de Loliinae. Nuestros resultados sugieren que el ancestro común de las Loliinae apareció durante el Oligoceno tardío, lo que supone un adelanto con respecto a las últimas estimaciones publicadas. Por su parte, la mayor parte de las divergencias estimadas en los principales linajes de Loliinae ocurrieron entre el Mioceno medio y el inferior. Los escenarios biogeográficos estimados sugieren la existencia de recurrentes dispersiones a larga distancia (Long Distance Dispersions; LDD), neocolonizaciones y recolonizaciones, principalmente desde el hemisferio norte al hemisferio sur, pero también en

dirección opuesta, y entre áreas situadas en el hemisferio sur. Nuestras inferencias indican que las Loliinae de hoja ancha ("broad-leaved Loliinae") presentan un ancestro común originario del Mediterraneao occidental. Sin embargo, las Loliinae de hoja fina ("fine-leaved Loliinae") se originaron de forma independiente en el Mediterráneo occidental y en Sudamérica.

INTRODUCTION

STATE OF THE ART AND JUSTIFICATION FOR THIS DOCTORAL DISSERTATION

1. Importance of the grass family

If you can confidently say that you know the grass family (Poaceae), then you are misled. Even a specialist with a lifelong devotion will only just have begun to understand the diversity that exists within this family. Poaceae is the fifth-most species-rich angiosperm family, ranking only behind Asteraceae, Fabaceae, Orchidaceae and Rubiaceae. Overall there are about 800-900 genera and more than 11,000 species of grasses (TZVELEV, 1989; WATSON AND DALLWITZ, 1992; GPWG, 2012). These figures are approximate, because the limits of some genera and the number of species included in them are still controversial. Although there are other angiosperm families with more species and more genera, the Poaceae exceed all other families in one important trait: ecological dominance. Grasses dominate several of the natural and artificial landscapes of the world: steppes, temperate grasslands, and tropical-subtropical savannas (Ref. BOUCHENAK-KHELLADI *ET AL.*, 2008). Approximately one-third of the world's dry land is covered by some species of Poaceae. The subfamilies and major tribes of grasses are distributed throughout the world in broad latitudinal belts (HARTLEY, 1950; CROSS, 1980). Unlike other large families of angiosperms, which occur mostly in tropical regions (e.g., Rubiaceae, Euphorbiaceae) or in extratropical regions (e.g. Brassicaceae, Apiaceae), grasses are rather uniformly distributed on all continents and in all climatic zones. In Arctic regions they often have more species than any other family, and they are among the few species of flowering plants inhabiting the mostly ice-covered Antarctic (TZVELEV, 1989). There is also a high percentage of grasses in the flora of mountainous regions, where they extend upwards to the highest elevations inhabitable by flowering plants. The rich variety of ecological niches in mountains has fostered their micro-speciation (CLAYTON AND COPE, 1980A,B; TZVELEV, 1989).

Economically, the Poaceae are by far the most valuable of all plant families, since they provide food and habitat for humans and several animals upon which humans have come to depend (STRÖMBERG, 2004). The most important food plants of mankind are cereals such as wheat, rice, and maize. Grasses are also very important as hay and fodder plants (KELLOGG, 2001). Many food crops (maize, oats, barley, rice, etc.) and numerous species of various genera cultivated in hay fields and pastures (for example, meadow and red fescues, meadow timothy,

smooth brome-grass, and others) belong to this category. Many bamboos and some other tall grasses make good construction material, as well as providing raw material for paper and other products. The increasing economic use of grasses, the introduction of new fodder plants, and the necessity of continued selection in cultivated species make understanding the natural system of this family exceptionally important (TZVELEV, 1989). Their almost ubiquitous biogeographical presence, and their pervasive economic importance since the very beginning of human civilization mean that grasses have traditionally received abundant scientific attention, and widespread interest in their evolution and classification (MATHEWS *ET AL.*, 2000).

2. Taxonomy of the grass family

The grasses are among the most thoroughly studied and best characterized of the plant families, due largely to their economic significance. The earliest world compilation of grass genera, understandable in today's nomenclature, is the entry for the family in the *Species Plantarum* of LINNÉ (1753). The scientific effort to distinguish properly, and ultimately to classify the grass family into natural groups, continued with systematic attempts to group species according to their morphology based on macroscopic characters such as spikelets and whole inflorescences (BROWN, 1810; Fig. 1). Brown's division of the family was subsequently formalized by DUMORTIER (1823), BENTHAM (1878), BENTHAM AND HOOKER (1883), HACKEL (1887), GOEBEL (1895), and SCHUSTER (1910), among others. These authors described many tribes and subtribes of grasses according to the rules of nomenclature of their time. The next stage came in the 1930's, when the classical morphological approach to grass systematics was supplemented and complemented by the inclusion of numerous cryptic (micromorphological) characters. These advances allowed for the resolution of many, but not all, of the taxonomic problems detected up until that point. Some of the more notable taxonomic investigations of the grasses were based on microscopic characters such as cytology (AVDULOV, 1931), embryology (REEDER, 1953), anatomy (PRAT, 1932, 1936; REEDER, 1957; TATEOKA *ET AL.*, 1959), and physiology (HATTERSLEY AND WATSON, 1992).

Figure 1. Robert Brown

Systematists have traditionally used morphological characters to determine the evolutionary history of a group of organisms. These morphological characters are particularly difficult to study in the Poaceae, and require extensive developmental and anatomical investigation to establish appropriate comparisons. The plethora of grass taxonomic studies

published during the last 262 years demonstrate that it is possible to classify the grasses based only on their combinations of morphological characters. However, authors such as TZVELEV (1989) have indicated that the importance of morphological characters in the grasses is slightly overestimated. TZVELEV hypothesized that phenological characters were adaptations to particular ecological conditions. Thus, in many cases it is highly probable that they evolved in parallel in many evolutionary lines, just as uniflorous spikelets or spike-like inflorescences evolved independently in many groups of grasses of diverse affinities from many-flowered spikelet and paniculate inflorescences. These morphological conflicts explain why two of the most recently published systematic studies of the grasses based on morphological, anatomical, physiological and ecological information (CLAYTON AND RENVOIZE, 1986; WATSON AND DALLWITZ, 1992; Fig. 2) differ in the number of genera reported to be part of the family by 134. WATSON AND DALLWITZ comment on this situation in the introduction to their book, where they remark that it is easier to coalesce existing descriptions than to subdivide them when taxonomists don't agree on generic limits. Consequently, the phenological information acquired during the last two and a half centuries in the grasses has often been blurred and its biological implications uncertain, involving bidirectional character evolution, hybridization and polyploidy (DUVALL AND MORTON, 1996).

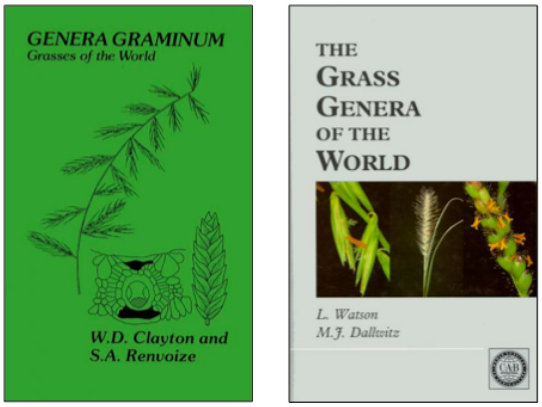


Figure 2. Cover of the Genera Graminum (CLAYTON & RENVOIZE, 1986) and The Grass Genera of the World (WATSON &DALLWITZ, 1992).

Since morphology alone fails to unambiguously resolve grass systematic relationships, its combination with modern ideas based on molecular evidence is especially useful. These ideas gave fresh impetus to a traditional discipline, though our understanding of them and their bearing on classification is far from complete. It has long been evident that molecular sequences contain useful information about evolutionary history, but until the appearance of the polymerase chain reaction (PCR) method in the mid-1980s, taxonomy studies based on DNA information were not widespread (CLEGG, 1993). Since then, molecular data have provided numerous unexpected and robust phylogenetic hypotheses at different levels in the Poaceae. HAMBY AND ZIMMER (1988) and DOEBLEY *ET AL.* (1990) published the first molecular phylogenies for the family, based on nuclear and plastid markers. The first extensive application of molecular data to grass phylogeny was undertaken by DAVIS AND SORENG (1993), using plastid DNA restriction site variation for 31 taxa representing the six subfamilies

of grasses recognized by CLAYTON AND RENVOIZE (1986). SORENG AND DAVIS (1998) combined a structural data set (including morphological, anatomical, chromosomal, and biochemical characters as well as structural features of the chloroplast genome) and an expanded chloroplast restriction site data set to analyze phylogenetic relationships within the Poaceae. The Grass Phylogeny Working Group (GPWG, 2001; Fig. 3) performed the most extended analyses on the grass family at that time, based on a combination of morphological and molecular evidence. This research yielded robust support for the major clades within the Poaceae, and supplied the basis for the first family-wide subfamilial classification based on an explicit phylogenetic hypothesis. Other remarkable grass taxonomy works based on molecular evidence were those of MATHEWS *ET AL*. (2000); DUVALL *ET AL*. (2007); HODKINSON *ET AL*. (2007); BOUCHENAK-KHELLADI (2008); and EDWARDS AND SMITH (2010).

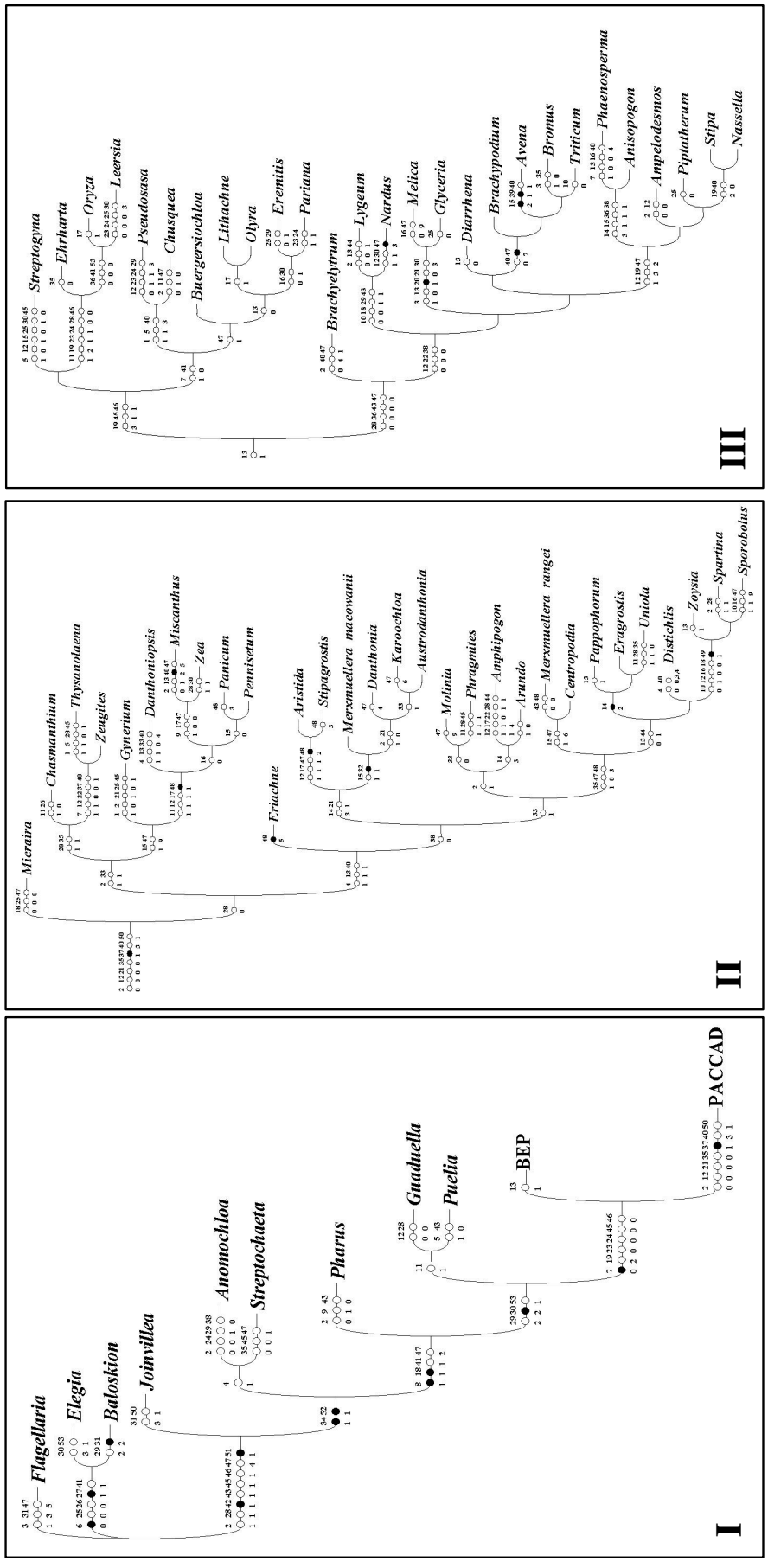


Figure 3. Most parsimonius phylogenetic trees reconstructed for the grasses and their close relatives published by the GPWG (2001). I. Overall cladogram; II. PACCAD (Panicoideae, Arundinoideae, Chloridoideae, Centothecoideae, Aristidoideae, and Danthonioideae) clade; III. BEP (Bambusoideae, Ehrhartoideae, and Pooideae) clade. Character number is indicated above the branch, and the state to which the character changes below. Filled circles represent unique occurrences of character state: open circles represent homoplasies.

Understanding the evolution of a family as large as the Poaceae requires comprehensive and robust phylogenetic trees (HODKINSON AND PARNELL, 2007; KELLOGG, 2000). Although some advances have been made in this area, we are still far from a complete "Tree of Life" for the Poaceae. The most comprehensive and robust molecular phylogeny for the grasses to date (Fig. 4) was published by GPWG (2012). This family-wide phylogenetic study was based on 531 taxa (5% of the recognized grass species) sequenced over three genetic markers from the chloroplast genome: *rbc*L, *ndh*F, and the region encompassing the *mat*K coding gene and the *trn*K introns (*trn*K/*mat*K). This phylogeny identified three species-poor lineages that were successively sisters to all other grasses (Anomochlooideae, Pharoideae and Puelioideae), and placed the bulk of grass diversity into two main strongly supported clades, known by their acronyms as BEP (Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae; 0.98 PP/ 83 BP) and PACMAD (Panicoideae, Aristidoideae, Chloridoideae, Micrairoideae, Arundinoideae, and Danthonioideae; 1.00 PP/ 100 BP) or PACCMAD (including Centothecoideae). The branching order of the six PACCMAD subfamilies is resolved with strong support for Panicoideae (1.00 PP/ 100 BP), the sister taxon relationship of Micrairoideae + Arundinoideae (1.00 PP/ 97 BP), and the Danthonioideae + the Centropodia clade-Chloridoideae (1.00 PP/ 95 BP). Strong support was found for the sister relationship of Bambusoideae / Pooideae (1.00 PP/ 65 BP) in the BEP clade, while Ehrhartoideae appeared strongly clustered with and sister to these two subfamilies (1.00 PP/ 83 BP).

The number of grass species investigated has been continuously increasing over the last two decades thanks to numerous taxonomically motivated sequencing studies of subfamilies, tribes and even smaller groups (among others in the temperate grasses, TORRECILLA *ET AL.*, 2003, 2004; QUINTANAR *ET AL.*, 2007; INDA *ET AL.*, 2008; SCHNEIDER *ET AL.*, 2009; SUNGKAEW *ET AL.*, 2009; PETERSON *ET AL.*, 2010; SALARIATO *ET AL.*, 2010; TANG *ET AL.*, 2010). Most studies focus on specific groups below the subfamily level, usually using a broad variety of nuclear and plastid noncoding markers that are frequently difficult to align between distant species. Although it is still difficult to construct a supermatrix that would allow a simultaneous approach to a high number of grasses (e.g. SALAMIN *ET AL.*, 2002; EDWARDS AND SMITH, 2010), this intense molecular research has unveiled a number of discrepancies in phylogenetic reconstruction between genomes. These conflicts have been interpreted in different ways, e.g. gene duplications/deletions, lineage sorting, reticulation or introgression due to hybridization (e.g. RIESEBERG AND SOLTIS, 1991; DOYLE *ET AL.*, 1992; SORENG AND DAVIS, 2000). Potential conflict may also arise within a single dataset. For example, BUCKLER *ET AL.* (1997) detected

divergent paralogues in ribosomal DNA, and GAUT *ET AL*. (2000) detected pseudogene copies of ITS sequences in the genus *Lolium*.

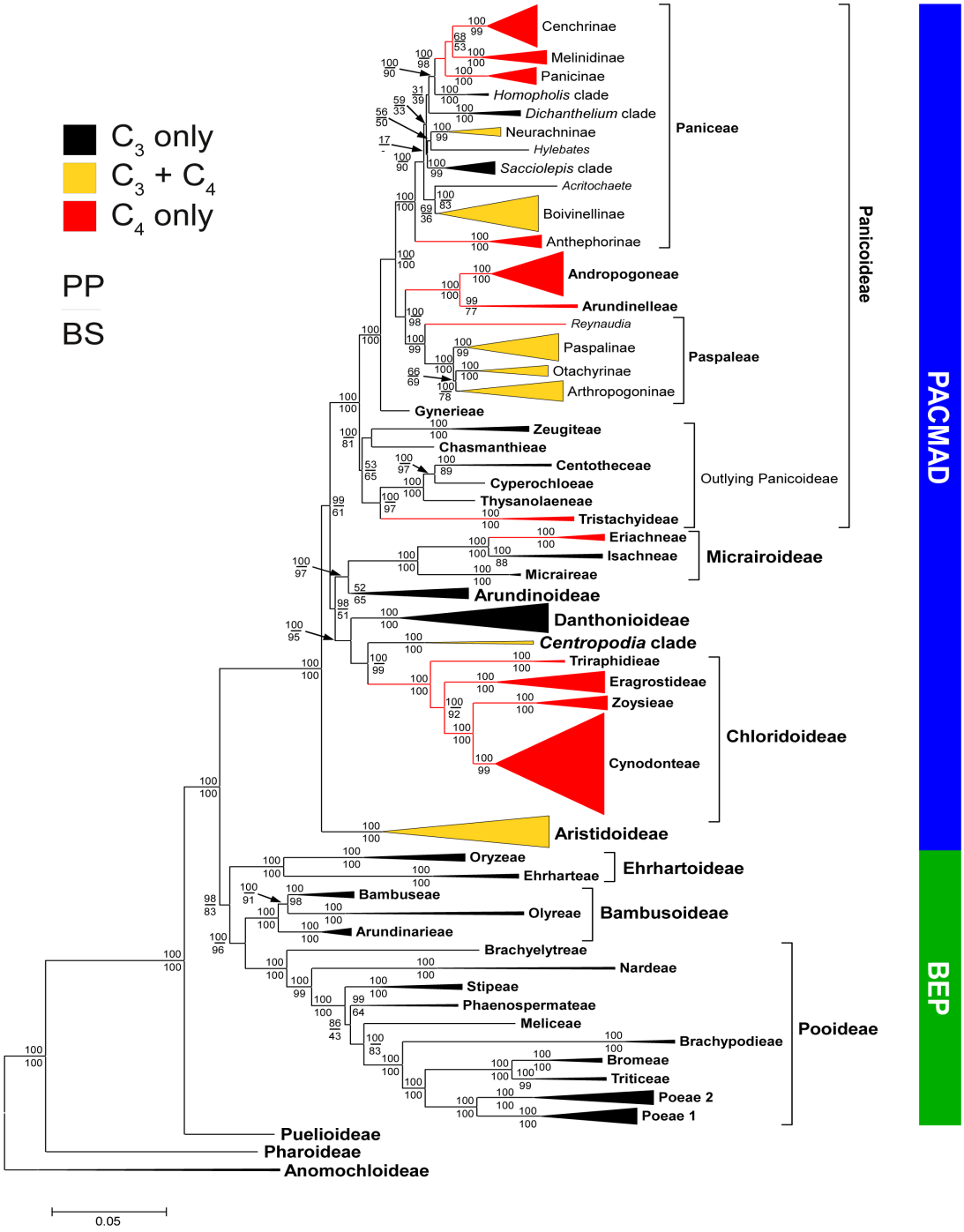


Figure 4. Bayesian consensus tree of the grass family published by GPWG (2012). Posterior probabilities are indicated above branches and likelihood bootstraps below. C_4 taxa are indicated in red.

The phylogenetic utility of cpDNA is largely limited by its slow rate of evolution and uniparental inheritance. As a result, cpDNA alone is not sufficient for phylogenetic

reconstruction at low taxonomic levels or for plant groups with histories of hybridization (OLMSTEAD AND PALMER, 1994). Nuclear markers (nrDNA), on the other hand, are biparentally inherited and usually have higher rates of mutation than cpDNA markers, so they can be used to increase phylogenetic resolution at tribal and subtribal levels. However, nuclear ribosomal (e.g. ITS) markers are unreliable for the reconstruction of hybrid speciation because the nuclear sequences derived from both parents are prone to a quick homogenization towards one of the homologues though concerted evolution and array expansion/contraction. Thus, the detection of the paralogous parental sequences could be difficult, or even impossible (BUCKLER *ET AL.*, 1997; ALVAREZ AND WENDEL, 2003; SMALL *ET AL.*, 2004). Alternatively, given the potential pitfalls of the cpDNA and nrDNA regions for inferring phylogenies, some authors have recommended routine use of low-copy nuclear genes (LCNG) or the use of the mitochondrial genome (SANG, 2002; SMALL *ET AL.*, 2004; NIETO FELINER AND ROSSELLÓ, 2007).

3. Low-copy nuclear genes

The use of low-copy nuclear genes (LCNGs) for phylogenetic reconstruction has both advantages and limitations. Among the advantages, LCNGs in plants present potentially higher mutation rates than organelle genes, and thus LCNGs are expected to contain more variable sites for reconstructing the evolution of genes and organisms at all taxonomic levels (GAUT, 1998; SENCHINA *ET AL.,* 2003; NIETO FELINER AND ROSSELLÓ, 2007). LCNGs are especially interesting at the interspecific and intraspecific levels, where universal markers such as cpDNA and nrDNA are poorly resolved, weakly supported, and/or incongruent with each other. Another desirable property of LCNGs is their biparental inheritance, because it provides information on the maternal and paternal parentage, which is crucial information when the evolutionary history of a group involves reticulation events (NIETO FELINER AND ROSSELLÓ, 2007). LCNGs are particularly effective in identifying the origins of hybrid or allopolyploid taxa because they are less frequently subject to concerted evolution and recombination than the multi-copy nuclear genes. Some remarkable studies based on LCNG are those of DOYLE *ET AL.* (1999); GE *ET AL.* (1999); SANG AND ZHANG (1999); MASON-GAMER (2001); SANG (2002); and SMALL *ET AL.* (2004).

The limitations of the LCNGs, however, need to be considered. LCNGs can be more difficult to amplify because of their low copy number, and because online databases often contain fewer comparable sequences from which amplification primers can be designed. The smaller sequence database also narrows the phylogenetic context within which new data sets

can be analyzed, and makes it more difficult to assemble the crucial copy-number information that would prevent misinterpretation of unsuspected variation among paralogues. Sequencing of LCNGs requires a larger amount of high-quality DNA compared to cpDNA and nrDNA, and usually an additional step of cloning (SANG, 2002). Cloning efforts, which are infrequent in routine phylogenetic projects, may unveil divergent intragenomic copies (orthologous sequences) derived from multiple genome donors (NIETO FELINER AND ROSSELLÓ, 2007). The LCNGs are often avoided as a source of phylogenetic information because identification of strictly orthologous sequences may be confounded by the presence of paralogous sequences if: (i) related sequences within a genome are evolving in concert (SANDERSON AND DOYLE, 1992); or (ii) we fail to sample all members of a gene family, even if they are evolving independently (ALVAREZ AND WENDEL, 2003; NIETO FELINER AND ROSSELLÓ, 2007).

Since LCNGs come at a higher cost with regard to development than either cpDNA or rDNA markers, they remain under-represented in phylogenetic studies in plants. So the real utility of LCNGs for phylogenetic reconstruction is less encouraging than their potential advantages would predict. SMALL *ET AL*. (2004) have said that their use is in its "relative infancy". In spite of the difficulties, a variety of phylogenies based on LCNGs are available on the grass family. Two remarkable studies are those published by MASON-GAMER *ET AL*. (2001), based on the phylogeny of the *waxy* gene, and MASON-GAMER *ET AL*. (2005) based on *ß-amylase* sequences. Following the steps of Mason-Gamer's research on the *ß-amylase* gene, the first and second chapters of this dissertation will be based on the nuclear sequences of the low-copy *ß-amylase* region. The first chapter of this dissertation will look at the molecular evolution and phylogeny of the *ß-amylase* gene in the Poaceae. Specifically, we analyzed the existence of paralogy, recombination, and phylogenetic bias within the coding and non-coding positions of the *ß-amylase* gene. The second chapter will analyze the occurrence of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) and their potential footprints in representatives of the whole grass family, but with special attention to the Pooideae subfamily, as it was assessed within an explicit phylogenetic framework.

4. Miniature Inverted repeat Transposable Elements (MITEs)

Closely associated with the introns of the LCNGs are the Miniature Inverted repeat Transposable Elements (MITEs) (LU *ET AL*., 2012, and references therein). MITEs are a heterogeneous group of small—from a few dozen to a few hundred base pairs—non-autonomous mobile elements. Numerically, they are the most abundant of all types of

transposable elements in plants, with copy numbers in the thousands (FESCHOTTE *ET AL.*, 2002A, 2003; JIANG *ET AL.*, 2004). MITEs likely originated from defective remnants of a subset of much longer autonomous class II DNA transposons (FESCHOTTE AND WESSLER, 2002; FESCHOTTE *ET AL.*, 2002A,B, 2003, 2005; ZHANG *ET AL.*, 2004; MENZEL *ET AL.*, 2006) that lost their transposase-encoding ORFs due to internal deletions (FLAVELL *ET AL.*, 1994). MITEs have no coding capacity, and thus their mobility depends on the activity *in trans* of transposases encoded by cognate full-length autonomous class II transposons (JIANG *ET AL.*, 2003, 2004; FESCHOTTE *ET AL.*, 2005) that recognize a common Terminal Inverted Repeated (TIR) sequence. These specific transposases catalyze both the insertion and the excision of their associated transposons (FESCHOTTE *ET AL.*, 2002A,B; JIANG *ET AL.*, 2003; NAKAZAKI *ET AL.*, 2003). Thus, MITEs should be expected to excise with a certain frequency, and—as is the case in class II transposons—leave footprints upon excision. Neither substitutions nor deletions lead to compensatory changes; hence, the highly stable secondary structure of the elements may gradually be reduced (PETERSEN AND SEBERG, 2000; FESCHOTTE *ET AL.*, 2002B). Although the dynamics underlying the acquisition, loss, and high copy number of MITEs are obscure, MASON-GAMER (2007) described their presence at the fourth intron of the *ß-amylase* gene in various Triticeae genera, revealing a complex history of gains and losses of MITEs in the wheat tribe.

5. Mitochondrial genome

Five hundred million years of land plant evolution resulted in few alterations to the chloroplast genome but in considerable changes to the mitochondrial genome (mtDNA) (KNOOP, 2004). The angiosperms' mitochondrial genome is among the least understood, in part because it is the largest and most complex of those of the eukaryotes. Furthermore, mtDNA is generally the most slowly evolving of the three plant cell genomes at the sequence level, with unique exceptions in only some plant lineages (WOLFE *ET AL.*, 1987; EYRE-WALKER AND GAUT, 1997). Despite the large size of the current mitochondrial genome in angiosperms, the gene content is much less than that of a eubacterial genome, (a possible ancestral symbiont), suggesting the occurrence of gene transfer events during the course of cell evolution (GRAY *ET AL.*, 1992). MARIENFELD *ET AL.* (1999) and KOULINTCHENKO *ET AL.* (2003) consider that only 11-18% of mtDNA correspond to protein or structural genes. This means that non-coding regions and intergenic spacer regions occupy a majority of the mitochondrial genome, together with pseudogenes, pieces of foreign DNA of unknown origin, many small repeated sequences, and regulatory elements such as promoter and terminator elements (ADAMS AND PALMER, 2003;

KUBO AND MIKAMI, 2007; KUBO AND NEWTON, 2008). In strong contrast, animal (metazoa) mtDNA is smaller and has intron-less genes, so it has become more compact since its endosymbiotic origin. Rearrangements in animal mitochondrial genomes are exceptionally rare events that are unlikely to have occurred independently in separate lineages. Thus, mtDNA sequences have long been an important source of data in animal phylogenetics, equivalent in importance to the use of plastid DNA data in plant studies (BOORE, 1999).

Nuclear and chloroplast plant genomes evolve principally by point mutations; however, the plant mitochondrial genome has a very low substitution rate, and its evolution is characterized by frequent structural rearrangements (legitimate and illegitimate recombinations) (PALMER AND HERBON, 1988). Since the successive genomic rearrangements result in the generation of subgenomic molecules (loop outs) or isomeric forms (flip flops), angiosperm mitochondrial genomes have lost and acquired sequences through this process, as initially pointed out by ALBERT *ET AL.* (1998) and more recently by KUBO AND MIKAMI (2007). Thus, these rearrangements can affect gene expression, producing polymorphism between the genomes, or even disrupt the continuity of genes (KUBO AND NEWTON, 2008). Recombination via long and short repeated sequences can be found in mtDNA. The difference between the long and short repeats is not only the length but also the activity of recombination. Whereas many of the long repeats appear to actively and frequently recombine, the short repeats are usually inactive, but they seem to play a central role in the evolution of angiosperm mitochondrial genomes (ANDRE *ET AL.*, 1992; CONKLIN AND HANSON, 1994; KNOOP, 2004; KUBO AND NEWTON, 2008). The mitochondrial genomes are presumed to generate chimeric molecules in two different contexts: (i) intramitochondrial recombination (NEWTON *ET AL.*, 2004); and (ii) recombination between native and foreign mitochondrial genes through horizontal gene transfer (HGT) (BERGTHORSSON *ET AL.*, 2003; BARKMAN *ET AL.*, 2007; WEILONG AND PALMER, 2009). Much of the mitotype-specific sequence has been determined to originate from the chloroplast, the nucleus or from plasmids resident in the mitochondria; however, DNA whose origin is unknown can also be found in the mtDNA genomes of grasses (ALLEN *ET AL.*, 2007). The third chapter of this dissertation evaluates the phylogenetic signal of the mitochondrial *rps3* gene and the potential existence of Horizontal Gene Transfer (HGT) and other evolutionary events, like pseudogenization, recombination, and gene conversions, within a broad representation of the main Poaceae lineages

6. The mitochondrial genome in phylogenetic studies

The reluctance among botanists to use mitochondrial data was induced primarily by the pronounced structural diversity of plant mitochondrial genomes, caused at least in part by their ability to recombine. Structural instability coupled with a low level of sequence diversity rendered the use of restriction endonuclease site variability impractical in the early days of molecular systematics. By the time systematists turned to DNA sequence data, the focus on plastid and nuclear genomes had already been established (PALMER, 1992; HIESEL *ET AL.*, 1994; PETERSEN *ET AL.*, 2006). However, in recent years plant systematists have begun exploring mitochondrial gene sequences as a new source of useful phylogenetic characters. Mitochondrial genes have been used mainly at high taxonomic levels due to the conserved nature of the coding sequences. Hence, an increasing amount of mitochondrial sequence data is being collected: e.g. for the phylogeny of monocots (CHASE *ET AL.*, 2006), Asparagales (PIRES *ET AL.*, 2006) and Liliales (FAY *ET AL.*, 2006).

As previously described for the low copy nuclear genes, the use of mtDNA for phylogeny reconstructions has both advantages and limitations. Primary among the advantages is that mitochondria belong to a separate linkage group from plastids, and hence provide independent phylogenetic evidence. Further, mitochondrial sequences may be the only data that are able to place achlorophyllous taxa in phylogenies based on data derived from the organelle genomes (PETERSEN *ET AL.*, 2006). The limitations of the mtDNA, on the other hand, need to be considered. The slow substitution rates described for mtDNA in angiosperms may result in little phylogenetic information on specific groups below the family or subfamily level. Features of the plant mitochondrial genome such as RNA editing, gene duplication (paralogous sequences), rearrangements that lead to gene conversion, and gene transfer may potentially create problems for phylogenetic reconstructions, and likely explain the incongruence found between phylogenies based on mitochondrial and plastid data sets (BERGTHORSSON *ET AL.*, 2003; PETERSEN *ET AL.*, 2006). Gene conversion events, when properly recognized as such, represent the so-called "rare genomic changes", which have the potential to serve as individual phylogenetic characters of notable importance (ROKAS AND HOLLAND, 2000). However, when unrecognized (as has been the case heretofore), these "single characters" will be improperly overweighed. Overweighting because of unrecognized conversion in combination with the very low substitution rates that characterize most plant mitochondrial genomes could cause serious phylogenetic error (WEILONG AND PALMER, 2009). Furthermore, the occurrence of gene conversions blurs the boundaries between species even more than hybridization (DAVIES,

1996). Consequently, accepting that plant mitochondria exchange genes relatively frequently, caution is necessary when interpreting plant phylogenies from one or a few mitochondrial genes, as they may not reflect the underlying organismal phylogeny. Fortunately, the majority of molecular phylogenetic studies in plants have used chloroplast and nuclear sequences, which seem essentially immune to frequent structural rearrangements and HGT (BERGTHORSSON *ET AL.*, 2003; RICHARDSON AND PALMER, 2007).

7. Horizontal Gene Transfer

Our concept of a universal tree of life rests on the assumption that most or all of the components of an organism share a common evolutionary heritage: the transmission of traits from parents to offspring, also known as vertical inheritance. This assumption is often violated at the molecular level due to the process known as horizontal gene transfer (HGT). HGT, understood in this dissertation as the transfer of genes or DNA sequences across reproductively isolated species, suggests that there is no fundamental barrier to gene flow between non-mating species. In most known cases the process is unidirectional, and it rarely involves reciprocal exchanges of DNA. The analysis of complete genome sequences has led to suggestions that the extent of HGT is a major evolutionary force in the prokaryotic world (e.g. DAVIES, 1996; EISEN, 2000; KOONIN *ET AL.*, 2001; NAKAMURA *ET AL.*, 2004). However, in recent years we have seen a small explosion of studies reporting cases of horizontal transfer of genes in the mitochondrial genome of plants (e.g. DAVIS AND WURDACK, 2004; RICHARDSON AND PALMER, 2007) but also within the nuclear genome (MASON-GAMER, 2005; MINAYA *ET AL.* 2013). Many transfers are quite recent in evolutionary terms, being restricted to a single genus of recipient plants, or even a limited number of species within a genus (BERGTHORSSON *ET AL.*, 2003; MOWER *ET AL.*, 2004). Interestingly, as far as we know, not a single case of HGT has been documented in chloroplast or in animal mitochondrial genomes, despite their far greater sampling.

HGT should really be considered a multi-step process. First, the sequence to be transferred evolves within its donor lineage. At some point, the sequence is then transferred to another lineage either via a vector or by direct or indirect DNA exchange. The ways that sequences from distant relatives could come into contact with one another to transmit information is currently a matter of speculation and could include illegitimate pollination, parasitic or epiphytic plant growth on host plants (RICHARDSON AND PALMER, 2007) or other unknown vectors. Then, the foreign sequence must also be in or get into the recipient genome

in a format that allows for long-term maintenance and replication. Most of the known cases of plant HGT seem to have occurred by illegitimate (non-homologous) recombination of the donor gene into the recipient genome (WEILONG AND PALMER, 2009), though this hypothesis has yet to be confirmed. Finally, once this extra, unneeded DNA is incorporated, it is subject to rearrangement and eventually loss, so that it becomes unrecognizable with the passage of time (KUBO AND NEWTON, 2008). The second and third chapters of this dissertation will describe potential cases of HGT within the nuclear and mitochondrial genomes in the grass family, with special focus on the Pooideae subfamily.

8. Diversification rates and chromosome evolution models

The extraordinary contemporary species richness and ecological predominance of the angiosperms are even more remarkable when considering the relatively recent onset of their evolutionary diversification. We examine the evolutionary diversification of the Pooideae subfamily by estimating its rate of diversification. By making an integrative and critical use of the paleobotanical record, we obtain reasonably secure approximations for the age of a large set of Pooideae clades. Diversification was modeled as a stochastic, time-homogeneous birth-and-death process that depends on the diversification rate (r) and the relative extinction rate (ε). In species-rich clades, diversity is heterogeneously distributed; there are some branches that have diversified profusely, while others have given rise to only a few species (MAGALLON AND SANDERSON, 2001). Diversification rates in angiosperm clades have been measured using very different methods, and the obtained results have reflected a great disparity. Speciation rates in the Poales, for example, have been estimated to be between r=0.0611 spp. My^{-1} and r=1,013 spp. My^{-1}, depending on the parameters of the analysis (MAGALLÓN AND CASTILLO, 2009).

A great variety of different traits, alone or in combination, have been advocated as drivers of diversification or as correlates of species-richness. Biotic pollination and an herbaceous life form have been found to be positively correlated with species richness or high diversification rate (e.g., BURGER, 1981; STEBBINS, 1981; ERIKSSON AND BREMER, 1992; RICKLEFS AND RENNER, 1994, 2000; DODD *ET AL.*, 1999; SILVERTOWN *ET AL.*, 2000). Other positively correlated traits include the rate of chromosomal evolution (LEVIN AND WILSON, 1976), the rate of molecular substitution (BARRACLOUGH *ET AL.*, 1996; BARRACLOUGH AND SAVOLAINEN, 2001, but see DAVIES *ET AL.*, 2004), defense against predators (EHRLICH AND RAVEN, 1964; FARRELL *ET AL.*, 1991), floral zygomorphy (SARGENT, 2004), environmental energy (DAVIES *ET AL.*, 2004) and the capacity of species to adopt new life history attributes

(RICKLEFS AND RENNER, 1994, but see DODD *ET AL.,* 1999; RICKLEFS AND RENNER, 2000; SILVERTOWN *ET AL.,* 2000). Other traits such as dioecy (HEILBUTH, 2000) and age at maturity among woody species (VERDÚ, 2002) have been found to be negatively correlated with angiosperm species-richness/ diversification.

Information on chromosome numbers and diversity of taxonomic groups is commonly used in phylogenetic analyses in order to investigate aspects such as the *tempo* and *mode* of diversification in a lineage and the impact of chromosome changes on diversification (ESCUDERO *ET AL.*, 2012). Chromosome evolution has been demonstrated to have profound effects on diversification rates and speciation in angiosperms (SOLTIS *ET AL.*, 2009). While variation in chromosome number is widespread among plants, its role in species diversification has long been debated (STEBBIN, 1971; SOLTIS AND SOLTIS, 2000; SOLTIS ET AL., 2009). Transitions in chromosome number comprise the multiplication of a whole chromosome set (COHLAN *ET AL.*, 2005), which may entail: (i) a whole genome duplication; (ii) an increase by half of the genome; (iii) demipolyploidy, when one homeologous set of chromosomes is duplicated in an already existing polyploid (MAYROSE *ET AL.,* 2010); or (iv) changes in single chromosome number resulting from fission, fusion, duplication or deletion of a single or a few chromosomes. Demipolyploidy (*sensu* MAYROSE *ET AL.*, 2010) is thought to occur through the crossing of gametes of different ploidy levels (e.g., a tetraploid crossed with a diploid, followed or preceded by a doubling of the genome, yields a hexaploid, which is inferred as a 4X to 6X demiploid transition. Interestingly, according to ESCUDERO *ET AL.* (2014), there has been comparatively little attention to the evolutionary role of gains and losses of single chromosomes, which may or may not entail changes in the DNA content (then called aneuploidy or dysploidy, respectively).

Researchers have long argued about the prevalence of polyploidy in flowering plants as well as its evolutionary and genomic consequences (SOLTIS *ET AL.,* 2009). Although it is clear that chromosome transitions, especially polyploidy in the broad sense, are widespread in angiosperm evolution (e.g. SOLTIS *ET AL.*, 2009; MAYROSE *ET AL.*, 2011), their evolutionary significance remains obscure. Diverse works have demonstrated that polyploidy is ubiquitous in angiosperms and has played an important role in many lineages, with evidence of several rounds of both ancient (paleopolyploid) and recent (neopolyploid) polyploidizations (LEITCH *ET AL.*, 2004; VAN DE PEER *ET AL.*, 2009; WOOD *ET AL.*, 2009; JIAO *ET AL.*, 2011; MAYROSE *ET AL.*, 2011; VAN DE PEER, 2011). However, what is ancient and recent varies among authors (e.g. neopolyploidy may be defined as a polyploidization event within a genus or within a species

(GUERRA, 2008). In addition, there are still conflicting opinions on whether there is a positive or negative relationship between ancient polyploid events and diversification rates. Some studies have detected a relationship between chromosome transitions and diversification (e.g. SOLTIS *ET AL.*, 2009), whereas others support the traditional view that chromosome transitions, mainly polyploidy, are evolutionary dead-ends (STEBBINS, 1971; SOLTIS AND SOLTIS, 2003). Polyploidizations are rather frequent, but newly formed polyploid lineages generally fail to persist, which would explain their low diversification rates and the biased distribution of polyploidy towards terminal branches of the plant tree of life (FAWCETT *ET AL.*, 2009; MAYROSE *ET AL.*, 2011). Whereas neopolyploids are widespread among angiosperms (cf. WOOD *ET AL.*, 2009; MAYROSE *ET AL.*, 2011), paleopolyploidization events are comparatively very rare (FAWCETT *ET AL.*, 2009; JIAO *ET AL.*, 2011; MAYROSE *ET AL.*, 2011). Thus, although polyploidy may initiate rapid diversification within distinct evolutionary lineages (i.e. seed plants and angiosperms; JIAO *ET AL.*, 2011), it is generally associated with decreased diversification rates and higher extinction risk (WOOD *ET AL.*, 2009; MAYROSE *ET AL.*, 2011; but see SOLTIS *ET AL.*, 2014). Polyploid lineages may have succeeded in becoming established during past periods of great environmental upheaval and mass extinction events, which created new ecological niches and disturbed habitats (FAWCETT *ET AL.*, 2009; SOLIS AND BURLEIGH, 2009; FAWCETT AND VAN DE PEER, 2010). The surviving polyploid lineages could have an enhanced potential for diversification due to their genetic, genomic and epigenetic features (e.g. SOLTIS AND SOLTIS, 2000; OTTO AND WHITTON, 2000; VAN DE PEER, 2011), thus attaining long-term evolutionary success (FAWCETT *ET AL.*, 2009; FAWCETT AND VAN DE PEER, 2010; MAYROSE *ET AL.*, 2011; VAN DE PEER, 2011). A parallel situation may currently be reflected in the high proportion of polyploid species in harsh environments such as high altitudes or latitudes (STEBBINS, 1984; BROCHMANN *ET AL.*, 2004).

Grass genome evolution models (e.g. SALSE *ET AL.*, 2008) have inferred a chromosome base number x=5 for the ancestor of the grasses (c. 90 Ma), followed by paleoduplication (paleopoliploidization) to x=10 and additional breakage plus fusion to x=12 (c. 70-50 Ma). From this putative diploidized paleopolyploid ancestor, distinctly inferred genomic arrangements led to current genomes that are relatively conserved (e.g. *Oryza*, x=12), to other more severely reduced (e.g. *Triticum*, x=7), or to alternative rearrangements that resulted in different genomic reductions (e.g. *Sorghum*, x=10; *Zea*, x=10) (cf. SALSE *ET AL.*, 2008). In the highly diverse Pooideae subfamily, x=7 is considered to be the commonest chromosome base number in different groups (e.g. Triticeae, SALSE *ET AL.*, 2008; Poodae, HSIAO *ET AL.*, 1995). A

decreasing aneuploid series was proposed in the Pooideae (CATALAN *ET AL.*, 1997), ranging from x=13, 10 in the more ancestral Lygeae (and Brachyelytreae, x=11; SAARELA *ET AL.*, 2003), through x= 12, 11, 10, 9, 8 in the successively diverged Stipeae, Meliceae and Brachypodieae, to x=7 in the more recently evolved core pooids (Poodae + Triticeae s. l., including the former tribes Poeae and Aveneae and Triticeae and Bromeae). A strong reduction to very small chromosome base numbers has also been documented in some pooid lineages [e.g. x=5 Brachypodieae, x=5 Anthoxanthiinae (PIMENTEL *ET AL.*, 2013)] including one of the lowest numbers in the angiosperms [x=2 *Zingeria, Colpodium* (KIM *ET AL.*, 2009)].

The fourth chapter of this dissertation will explore the diversification rates and the speciation-extinction rates within the main Pooideae lineages. We used MEDUSA (ALFARO *ET AL.*, 2009) as implemented in the R package Geiger (HARMON *ET AL.*, 2008; R DEVELOPMENT CORE TEAM, 2011). Furthermore, we investigated the patterns of chromosome number evolution in the main Pooideae lineages by combining two coding (*matK*, *ndhF*) and three non-coding (*trn*TL, *trn*LF, *trnH-psbA*) plastid regions with cytogenetic information taken from our own records and the literature (*e.g.* www.tropicos.org, Missouri Botanical Garden; WATSON AND DALLWITZ, 1992). We used the software CHROMEVOL v. 2.0 (MAYROSE *ET AL.*, 2010), which implements a likelihood-based method for tracking the pattern of haploid chromosome number change along a phylogeny (MAYROSE, 2014). Finally, we combine the phylogenetic information of the main Pooideae lineages with the major climatic and biome changes that occurred on Earth throughout the Cenozoic in order to gain insight into the evolution of the Pooideae.

9. Biogeography of the Loliinae subtribe

The fifth chapter of this dissertation uses chloroplast and nuclear DNA sequences to investigate another important aspect of grass evolution: the molecular-based dating and biogeography of the Loliinae subtribe (Pooideae). Loliinae is one of the largest groups of temperate grasses and is distributed throughout all the continents except Antarctica. Previous biogeographical analyses confirmed the Mediterranean origin of this subtribe, which has usually been associated with the geological and climatic changes of the Late Tertiary and Quaternary periods (BALFOURIER *ET AL.*, 2000; FJELLHEIM *ET AL.*, 2006; CATALAN, 2006; INDA *ET AL.*, 2008). However, dating and estimation of the biogeographical patterns of the Loliinae lineages have been hampered until very recently by the lack of reliable fossil records, the absence of comprehensive phylogenetic trees, and the need for development of a statistical

model that would allow investigators to take dating into account and create biogeographical models of this important lineage of grasses. The lack of reliable fossil records for estimating the absolute divergence time of members of the Pooideae subfamily was palliated, at least in part, by the compendium published by STRÖMBERG (2011). The absence of comprehensive Loliinae phylogenetic trees has been partially resolved during the last decade thanks to the more complete molecular phylogenies published (e.g. CATALAN *ET AL.*, 2004; TORRECILLA *ET AL.*, 2003; 2004; CATALAN, 2006) and the broad range of sampling performed in the Equatorial and Southern Hemisphere regions (Africa, Asia, South America, Australia and Southern Pacific regions) (e.g. CHEN *ET AL.*, 2003; SORENG *ET AL.*, 2003; NAMAGANDA *ET AL.*, 2006; STANCÍK AND PETERSON, 2007; NAMAGANDA AND LYE, 2008). Dating methods have evolved from the strict-molecular clock (e.g. ZUCKERKANDL AND PAULIN, 1965) to more realistic methods that use a "relaxed clock" approach, which models the rate variation among lineages (e.g. Penalized Likelihood, SANDERSON, 2002; Multidivtime, THORNE AND KISHINO, 2002; Relaxed Phylogenetics, DRUMMOND *ET AL.*, 2006). Similarly, in recent years researchers have been able to reconstruct past biogeographical ranges in a way that strives to integrate information from phylogenies, fossils, molecular dating, the geologic record, and paleogeographic and paleoclimatic reconstructions. The integration of all this information and its computer processing in a reasonable time is only now possible thanks to the recent development of parametric models in historical biogeographical inference such as dispersal-extinction-cladogenesis (DEC) analysis (more information in REE AND SANMARTÍN, 2009).

10. Literature cited

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29, 380-395.

Albert, B., Godelle, B., Gouyon, P.H., 1998. Evolution of the plant mitochondrial genome: dynamics of duplication and deletion of sequences. *Journal of Molecular Evolution* 46, 155-158.

Allen, J.O., Fauron, C.M., Minx, P., Roark, L., Oddiraju, S., Lin, G.N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S.W., Newton, K.J., 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 117, 1173-1192.

Alfaro, M.E., Santinia, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G., Harmong, L.J., 2009. Nine exceptional radiations plus high turnover explain

species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences USA* 106, 13410-13414.

Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29, 417-434.

Andre, C., Levy, A., Walbot, V., 1992. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends in Genetics* 8, 128-132.

Avdulov, N.P., 1931. Kario-sistematicheskoye issledovaniye semeystva zlakov. *Trudy Prikl. Bot. Prilozheniye* 44: 1-352 [Karyosystematic studies in the grass family. Supplement 44 to *The Bull. Appl. Bot. Genet. Pl. Breed.*, Leningrad. Russian text, pp. 1-352; German summary, pp. 353-425; index pp. 426. English translation published by the Smithsonian Institution and the National Scientific Documentation Centre, New Delhi. 1975]

Balfourier, F., Imbert, C., Charmet, G., 2000. Evidence for phylogeographic structure in *Lolium* species related to the spread of agriculture in Europe. A cpDNA study. *Theoretical and Applied Genetics* 101, 131-138.

Barkman, T.J., McNeal, J.R., Lim, S., Coat, G., Croom, H.B., Young, N.D., dePaphilis, C.W., 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 7, 248-263.

Barraclough, T.G., Harvey, P.H. and Nee, S., 1996. Rate of *rbcL* gene sequence evolution and species diversification in flowering plants (angiosperms). *Proceedings of the Royal Society of London, Biological Sciences* 263, 589-591.

Barraclough, T.G., Savolainen, V., 2001. Evolutionary rates and species diversity in flowering plants. *Evolution* 55, 677-683.

Bentham, G., 1878. *Flora Australiensis* 7, 449-670.

Bentham, G., Hooker, J.D., 1883. Gramineae. pp. 1074-1215. *Genera Plantarum*, vol. 3. Reeve, London.

Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D., 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424, 197-201.

Boore, J., 1999. Animal mitochondrial genomes. *Nucleic Acids Research* 27, 1767-1780.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses

(Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Brochmann, C., Brysting, A.K., Alsos, I.G., Borgen, L., Grundt, H.H., Scheen, A., C., Elven, R., 2004. Polyploidy in arctic plants. *Biological Journal of the Linnean Society* 82, 521-536.

Brown, R., 1810. *Prodomus florae Novae Hollandiae et insulae Van-Diemen*, Vol. 1. London.

Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821-832.

Burger, W. C., 1981. Why are there so many kinds of flowering plants? *Bioscience* 31, 572-581.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., 2006. Phylogeny and evolution of *Festuca* L. and related genera of subtribe Loliinae (Poeae, Poaceae). In: A.K. Sharma, A. Sharma (Eds.), *Plant Genome. Biodiversity and Evolution*. Science Publishers, Enfield, NH, US, pp. 255-303.

Chase, M.W., Fay, M.E., Devey, D.S., Maurin, O., Ronsted, N., Davies, T.J., Pillon, Y., Petersen, G., Seberg, O., Tamura, M.N., Asmussen, C.B., Hilu, K., Borsch, T., Davis, J.I., Stevenson, D.W., Pires, J.C., Givnish, T.J., Systsma, K.J., McPherson, M.A., Graham, S.W., Rai, H.S., 2006. Multigene analyses of monocot relationships: a summary. In: J.T. Columbus, E.A. Friar, J.M. Porter, L.M. Prince, M.G. Simpson (Eds.), *Monocots: comparative biology and evolution (excluding Poales)*. Rancho Santa Ana Botanic Garden, Claremont, California, USA, pp. 63-75.

Chen, X., Aiken, S.G., Dallwitz, M.J., Bouchard, P., 2003. Systematic studies of *Festuca* (Poaceae) occurring in China compared with taxa in North America. *Canadian Journal of Botany* 81, 1008-1028.

Clayton, W.D., Cope, T.A., 1980a. The chorology of Old World species of Gramineae. *Kew Bulletin* 35, 135-171.

Clayton, W.D., Cope, T.A., 1980b. The chorology of North American species of Gramineae. *Kew Bulletin* 35, 567-576.

Clayton, W.D., Renvoize, S.A., 1986. Genera Gramineum: Grasses of the World. *Kew Bulletin*. Addit. Series XIII, Royal Botanic Gardens, Kew.

Clegg, M., 1993. Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences, U.S.A.* 90, 363-367.

Conklin, P.L., Hanson, M.R., 1994. Recombination of plant mitochondrial genomes. In: J. Paszkowski (Ed.), *Homologous Recombination and Gene Silencing in Plants*. Kluwer. Academic Publishers, Netherlands, pp. 61-81.

Cross, R.A., 1980. Distribution of subfamilies of Gramineae in the Old World. *Kew Bulletin* 35, 279-289.

Cohlan, A.E., Eichler, E., Olivera, S.G., Paterson, A.H., Stein, L., 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends in Genetics* 21: 673-682.

Davis, JI., Soreng, RJ., 1993. Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. *American Journal of Botany* 80, 1444-1454.

Davis, C.C., Wurdack, K.J., 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* 305, 676-678.

Davies, J., 1996. Origins and evolution of antibiotic resistance. *Microbiologia* 12, 9-16.

Davies, T.J., Savolainen, V., Chase, M.W., Moat, J., Barraclough, T.G., 2004 . Environmental energy and evolutionary rates in flowering plants. *Proceedings of the Royal Society of London, Biological Sciences* 271, 2195-2200.

Doebley, J., Durbin, M., Golenberg, E.M., Clegg, M.T., Ma, D.P., 1990. Evolutionary analysis of the large subunit of Carboxylase (*rbc*L) nucleotide sequence data among the grasses (Poaceae). *Evolution* 44, 1097-1108.

Dodd, M.E., Silvertown, J., Chase, M.W., 1999. Phylogenetic analysis of trait evolution and species diversity variation among angiosperm families. *Evolution* 53, 732 – 744.

Doyle, J.J., Davis, J.I., Soreng, R.J., Garvin, D., Anderson, M.J., 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). *Proceedings of the National Academy of Sciences, U.S.A.* 89, 7722-7726.

Doyle, J.J., Doyle, J.L., Brown, A.H.D., 1999. Origins, colonization, and lineage recombination in a widespread perennial soybean polyploid complex. *Proceedings of the National Academy of Sciences of the United States of America* 96, 10741-10745.

Drummond, A.J., Ho, S.Y.W., Philllips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4, 699-710.

Dumortier, B.C.J., 1823. *Observations sur les Graminées de la flore Belgique*. J. Casterman, Tournay.

Duvall, M.R., Morton, B.R., 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbc*L sequence data. *Molecular Phylogenetics and Evolution* 5, 352-358.

Duvall, M.R., Davis, J.I., Clark, L.G., Noll, J.D., Goldman, D.H., Sánchez-Ken, J.G., 2007. Phylogeny of the grasses (Poaceae) revisited. *Aliso* 23, 237-247.

Edwards, E.J., Smith, S.A., 2010. Phylogenetic analyses reveal the shady history of C4 grasses. *Proceedings of the National Academy of Sciences, U.S.A.* 107, 2532-2537.

Ehrlich, P.R. Raven, P.H., 1964. Butterflies and plants: A study in coevolution. *Evolution* 18, 586-608.

Eisen, J.A., 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Current Opinion in Genetics and Development* 10, 606-611.

Eriksson, O., Bremer, B., 1992. Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. *Evolution* 46, 258-266.

Escudero, M., Hipp, A.L., Waterway, M.J., Valente, L.M., 2012. Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (Carex, Cyperaceae). *Molecular Phylogenetics and Evolution* 63, 650-655.

Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A.L., Pimentel, M., Jiménez-Mejías, P., Valcárcel, V., Vargas, P., Luceño, M., 2014. Karyotypic Changes through Dysploidy Persist Longer over Evolutionary Time than Polyploid Changes. *PLoS ONE 9*, e85266.

Eyre-Walker, A., Gaut, B.S., 1997. Correlated Rates of synonymous site evolution Across plant genomes. *Molecular Phylogenetics and Evolution* 14, 455-460.

Farrell, B., Dussourd, D.E., Mitter, C., 1991. Escalation of plant defense: Do latex and resin canals spur plant diversifi cation? *American Naturalist* 138, 881-900.

Fay, M.F., Chase, M.W., Ronsted, N., Devey, D.S., Pillon, Y., Pire, J.C., Petersen, G., Seberg, O., Davis, J.I., 2006. Phylogenetics of Liliales: summarized evidence from combined analyses of five plastid and one mitochondrial loci. In: J.T. Columbus, E.A. Friar, J.M. Porter, L.M. Prince, M.G. Simpson (Eds.), *Monocots: comparative biology and evolution (excluding Poales)*. Rancho Santa Ana Botanic Garden, Claremont, California, USA., pp. 559-565.

Fawcett, J.A., Maere, S., Van de Peer, Y., 2009. Plants with double genomes might have had a

better chance to survive the cretaceous-tertiary extinction event. *Proceedings of the National Academy of Sciences, U.S.A.* 106, 5737-5742.

Fawcett, J.A., Van de Peer, Y., 2010. Angiosperm polyploids and their road to evolutionary success. *Trends in Evolutionary Biology* 2, 16-21.

Feschotte, C., Jiang, N., Wessler, S.R., 2002a. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3, 329-341.

Feschotte, C., Zhang, X., Wessler, S.R., 2002b. Miniature inverted-repeat transposable elements MITEs and their relationship with established DNA transposons. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambovitz (Eds.), *Mobile DNA II*. American Society of Microbiology Press, Washington DC, pp. 1147-1158.

Feschotte, C., Wessler, S.R., 2002. Mariner-like transposases are widespread and diverse in flowering plants. *Proceedings of the National Academy of Sciences, U.S.A.* 99, 280-285.

Feschotte, C., Lakshmi, S., Wessler, S.R., 2003. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway miniature inverted repeat transposable elements MITEs. *Genetics* 165, 747-758.

Feschotte, C., Osterlund, M.T., Peeler, R., Wessler, S.R., 2005. DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. *Nucleic Acids Research* 33, 2153-2165.

Fjellheim, S. Rognli, O.A., Fosnes, K., Brochmann, C., 2006. Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography* 33, 1470-1478.

Flavell, A.J., Pearce, S.R., Kumar, A., 1994. Plant transposable elements and the genome. *Current Opinion in Genetics and Development* 4, 838-844.

Gaut, B.S., 1998. Molelcular clocks and nucleotide substitution rates in higher plants. *Evolutionary Biology* 30, 93-120.

Gaut, B.S., Tredway, L.P., Kubik, C., Gaut, R.L., Meyer, W., 2000. Phylogenetic relationships and genetic diversity among members of the *Festuca-Lolium* complex (Poaceae) based on ITS sequence data. *Plant Systematics and Evolution* 224, 33-53.

Ge, S., Sang, T., Lu, B-R., Hong, D-Y., 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences, U.S.A.* 96, 14400-14405.

Goebel, K., 1895. Ein Beitrag zur Morphologie der Gräser. *Flora* 81, 17-19.

G.P.W.G., The Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88, 373-457.

G.P.W.G., The Grass Phylogeny Working Group, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Gray, M.W., Hanic-Joyce, P.J., Covello, P.S., 1992. Transcription, processing and editing in plant mitochondria. *Annual Review of Plant Physiology and Plant Molecular Biology* 43, 145-175.

Guerra, M., 2008. Chromosome numbers in plant cytotaxonomy: Concepts and implications. *Cytogenetic and Genome Research* 120, 339-350.

Hackel, F., 1887. Gramineae. In: A. Engler, K. Prantl (Eds.), *Die natürlichen Pflanzenfamilien*. Engelmann, Leipzig.pp. pp. 1-97.

Hamby, R.K., Zimmer, E.A., 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). *Plant Systematics and Evolution* 160, 29-37.

Hartley, W., 1950. The global distribution of tribes of the Gramineae in relation to historical and environmental factors. *Australian Journal of Agricultural Research* 1, 355-373.

Hattersley, P.W., Watson, L., 1992. Diversification of photosynthesis. In: G.P. Chapman (Ed.), *Grass Domestication and Evolution*. Cambridge Univ. Press, Cambridge, U.K., pp. 38-116.

Harmon, L.J., Weir, J.T., Brock, Ch.D., Glor, R.E., Challenger, W., 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics Applications notes* 24, 129-131.

Heilbuth, J.C., 2000. Lower species richness in dioecious clades. *American Naturalist* 156, 221-241.

Hiesel, R., Von Haeseler, A., Brennicke, A., 1994. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proceedings of the National Academy of Sciences, U.S.A.* 91, 634-638.

Hodkinson, T.R., Parnell, J.A.N., 2007. Introduction to the systematics of species rich groups. In: T.R. Hodkinson, J.A.N. Parnell (Eds.), *Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa*. CRC Press, Boca Raton, FL, USA, pp. 3-20.

Hodkinson, T.R., Salamin, N., Chase, M.W., Bouchenak-Khelladi, Y., Renvoize, S.A., Savolainen, V., 2007. Large trees, supertrees and diversification of the grass family. *Aliso* 23, 248-258.

Hsiao, C., Chatterton, N.J., Asay, K.H., Hensen, K.B., 1995. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theoretical and Applied Genetics* 90, 389-398.

Inda, L.A., Segara-Moragues, J.G., Muller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., Wessler, S.R., 2003. An active DNA transposon family in rice. *Nature* 421, 163-167.

Jiang, N., Feschotte, C., Zhang, X., Wessler, S.R., 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements MITEs. *Current Opinion in Plant Biology* 7, 115-119.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens, M., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-100.

Kellogg, E.A., 2000. The grasses: A case study in macroevolution. *Annual Review of Ecology, Evolution and Systematics* 31, 217-238.

Kellogg, E.A., 2001. Evolutionary history of the grasses. *Plant Physiology* 125, 1198-1205.

Kim, E.S., Bolsheva, N.L., Samatadze, T.E., Nosov, N.N., Nosova, I.V. Zelenin, A.V., Punina, E.O., Muravenko, O.V., Rodionov, A.V., 2009. The unique genome of two-chromosome grasses *Zingeria* and *Colpodium*, its origin, and evolution. Russian Journal of Genetics, 11, 1329-1337.

Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics* 46, 123-139.

Koonin E.V., Makarova, K.S., Aravind, L., 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology* 55, 709-742.

Koulintchenko, M., Konstantinov, Y., Dietrich, A., 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. *European Molecular Biology Organization* 22, 1245-1254.

Kubo, T., Mikami, T., 2007. Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum* 129, 6-13.

Kubo, T., Newton, K.J., 2008. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8, 5-14.

Leitch, I.J., Bennett, M.D., 2004. Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* 82, 651-663.

Levin, D.A., Wilson, A.C., 1976. Rates of evolution in seed plants: Net increase in diversity of chromosome numbers and species numbers through time. *Proceedings of the National Academy of Sciences, USA* 73, 2086-209.

Linné, C.V., Wildenow, K.L., Link, H.F., 1753. *Species plantarum: exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. Vol. 1. Estocolmo

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., Kuang, H., 2012. Miniature Inverted-Repeat Transposable Elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Molecular Biology and Evolution* 29, 1005-1017.

Magallon, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762-1780.

Magallon, S., Castillo, A., 2009. Angiosperm diversification through time. *American Journal of Botany* 96, 349-365.

Marienfeld, J., Unseld, M., Brandt, P., Brennicke, A., 1999. The mitochondrial genome of the flowering plant *Arabidopsis thaliana* is a composition of native and immigrant information. *Trends in Plant Science* 4, 495-502.

Mason-Gamer, R.J., 2001. Origin of North American *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Systematic Botany* 26, 757-768.

Mason-Gamer, R.J., 2005. The *ß-amylase* genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). *American Journal of Botany* 92, 1045-1058.

Mason-Gamer, R.J., 2007. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. *BMC Evolutionary Biology* 7, 92-102.

Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *American Journal of Botany* 87, 96-107.

Mayrose, I., Barker, M.S., Otto, S.P., 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* 59, 132-144.

Mayrose, I., Zhan, S.H., Rothfelds, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., Otto, S.P., 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333, 1257.

Mayrose, I., 2014. *Chromevol version. 2.0 Manual*. Tel Aviv University. Available at http://www.tau.ac.il/~itaymay/cp/chromEvol/chromEvol_v2.0_manual.pdf

Menzel, G., Dechyeva, D., Keller, H., Lange, C., Himmelbauer, H., Schmidt, T., 2006. Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. *Cromosome research* 14, 831-844.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics Evololution* 68, 106-118.

Mower, J.P., Stefanovic, S, Young, G.J., Palmer, J.D., 2004. Gene transfer from parasitic to host plants. *Nature* 432, 165-166.

Nakamura, Y., Itoh, T., Matsuda, H., Gojobori, T., 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature Genetics* 36, 760-766.

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., Tanisaka, T., 2003. Mobilization of transposons in the rice genome. *Nature* 421, 170-172.

Namaganda, M., Lye, K.A., Friebe, B., Heum, M., 2006. AFLP-based differentiation of tropical African *Festuca* species compared to the European *Festuca* complex. *Theorical and Applied Genetics* 113, 1529-1538.

Namaganda, M., Lye, K.A., 2008. A taxonomic comparison between tropical African and related European broad-leaved species of *Festuca* L. (Poaceae). *South African Journal of Botany* 74, 295-305.

Newton, K.J., Gabay-Laughnan, S., dePaepe, R., 2004. Mitochondrial mutations in plants. In: D.A. Day, A.H. Millar, J. Whelan (Eds.), *Plant Mitochondria: From Genome to Function*. Kluwer, Dordrecht, The Netherlands, pp. 121-142.

Nieto Feliner, G., Rosselló, J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA *ITS* in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44, 911-919.

Olmstead, R.G., Palmer, J.D., 1994. Chloroplast DNA systematics: a review of methods. *American Journal of Botany* 81, 1205-1224.

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34, 401-437.

Palmer, J.D., Herbon, L.A., 1988. Plant mitochondrial-DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28, 87-97.

Palmer, J.D., 1992. Mitochondrial DNA in plant systematics: applications and limitations. In: P.S. Soltis, D.E. Soltis, J.J. Doyle (Eds.), *Molecular Systematics of plants*. Chapman and Hall, New York, USA, pp. 36-39.

Petersen, G., Seberg, O., 2000. Phylogenetic evidence for excision of Stowaway miniature inverted-repeat transposable elements in Triticeae (Poaceae). *Molecular Biology and Evolution* 17, 1589-1596.

Petersen, G., Seberg, O., Davis, J.T., Goldman, D. H., Stevenson, W., Campbell, L.M., Michelangeli, F.A., Specht, C.D., Chase, M.W. Fay, M. F., Pires, J.C., Freudenstein, J.V., Hardy, C.R., Simmons, M.P., 2006. Mitochondrial data in monocot phylogenetics. *Aliso* 22, 52-62.

Peterson, P.M., Romaschenko, K., Johnson, G., 2010. A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees. *Molecular Phylogenetics and Evolution* 55, 580-598.

Pimentel, M., Sahuquillo, E., Torrecilla, Z., Popp, M., Catalan, P., Brochmann, C., 2013. Hybridization and long-distance colonization at different time scales: towards resolution of long-term controversies in the sweet vernal grasses (*Anthoxanthum*). *Annals of Botany*, 1-16.

Pires, J.C., Maureira, I.J., Givnish, T.J., Sytsma, K.J., Seberg, O., Petersen, G., Davis, J.I., Stevenson, D.W., Rudall, P.J., Fay, M.F., Chase, M.W., 2006. Phylogeny, genome size, and chromosome evolution of Asparagales. In: J.T. Columbus, E.A. Friar, J.M. Porter, L.M. Prince, M.G. Simpson (Eds.), *Monocots: comparative biology and evolution (excluding Poales)*. Rancho Santa Ana Botanic Garden, Claremont, California, USA., pp. 287-304.

Prat, H., 1932. L'epiderme des graminées: étude anatomique et systematique. *Annales des Sciences Naturelles; Botanique, série 9* 14, 117-324.

Prat, H., 1936. La systématique des Graminées. *Annales des Sciences Naturelles; Botanique, série 9* 18, 165-258.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trn*T-F and nuclear *ITS* sequences. *American Journal of Botany* 94, 1554-1596.

R Development Core Team, 2011. R: *A Language and Environment for Statistical Computing*. *R* Foundation for Statistical Computing, Vienna, Austria.

Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* 35, 428-449.

Reeder, J.R., 1953. The embryo of *Streptochaeta* and its bearing on the homology of the coleoptile. *American Journal of Botany* 49, 77-80.

Reeder, J.R., 1957. The embryo in grass systematics. *American Journal of Botany* 44, 756-769.

Richardson A.O., Palmer, J.D., 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* 58, 1-9.

Ricklefs, R.E., Renner, S.S., 1994. Species richness within families of flowering plants. *Evolution* 48, 1619-1636.

Ricklefs, R. E., Renner, S.S., 2000. Evolutionary flexibility and flowering plant familial diversity: A comment on Dodd, Silvertown, and Chase. *Evolution* 53, 1061-1065.

Rieseberg, L., Soltis, D., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* 5, 65-84.

Rokas, A., Holland, P.W., 2000. Rare genomic changes as a tool for phylogenetics. *Trends in Ecology and Evolution* 15, 454-459.

Salamin, N., Hodkinson, T.R., Savolainen, V., 2002. Building supertrees: an empirical assessment using the grass family (Poaceae). *Systematic Biology* 51, 136-150.

Saarela, J.M., Peterson, P.M., Soreng, R.J., Chapman, R.E., 2003. A Taxonomic Revision of the Eastern North American and Eastern Asian Disjunct Genus *Brachyelytrutn* (Poaceae): Evidence from Morphology, Phytogeography and AFLPs. *Systematic Botany* 28: 674-692.

Salariato, D.L., Zuloaga, F.O., Giussani, L.M., Morrone, O., 2010. Molecular phylogeny of the subtribe Melinidinae (Poaceae: Panicoidae: Paniceae) and evolutionary trends in the homogenization of inflorescences. *Molecular Phylogenetics and Evolution* 56, 355-369.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Masood-Quraishi, U., Calcagno, T., Cooke, R., Delseny, M., Feuillet, C., 2008. Identification and characterization of

shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20, 11-24.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Systematic Biology* 41, 4-17.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19, 101-109.

Sang, T., Zhang, D., 1999. Reconstructing hybrid speciation using sequences of low-copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Botany* 24, 148-163.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in Plant Phylogenetics. *Biochemistry and Molecular Biology* 37, 121-147.

Sargent, R.D., 2004. Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society of London, Biological Sciences* 271, 603-608.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *mat*K gene-3' *trn*K exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

Schuster, J., 1910. Über die Morphologie der Grasblüte. *Flora* 100, 213-266.

Senchina, D.S., Alvarez, I., Cronn, R.C., Liu, B., Rong, J., Noyes, R.D. Paterson, A.H., Wing, R.A., Wilkins, T.A., Wendel, J.F., 2003. Rate variation among nuclear genes and the age of polyploidy in Gossypium. *Molecular Biology and Evolution* 20, 633-643.

Silvertown, J., McConway, K.J., Dodd, M.E., Chase, M.W., 2000. "Flexibility" as a trait and methodological issues in species diversity variation among angiosperm families. *Evolution* 54, 1066-1068.

Soltis, P.S., Soltis, D.E., 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences, U.S.A.* 97, 7051–7057.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17, 145-170.

Soltis, D.E., Soltis, P.S., 2003. The Role of Phylogenetics in Comparative Genetics. *Plant Physiology* 132, 1790-1800.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, Ch., Sankoff, D., dePamphilis, C.W., Wall, P.K., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96, 336-348.

Soltis, D.E., Burleigh, J.G., 2009. Surviving the K-T mass extinction: New perspectives of polyploidization in angiosperms. *Proceedings of the National Academy of Sciences, U.S.A.* 106, 5455-5456.

Soltis, D.E., Segovia-Salcedo, M.C., Jordon-Thaden, I., Majure,, L., Miles, N.M., Mavrodiev, E.V., Mei, W., Cortez, M.B., Soltis, P.S., Gitzendanner, M.A., 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose *et al.* (2011). *New Phytologist* 202, 1105-1117.

Soreng, R.J., Davis, J.I., 1998. Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Botanical Review* 64, 1-85.

Soreng, R.J., Davis, J.I., 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: Misclassification versus reticulation. In: S.W.L. Jacobs, J. Everett (Eds.), *Grasses: Systematics and Evolution*. CSIRO, Melbourne, pp. 61-74.

Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O. (Eds.), 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Stancík, D., Peterson, P.M., 2007. A revision of *Festuca* (Poaceae: Loliinae) in South American páramos. *Contributions from the US National Herbarium* 56, 1-184.

Stebbins, G.L., 1971. *Chromosomal evolution in higher plants*. London: Addison-Wesley.

Stebbins, G.L., 1981. Why are there so many species of flowering plants? *Bioscience* 31, 573-577.

Stebbins, G.L., 1984. Polyploidy and the distribution of the arctic-alpine flora: New evidence and a new approach. *Botanica Helvetica* 94, 1-13.

Strömberg, C.A.E., 2004. The Evolution of grass-dominated ecosystems during the late Cenozoic. *Palaeogeography, Palaeoclimatology, Palaeoecology* 207, 199-201.

Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. *Annual review of Earth and Planetary Sciences* 39, 517-544.

Sungkaew, S., Stapleton, C.M.A., Salamin, N., Hodkinson, T.R., 2009. Non-monophyly of woody bamboos (Bambuseae, Poaceae): A multi-gene region phylogenetic analysis of Bambusoideae s.s. *Journal of Plant Research* 122, 95-108.

Tang, L., Zou, X.-H., Achoundong, G., Potgieter, C., Second, G., Zhang, D.-Y., Ge, S., 2010. Phylogeny and biogeography of the rice tribe (Oryzeae): Evidence from combined

analysis of 20 chloroplast fragments. *Molecular Phylogenetics and Evolution* 54, 266-277.

Tateoka, T., Inoue, S., Kawano, S., 1959. Systematic significance of bicellular microhairs of leaf epidermis. *Botanical Gazette* 121, 80-91.

Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* 51, 689-702.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of *ITS* and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Tzvelev, N.N., 1989. The system of grasses (Poaceae) and their evolution. *Botanical Review* 55, 141-204.

Van de Peer, Y., 2011. A mystery unveiled. *Genome Biology* 12, 113.

Van de Peer, Y., Maere, S., Meyer, A., 2009. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* 10, 725-732.

Verdú, M., 2002. Age at maturity and diversification in woody angiosperms. *Evolution* 56, 1352-1361.

Watson, L., Dallwitz, M.J., 1992 onwards. *The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references*. Version: 12th August 2014. http://delta-intkey.com

Weilong, H., Palmer, J., 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, tran-compartmentally chimeric mitocondrial genes. *Proceedings of the National Academy of Sciences, U.S.A.* 29, 16728-16733.

Wolfe, K.H., Li, W., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, U.S.A.* 84, 9054-9058.

Wood, T.E., Takebayashi, N., Barker, M.S., Mayrose, I., Greenspoon, P.B., Rieseberg, L.H. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the*

National Academy of Sciences, U.S.A. 106, 13875-13879.

Zhang, X., Jiang, N., Feschotte, C., Wessler, S.R., 2004. PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics* 166, 971-986.

Zuckerkandl, E., Paulin, L., 1965. Evolutionary divergence and convergence in proteins. In: V. Bryson, H.J. Vogel (Eds.), *Evolving Genes and Proteins*. Academic Press, New York, pp. 97-166.

OBJECTIVES OF THIS DISSERTATION

The main goal of the investigation in this thesis is to explore the evolutionary processes in the grass family (Poaceae), with an emphasis on the Pooideae subfamily. To this end, analyses of nuclear, chloroplast, and mitochondrial genes are used, allowing for the exploration of the processes of duplication, recombination, pseudogenization, horizontal gene transfer, gene conversion, and polyploidization. These studies, along with the construction of a robust molecular phylogeny and the estimation of divergence times and historical biogeographical patterns, come together to explain, from the micro- to the macro-evoltionary level, the appearance of new lineages of temperate grasses and their historical patterns of dispersion and distribution.

The specific aims of this thesis are:

1.- To explore and evaluate the extent of purifying selection, pseudogenization, paralogy, homeology, recombination and misplacements of the LCNG *ß-amylase* and mitochondrial *rps3* sequences within a broad representation of the main Pooideae lineages, with special emphasis on the Poeae-Aveneae group.

2.- To study the patterns of distribution and the evolutionary dynamics of the MITEs found in the fourth intron of the *ß-amylase* gene across the major lineages of Poaceae, and to identify the occurrence of similar transposable elements in other loci of the grasses' genomes.

3.- To further the understanding of phylogenetic relationships and broaden the scope of taxa studied in the major subfamilies and tribes of Poaceae, with special attention to the large subfamily Pooideae, using nuclear and plastid genes.

4.- To estimate the diversification rate dynamics, the patterns of chromosome number evolution, and the times of divergence operating in the Pooideae lineages and their potential relationships with historical climate changes.

5.- To investigate the main migration patterns in the broad- and the fine-leaved Loliinae, expanding the taxonomic and geographical coverage of this group in the Southern Hemisphere.

CHAPTER I

EVOLUTION OF THE *BETA-AMYLASE* GENE IN THE TEMPERATE GRASSES: NON-PURIFYING SELECTION, RECOMBINATION, SEMIPARALOGY, HOMEOLOGY AND PHYLOGENETIC SIGNAL

1. Abstract

- Low-copy nuclear genes (LCNGs) have complex genetic architecture and evolutionary dynamics. However, unlike multicopy nuclear genes, LCNGs are rarely subject to gene conversion or concerted evolution and present higher mutation rates than organellar or nuclear ribosomal DNA markers, so they have great potential for improving the robustness of phylogenetic reconstructions at all taxonomic levels.
- In this study, we evaluate the phylogenetic value of the LCNG *ß-amylase* and the potential existence of pseudogenization, paralogy, homeology, recombination and misplacement events within a broad representation of the main Pooideae lineages using Premature Terminal Codons (PTCs) detection, diploid/polyploid-based topologies, Branch-Specific Branch-Site REL (Random Effects Likelihood) modelling and Recombination Detection (RDP4) approaches. A multigenic (ITS, *mat*K, *ndh*F, *trn*TL and *trn*LF) organismal tree of the study group provided a framework for assessing the evolutionary and phylogenetic patterns observed in the *ß-amylase* phylogeny.
- Eight accessions showed complete absence of purifying selection, suggesting the existence of putative pseudogenic copies or other relaxed selection pressures; resolution of *Vulpia alopecuros* 2x clones indicated its potential (semi) paralogy, and homeologous copies of *Festuca simensis, F. fenas* and *F. arundinacea* tracked the Mediterranean origin of these allopolyploid *Schedonorus* species. Two recombination events were found within early-diverged Pooideae lineages and five were associated with representatives of the PACCMAD clade.
- The atypical phylogenetic patterns described for 37 grass species (26% of the sampled species) highlight the frequent occurrence of alternative distorting evolutionary events, which recomends caution before using this LCNG as phylogenetic marker. However, once the pitfalls are identified and removed, the phylogenetic reconstruction of the grasses based on the *ß-amylase* exon+intron positions is optimal at all taxonomic levels.

2. Introduction

Phylogenies based on molecular data help us to understand a variety of genomic evolutionary phenomena such as pseudogenization, gene duplication (paralogy), lineage sorting, recombination, transposable elements and horizontal gene transfer (MASON-GAMER AND KELLOGG, 1997; SORENG AND DAVIS, 2000; CATALAN *ET AL.*, 2004; MINAYA *ET AL.*, 2013). These evolutionary events can suggest a phylogenetic history that is more reticulate than bifurcating (cf. KELLOGG, 2006), disabusing of the premise that analysis of DNA gene trees would yield the true phylogenies of the organisms (c.f. LINDER AND RIESEBERG, 2004; FELINER AND ROSSELLO, 2007).

The inclusion of a mixture of orthologous and paralogous sequences in phylogenetic analyses can produce robust, yet erroneous, hypotheses of relationships (SANDERSON AND DOYLE, 1992; WENDEL AND DOYLE, 1998). Although a number of criteria have been suggested as evidence of orthology (e.g. DOYLE, 1991; DOYLE AND DOYLE, 1999; CRONN *ET AL.*, 2002), the simplest approach to identifying orthologs is through sequence similarity and phylogenetic analyses. Orthologs are expected to be more similar and phylogenetically more closely related to each other than to any paralogs, and consequently, orthologs must form monophyletic clades and conform to generally accepted species trees (BUCKLER *ET AL.*, 1997; SMALL *ET AL.*, 2004). Like divergent paralogues, recombinants can also compromise phylogeny reconstruction (SANDERSON AND DOYLE, 1992; BUCKLER *ET AL.*, 1997). Recombination, defined here as the process by which alleles with dissimilar nucleotide sequences interact and exchange sequence information (POSADA AND CRANDALL, 2001; CROMIE AND SMITH, 2007), violates the single-tree assumption underlying the evolution of the sequences under study. Recombination generates mosaic genes, where different regions have different phylogenetic histories (DOYLE, 1996). For all these reasons, the accurate detection and elimination of paralogous and recombinant sequences before phylogenetic reconstruction is highly relevant (c.f. BUCKLER *ET AL.*, 1997; POSADA AND CRANDALL, 2001).

The extent to which selection contributes to nucleotide substitutions has been one of the outstanding problems in understanding molecular sequence evolution since the proposal of the neutral theory (KIMURA, 1968). Here we investigate selection on a site-by-site approach, using the generalized branch-site methods of KOSAKOVSKY-POND *ET AL.* (2011) and fitting a distribution of substitution rates across sites (random effect model = REL; NIELSEN AND YANG, 1998), thus inferring the rate at which individual sites evolve given this distribution (KOSAKOVSKY-POND AND FROST, 2005). Theory predicts that the number of substitutions

leading to adaptive amino acid changes (non-synonymous mutations, d_N) should be significantly larger than the number of substitutions at synonymous sites (d_S) (MIYATA AND YASUNAGA, 1980; LI, 1997). Conversely, if selection consistently weeds out amino acid changes, then d_N should be significantly smaller than d_S. Thus, whereas an estimated d_N/d_S value near 1 suggests neutrality, a d_N/d_S greater than or less than 1 indicates positive or negative directional selection, respectively (reviewed by YANG AND BIELAWSKI, 2000; PRESTON AND KELLOGG, 2006; BRUNNER *ET AL.*, 2009).

The grass family (Poaceae) is one of the largest and most diverse families of the angiosperms. It contains approximately 10,000 species and 600-700 genera (CLAYTON AND RENVOIZE, 1986; WATSON AND DALLWITZ, 1992). Previous evolutionary studies of the Poaceae (e.g., GPWG, 2001, 2012; DAVIS AND SORENG, 2007; BOUCHENAK-KHELLADI *ET AL.*, 2008; SCHNEIDER *ET AL.*, 2009) have provided valuable insights into the evolutionary structure of the grass genomes and their phylogeny. However, major phylogenetic relationships remain to be solved, especially at the tribal and generic levels. Low copy nuclear genes (LCNG), unlike multicopy nuclear genes (e.g. ribosomal DNA genes: ITS, ETS; ALVAREZ AND WENDEL, 2003; FELINER AND ROSSELLO, 2007), are rarely subject to gene conversion or concerted evolution. Furthermore, they show higher sequence variation than organellar or rDNA genes, and have a great potential for improving the robustness of phylogenetic reconstructions at all taxonomic levels (cf. SANDERSON AND DOYLE, 1992; SANG AND ZHANG, 1999; SANG, 2002; SMALL *ET AL.*, 2004; HUGHES *ET AL.*, 2006). The real utility of LCNG for phylogenetic reconstruction is, however, less encouraging than their potential advantages would predict. This is because of their complex genetic architecture and evolutionary dynamics, and the fact that LCNGs experience gene duplication and recombination (WENDEL AND DOYLE, 1998; SMALL *ET AL.*, 2004; FELINER AND ROSSELLO, 2007).

The *ß-amylase* gene (1-4-alpha-glucan maltohydrolase) has been scarcely explored as a phylogenetic marker in angiosperms (RAJAPAKSE *ET AL.*, 2004); however a few studies have used it for phylogenetic reconstruction of some grass lineages (e.g. WANG *ET AL.*, 1997; ZIEGLER, 1999; MASON-GAMER, 2005). This background drove our interest in characterizing the molecular evolution and phylogeny of the LCNG *ß-amylase* in the temperate grasses. The specific aims of this study were to: (i) explore the extent of purifying selection across the studied *ß-amylase* grass sequences; (ii) evaluate pseudogenization, paralogy, homeology, recombination and misplacements of the LCNG *ß-amylase* within a broad representation of the main Pooideae lineages, with special emphasis on the Poeae-Aveneae; (iii) examine

phylogenetic bias among different sets of coding and non-coding positions [exon + intron positions, exons, non-synonymous, synonymous, and synonymous positions + introns] by comparing each phylogenetic inference with a multigenic organismal tree of the Poaceae; and (iv) reconstruct the grass phylogenetic relationships based on the best-resolved and most congruent data set.

3. Material and methods

3.1 Taxon sampling and DNA analysis

The taxon sampling was designed to represent the main Pooideae lineages and close outgroups, with special emphasis on the Poeae-Aveneae (see the appendix of chapter II; Table S1, page 131). Sampling included 142 species classified in 88 genera and 7 subfamilies (GPWG, 2001, 2012; SORENG *ET AL.*, 2007, and http://www.tropicos.org/): Centothecoideae: 1 species, Panicoideae: 6, Danthonioideae: 7, Chloridoideae: 10, Ehrhartoideae: 2, Bambusoideae: 1, and Pooideae: 115 (Lygeeae/Nardeae: 2, Stipeae: 3, Diarrheneae: 1, Brachypodieae: 3, Bromeae: 1, Triticeae: 25, Poeae – Aveneae: 80). Four of the six subfamilies of the outgroup PACCMAD clade are sampled in this study (all but Arundinoideae and Micrairoideae). Amplified sequences from 16 samples were cloned: *Brachypodium distachyon* (9 clones), *Festuca abyssinica* (2), *F. altaica* (7), *F. arundinacea* (7), *F. capillifolia* (9), *F. elegans* (7), *F. fenas* (7), *F. hystrix* (5), *F. panciciana* (2), *F. paniculata* (10), *F. rubra* (5), *F. scabra* (5), *F. scariosa* (8), *F. simensis* (7), *Micropyropsis tuberosa* (2), *Vulpia alopecuros* (5). These taxa previously showed paralogous copies in other LCNG in Loliinae and close allies (GBSS; DÍAZ-PÉREZ *ET AL.*, 2014) and *Brachypodium* (CAL, DGAT, GI; CATALAN *ET AL.*, 2012). Amplified products from both diploid and polyploid species were cloned to check for potential *ß-amylase* paralogous and heterologous copies. The origins of the plants studied, the locations of voucher specimens, and the GenBank sequence accession numbers are listed in the appendix of chapter II; Table S1 (page 131). *Joinvillea ascendens* (Joinvilleaceae) was selected to root the tree because it is closely related to the grasses and clearly positioned in Poales (CAMPBELL AND KELLOGG, 1987; LINDER AND RUDALL, 1993).

DOYLE AND DOYLE'S (1987) CTAB method was used to isolate DNA from silica-gel-dried leaves and from fresh materials for most of the studied samples. For herbarium samples, DNA extracts were obtained using the DNAeasy® Plant Mini Kit (QIAGEN Ltd., West Crawley, UK) procedure. A 2370-nucleotide region (see Fig. S1 in the appendix of chapter II; page 130), extending from exons 2 to 5 of the low-copy nuclear *ß-amylase* gene was amplified

and sequenced using the primers 2a-for, 2b-for, 2c-for, 2d-for, 2g-for, 3a-for, 3a-bac, 4a-bac, 4b-for, 5a-bac and 5b-bac, following the procedures indicated in MASON-GAMER (2005). Procedures for DNA amplification, cloning and sequencing are indicated in the appendix of chapter II (page 126).

The *ß-amylase* sequences were aligned with Protein Multiple Sequence Alignment Software version 3.7 (MUSCLE) (EDGAR, 2004). The alignments were manually adjusted using MacClade 4.08 OS X (MADDISON AND MADDISON, 2008). Amino acid translations were used to guide the nucleotide alignments. While alignment was generally straightforward in the coding regions, there were some intron regions where patterns of length variation were complex enough to make homology assessment difficult. We also found a Miniature Inverted repeat Transposable Element (MITE) region in intron 4, which included highly variable and non-homologous transposable elements and their footprints (MINAYA *ET AL.*, 2013). Seven regions with ambiguous alignments and/or MITEs were excluded from subsequent analyses [see in the appendix of chapter II; Fig. S1; page 130): intron 2 (positions 160-319, 329-370); intron 3 (560-740, 784-821, 847-880, 976-1269); intron 4 (1602-2083)].

3.2 Organismal tree reconstruction

A multigenic organismal tree of the Poaceae was used to assess the phylogenetic distribution of paralogous copies, recombinant sequences, and the existence of phylogenetic bias between the *ß-amylase* reconstructions based on five different sets of positions (see below). The Poaceae organismal tree (cf. MINAYA *ET AL.*, 2013) was constructed using the multicopy ribosomal *ITS*1-5.8S-*ITS*2 region and the plastid *mat*K, *ndh*F, *trn*TL and *trn*LF regions from a collection of 142 selected grass taxa plus the close relative *Joinvillea* (Joinvillaceae) included as an outgroup. The ITS, *trn*TL and *trn*LF regions were amplified and sequenced using the primers and procedures from CATALAN *ET AL.* (2004) and QUINTANAR *ET AL.* (2007), the *ndh*F gene followed CATALAN *ET AL.* (1997), and the *mat*K DÖRING *ET AL.* (2007). The potential combinability of the data sets was assessed using the Partition Homogeneity (PH) test (Incongruence Length Difference of FARRIS *ET AL.*, 1994), implemented in PAUP* 4.0 beta 10 (SWOFFORD, 2002). The PH test was conducted through heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON to estimate if the separate and combined nuclear and plastid data sets were significantly different from random partitions of the same sizes.

Bayesian and Maximum parsimony (MP) based searches were performed for both the independent and the combined data sets using, respectively, MrBayes 3.2.2 (ALTEKAR *ET AL.*, 2004; RONQUIST *ET AL.*, 2011) and PAUP*. All gaps were treated as missing data. Procedures for the Bayesian and parsimony-based searches are indicated in the appendix of chapter II (page 126).

3.3 Detection of potential *ß-amylase* paralogues, misplacements, recombinants, and non-purifying sequences or pseudogenes

Paralogous copies of accessions were inferred using the tree-based method of ALTENHOFF AND DESSIMOZ (2012) from confirmed diploid sequences showing clones placed on two divergent positions of a phylogenetic tree. This method, which also allowed us to identify misplaced accessions, is based on a species tree (e.g., multigenic organismal tree of the Poaceae) and the *ß-amylase* gene tree reconciliation, which is the process of annotating all splits of a given gene tree as duplications or speciations, according to the placement and support of the cloned sequences from the studied accessions. Once the branchings of the gene tree have been classified as speciation or duplication events, it is simple to infer whether the given accession shows orthologous or paralogous sequences, based on where the branchings appear. Since the intron positions might cover some paralogous and/or misplaced accessions, the *ß-amylase* gene tree was based on a 588-bp matrix that included only the exons positions of 142 grass species (totalling 223 grass sequences, including all the direct sequences and the clones; Table 1, *ß-amylase* data set 1). We further used the species-overlap rule (GABALDON, 2008), a complementary method to corroborate the potential existence of paralogous copies in the data set, based on the analysis of the *ß-amylase* trees. According to GABALDON (2008), it is assumed that a node represents a duplication event if it is ancestral to a tree bifurcation (e. g. sister clades) that contains sets of sequences from the same species or accessions that overlap to some degree. Conversely, if the two partitions contain sets of species that are mutually exclusive, the node is considered to represent a speciation event.

Table 1. *β-amylase* data partitions used in this research. Each data set is enumerated and described according to their appearance in the main text

***β-amylase* data partition**	**nº grass species/sequences**	**Number of nucleotide positions**	**Characteristics**	**Analysis performed**
1.- Exon positions	142 / 223	588	Including all species and clones sampled	• Compared to the Organismal tree to find paralogous copies and misplaced sequences • Check the existence of PTCs among the coding positions • Bayesian phylogeny (Fig. S1)
2.- Exon positions	142 / 146	588	Including: • All species sampled. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01	Recombination (RDP4) (Table 3)
3.- Exon positions	65 / 69	588	Selection of representative sequences classified within the BEP clade, including: • Misplaced accessions and accessions under potential paralogy, homeology, or recombination. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01.	Branch-Site REL (DataMonkey) (Table 4)
4.- Exon + intron positions	142 / 146	850	Including: • All species sampled. • Cloned sequences collapsed into a consensus type when their p-distance was lower than 0.01.	• Bayesian phylogeny (Fig. 2) • NeighborNet partition network tree (Fig. 3)
5.- Exon + intron positions	105 / 105	850	• Excluding 37 species (41 accessions including the collapsed cloned sequences) under misplacement, potential pseudogenization, semi-paralogy, homeology, or recombination. • Cloned sequences were collapsed into a consensus type when their p-distance was lower than 0.01.	• SH test (Table 5) • Bayesian phylogeny (Fig. 5) • PH and SH tests • Comparison of the Bayesian topology (Fig. 4)
5a.- Exon positions	105 / 105	588		• PH and SH tests • Comparison of the Bayesian topology
5b.- Non-synonymous positions	105 / 105	392		• PH and SH tests • Comparison of the Bayesian topology
5c.- Synonymous positions	105 / 105	196		• PH and SH tests • Comparison of the Bayesian topology
5d.- Synonymous + intron positions	105 / 105	456		• PH and SH tests • Comparison of the Bayesian topology

The Recombination Detection Program (RDP4) version 4.16 (MARTIN *ET AL.*, 2010) was used to analyze potential recombination events within the *ß-amylase* sequences. The RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, LARD, and 3SEQ methods implemented in RDP4 were used to identify recombinant and parental sequences and to calculate probability scores for potential recombination events. We used the default settings in all cases. Only potential recombinant events detected by at least two methods were considered significant. The Bonferroni multiple comparison correction test was performed to diminish the expected number of false positive results. Masking of similar sequences was allowed to increase the power of the recombination detection methods. For these analyses we used a matrix with only exon positions. This matrix presented all Poaceae sequences sampled (142 grass species); however, to discard spurious variation originating from PCR (Polymerase Chain Reaction) artifacts, the cloned sequences from 16 species were collapsed into a single consensus sequence (or type) from closely related sequences following DÍAZ-PÉREZ *ET AL.* (2014). All intraspecific sequences with a p-distance lower than 0.01 base differences per site were collapsed into a single type using the program BioEdit v7.2.5 (HALL, 1999) (Table 1, *ß-amylase* data set 2; 146 sequences, 588 positions, exons 2 to 5).

Two complementary criteria were used to detect pseudogenic copies, and one of them also detected potential types of selection pressure on the *ß-amylase* sequences. First, we checked for the existence of Premature Terminal Codons (PTCs) among the coding positions of the 223 sampled *ß-amylase* sequences (Table 1, *ß-amylase* data set 1). According to AKHUNOV *ET AL.* (2013) some potential pseudogenic copies could accumulate mutations and eventually produce PTCs at the early stages of evolution.

Secondly, we considered that most of the coding positions in a functional protein are constrained by purifying selection, rather than positive/adaptive selection or neutral variation (KIMURA, 1968). Selection was investigated through the analysis of distribution of substitution rates across sites and branches (REL model; NIELSEN AND YANG, 1998; KOSAKOVSKY-POND AND FROST, 2005), where the distribution is represented by discrete classes defined by specific selection pressure ranges ($\omega=d_N/d_S$; *dN* and *dS* representing nonsynonymous/ synonymous substitutions per nonsynonymous/synonymous sites, respectively). REL models assume that substitutions along a branch of a phylogenetic tree is characterized by a continuous-time stationary Markov process, defined by its instantaneous rate matrix (Q). Similar to nucleotide substitution models, the likelihood of a change along a branch is given by the transition matrix. Considering that selective pressures may vary over both sites and time, branch-site REL

methods are able to estimate the proportion of sites with different ω values for all lineages of the tree (KOSAKOVSKI-POND *ET AL.*, 2011). In this case, each branch-site combination has instantaneous rate and transition matrices. We restricted our analysis to terminal lineages to evaluate recent pressures associated with sequences from a subset of 65 BEP species, including all putative sequences under likely paralogy, homeology, recombination and misplacement (Table1, *ß-amylase* data set 3; 69 sequences, 588 positions, exons 2 to 5). This was done to detect sequences with high proportions of sites under neutral or positive selection, indicating potential relaxed selection pressure (represented by $\omega = dN/dS \geq 1$) and evidence of pseudogenization. We performed the Branch-Specific Branch-Site REL (BS-REL; REL = Random Effect Likelihood) model, as suggested in KOSAKOVSKY-POND *ET AL.* (2008), to estimate *ß-amylase* lineage-specific proportion of sites (Pr1, Pr2 and Pr3) showing rates circumscribed to three discrete classes ($\omega1 \leq \omega2 \leq 1 \leq \omega3$), respectively. We considered that classes ω1 and ω2 can be less than or equal to 1, and are associated with the neutral model, specifying, respectively, stronger and weaker purifying selection, whereas ω3 (unconstrained class) is related to adaptive selection if it is greater than one. According to the estimated ω1, ω2, ω3 values, we assigned each of these classes and its proportion to one out of five selection scenarios: (I) $\omega = 0 - 0.6$ (strong negative selection), (II) $\omega = 0.61 - 0.9$ (weak negative selection), (III) $\omega = 0.91 - 1.5$ (neutral selection), (IV) $\omega = 1.51 - 5$ (weak positive selection), and (V) $\omega > 5$ (strong positive selection). BS-REL was performed using the maximum likelihood-based tools implemented in the WEB interface DataMonkey (POND AND FROST, 2005).

3.4 Reconstruction of *ß-amylase* phylogeny

A Bayesian tree and a NeighborNet partition network (BRYANT AND MOULTON, 2004; HUSON *ET AL.*, 2010) were computed based on a *ß-amylase* 850-bp data matrix that included the exon + unambiguous intron positions and 146 sequences (from 142 species) after collapsing the clone sequences into a consensus type (Table 1, *ß-amylase* data set 4). The Bayesian phylogeny and the NeighborNet partition network tree provide a graphical representation of phylogenetic relationships considering the putative occurrence of incompatible and ambiguous signals (HUSON AND BRYANT, 2006) in the *ß-amylase* data. These signals could have originated from incomplete lineage sorting, horizontal gene transfer, recombination or hybridization events. Within the NeighborNet partition network analysis, a pairwise *p-distance* matrix representing the proportion of base differences between sequences was computed first. Then, the

NeighborNet algorithm was employed to generate circular partitions which were subsequently processed by the *Equal Angle* algorithm (BRYANT AND MOULTON, 2004; HUSON *ET AL.*, 2010). Partition statistical support was generated through 1,000 bootstrap pseudoreplicates. All analyses were performed using *Splitstree* version 4.12.3 (HUSON AND BRYANT, 2006).

Once we had detected and excluded pseudogenes, paralogs, recombinants, or misplaced sequences (see results for more details), we reconstructed the *ß-amylase* phylogeny based on a 850-bp matrix that included the exons plus the unambiguous intron positions from 105 species, including the different types of collapsed clones (Table 1, *ß-amylase* data set 5). Because one of the aims of this study is to determine the existence of phylogenetic bias among *ß-amylase* reconstructions based on different sets of coding and non-coding positions, the data matrix (*ß-amylase* data set 5 in Table 1) was subdivided into four additional data matrices: (5a) only coding positions (588 bp); (5b) only non-synonymous positions (392 bp); (5c) only synonymous positions (196 bp); and (5d) synonymous positions plus unambiguous intron positions (456 bp) (Table 1, *ß-amylase* data sets 5a-5d).

The Partition Homogeneity (PH) test (Incongruence Length Difference [ILD] test of FARRIS *ET AL.*, 1994), and the more conservative Shimodaira–Hasegawa (SH) nonparametric bootstrap test (SHIMODAIRA AND HASEGAWA, 1999) were conducted on the fifth *ß-amylase* data set (Table 1, *ß-amylase* data set 5) and its four subdivisions (Table 1: *ß-amylase* data sets 5a, 5b, 5c, 5d). Both tests were implemented in PAUP* v4.10b (SWOFFORD, 2002). The PH test analyzed the potential combinability of subsets 5a to 5d. The PH test was evaluated through heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON to estimate if the separate and combined data sets were significantly different from random partitions of the same size. The SH test was done using PAUP*v4.10b with RELL (resampling estimated log-likelihood) optimization and 1000 bootstrap replicates. The SH test was used to determine whether the topology of the Bayesian tree derived from the data set of exon plus unambiguous intron positions (Table 1, *ß-amylase* data set 5), which corresponds to the most likely tree (see results), was significantly different from the Bayesian trees derived from its four subdivisions (5a to 5d). Since comparing more than two trees leads to a multiple comparison problem that cannot be solved by Bonferroni corrections (cf. SHIMODAIRA AND HASEGAWA, 1999), we decided to run the SH test comparing only two trees each time.

The independent Bayesian searches conducted for each of seven separate *ß-amylase* data sets (Table 1, *ß-amylase* data sets 1, 4, 5, 5a, 5b, 5c, 5d) were run with MrBayes 3.2.2 (ALTEKAR *ET AL.*, 2004; RONQUIST *ET AL.*, 2011), using *Joinvillea ascendens* to root the trees.

Tests of goodness of fit for alternative nucleotide substitution models were performed through the Hierarchical Likelihood Ratio Test (hLRTs), AIC and BIC tests in jModelTest 2 (GUINDON AND GASCUEL, 2003; DARRIBA *ET AL.*, 2012). The estimated GTR+I+gamma model was fixed for all Bayesian inferences performed in this study.

4. Results

4.1 The organismal tree of the grasses

The multigenic organismal tree of the Poaceae (Fig. 1) provided a framework for evaluating the potential existence of paralogous copies and misplacements among the *ß-amylase* grass sequences analyzed and for assessing potential phylogenetic bias among the four *ß-amylase* subdivisions (Table 1: *ß-amylase* data sets 1, 4, 5, 5a, 5b, 5c, 5d). The multigenic data set consisted of nuclear ITS (134 sequences/ 609 nucleotides) and plastid *mat*K (120/ 1259), *ndh*F (124/ 710), *trn*TL (116/ 789) and *trn*LF (127/ 975) sequences. GenBank accession numbers for all sequences are indicated in the appendix of chapter II; Table S1 (page 131).

The independent phylogenetic searches conducted with the five separate ITS, *mat*K, *ndh*F, *trn*TL and *trn*LF data sets yielded highly congruent topologies. The PH test detected incongruence among the plastid (cpDNA) data sets ($P< 0.01$), except for *mat*K – *trn*TL ($P= 0.4$) and *ndh*F – *trn*TL ($P= 0.3$). Partition incongruence was also found between the plastid and nuclear data ($P= 0.01$). Despite this, the separate topologies were similar to the tree based on the combined nuclear and plastid datasets, with incongruence observed mostly among some terminal tips. Furthermore, those topological conflicts were not well supported. Since the incongruence did not affect the main results and conclusions of this research, we chose to conduct phylogenetic analyses on the combined dataset of the five sequenced genes. The BI tree (Fig. 1) and MP tree (Fig. 1; bootstrap support values) from the combined data set were highly congruent to each other and with those obtained for the grasses by previous authors. The MP statistics of the phylogenetic reconstructions of this chapter are shown in Table 2.

Figure 1. Poaceae organismal tree. Bayesian Inference estimation based on nuclear (ITS) and plastid (*trn*TL, *trn*LF, *ndh*F, *mat*K) gene sequences (Table S1, page 131). Bayesian posterior probabilities of ≥ 90% and MP bootstrap support values of ≥ 75% are shown above and below the nodes, respectively. Species assignment to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure following SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2003, 2004); CATALAN *ET AL.* (2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). Identical criteria were followed in the remaining phylogenetic inferences presented in this chapter.

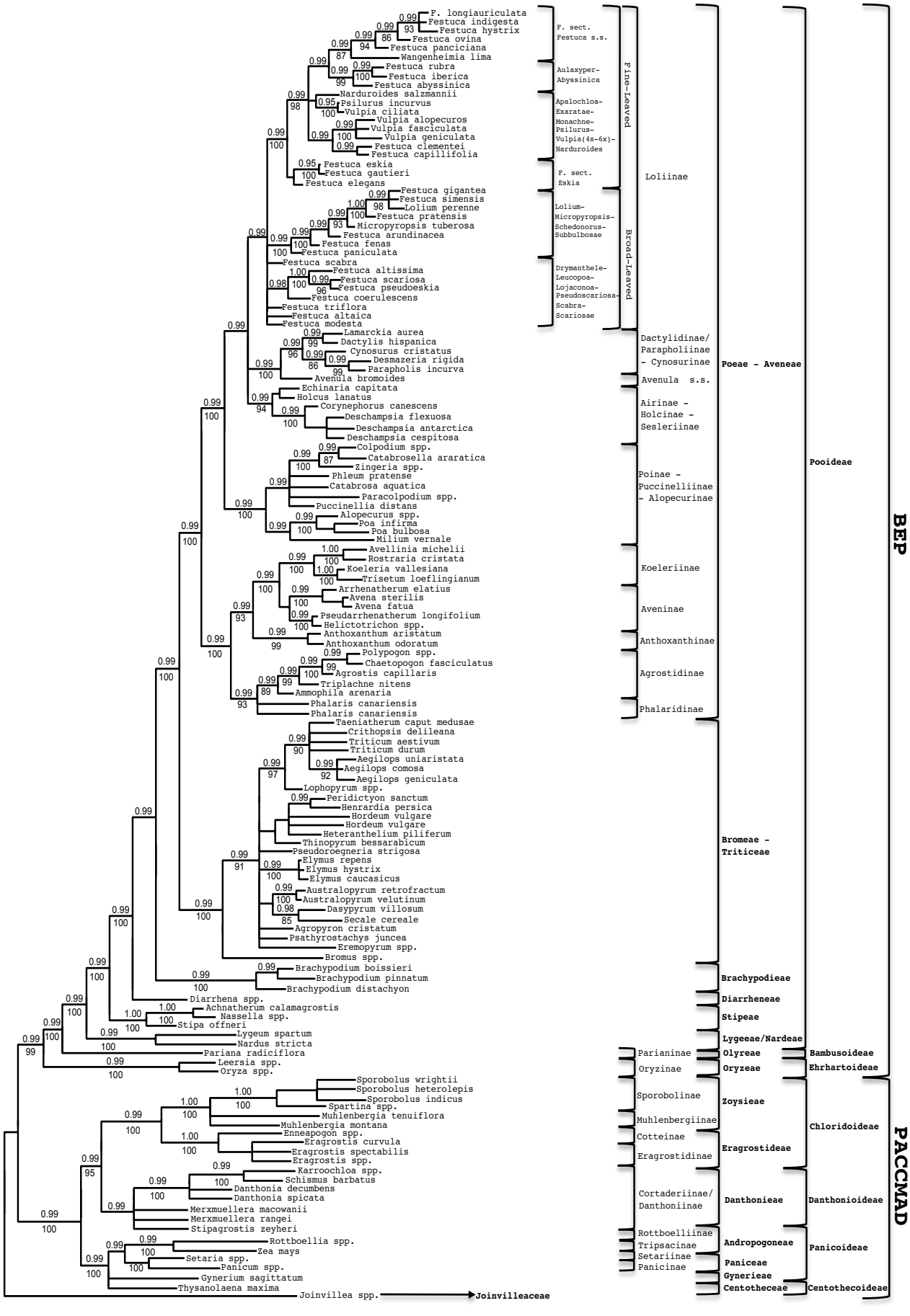

F. longiauriculata
Festuca indigesta
Festuca hystrix
Festuca ovina
Festuca panciciana
Wangenheimia lima
Festuca rubra
Festuca iberica
Festuca abyssinica
Narduroides salzmannii
Psilurus incurvus
Vulpia ciliata
Vulpia alopecuros
Vulpia fasciculata
Vulpia geniculata
Festuca clementei
Festuca capillifolia
Festuca eskia
Festuca gautieri
Festuca elegans
Festuca gigantea
Festuca simensis
Lolium perenne
Festuca pratensis
Micropyropsis tuberosa
Festuca arundinacea
Festuca fenas
Festuca paniculata
Festuca scabra
Festuca altissima
Festuca scariosa
Festuca pseudoeskia
Festuca coerulescens
Festuca triflora
Festuca altaica
Festuca modesta
Lamarckia aurea
Dactylis hispanica
Cynosurus cristatus
Desmazeria rigida
Parapholis incurva
Avenula bromoides
Echinaria capitata
Holcus lanatus
Corynephorus canescens
Deschampsia flexuosa
Deschampsia antarctica
Deschampsia cespitosa
Colpodium spp.
Catabrosella araratica
Zingeria spp.
Phleum pratense
Catabrosa aquatica
Paracolpodium spp.
Puccinellia distans
Alopecurus spp.
Poa infirma
Poa bulbosa
Milium vernale
Avellinia michelii
Rostraria cristata
Koeleria vallesiana
Trisetum loeflingianum
Arrhenatherum elatius
Avena sterilis
Avena fatua
Pseudarrhenatherum longifolium
Helictotrichon spp.
Anthoxanthum aristatum
Anthoxanthum odoratum
Polypogon spp.
Chaetopogon fasciculatus
Agrostis capillaris
Triplachne nitens
Ammophila arenaria
Phalaris canariensis
Phalaris canariensis
Taeniatherum caput medusae
Crithopsis delileana
Triticum aestivum
Triticum durum
Aegilops uniaristata
Aegilops comosa
Aegilops geniculata
Lophopyrum spp.
Peridictyon sanctum
Henrardia persica
Hordeum vulgare
Hordeum vulgare
Heteranthelium piliferum
Thinopyrum bessarabicum
Pseudoroegneria strigosa
Elymus repens
Elymus hystrix
Elymus caucasicus
Australopyrum retrofractum
Australopyrum velutinum
Dasypyrum villosum
Secale cereale
Agropyron cristatum
Psathyrostachys juncea
Eremopyrum spp.
Bromus spp.
Brachypodium boissieri
Brachypodium pinnatum
Brachypodium distachyon
Diarrhena spp.
Achnatherum calamagrostis
Nassella spp.
Stipa offneri
Lygeum spartum
Nardus stricta
Pariana radiciflora
Leersia spp.
Oryza spp.
Sporobolus wrightii
Sporobolus heterolepis
Sporobolus indicus
Spartina spp.
Muhlenbergia tenuiflora
Muhlenbergia montana
Enneapogon spp.
Eragrostis curvula
Eragrostis spectabilis
Eragrostis spp.
Karroochloa spp.
Schismus barbatus
Danthonia decumbens
Danthonia spicata
Merxmuellera macowanii
Merxmuellera rangei
Stipagrostis zeyheri
Rottboellia spp.
Zea mays
Setaria spp.
Panicum spp.
Gynerium sagittatum
Thysanolaena maxima
Joinvillea spp.
Joinvilleaceae
F. sect. Festuca s.s.
Aulaxyper-Abyssinica
Apalochloa-Exaratae-Monachne-Psilurus-Vulpia(4x-6x)-Narduroides
F. sect. Eskia
Lolium-Micropyropsis-Schedonorus-Subbulbosae
Drymanthele-Leucopoa-Lojaconoa-Pseudoscariosa-Scabra-Scariosae
Fine-Leaved
Broad-Leaved
Loliinae
Dactylidinae/ Parapholiinae - Cynosurinae
Avenula s.s.
Airinae - Holcinae - Sesleriinae
Poinae - Puccinelliinae - Alopecurinae
Koeleriinae
Aveninae
Anthoxanthinae
Agrostidinae
Phalaridinae
Poeae - Aveneae
Bromeae - Triticeae
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae
Parianinae
Oryzinae
Sporobolinae
Muhlenbergiinae
Cotteinae
Eragrostidinae
Cortaderiinae/ Danthoniinae
Rottboelliinae
Tripsacinae
Setariinae
Panicinae
Olyreae
Oryzeae
Zoysieae
Eragrostideae
Danthonieae
Andropogoneae
Paniceae
Gynerieae
Centotheceae
Pooideae
Bambusoideae
Ehrhartoideae
Chloridoideae
Danthonioideae
Panicoideae
Centothecoideae
BEP
PACCMAD
0.02 substitution per site

Table 2. Data sets used for phylogenetic reconstructions and tree statistics obtained in Maximum Parsimony (MP) analyses. CI: consistency index (KLUGE AND FARRIS, 1969) excluding uninformative characters; RI: retention index (FARRIS, 1989).

Data set-based phylogenies (see Table 1)	Molecules used	nº grass sequences	Number of nucleotide positions	MP tree length [steps]	Parsimony-informative characters	CI	RI
Poaceae organismal tree	ITS, *mat*K, *ndh*F, *trn*TL, *trn*LF	142	4721	10310	1382 (40.98%)	0.4368	0.7598
1.- Exon positions	*ß-amylase*	223	588	1742	289 (48.9%)	0.3226	0.7633
4.- Exon + intron positions	*ß-amylase*	146	850	2561	452 (53.17%)	0.3838	0.7432
5.- Exon + intron positions	*ß-amylase*	105	850	3085	465 (54.70%)	0.3420	0.6580
5a.- Exon positions	*ß-amylase*	105	588	1202	239 (40.64%)	0.3927	0.6073
5b.- Non-synonymous positions	*ß-amylase*	105	392	543	133 (33.16%)	0.4254	0.8122
5c.- Synonymous positions	*ß-amylase*	105	196	1026	157 (79.36%)	0.2907	0.7054
5d.- Synonymous + intron positions	*ß-amylase*	105	456	2378	325 (71.27%)	0.4075	0.6903

The 142 Poaceae species fell into the two traditionally described PACCMAD and BEP clades (e.g. GPWG, 2001, 2012; SANCHEZ-KEN *ET AL.*, 2007; BOUCHENAK-KHELLADI *ET AL.*, 2008). The PACCMAD clade split into two strongly supported sister lineages, Centothecoideae-Panicoideae and Danthonioideae-Chloridoideae (e.g., SORENG AND DAVIS, 1998; BOUCHENAK-KHELLADI *ET AL.*, 2008, and references therein). The BEP clade included the Ehrhartoideae, Bambusoideae, and Pooideae, which were sorted as described by DAVIS AND SORENG (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), SCHNEIDER *ET AL.* (2009) and TRIPLETT AND CLARK (2010). The Pooideae clade showed the successive divergences of the more ancestral Lygeeae/Nardeae, Stipeae, Diarrheneae, and Brachypodieae lineages, and the separation of the more recently evolved core pooids (Bromeae – Triticeae and Poeae – Aveneae lineages). The divergences within the Poeae – Aveneae clade agreed with those proposed by DÖRING *ET AL.* (2007), QUINTANAR *ET AL.* (2007), and SCHNEIDER *ET AL.* (2009). The successively enlarged phylogenies of Loliinae obtained by CATALAN *ET AL.* (2004, 2006), and INDA *ET AL.* (2008) were confirmed in our analyses; the Loliinae split into two main groups, the broad-leaved and fine-leaved Loliinae.

4.2 Estimates of recombination, pseudogenization, and misplacement within the LCN *ß-amylase* gene

Highly similar clones were grouped into consensus types according to the p-distance (<0.01) criterion (DÍAZ-PÉREZ *ET AL.*, 2014). Clones from twelve out of 16 species were grouped into a single type, suggesting that they represent the same orthologous copy (Fig. 2). However, clones from each of the remaining four species (*Brachypodium distachyon*, *Festuca fenas, F. arundinacea* and *Vulpia alopecuros*) were classified into different types (p-distance>0.01). Diploid *Brachypodium distachyon* clones 2 and 3 vs. clones 1 and 4-9 were phylogenetically close enough to be considered allelic copies. Diploid *Vulpia alopecuros* clones 1-3 vs. 4-5 were successively divergent in misplaced positions (Fig. 2) within the Aulaxyper clade [(*V. alopecuros* clone 4-5 [0.89 BS (Bootstrap Support) / - PP (Posterior Probability)], (*V. alopecuros* clone 1-3, *F. iberica, F. rubra* clone 1-5 [1.00 BS / 97 PP]))], rather than within the expected Loretia (*V. fasciculata* and *V. geniculata*) clade (cf. Fig. 1). Though the Loretia group appeared in the *ß-amylase* reconstruction with low support, we considered that the strongly grouped *V. alopecuros* clones within the Aulaxyper clade likely indicate the existence of a "semi-paralogous" copy (concept used here to point out the detection of only one paralogous copy [Aulaxyper-type] but not the other [Loretia-type]). Related

tetraploid *Festuca fenas* and hexaploid *F. arundinacea* clones grouped into two lineages, each consisting of one representative of each species, suggesting two homeologous copies (Fig. 2).

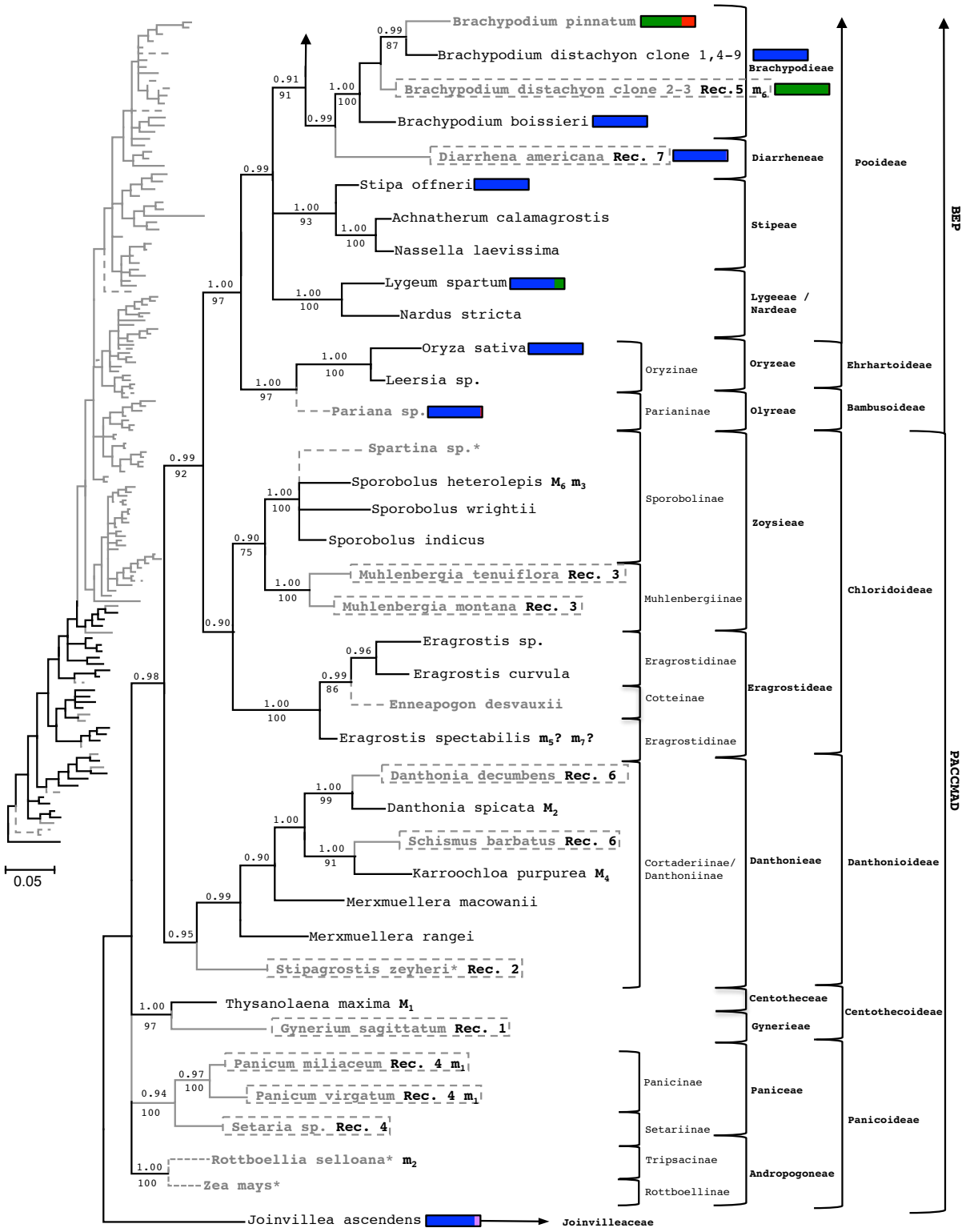


Figure 2. *(Continued)*

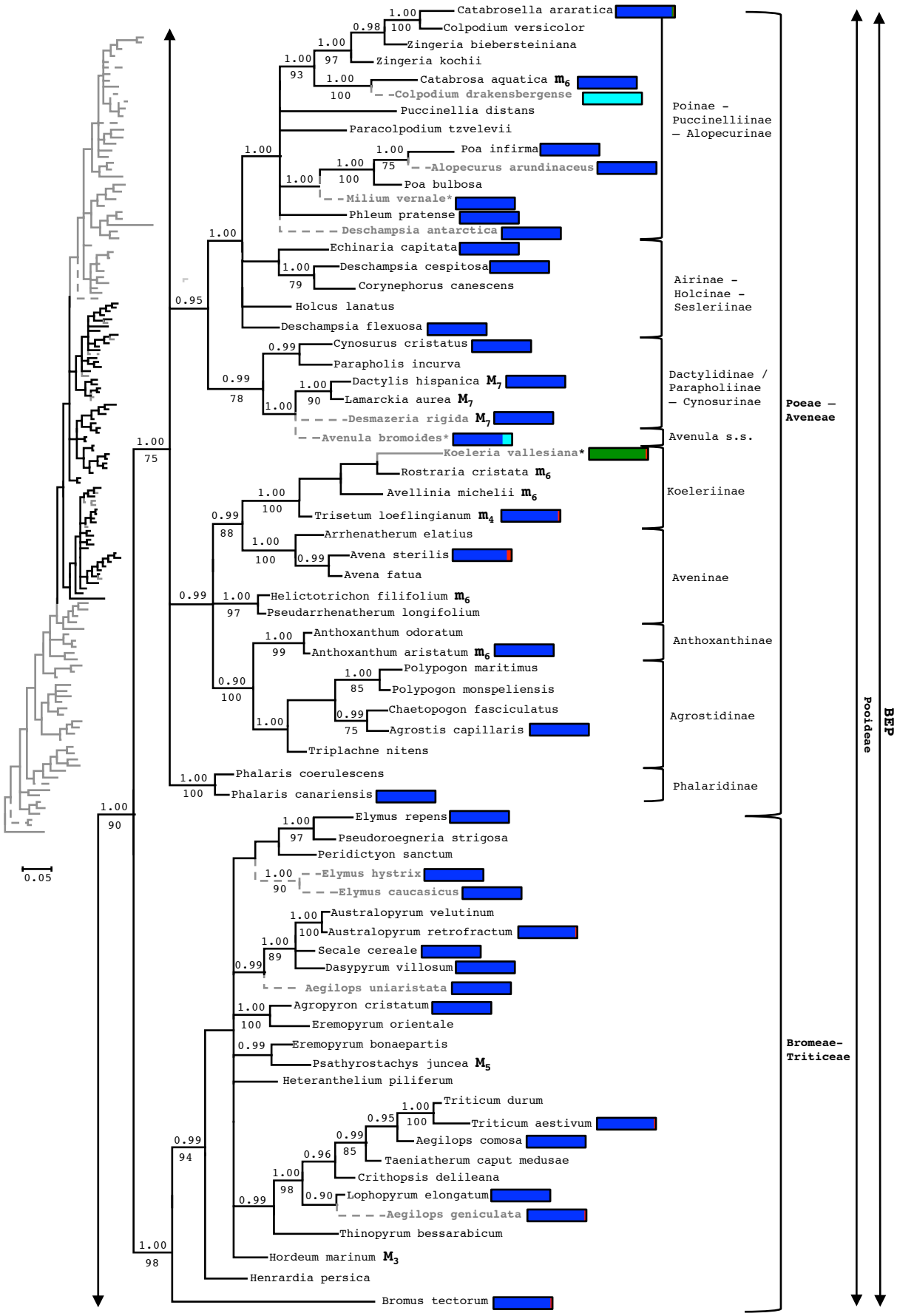


Figure 2. *(Continued)*

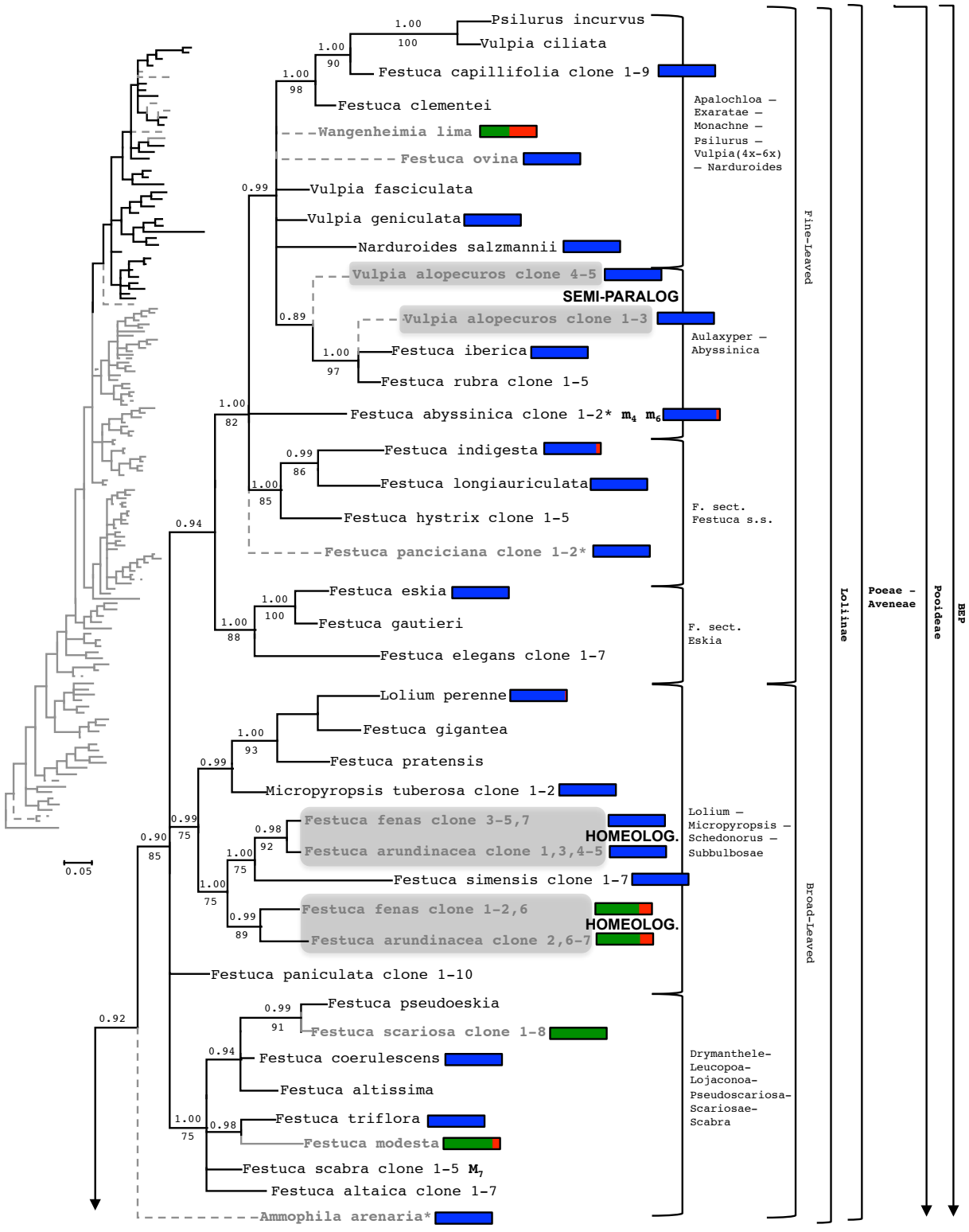


Figure 2. *(Continued)*

Figure 2. Bayesian Majority Rule consensus tree of Poaceae based on the complete *ß-amylase* exon + unambiguous intron data set. Bayesian posterior probabilities ≥ 90% and bootstrap support ≥ 75% are shown above and below the branches, respectively. Species assignment to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure. Dashed boxes and bold gray species names indicate recombination events detected in the exon-only data set (Table 1, *ß-amylase* data set 2) by two or more methods implemented in the RDP4 program; Rec. n: indicates the number of recombination events (see Table 3). M_n: major parents in the recombination n; m_n: minor parents in the recombination n; m_n?: most likely parents in the recombination n. Bold gray species names and gray branches indicate a high proportion of neutral, positive selected sites, and a complete absence of purifying selective positions, based on BS-REL method (see Table 4). Colors on the horizontal bars indicate strong negative selection (dark blue), negative selection (light blue), neutral selection (green), positive selection (pink), and strong positive selection (red). Grey boxes indicate homeologous clones of the allopolyploid species *Festuca fenas* and *F. arundinacea* and semi-paralogous clones of *Vulpia alopecuros*. Bold gray species names and gray dotted branches indicate that these accessions are misplaced when compared with their placement in the Poaceae organismal tree (Fig. 1). Asterisks (*) indicate placements different in the phylogeny based on the exon-only data set (Table 1, *ß-amylase* data set 1; Fig. S1; page 85) versus exon+intron data set (Table 1, *ß-amylase* data set 4; this figure).

Seven recombination events were detected by two or more algorithms ($p<0.05$) implemented in the program RDP4 in the exon-only *ß-amylase* data set (Table 3; Fig. 2). The recombinant sequences were from: (i) *Gynerium sagittatum* (Gynerieae), (ii) *Stipagrostis zeyheri* (Cortaderiinae / Danthoniinae); (iii) *Muhlenbergia tenuiflora* and *M. montana* (Muhlenbergiinae); (iv) *Setaria* sp. (Setariinae) and *Panicum miliaceum* and *P. virgatum* (Panicinae); (v) *Brachypodium distachyon* clones 2 and 3 (Brachypodieae); (vi) *Danthonia decumbens* and *Schismus barbatus* (Cortaderiinae / Danthoniinae); and (vii) *Diarrhena americana* (Diarrheninae). Five of the recombination events occurred in the PACCMAD clade, whereas two were associated with early-diverged Pooideae. No recombination events were detected in the core Pooideae clade. Most recombination events showed phylogenetically widely divergent minor and major parents involving different subtribes (Table 3; Fig. 2), suggesting ancient hybridization events, although we cannot discard a putative wrong assignment of parents given the limited sampling sizes associated with the PACCMAD and the early divergent Pooideae lineages.

Table 3. Summary of recombination events identified in the *β-amylase* exon data set (588 positions: exons 2 to 5) by two or more methods implemented in the program RDP4. These events were detected through 142 species (146 sequences) of Poaceae that included the consensus sequences (p-distance<0.01) of 16 cloned species (Table 1, *β-amylase* data set 2). Number of Rec. Event: Each recombination event is enumerated according to its statistical importance in the RDP4 analysis. Rec. sp.: Species under recombination.

Number of Rec. Event	Rec. sp.	Parental sequence(s)		Probability of significant tests for different detection methods in RDP4			
		Major	Minor	MaxChi	Chimaera	SiScan	3Seq
1.	*Gynerium sagittatum*	*Thysanolaena maxima*	*Panicum miliaceum* *Panicum virgatum*	1.36E-4	-	-	3.24E-2
2.	*Stipagrostis zeyheri*	*Dantohia spicata*	*Rottboellia selloana*	1.14E-3	4.29E-2	-	-
3.	*Muhlenbergia montana* *Muhlenbergia tenuifolia*	*Hordeum marinum*	*Sporobolus heterolepis*	2.34E-3	-	7.51E-5	-
4.	*Setaria* sp. *Panicum miliaceum* *Panicum virgatum*	*Karroochloa purpurea*	*Trisetum loeflingianum* *Festuca abyssinica* 1-2	-	2.98E-2	1.46E-4	-
5.	*Brachypodium distachyon* clone 2-3	*Psathyrostachys juncea*	Unknown (*Eragrostis spectabilis*)	4.39E-2	1.42E-2	-	-
6.	*Danthonia decumbens* *Schismus barbatus*	*Sporobolus heterolepis*	*Anthoxanthum aristatum* *Avellinia michelii* *Catabrosa aquatica* *Festuca abyssinica* *Helictotrichon filifolium* *Rostraria cristata*	8.17E-4	-	3.96E-6	-
7.	*Diarrhena americana*	*Dactylis hispanica* *Desmazeria rigida* *Festuca scabra* *Lamarckia aurea*	Unknown (*Eragrostis spectabilis*)	-	2.58E-2	8.89E-3	-

The criterion used to detect potential PTCs (AKHUNOV *ET AL.*, 2013) among the *ß-amylase* coding sequences did not retrieve any evidence of pseudogenic sequences. However, the Branch-Site REL (BS-REL) analysis detected different sequence positions under strongly negative, negative, neutral, positive and strongly positive selection within exons (Table 4; Fig. 2). Strong negative selection [$0 \leq \omega \leq 0.9$; range I (Table 4), dark blue bar (Fig. 2); 85.25% average value] was predominant across sequences and positions, followed by neutral selection [($0.91 \leq \omega \leq 1.5$; range III (Table 4), green bar (Fig. 2); 9.88%], and then by low percentages of strongly positive selection [$\omega > 5$; range V (Table 4), red bar (Fig. 2); 2.78%], negative selection [$0.9 \leq \omega \leq 0.61$; range II (Table 4), dark blue bar (Fig. 2); 1.7%] and positive selection ($0.5 \leq \omega \leq 1.51$; range IV (Table 4), pink bar (Fig. 2); 0.23%]. Fifty-eight out of 69 sequences tested with the BS-REL method showed ω values above the negative selection average (85.4%), most of them (57 sequences) with 0% neutral sites; whereas one sequence showed a considerable proportion of sites with weak negative selection pressure [*Colpodium drakensbergense* (100%); Table 4, Fig. 2]. Seven sequences, however, showed a high proportion of neutral sites, low proportion of positive selected sites, and a complete absence of positions under purifying selection [*Brachypodium pinnatum* (72% neutral sites / 28% positively selected sites), *F. arundinacea* clones 2, 6, and 7 (75% / 25%), *F. modesta* (85% / 15%), *F. fenas* clones 1, 2, and 6 (75% / 25%), *F. scariosa* clones 1-8 (100% / -), *Koeleria vallesiana* (95% / 5%), *Wangenheimia lima* (50% / 50%); Table 4, Fig. 2]. These figures suggest the existence of relaxed selection rates in their functional protein or putative pseudogenic copies.

Table 4. Summary of the Branch-Site REL (BS-REL) results analyzed in the WEB interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at such classes along each branch. $\omega = d_N/d_S$, where *dN* and *dS* are the number of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) ω= 0 – 0.6, Strongly Negative Selection (dark blue bar in Fig. 2); (II) ω= 0.61 – 0.9, Weak Negative Selection (light blue bar in Fig. 2); (III) ω= 0.91 – 1.5, Neutral Selection (green bar in Fig. 2); (IV) ω= 1.51 – 5, Weak Positive Selection (pink bar in Fig. 2); and (V) ω > 5, Strongly Positive Selection (red bar in Fig. 2). These analyses were done through 69 sequences (from 65 species) of Poaceae that included the consensus sequences (p-distance<0.01) of 16 cloned species (Table 1, *β-amylase* data set 3).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
Aegilops comosa	0	0	0	0	0	1	100%	0	0	0	0
Aegilops geniculata	0.06	0.96	0.06	0.03	10000	0.01	99%	0	0	0	1%
Aegilops uniaristata	0	0	0	0	0	1	100%	0	0	0	0
Agropyron cristatum	0	0	0	0	0	1	100%	0	0	0	0
Agrostis capillaris	0	0	0	0	0	1	100%	0	0	0	0
Alopecurus arundinaceus	0	0	0	0	0	1	100%	0	0	0	0
Ammophila arenaria	0	0	0	0	0	1	100%	0	0	0	0
Anthoxanthum aristatum	0	0.96	0	0.03	20.67	0.01	99%	0	0	0	1%
Australopyrum retrofractum	0	0.96	0	0.02	25.39	0.02	99%	0	0	0	2%
Avena sterilis	0	0.94	8.18E-07	0	8.42	0.06	94%	0	0	0	6%
Avenula bromoides	0.29	0.83	0.65	0.17	1.11	0	83%	17%	0	0	0
Brachypodium distachyon clone 2-3	1	0.97	1	0.02	834.46	0.01	0	0	99%	0	1%
Brachypodium distachyon clone 1,4-9	0	0	0	0	0	1	100%	0	0	0	0
Brachypodium boissieri	0.31	0.94	0.31	0.06	1.01	0	100%	0	0	0	0
Brachypodium pinnatum	1	0.72	1.00	4.57E-06	614.23	0.28	0	0	72%	0	28%
Bromus tectorum	0.37	0.97	0.81	0	33.03	0.03	97%	0	0	0	3%
Catabrosa aquatica	0	0	0	0	0	1	100%	0	0	0	0
Catabrosella araratica	0.06	0.98	0.91	0.00	38.01	0.02	98%	0	0.2%	0	1.8%
Colpodium drakensbergense	0.65	1	0.55	0	1.08	0	0	100%	0	0	0
Cynosurus cristatus	0	0	0	0	0	1	100%	0	0	0	0
Dactylis hispanica	0	0	0	0	0	1	100%	0	0	0	0
Dasypyrum villosum	0	0	0	0	0	1	100%	0	0	0	0
Deschampsia antarctica	0	0	0	0	0	1	100%	0	0	0	0
Deschampsia cespitosa	0	0	0	0	0	1	100%	0	0	0	0
Deschampsia flexuosa	0	0	0	0	0	1	100%	0	0	0	0
Desmazeria rigida	0.28	0.99	0.28	0.00	137.74	0.01	99%	0	0	0	1%
Diarrhena americana	0	0.86	0	0.02	3.80	0.13	98%	0	0	2%	0

Table 4. Summary of the Branch-Site REL (BS-REL) results analyzed in the WEB interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at such classes along each branch. $\omega = d_N/d_S$, where *dN* and *dS* are the number of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) ω= 0 – 0.6, Strongly Negative Selection (dark blue bar in Fig. 2); (II) ω= 0.61 – 0.9, Weak Negative Selection (light blue bar in Fig. 2); (III) ω= 0.91 – 1.5, Neutral Selection (green bar in Fig. 2); (IV) ω= 1.51 – 5, Weak Positive Selection (pink bar in Fig. 2); and (V) ω > 5, Strongly Positive Selection (red bar in Fig. 2). These analyses were done through 69 sequences (from 65 species) of Poaceae that included the consensus sequences (p-distance<0.01) of 16 cloned species (Table 1, *ß-amylase* data set 3).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
Echinaria capitata	0	0	0	0	0	1	100%	0	0	0	0
Elymus caucasicus	0	0	0	0	0	1	100%	0	0	0	0
Elymus hystrix	0	0	0	0	0	1	100%	0	0	0	0
Elymus repens	0.171564	1	0.248175	0	0.24	0	100%	0	0	0	0
Festuca abyssinica clone 1-2	0	0.94	0	0.00	6.54	0.06	94%	0	0	0	6%
Festuca arundinacea clone 2,6-7	1	0.75	1.00	0	1468.99	0.25	0	0	75%	0	25%
Festuca arundinacea clone 1,3,4-5	0	0	0	0	0	1	100%	0	0	0	0
Festuca capillifolia clone 1-9	0	0	0	0	0	1	100%	0	0	0	0
Festuca coerulescens	0	0	0	0	0	1	100%	0	0	0	0
Festuca eskia	0	0.99	0.88	0	49.19	0.01	99%	0	0	0	1%
Festuca iberica	0	0	0	0	0	1	100%	0	0	0	0
Festuca indigesta	0.00	0.90	0.89	0.00	8.76	0.10	90%	0	0	0	10%
Festuca longiauriculata	0	0	0	0	0	1	100%	0	0	0	0
Festuca modesta	1	0.85	0.99	1.85E-07	754.11	0.15	0	0	85%	0	15%
Festuca ovina	0	0	0	0	0	1	100%	0	0	0	0
Festuca triflora	0.34	0.75	0.39	0.25	0.98	0	100%	0	0	0	0
Festuca fenas clone 1-2,6	1	0.75	1.00	3.40E-05	1098.20	0.25	0	0	75%	0	25%
Festuca fenas clone 3-5,7	0	0	0	0	0	1	100%	0	0	0	0
Festuca panciciana clone 1-2	0	0	0	0	0	1	100%	0	0	0	0
Festuca scariosa clone 1-8	1	0.67	0.86	0	1.17	0.33	0	0	100%	0	0
Festuca simensis clone 1-7	0	0	0	0	0	1	100%	0	0	0	0
Joinvillea ascendens	0	0.86	0	0.00	4.85	0.14	86%	0	0	14%	0
Koeleria vallesiana	1	0.94	1	0.01	2541.96	0.05	0	0	95%	0	5%
Lolium perenne	0	0.96	0	0.01	15.34	0.03	97%	0	0	0	3%
Lophopyrum elongatum	0	0	0	0	0	1	100%	0	0	0	0
Lygeum spartum	0.06	0.79	0.96	0	1.46	0.21	79%	0	21%	0	0
Milium vernale	0	0	0	0	0	1	100%	0	0	0	0

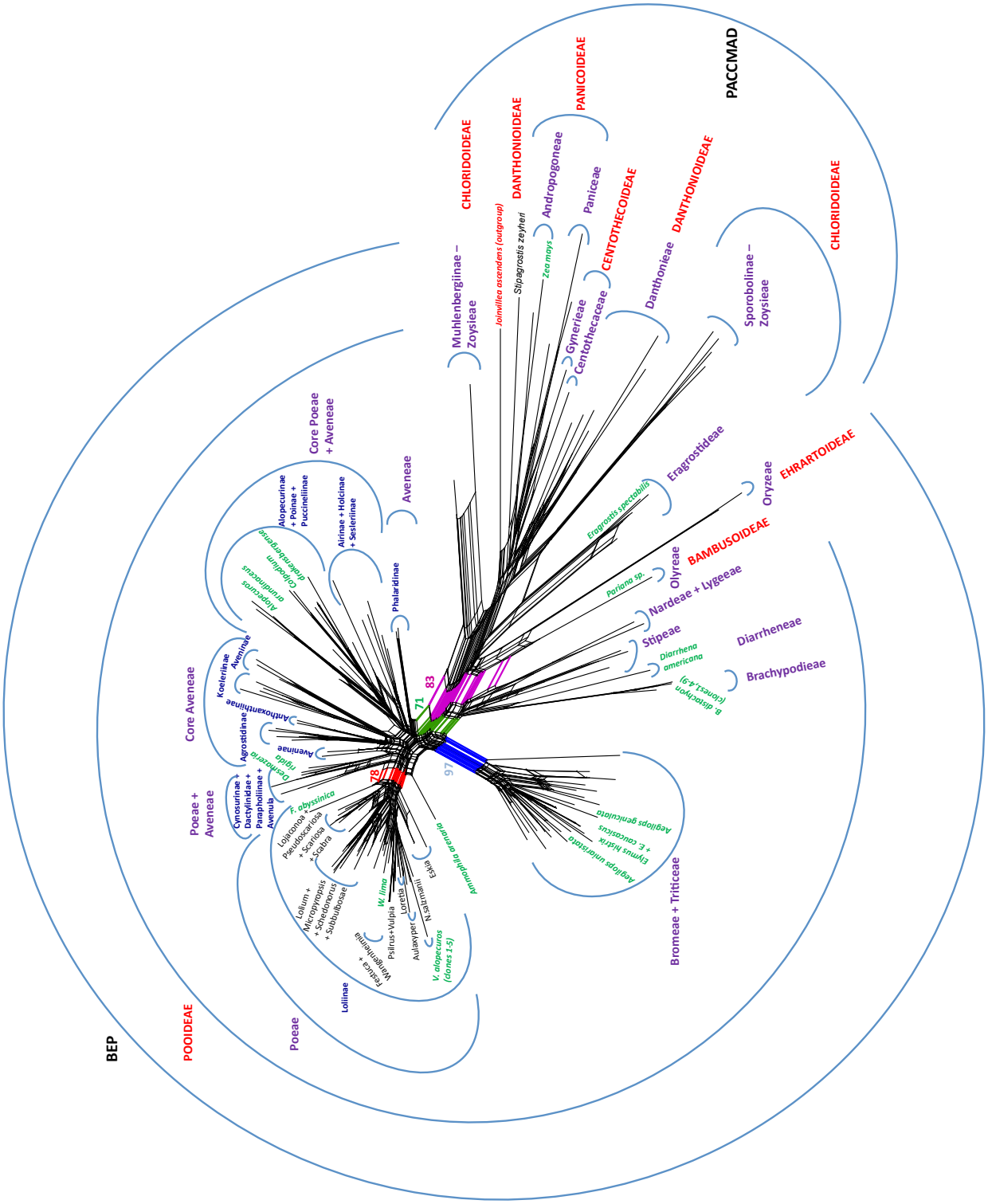
Table 4. Summary of the Branch-Site REL (BS-REL) results analyzed in the WEB interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at such classes along each branch. $\omega = d_N/d_S$, where *dN* and *dS* are the number of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) $\omega = 0 - 0.6$, Strongly Negative Selection (dark blue bar in Fig. 2); (II) $\omega = 0.61 - 0.9$, Weak Negative Selection (light blue bar in Fig. 2); (III) $\omega = 0.91 - 1.5$, Neutral Selection (green bar in Fig. 2); (IV) $\omega = 1.51 - 5$, Weak Positive Selection (pink bar in Fig. 2); and (V) $\omega > 5$, Strongly Positive Selection (red bar in Fig. 2). These analyses were done through 69 sequences (from 65 species) of Poaceae that included the consensus sequences (p-distance<0.01) of 16 cloned species (Table 1, *ß-amylase* data set 3).

Species	**ω_1**	**Pr1**	**ω_2**	**Pr2**	**ω_3**	**Pr3**	**I**	**II**	**III**	**IV**	**V**
Micropyropsis tuberosa clone 1-2	0	0	0	0	0	1	100%	0	0	0	0
Narduroides salzmannii	0.04	0.78	0.07	0.22	3.23	0	100%	0	0	0	0
Oryza sativa	0	0	0	0	0	1	100%	0	0	0	0
Pariana sp.	0.05	0.97	0.05	0.00	8.41	0.03	97%	0	0	0	3%
Phalaris canariensis	0.38	0.39	0.38	0.19	0.38	0.42	100%	0	0	0	0
Phleum pratense	0	0	0	0	0	1	100%	0	0	0	0
Poa infirma	0	0	0	0	0	1	100%	0	0	0	0
Secale cereale	0	0	0	0	0	1	100%	0	0	0	0
Stipa offneri	0	0	0	0	0	1	100%	0	0	0	0
Trisetum loeflingianum	0	0.97	0	0.00	7.04	0.03	97%	0	0	0	3%
Triticum aestivum	0	0.97	0	0.02	60.27	0.01	99%	0	0	0	1%
Vulpia alopecuros clone 4-	0	0	0	0	0	1	100%	0	0	0	0
Vulpia alopecuros clone 1-3	0	0	0	0	0	1	100%	0	0	0	0
Vulpia geniculata	0	0	0	0	0	1	100%	0	0	0	0
Wangenheimia lima	1	0.50	0.88	0	4299.40	0.50	0	0	50%	0	50%
Average percentage							85.25%	1.7%	9.88%	0.23%	2.78%

The Bayesian phylogenetic inference (BI) based on the *ß-amylase* exon and exon + intron data sets (Table 1, *ß-amylase* data sets 1 and 4, and Figs. 2 and S1 (see supplementary figure of chapter I in page 85), respectively) presented 28 misplaced species when assessed in comparison to the Poaceae organismal tree (Fig. 1). Eight of these misplacements (*Festuca paniciana*, *F. abyssinica*, *Koeleria vallesiana, Milium vernale, Rottboellia selloana, Stipagrostis zeyherii, Spartina* sp. and *Zea mays*) were only observed in the exon based-tree (Fig. S1, page 85), fixing their expected positions (or remaining unresolved) in the exon+intron phylogeny (Fig. 2). Two of the misplaced sequences were associated with neutral selection (*Koeleria vallesiana* and *Wangenheimia lima*; Table 4) while one misplacement (*Stipagrostis zeyherii*) was associated with a recombination event (Table 3). The remaining misplaced sequences, which showed moderate to high support in the *ß-amylase* phylogenies (Figs. 2, S1, page 85), were *Enneapogon desvauxii* (PACCMAD), *Pariana* sp. (Bambusoideae), *Aegilops geniculata*, *A. uniaristata*, *Elymus caucasicus* and *E. hystrix* (Triticeae), and *Avenula bromoides*, *Desmazeria rigida*, *Deschampsia antarctica*, *Alopecuros arundinaceus*, *Colpodium drakensbergense*, *Ammophila arenaria*, *Vulpia alopecuros* and *Festuca ovina* (Poeae-Aveneae).

The NeighborNet graph partitioned data into five statistically supported groups according to splits with moderate-to-high bootstrap support (BS >70%) (Fig. 3). These groups were: (i) the Poeae Loliinae subtribe (BS=78%); (ii) the remaining subtribes of Aveneae + Poeae (BS= 71% to 97%); (iii) Bromeae + Triticeae (BS= 97%); (iv) Olyreae + Nardeae + Lygeeae + Stipeae + Diarrheneae + Brachypodieae (BS= 71 to 83%); and (iv) PACCMAD + BEP early diverged lineages Oryzeae + Olyreae (BS=83%).

Figure 3. NeighborNet partition network tree. Graphical representation of phylogenetic relationships considering the putative occurrence of incompatible and ambiguous signals in the grass *ß-amylase* exon + intron sequences (Table 1, *ß-amylase* data set 4). Major tribal and subtribal lineages correspond to those indicated in Figs. 1 and 2. Green accessions represent misplaced lineages. Bootstrap support values are indicated for the four main splits: Pooideae (83%, purple); core pooids (71%, green); Triticeae – Bromeae (97%, blue); Loliinae (78%, red).


BEP
POOIDEAE
Poeae
Lolliinae
Festuca + Wangenheimia
Psilurus+Vulpia
Aulaxyper
Coleb
Lamarckia
V. alopecuros (clones 1-5)
W. limo
Lolium + Micropyropsis + Scheludorus + Subbulbosae
Lolgeum + Pseudoscarrosa + Scarrosa + Scabra
Ammophila arenaria
Poeae + Aveneae
Cynosurinae + Dactylinidae + Paraphollinae + Avenula
T. abyssinica
Dactylocteni
Agrostidinae
Aveninae
Rhyncho
Anthoxanthinae
Koelerinae
Meninae
Core Aveneae
Alopecurinae + Poinae + Puccinelliinae
Airinae + Holcinae + Sesleriinae
Aveneae
Core Poeae + Aveneae
Phalaridinae
Alloteropsis
Cinna
Chloridium
Cinnastrageae
Arundinoideae
Aegilops uniaristata + E. caucasicus + Elymus hystrix
Aegilops geniculata
Bromeae + Triticeae
Ioinvillea ascendens (outgroup)
Stipagrostis zeyheri
Zea mays
Andropogoneae
Paniceae
PANICOIDEAE
CHLORIDOIDEAE
DANTHONIOIDEAE
CENTOTHECOIDEAE
Gynerieae
Centothecaceae
Danthonieae
DANTHONIOIDEAE
PACCMAD
Muhlenbergiinae – Zoysieae
Sporobolinae – Zoysieae
CHLORIDOIDEAE
Eragrostis spectabilis
Eragrostideae
Oryzeae
EHRARTOIDEAE
Parlana sp.
Olyreae
BAMBUSOIDEAE
Nardeae + Lygeeae
Stipeae
Diarrhena americana
B. distachyon (clones1-4,9)
Diarrheneae
Brachypodieae
83
71
78
91
16

4.3 *Beta-amylase* phylogeny

The PH test did not detect incongruence (Table 5; $p>0.05$) among the exon, non-synonymous, synonymous and synonymous + intron positions data sets (Table 1, *ß-amylase* data sets 5a-5d) after removal of the 37 conflicting species (41 sequences including the collapsed clones) with likely pseudogenization, semi-paralogy, homeology, recombination, or misplacement. The SH pairwise topological tests (Table 5) used the exons + unambiguous introns tree as a reference (Table 1, *ß-amylase* data set 5); this tree was selected as the optimal *ß-amylase* tree since it was better resolved and more robust than any of the alternative reconstructions (Figs. 4A-D) and also showed the best fit to the organismal tree (Fig. 1), and to previously published phylogenies of the grass family. The SH results indicated that all tested topologies were equally well supported by the data (Table 5; $p<0.05$*). A visual comparison of the topologies (Figs. 4A-D) showed that branching order was, in general, poorly resolved in all the alternative *ß-amylase* trees. The Bayesian trees based on exons (Fig. 4A) and synonymous plus intron positions (Fig. 4D) had higher resolution than did the trees based on either non-synonymous (Fig. 4B) or synonymous (Fig. 4C) positions. Surprisingly, the Bromeae – Triticeae lineages were better resolved in the synonymous positions-based tree (Fig. 4C), and the fine-leaved Loliinae clade was better resolved in the synonymous + intron (Fig. 4D) positions-based tree. Branch lengths suggest that the synonymous + intron positions (Fig. 4D) have evolved considerably faster in the Bromeae – Triticeae lineages than in the Loliinae subtribe. However, non-synonymous positions (Fig. 4B) evolved faster in the fine-leaved Loliinae.

Table 5. Results of the Partition Homogeneity (PH) test [H_0: the compared data sets are incongruent] and the Shimodaira-Hasegawa test (SH) tests [H_0: the topology of the most likely tree (Table 1, *ß-amylase* data set 5: phylogeny with the highest likelihood score) is similar to the compared topology (Table 1, *ß-amylase* data sets 5a-5d; Figs. 4a to 4d)]. Exon pos.: Exon positions; Non-synony. pos.: non-synonymous positions; Synony. pos.: Synonymous positions; Synony. + intron pos.: Synonymous + intron positions; LnL: likelihood score; Diff. LnL: difference of -LnL between the compared trees; * the null hypotheses is accepted. These analyses were done using 105 sequences (from 105 species) of Poaceae that included the consensus sequences (p-distance<0.01) of 16 cloned species (Table 1, *ß-amylase* data set 5).

Partition Homogeneity (PH) test

***ß-amylase* subdivisions**	**(5a) Exon positions**	**(5b) Non-synonymous positions**	**(5c) Synonymous positions**	**(5d) Synonymous + intron positions**
(5a) Exon positions	-	1.00	1.00	0.96
(5b) Non-synonymous positions		-	0.86	0.32
(5c) Synonymous positions			-	1.00
(5d) Synonymous + intron positions				-

Shimodaira-Hasegawa test (SH) test

Bayesian topology compared	**-LnL**	**Diff. -LnL**	**p-value**
(5) Exon+intron positions	13266.38892	(Best)	-
(5d) Synonymous + intron positions	13352.79645	86.40752	0.013*
(5a) Exon positions	13423.06073	156.67181	<0.001*
(5b) Non-synonymous positions	14233.37348	966.98456	<0.001*
(5c) Synonymous positions	14455.69237	1189.30345	<0.001*

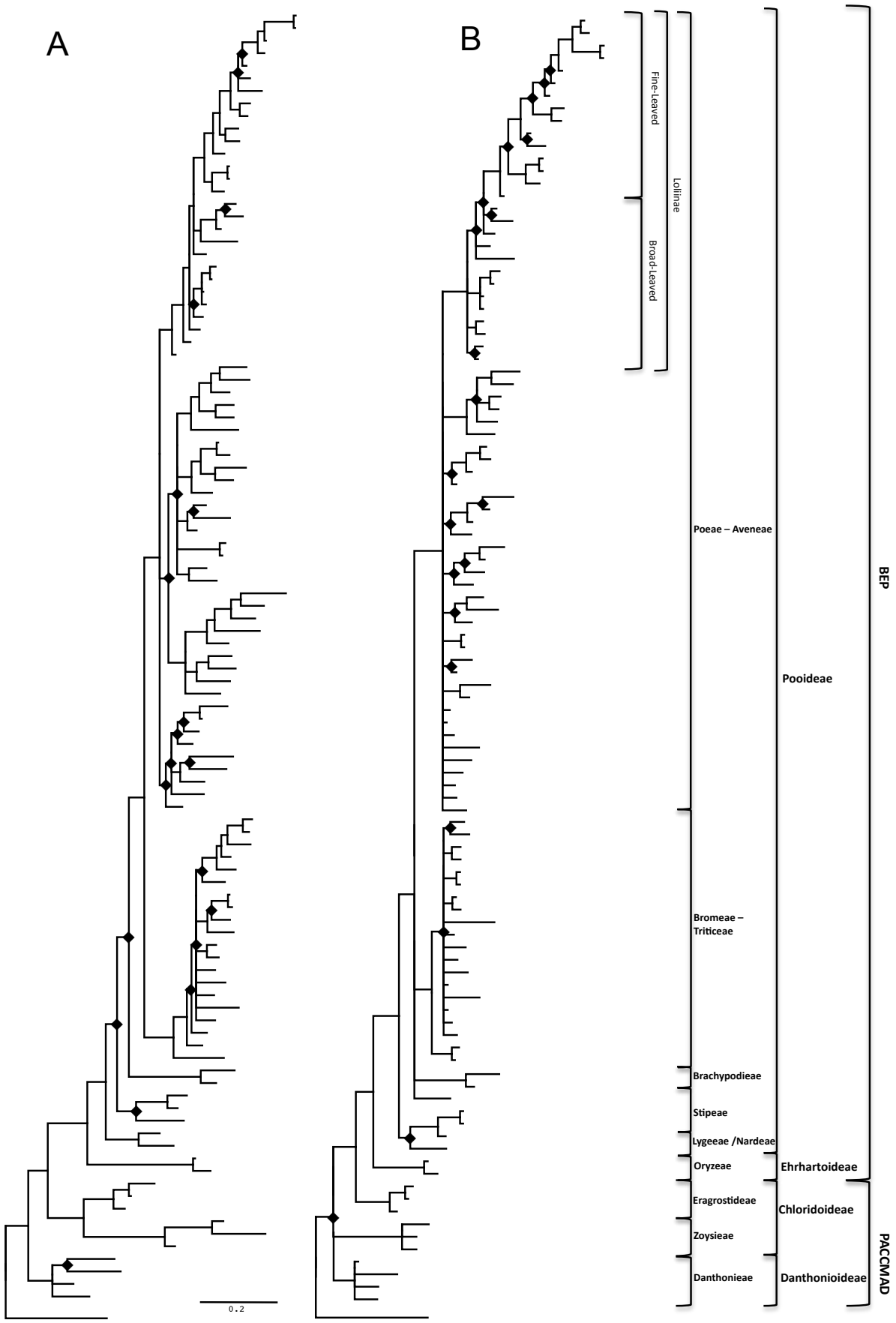


Figure 4. *(Continued)*

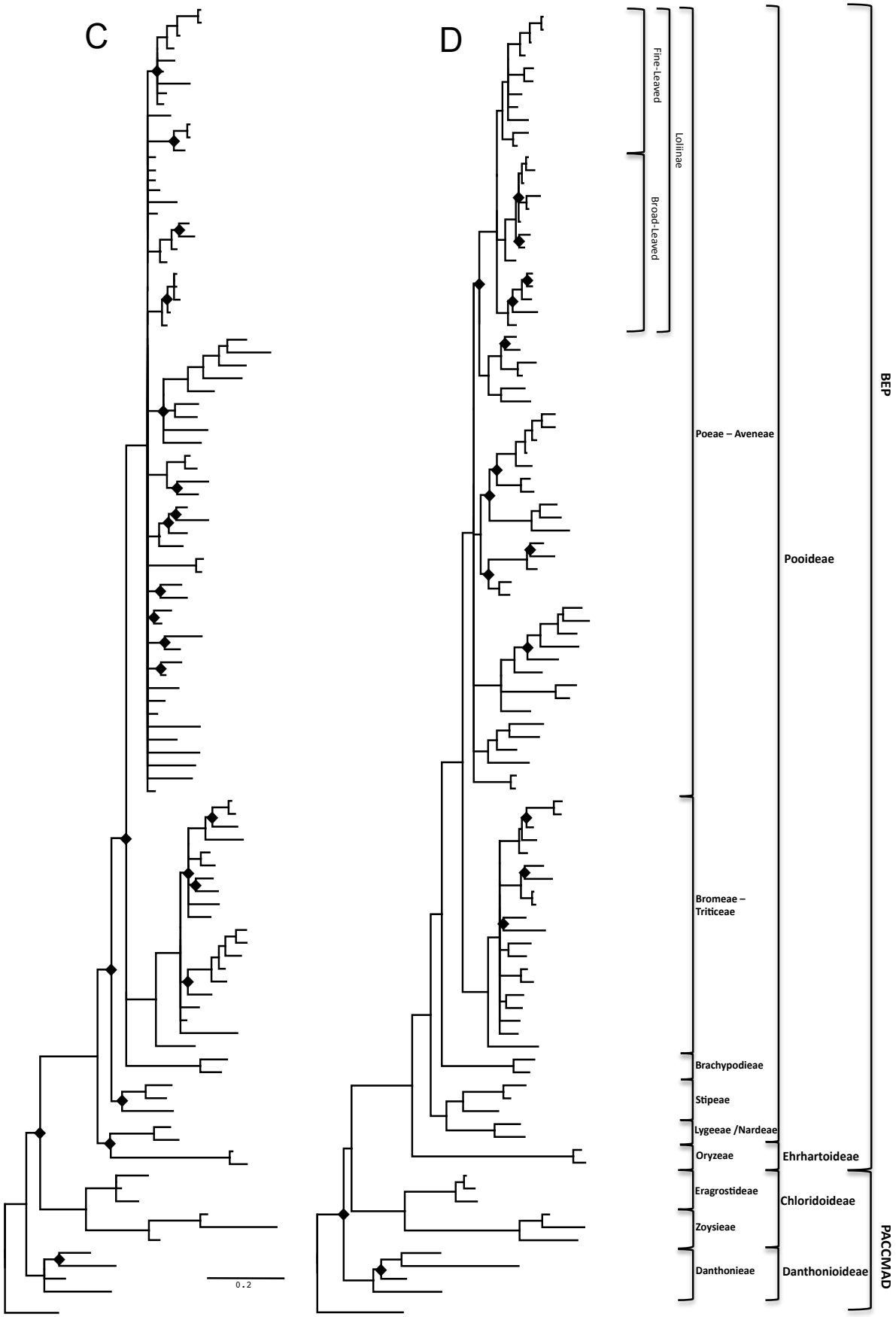


Figure 4. *(Continued)*

Figure 4. Alternative Bayesian topologies of grasses based on different partitions of the *ß-amylase* gene. (A) exon positions; (B) non-synonymous positions; (C) synonymous positions; and (D) synonymous positions plus aligned unambiguous introns after removing the 37 grass species (41 accessions) with likely pseudogenization, recombination and/or misplacement (Table 1, *ß-amylase* data sets 5a-5d, respectively). Diamonds indicate clades with low bootstrap (<75%) or posterior probability (<90%) support. Lineage assignments to tribes, subfamilies and clades are presented on the right side of the figures.

The most plausible *ß-amylase*-based phylogeny of the grasses is the one based on exon plus unambiguous intron positions (Fig. 5), which contained 105 sequences (from 105 grass species) sequences, including the collapsed clones, after removal of the 41 conflicting sequences (Table 1, *ß-amylase* data set 5). The samples were sorted into the two traditionally described BEP (1.00 PPS/ 100 BPS) and PACCMAD grass lineages. The paraphyletic PACCMAD clade showed the successive divergences of Centotheoideae (*Thysanolaena*), the monophyletic Danthonioideae (1.00/82), and Chloridoideae (1.00/88), in agreement with with the organismal tree (Fig. 1).

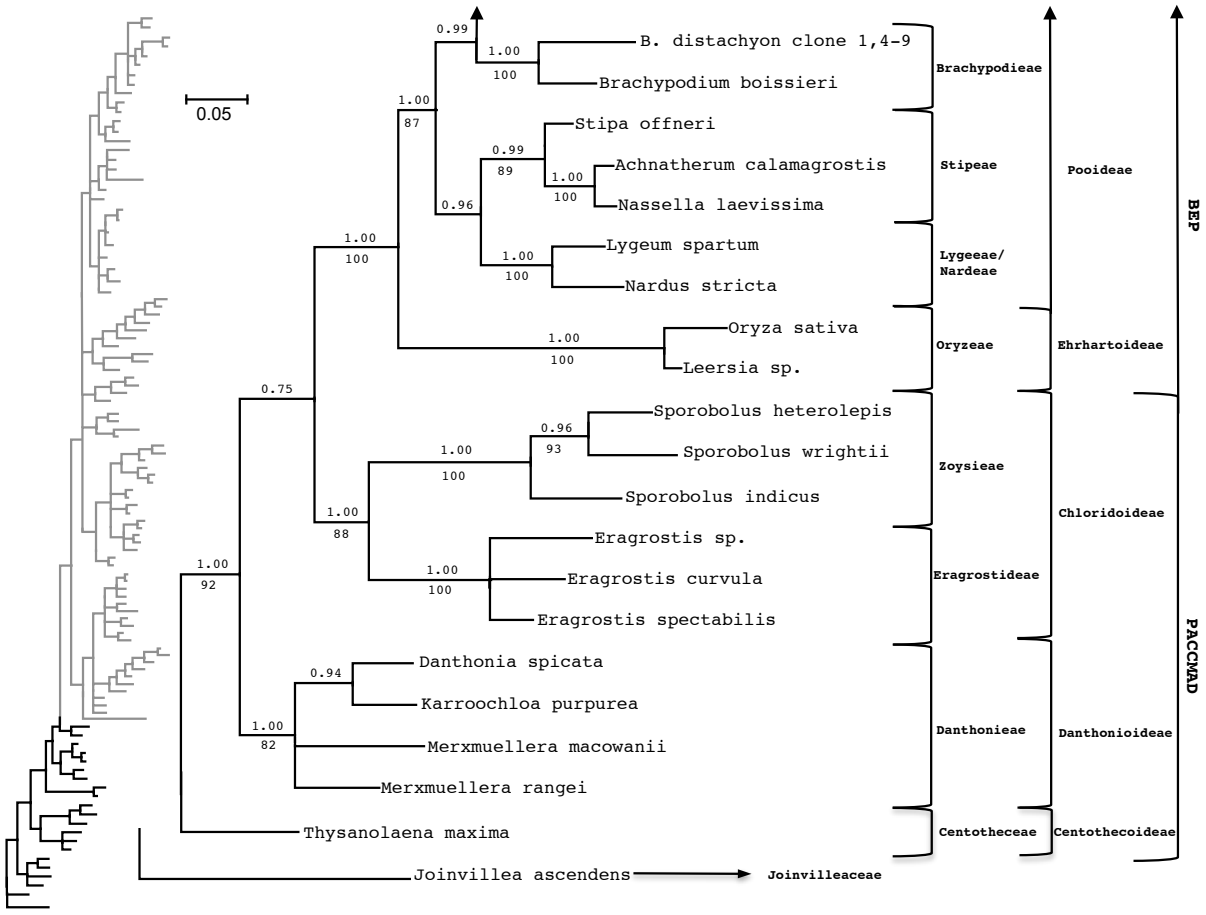


Figure 5. *(Continued)*

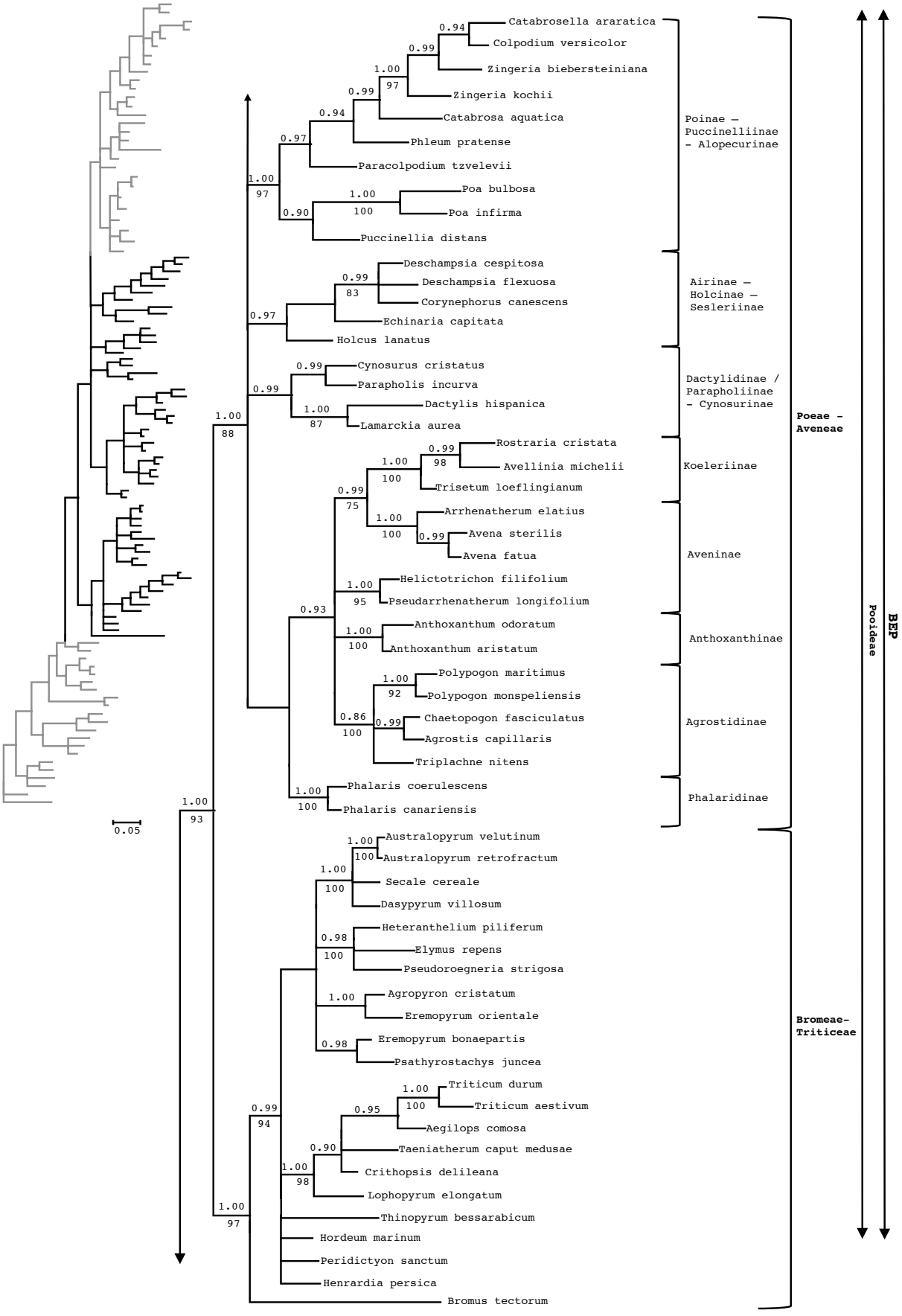


Figure 5. *(Continued)*

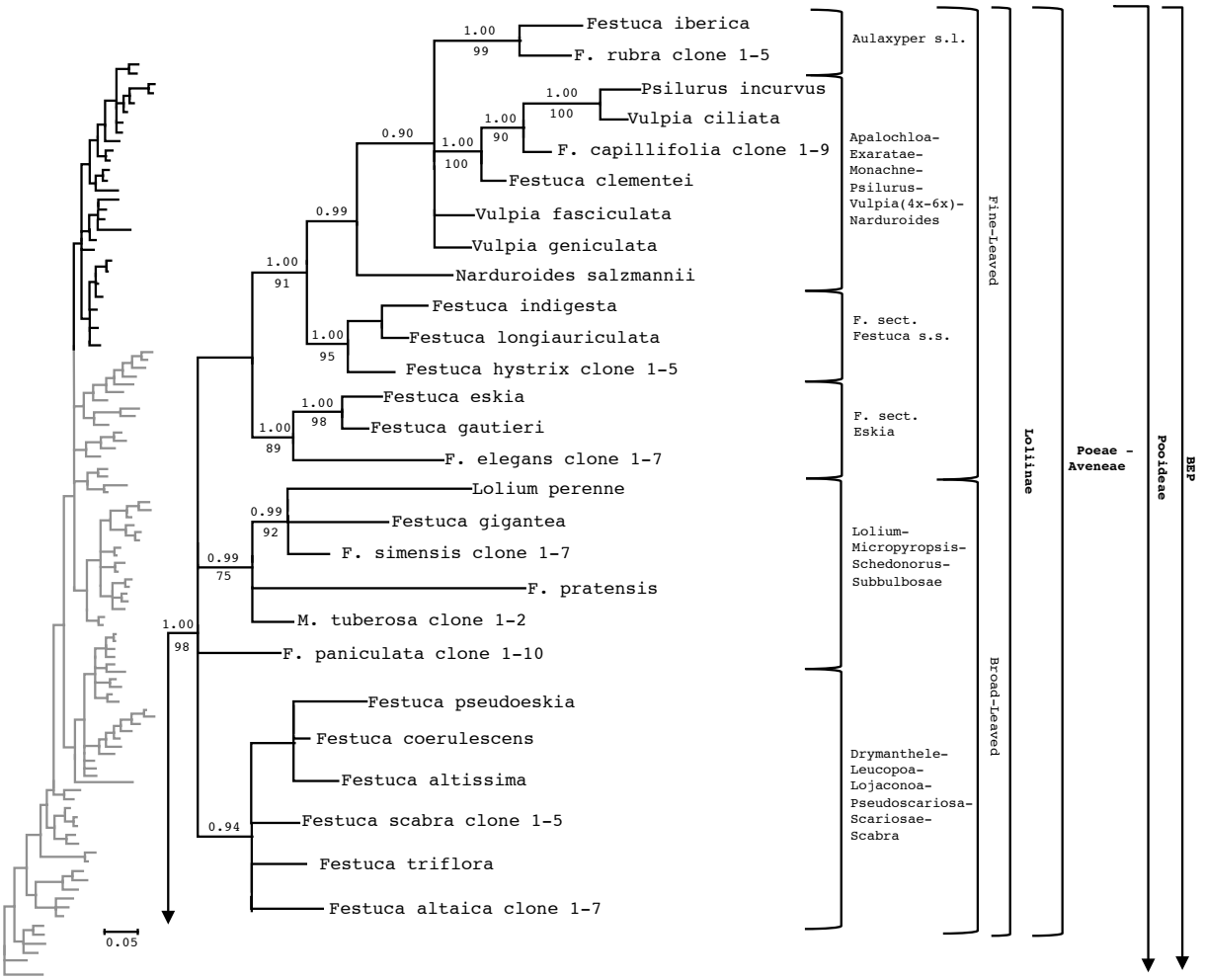


Figure 5. Bayesian Majority Rule consensus tree of Poaceae based on the *ß-amylase* exon + unambiguous intron data set with collapsed clone sequences (Table 1, *ß-amylase* data set 5; 850 nucleotide positions, 105 grass species, 105 sequences). Bayesian posterior probabilities ≥ 90% and bootstrap support ≥ 75% are shown above and below the nodes, respectively. Species assignment to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure.

The BEP clade (1.00/100) showed a well-supported sister relationship of Ehrhartoideae (*Oryza*, *Leersia*; 1.00/100) to Pooideae (1.00/87). Divergence within Pooideae consisted of the early split of the sisters Lygeeae/ Nardeae (1.00/100), Stipeae (0.99/89), Brachypodieae (1.00/100) and the core Pooideae (1.00/93). The resolution of the Bromeae – Triticeae clade was similar to that found in the *ß-amylase* exon-based tree (Fig. S1; page 85). The Poeae s. l. lineages collapsed in a large polytomy (1.00/88). The Poeae – Aveneae lineage was divided into six subtribes: (i) Phalaridinae (*Phalaris*; 1.00/100); (ii) Agrostidinae + Anthoxanthinae + Aveninae + Koeleriinae (0.93/-); (iii) Dactylidinae / Parapholiinae – Cynosurinae (0.99/-); (iv) Airinae – Holcinae – Sesleriinae (0.97/-); (v) Poinae – Puccinelliinae – Alopecurinae (1.00/97); and (vi) Loliinae (1.00/ 98). The Loliinae clade collapsed into a basal polytomy of two broad-leaved lineages and the fine-leaved fescues. The broad-leaved group included representatives of

(i) Drymanthele – Leucopoa – Lojaconoa – Psudoscariosa – Scariosa – Scabra clade (0.94/-); and (ii) Lolium – Micropyropsis – Schedonorus – Subbulbosae clade (0.99 / 98). The fine-leaved Loliinae split into three main clades: (i) the most basal Eskia (0.99/89); (ii) Festuca (1.00/95) and (iii) Apalochloa – Exaratae – Monachne – Psilurus – Vulpia (4x – 6x) – Narduroides + Aulaxyper s.l. (1.00/ 99).

5. Discussion

5.1 Evolution of the *ß-amylase* exon sequences in the grasses: a story of relaxation of selection pressure, pseudogenization, semi-paralogy, homeology, recombination, and misplacement events

Our study has shown the complex evolutionary dynamics of the LCNG *ß-amylase* in grasses through the detection of 41 sequences (from 37 grass species) with likely pseudogenization, paralogy, homeology, recombination and/or misplacement (Fig. 2). Our results also show that the identification and removal of such problematic sequences is key for using this LCNG to disentangle the phylogenetic history of the Poaceae (Fig. 5).

Our failure to detect PTCs as evidence of pseudogenic copies was corroborated using the amino acid translations to guide the alignment of these coding sequences. However, since exon 1 and a large part of exon 5 were not sequenced, this criterion alone does not conclusively rule out any potential presence of pseudogenes. Using a complementary method to detect pseudogenization, we interpreted sequences with a large proportion of sites under neutral or positive selection as indicative of relaxation of selection pressure (cf. KOSAKOVSKY-POND *ET AL.*, 2008, 2011). Thus, the eight sequences with a high proportion of neutral sites and/or positively selected sites, and a complete absence of purifying selective positions, are probably pseudogenes. The distribution of these copies among three separate temperate pooid lineages [Brachypodieae (*Brachypodium distachyon* clone 2-3*, B. pinnatum*); Aveneae - Koeleriinae (*Koeleria vallesiana*) and Poeae – Loliinae (*Festuca modesta, F. scariosa, F. fenas, F. arundinacea* clone 2, 6-7*, Wangenheimia lima*); Table 4, Fig. 2] indicates that pseudogenization and/or relaxation of selection pressure in *ß-amylase* sequences could be largely extended across the Poaceae, and affects both orthologous and homeologous copies.

Differentiating orthologous and paralogous DNA sequences is essential for phylogenetic inference (SANDERSON AND DOYLE, 1992; WENDEL AND DOYLE, 1998). However, distinguishing paralogous copies which diverged after a duplication event from multiple sequences related by descent (orthologs) can be a challenge (FU AND DOONER, 2002; SMALL *ET*

AL., 2004). Several publications have demonstrated paralogous copies in LCNGs such as the *ß-amylase* (KREIS *ET AL.*, 1987; WANG *ET AL.*, 1997; LI *ET AL.*, 2002; MASON-GAMER, 2005) and GBSSI (DÍAZ-PÉREZ *ET AL.*, 2014) genes in the grasses. In our study, a single case of *ß-amylase* semi-paralogy has been found in the fine-leaved Loliinae *Vulpia alopecuros*. The atypical nesting of the diploid *V. alopecuros* (clones 1-3 and clones 4 and 5) within the Aulaxyper clade (Figs. 2, S1; page 85) reflects a close relationship of this Loretia-type species (see organismic tree, Fig. 1) with the *Festuca rubra* group (Aulaxyper), supporting previous findings by DÍAZ-PÉREZ *ET AL.* (2014). These authors, in their LCN GBSSI-based study, recovered strong phylogenetic relationships between members of the Loretia clade (*Vulpia fasciculata*, *V. membranacea*, *V. fontqueri*, *V. geniculata*) and different Aulaxyper taxa, supporting recent gene flow between the two clades, which favored the occurrence of the intergeneric *x Festulpia* hybrids and introgressed plants. The two *V. alopecuros* Aulaxyper-type copies of *ß-amylase* clones likely represent a fraction of a semi-paralogous sequence, even though the orthologous sequences of *V. alopecuros*, which should be nested within the Loretia clade, have not been found. This could be explained as a consequence of PCR bias favoring only one paralogous copy, or gene loss. Hence, the two types of clones of *V. alopecuros* likely represent the first semi-paralogous event of the *ß-amylase* gene observed within the subtribe Loliinae. Alternativelly, because gene flow might explain the unexpected placement of *V. alopecuros* clones, introgressive hybridization could also be invoked to explain the two types of clones of *V. alopecuros* (cf. HARPER *ET AL.*, 2011; CUI *ET AL.* 2015).

The broad-leaved Loliinae polyploids *F. fenas* and *F. arundinacea* presented two types of *ß-amylase* homeologous copies (Fig. 2). The resolution within the Lolium – Micropyropsis – Schedonorus lineage observed in our study (Figs. 2, S1; page 85) confirms the origins of these allopolyploid Schedonorus species in the ancestral Western-Mediterranean region (INDA *ET AL.*, 2014). Several studies (e. g. cross-hybridizations, CHANDRASEKHARAN *ET AL.*, 1972; genome-mapping, HUMPHREYS *ET AL.*, 1995; phylogenetic analysis, INDA *ET AL.*, 2008, 2014) support the allohexaploid origin of *Festuca arundinacea* from diploid *F. pratensis* and allotetraploid *F. fenas* (syn. *F. arundinacea* var. *glaucescens* Boiss.). INDA *ET AL.* (2014) further suggested the putative contribution of *F. fenas* as the maternal progenitor of *F. arundinacea*. The *ß-amylase* phylogeny detected the existence of two divergent types of clones in *F. fenas* (*F. fenas* clones 1, 2, and 6 vs. 3-5 and 7), which could have been derived from each of its two putative homeologous genomes, and, in turn, were inherited by its descendant hybrid *F. arundinacea* (*F. arundinacea* clones 1, 3, 4, and 5 vs. 2, 6, and 7). By contrast, the absence of *F. arundinacea*

clones similar to the progenitor *F. pratensis* suggests that those paternal copies might have been lost in the allohexaploid or could have gone undetected with the current cloning sample size.

The *ß-amylase* exons + introns phylogeny (Fig. 2) reconstructed a robust sister relationship of the afromontane *F. simensis* (clones 1-7) to the group of *F. fenas* (clones 3-5 and 7) – *F. arundinacea* (clones 1, 3, 4, and 5), revealing for the first time the potential origin of one of the progenitors of the putative allotetraploid *F. simensis*. INDA *ET AL.* (2014) described a sister relationship of *F. simensis* with its potential diploid Eurasian ryegrass-type parent (*Lolium* clade) based on the nuclear ITS phylogeny, whereas their plastid *trn*TF phylogeny suggested a Eurasian Schedonorus maternal genome donor for *F. simensis*. Based on these hypotheses and our current findings, the tetraploid *F. simensis* would probably have resulted from a cross of a more ancestral putative diploid western Mediterranean *F. fenas*-type maternal parent and a more recently evolved Eurasian diploid *Lolium*-type paternal parent, followed by genome duplication and a likely Pliocene colonization of the afromontane belt (cf. INDA *ET AL.* 2014).

Recombination analyses of 146 *ß-amylase* sequences (from 142 grass species) detected seven recombination events; five were found in the less-sampled PACCMAD clade, and 2 in the basal Pooideae tribes (Table 3, Fig. 2). Although the major and minor parental sequences of the recombinants could not be unambiguously identified, the seven *ß-amylase* recombinants support the hypothesis that recombination occurs frequently in the grass family. This is explained at least in part by their allogamous pollination syndrome (cf. LIAN *ET AL.*, 2013), but could also be related to other unknown biological or evolutionary processes (cf. MINAYA *ET AL.* 2013). Most of the potential parental sequences for the detected recombinations are phylogenetically close. For example, the major and minor parental sequences detected in *Gynerium sagittatum* were *Thysanolaena maxima* (Centotheceae) and *Panicum miliaceum – P. virgatum* (Paniceae). Since nuclear recombination is commonly described among taxa of hybrid origin (MARQUES *ET AL.*, 2010), these observations could reveal the unexpected existence of hybridizations. Alternatively, these results might reflect traces of more ancestral paleohybridizations, as previously pointed out by MASON-GAMER (2008), SALSE *ET AL.* (2008), and MAHELKA AND KOPECKÝ (2010) in different grass groups. Two remarkable exceptions to the phylogenetically close parental sequences rule were observed in *Muhlenbergia* (Muhlengbergiinae), where the major parental sequence was detected in Triticeae (*Hordeum marinum*), and in *Panicum* (Panicinae), where the minor parental sequence was detected in

Koeleriinae (*Trisetum loeflingianum*) and Loliinae (*Festuca abyssinica*) (Table 3, Fig. 2). Because the putative parental lineages (Triticeae, Koeleriinae, and Loliinae) are highly divergent groups, each characterized by high rates of hybridization, polyploidy (cf. MASON-GAMER, 2005, 2008; CATALAN *ET AL.*, 2006), Horizontal Gene Transfer (HGT) (GHATNEKAR *ET AL.*, 2006), and the existence of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) (MASON-GAMER, 2005; MINAYA *ET AL.*, 2013), these evolutionarily distant recombinations could have occurred by a number of methods, including the combination of paleo- or neo-homeologous copies.

The misplacement of 28 species in the *ß-amylase* phylogenies (Figs. 2, S1; page 85) with respect to their placement in the organismal tree (Fig. 1) was caused by potential pseudogenization and recombination events in two and one instances, respectively. Since LCNGs are prone to a diversity of evolutionary fates, there is a variety of possible explanations for the remaining misplacements, such as incomplete lineage sorting, retention of alternative copies from the ancestral genome of polyploid species, horizontal gene transfer, differential gene copy losses (FORTUNE *ET AL.*, 2007; DÍAZ-PÉREZ *ET AL.*, 2014), or even small sample sizes in some lineages (i.e. PACCMAD and Bambusoideae).

5.2 Phylogenetic signals of *ß-amylase* sequences within the Poaceae

A visual comparison of the alternative topologies obtained based on the exon, non-synonymous, synonymous, and synonymous + intron positions data sets, after removal of the 41 conflicting accessions (37 species) (Fig. 4), corroborates the absence of incongruence among the four data sets predicted by the PH test (Table 5). In addition, these four topologies are similar to the phylogeny based on the exon + intron data set depicted in figure 5 (taken as the reference *ß-amylase* tree), which confirms the results of the SH test (Table 5). Resolution also varies along the trees; for example, the phylogeny recovered for the Triticeae is better resolved based on synonymous than non-synonymous positions (Fig. 4). The advantage of using synonymous positions in phylogenetic reconstructions becomes particularly clear at low taxonomic levels or among lineages originated by rapid and recent divergence events (cf. GAUT *ET AL.*, 1997; SMALL *ET AL.*, 1998; CRONN *ET AL.*, 2002; CHRISTIN *ET AL.*, 2007). Nevertheless, our results demonstrate that other grass lineages do not follow this pattern. The recently evolved Loliinae clade was less resolved when it was based on synonymous than non-synonymous positions (Fig. 4). Similarly, the Loliinae lineages presented shorter branch lengths in the Bayesian synonymous tree than in the non-synonymous reconstruction. In both cases, the

uncoupling of non-synonymous and synonymous evolutionary rates observed suggests that locus-specific selection coefficients can change widely among grass lineages. These observations highlight the importance of understanding how different parts of the *ß-amylase* molecule have different influence in the phylogenetic reconstruction of the grasses (cf. CHRISTIN *ET AL.*, 2007).

The most plausible hypothesis that summarizes the *ß-amylase* phylogeny of Poaceae after removing the 41 problematic accessions is the one based on exon plus unambiguous intron positions (Fig. 5). This reconstruction supports five subfamilies: Centothecoideae, Danthonioideae, Chloridoideae, Ehrhartoideae and Pooideae. The paraphyly of the PACCMAD group disagrees with several plastid- and nuclear gene-based phylogenies, which show a strongly supported monophyletic PACCMAD clade (e.g. DUVALL AND MORTON, 1996; SORENG AND DAVIS, 1998; MATHEW *ET AL.,* 2000). The paraphyly of the PACCMAD clade could reflect rapid radiation within this group, as described by GPWG (2001), obscuring the phylogenetic signal and making relationships difficult to reconstruct. Furthermore, the relationships among the major PACCMAD lineages have not yet been fully resolved to date (GPWG, 2012).

The monophyly of the BEP clade is strongly supported in the *ß-amylase* phylogeny (Figs. 2, 5). The Pooideae lineage, the most widely sampled subfamily in our study, displays the most complex and consistent internal structure. Although all the molecular analyses published to date have supported the monophyly of Pooideae, the relationships found among its constituent groups have varied widely. In particular, the early-diverging tribal lineages are frequently unresolved or only weakly supported (e.g. CATALAN *ET AL.,* 1997; SORENG AND DAVIS, 2000; GPWG, 2001). The basal divergence of the Lygeeae/Nardeae, suggested on morphological grounds (e.g. CLAYTON AND RENVOIZE, 1986; SORENG AND DAVIS, 1998) has been corroborated in our analysis, and is in agreement with other molecular studies (e.g. BOUCHENAK-KHELLADI *ET AL.,* 2008). However, the placement of the Stipeae as sister to the Lygeeae/Nardeae group conflicts with recent phylogenies that reconstruct Stipeae as a more recently evolved group (BOUCHENAK-KHELLADI *ET AL.,* 2008). A variety of molecular studies present Stipeae as a basal or subbasal pooid lineage but with a position that fluctuates according to the taxa sampled or the genome analyzed (e.g. HILU *ET AL.,* 1999; DÖRING *ET AL.,* 2007; BOUCHENAK-KHELLADI *ET AL.,* 2008).

The core Pooideae reconstruction shows a monophyletic Bromeae – Triticeae and a largely polytomic Poeae – Aveneae clade. For the Triticeae, our Bayesian inference supports

relationships based on a more restricted phylogenetic analysis of the *ß-amylase* gene (MASON-GAMER, 2005). It is, however, not in full agreement with phylogenies based on other data sets; the numerous molecular analyses have failed to converge on a straightforward estimate of the relationships among the diploid genera of this tribe (cf. KELLOGG *ET AL.,* 1996; MASON-GAMER AND KELLOGG, 1996; SEBERG AND FREDERIKSEN, 2001; PETERSEN AND SEBERG, 2002; and MASON-GAMER, 2005). The Poeae – Aveneae group includes the five main lineages of the core Aveneae group (Phalaridinae, Agrostidinae, Anthoxanthinae, Aveninae, and Koeleriinae), mostly agreeing with the phylogenies proposed by SORENG *ET AL.* (2003), QUINTANAR *ET AL.* (2007) and SAARELA *ET AL.* (2010). The unresolved position of the *Pseudarrhenatherum* / *Helictotrichon* lineage is consistent with both a nested relationship within Aveneae (e. g. QUINTANAR *ET AL.* 2007), and with placement outside of *Avena* and *Arrhenatherum* (COUDERC AND GUÉDÈS 1976; RÖSER *ET AL.* 2001).

The unresolved Poeae and former Aveneae p.p. lineages included four strongly supported lineages (Figs. 2, 5): the Dactylidinae / Parapholinae – Cynosurinae; the Airinae – Holcinae – Sesleriieae; the Poinae – Puccinelliinae – Alopecurinae; and the Loliinae groups. This phylogeny is more similar to nuclear-based phylogenies (e.g. RODIONOV *ET AL.,* 2005; QUINTANAR *ET AL.,* 2007) than to plastid-based phylogenies (e.g. SORENG *ET AL.,* 2007; GILLESPIE *ET AL.,* 2008). Our *ß-amylase* tree collapsed the broad-leaved Loliinae into a basal polytomy of two lineages, the Drymanthele – Leucopoa – Lojaconoa – Pseudoscarosa – Scariosae – Scabra and the Lolium – Micropyropsis – Schedonorus – Subbulbosae groups. This is partially in agreement with the molecular analyses of TORRECILLA AND CATALAN (2002), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2004, 2006) and INDA *ET AL.* (2008), which reconstructed the broad-leaved Loliinae as a monophyletic lineage. The *ß-amylase* exons+introns tree reconstructs a monophyletic fine-leaved Loliinae clade with a topology that is mostly in agreement with previous phylogenetic studies (CATALAN *ET AL.*, 2004, 2006; TORRECILLA *ET AL.,* 2004; INDA *ET AL.,* 2008).

The NeighborNet partition network analysis (Fig. 3) highlights the phylogenetic separation between the Loliinae and Aveneae subtribes. This separation was not observed in the Bayesian tree based on the LCNG GBSS (DÍAZ-PÉREZ *ET AL.*, 2014), where there was little resolution within the BEP clade. The genetic distinctiveness of the Triticeae – Bromeae partition from Poeae – Aveneae in the NeighborNet tree seems to indicate a more basal phylogenetic position for Triticeae - Bromeae. The more basal BEP lineages are found in the fourth and fifth groups of the NeighborNet tree, suggesting that the Oryzeae + Olyreae lineage,

placed in the fifth group of the NeighborNet tree, could be one of the first groups to diverge in the separation of the BEP and PACCMAD clades, as is also suggested by the GBSS Bayesian tree (DÍAZ-PÉREZ *ET AL.*, 2014).

The atypical phylogenetic patterns described for 37 grass species (26% of the sampled species; Fig. 2) and the detection of different types of MITEs (MINAYA ET AL., 2013) highlight the frequent occurrence of alternative distorting evolutionary events such as recombination, relaxed selection pressure, semiparalogy, incomplete lineage sorting, horizontal gene transfer, or differential gene copy losses in the *ß-amylase* gene, which recomends caution before using this LCNG as phylogenetic marker. However, once the pitfalls are identified and removed, the phylogenetic reconstruction of the grasses based on the *ß-amylase* exon+intron positions is optimal at all taxonomic levels.

6. Literature cited

Akhunov, E.D., Sehgal, S., Liang, H., Wang, S., Akhunova, A.R., Kaur, G., Li, W., Forrest, K.L., See, D., Simková, H., Ma, Y., Hayden, M.J., Luo, M., Faris, J.D., Dolezel, J., Gill, B.S., 2013. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiology* 161, 252-265.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC Bioinformatics* 20, 407-415.

Altenhoff, A.M., Dessimoz, C., 2012. Inferring orthology and paralogy. In: Anisimova, M. (Ed.), *Evolutionary Genomics: Statistical and Computational methods*, Vol. 1. Springer Science+Business Media, pp. 259-279.

Alvarez, I., Wendel, J.F., 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29, 417-434.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Bryant, D., Moulton, V., 2004. Neighbor-Net: an agglomerative method for the construction of Phylogenetic networks. *Molecular Biology and Evolution* 21, 255-265.

Brunner, P.C., Keller, N., McDonald, A., 2009. Wheat domestication accelerated evolution and triggered positive selection in the ß-Xylosidase enzyme of *Mycosphaerella graminicola*. *PlosOne 11, e7884*.

Buckler, E.S., Ippolito, A., Holtsford, T.P. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821-832.

Campbell, C.S., Kellogg, E.A., 1987. Sister group relationships of the Poaceae. In: Soderstrom, T.R., Hilu K.W., Campbell, C.S., and Barkworth, M.E. (Eds.), *Grass Systematics and Evolution*. Smithsonian Institution Press: Washington, DC, pp. 217-224.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndh*F gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, RG., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from *ITS* and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca, Lolium, Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: Bailey, J., Ellis, R.G. (Eds.), *Current taxonomic research on the British & European flora*. London: BSBI, pp. 45-70.

Catalan, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L.A., Langdon, T., Betekhtin, A., Siwinska, D., Pimentel, M., López-Alvarez, D., 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany* 109, 385-405.

Chandrasekharan, P., Lewis, E.J., Borrill, M., 1972. Studies in *Festuca* 2. Fertility, relationships between species of sections *Bovinae* and *Scariosae*, and their affinities with *Lolium*. *Genetica* 43, 375-386.

Christin, P., Salamin, N., Savolainen, V., Duvall, M.R., Besnard, G., 2007. C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology* 17, 1241-1247.

Clayton, W.D., Renvoize, S.A., 1986. *Genera Gramineum: Grasses of the World*. *Kew Bulletin*. Series XIII, Royal Botanic Gardens, Kew.

Couderc, J.M., Guédès , M., 1976. *Helictotrichon pallens* and *H. setifolia*. *Taxon* 25, 188.

Cromie, G.A., Smith, G.R., 2007. Branching out: meiotic recombination and its regulation. *Trends in Cell Biology* 17, 448-455.

Cronn, R.C., Small, R.L., Haselkorn, T., Wendel, J.F. 2002. Rapid diversification of the cotton genus (*Gossypium: Malvaceae*) related by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89, 707-725.

Cui, H., Sun, Y., Deng, J., Wang, M., Xia, G., 2015. Chromosome elimination and introgression following somatic hybridization between bread wheat and other grass species. *Plant Cell, Tissue and Organ Culture* 120, 203-210.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristic and parallel computing. *Nature Methods* 9, 772.

Davis, J.I., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in *GBSSI*. In: Columbus, J.T., Friar, E.A., Porter, J.M., Prince, L.M. and Simpson, M.G. (Eds.), *Monocots: comparative biology and evolution. Poales. Aliso* 23, 335-348.

Díaz-Pérez, A., Sharifi-Tehrani, M., Inda, L.A., Catalan, P., 2014. Polyphyly, gene-duplication and extensive allopolyploidy framed the evolution of the ephemeral *Vulpia* grasses and other fine-leaved Loliinae (Poaceae). *Molecular Phylogenetics and Evolution* 79, 92-105.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex (Pooideae, Poaceae). *Kew Bulletin* 62, 407-424.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Doyle, J.J., 1991. Evolution of higher-plant glutamine synthetase genes: tissue specificity as a criterion for predicting orthology. *Molecular Biology and Evolution* 8, 366-377.

Doyle, J.J., 1996. Homoplasy connections and disconnections: genes and species, molecules and morphology. In: Sanderson, M.J., Hufford, L. (Eds.), *Homoplasy: the recurrence of similarity in evolution*. Academic, San Diego, pp. 67-89.

Doyle, J.J., Doyle, J.L., 1999. Nuclear protein-coding genes in phylogeny reconstruction and homology assessment: some examples from Leguminosae. In: R.J. Gornall (Ed.), *Molecular systematics and plant evolution*. Taylor and Francis: London, pp. 229-254.

Duvall, M.R., Morton, B.R., 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbc*L sequence data. *Molecular Phylogenetics and Evolution* 5, 352-358.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.

Farris, J.S., 1989. The retention index and the rescaled consistency index. *Cladistics* 5, 417-419.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315-319.

Feliner, G.N., Rossello, J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44, 911-919.

Fortune, P.M., Schierenbeck, K.A., Ainouche, A.K., Jacquemin, J., Wendel, J.F., Ainouche, M.L., 2007. Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species (Poaceae). *Molecular Phylogenetics and Evolution* 43, 1040-1055.

Fu, H., Dooner, H.K., 2002. Intraspecific violation of genetic colinearity and its implications in maize. *Proceedings of the National Academy of Sciences of the United States of America* 99, 9573-9578.

Gabaldon, T., 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biology* 9, 235.

Gaut, B.S., Clark, L.G., Wendel, J.F., Muse, S.V., 1997. Comparisons of the molecular evolutionary process at *rbcL* and *ndhF* in the grass family (Poaceae). *Molecular Biology and Evolution* 14, 769-777.

Ghatnekar, L., Jaarola, M., Bengtsson, B.O., 2006. The introgression of a functional nuclear gene from *Poa* to *Festuca ovina*. *Proceedings of the Royal Society B* 273, 395-399.

Gillespie, L.J., Soreng, R.J., Bull, R.D., Jacobs, S.W.L., Refulio-Rodriguez, N., 2008. Phylogenetic relationships in subtribe Poinae (Poaceae, Poeae) based on nuclear *ITS* and plastid *trnT-trnL-trnF* sequences. *Botany* 86, 938-967.

G.P.W.G., The Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88, 373-457.

G.P.W.G., The Grass Phylogeny Working Group, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Guindon, S., Gascuel, O., 2003. A simple, fasta and accurate method to estimate large phylogenies by maximum-likelihood. *Systematics Biology* 52, 696-704

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41, 95-98.

Harper, J., Armstead, I., Thomas, A., James, C., Gasior, D., Bisaga, M., Roberts, L., King, I., King, J., 2011. Alien itrogression in the grasses *Lolium perenne* (perennial ryegrass) and *F. pratensis* (meadow fescue): the development of seven monosomic substitution lines and their molecular and cytological charaterization. *Annals of Botany* 107, 1313-1321.

Hilu, K.W., Alice, L.A., Liang, H., 1999. Phylogeny of Poaceae inferred from *mat*K sequences. *Annals of the Missouri Botanical Garden* 86, 835-851.

Hughes, C.E., Eastwood, R.J. Bailey, C.D., 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny reconstruction. *Philosophical Transactions of the Royal Society B* 361, 211-225.

Humphreys, M.W., Thomas, H.M., Morgan, W.G., Meredith, M.R., Harper, J.A., Thomas, H., Zwierzykowski, Z., Ghesquiére, M., 1995. Discriminating the ancestral progenitors of hexaploid *Festuca arundinacea* using genomic in situ hybridisation. *Heredity* 75, 171-174.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23, 254-267.

Huson, D.H., Rupp, R., Scornavacca, C., 2010. *Phylogenetic networks. Concepts, algorithms and applications*. Cambridge University Press.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Inda, L.A., Sanmartín, I., Buerki, S., Catalan, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (Festuca subgenus Schedonorus and Lolium). *Journal of Biogeography* 41, 600-614.

Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany* 21, 321-347.

Kellogg, E.A., 2006. Progress and challenges in studies of the evolution of development. *Journal of Experimental Botany* 57, 3505-3516.

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624-626.

Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18, 1-32.

Kosakovsky-Pond, S.L., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* 22, 1208-1222.

Kosakovsky-Pond, S.L., Poon, A.F.Y., Leigh Brown, A.J., Frost, S.D.W., 2008. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Molecular Biology and Evolution* 25, 1809-1824.

Kosakovsky-Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* 28, 3033–3043.

Kreis, M., Williamson, M., Buxton, B., Pywell, J., Hejgaard, J., Svendsen, I., 1987. Primary structure and differential expression of b-amylase in normal and mutant barleys. *European Journal of Biochemistry* 179, 517-525.

Lian, S., Lee, J., Cho, W.K., Yu, J., Kim, J. Choi, H., Kim, K., 2013. Phylogenetics and recombination analyses of tomato spotted wilt virus. *PlosOne* 8, e63380.

Li, C.D., Langridge, P., Zhang, X.Q., Eckstein, P.E., Rossnagel, B.G., Lance, R.C.M., Lefol, E.B., Lu, M.Y., Harvey, B.L., Scoles, G.J., 2002. Mapping on barley (*Hordeum vulgare* L.) *beta*-amylase alleles in which amino acid substitution determines *beta*-amylase isoenzyme type and the level of free *beta*-amylase. *Journal of Cereal Science* 35, 39-50.

Li, W.-H., 1997. *Molecular Evolution*. Sinauer, Sunderland, MA.

Linder, H.P., Rudall, P.J., 1993. The megagametophyte in *Anarthria* (Anarthriaceae, Poales) and its implications for the phylogeny of the Poales. *American Journal of Botany* 80, 1455-1464.

Linder, C.R., Rieseberg, L.H., 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91, 1700-1708.

Mahelka, V., Kopecký, D., 2010. Gene capture from across the grass family in the allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triticeae) as evidenced by ITS, GBSSI, and molecular cytogenetics. *Molecular Biology and Evolution* 27, 1370-1390.

Maddison, D.R., Maddison, W.P., 2008. *MacClade 4.08 version for OS X*. Sinauer Associates, Inc.

Marques, I., Feliner, G.N., Draper, D., Martins-Louçao, M.A., Fuertes-Aguilar, J., 2010. Unraveling cryptic reticulate relationships and the origin of orphan hybrid disjunct populations in Narcissus. *Evolution* 64, 2353-2368.

Martin D.P, Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP4: a flexible and fast computer program for analyzing recombination. *BMC Bioinformatics* 26, 2462-2463.

Mason-Gamer, R.J., Kellogg, E.A., 1996. Chloroplast DNA analysis of the monogenomic Triticeae: phylogenetic implications and genome-specific markers. In: P.P. Jauhar (Ed.), *Methods of genome analysis in plants*. CRC Press, Boca Raton, Florida, USA. pp. 301-325.

Mason-Gamer, R.J., Kellogg, E.A., 1997. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45, 524-545.

Mason-Gamer, R.J., 2005. The *b-amylase* genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). *American Journal of Botany* 92, 1045-1058.

Mason-Gamer, R.J., 2008. Allohexaploidy, introgression, and the complex phylogenetic history of Elymus repens (Poaceae). *Molecular Phylogenetics and Evolution* 47, 598-611.

Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *American Journal of Botany* 87, 96-107.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

Miyata, T., Yasunaga, T., 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* 16, 23-36.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positive selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929-936.

Petersen, G., Seberg, O., 2002. Molecular evolution and phylogenetic application of *DMC1*. *Molecular Phylogenetics and Evolution* 22, 43-50.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531-2533.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences, U.S.A.* 98, 13757-13762.

Preston, J.C., Kellogg, E.A., 2006. Reconstructing the evolutionary history of paralogous APETALA1/ FRUITFULL-Like genes in grasses (Poaceae). Genetics 174, 421-437.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trn*T-F and nuclear *ITS* sequences. *American Journal of Botany* 94, 1554-1596.

Rajapakse, S., Nilmalgoda, S.D., Molnar, M., Balnard, R., Ballard, R.E., Austin, D.F., Bohac, J.R., 2004. Phylogenetic relationships of the sweet potato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear b-amylase sequences. *Molecular Phylogenetics and Evolution 30*, 623-632.

Rodionov, A.V., Tyupa, N.B., Kim. E.S., Machs, E.M., Loskutov, I.G., 2005. Genome composition of the autotetraploid oat species *Avena macrostachya* inferred from comparative analysis of the ITS1 and ITS2 sequences: on the oat karyotype evolution during the early events of the *Avena* species divergence. *Russian Journal of Genetics* 41, 518-528.

Ronquist, F., Huelsenbeck, J.P., Teslenko M., 2011. MrBayes version 3.2 manual: tutorials and model summaries (http://mrbayes.sourceforge.net/mb3.2_manual.pdf).

Röser, M., Winterfeld, G., Grebenstein, B., Hemleben, V., 2001. Molecular diversity and physical mapping of 5S rDNA in wild and cultivated oat grasses (Poaceae: Aveneae). *Molecular Phylogenetics and Evolution* 21, 198-217.

Saarela, J.M., Liu, Q., Peterson, P.M., Soreng, R.J., Paszko, B., 2010. Phylogenetics of the grass Aveneae-type plastid DNA clade (Poaceae: Pooideae, Poeae) based on plastid and nuclear ribosomal DNA sequence data. In: O. Seberg, G. Petersen, A.S. Barfod (Eds.), *Diversity, Phylogeny, and Evolution in the monocotyledons:* proceedings of the fourth international conference on the comparative biology of the monocotyledons and the fifth international symposium of grass systematics and evolution, pp. 557-585.

Salse, J., Bolot, S., Throude, M.I., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuilleta, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20, 11-24.

Sanchez-Ken, J.G., Clark, L.G., Kellogg, E.A., Kay, E.E., 2007. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Systematic Biology* 54, 183-196.

Sanderson, M.J., Doyle, J.J., 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. *Systematic Biology* 41, 4-17.

Sang, T., Zhang, D., 1999. Reconstructing hybrid speciation using sequences of low-copy nuclear genes: hybrid origins of five *Paeonia* species based on *Adh* gene phylogenies. *Systematic Botany* 24, 148-163.

Sang, T., 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Biochemistry and Molecular Biology* 37, 121-147.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *mat*K gene-3' *trn*K exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

Seberg, O., Frederiksen, S., 2001. A phylogenetic analysis of the monogenomic Triticeae (Poaceae) based on morphology. *Botanical Journal of the Linnean Society* 136, 75-97.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16, 1114-1116.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogenetic reconstruction in a recently diverged plant group. *American Journal of Botany* 85, 1301-1315.

Small, R.L., Cronn, R.C., Wendel, J.F., 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Systematic Botany* 17, 145-170.

Soreng, R.J., Davis, J.I., 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction character sets. *Botanical Review* 64, 1-85.

Soreng, R.J., Davis, J.I., 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. In: S.W.L. Jacobs, J. Everett (Eds.), *Grasses: Systematics and Evolution*. CSIRO, Melbourne. pp. 61-74 .

Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., (Eds.), 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Swofford, D.L., 2002. *PAUP*: Phylogenetic analysis using parsimony and other methods. Version 4b10*. Sunderland (MA): Sinauer Associates.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear *ITS* sequences. *Systematic Botany* 27, 241-251.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Triplett, J.K., Clark, L.G., 2010. Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on Arundinaria and allies. *Systematic Botany* 35, 102-120.

Wang, S.M., Lue, W.L., Wu, S.Y., Huang, H.W., Chen, J.C., 1997. Characterization of a maize *ß-amylase* cDNA clone and its expression during seed germination. *Plant Physiology* 113, 403-409.

Watson, L., Dallwitz, M.J., 1992 onwards. *The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references*. Version: 12th August 2014. http://delta-intkey.com

Wendel, J.F., Doyle, J.J., 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In: D.E. Soltis, P.S. Soltis, J.J. Doyle (Eds.), *Molecular systematics of plants II. DNA sequencing*. Kluwer Academic Publishing.

Yang, Z., Bielawski, J.P., 2000. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* 19, 908-917.

Ziegler, P., 1999. Cereal beta-amylase. *Journal of Cereal Science* 29, 195-204.

7. Supplementary figure

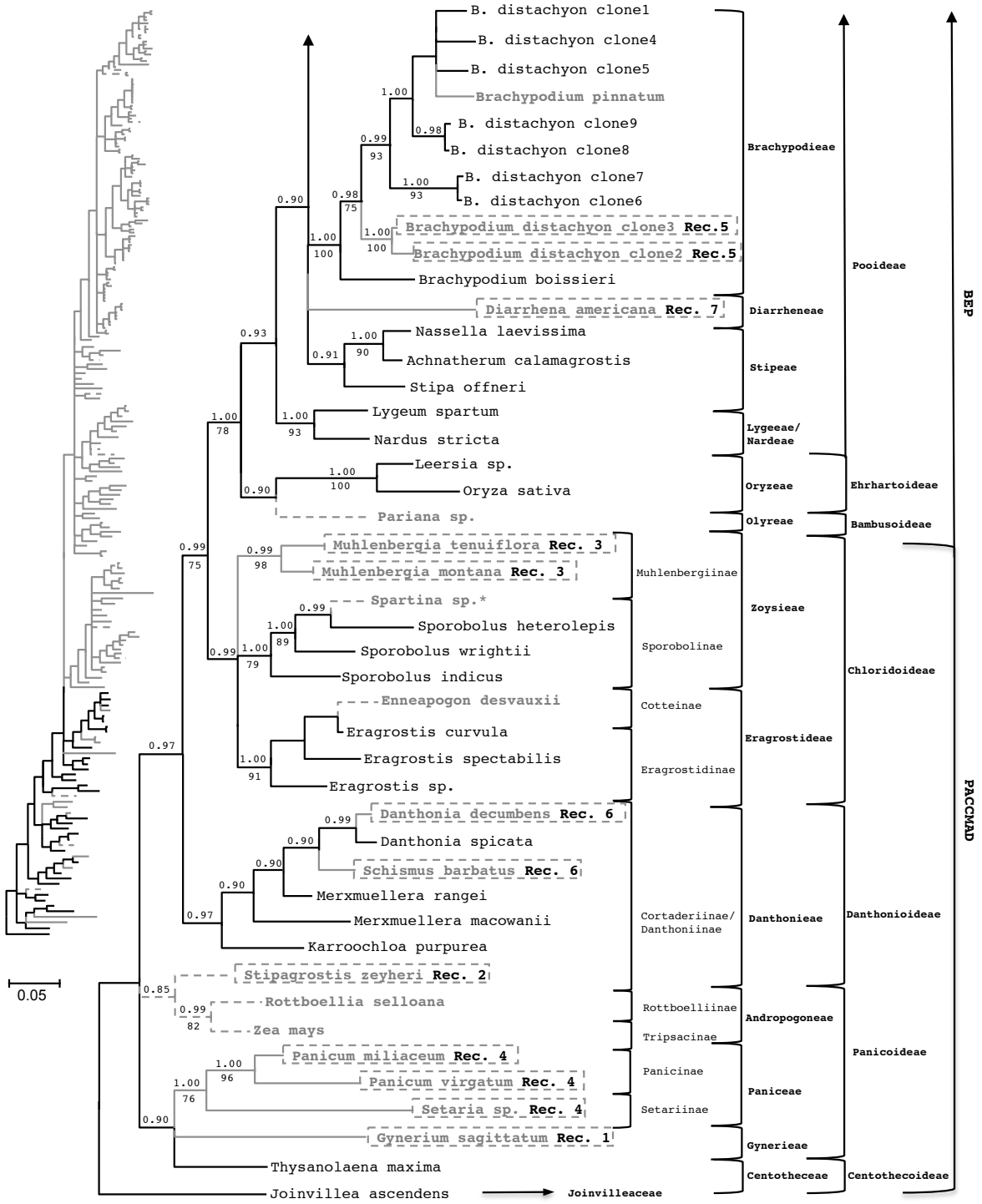


Figure S1. *(Continued)*

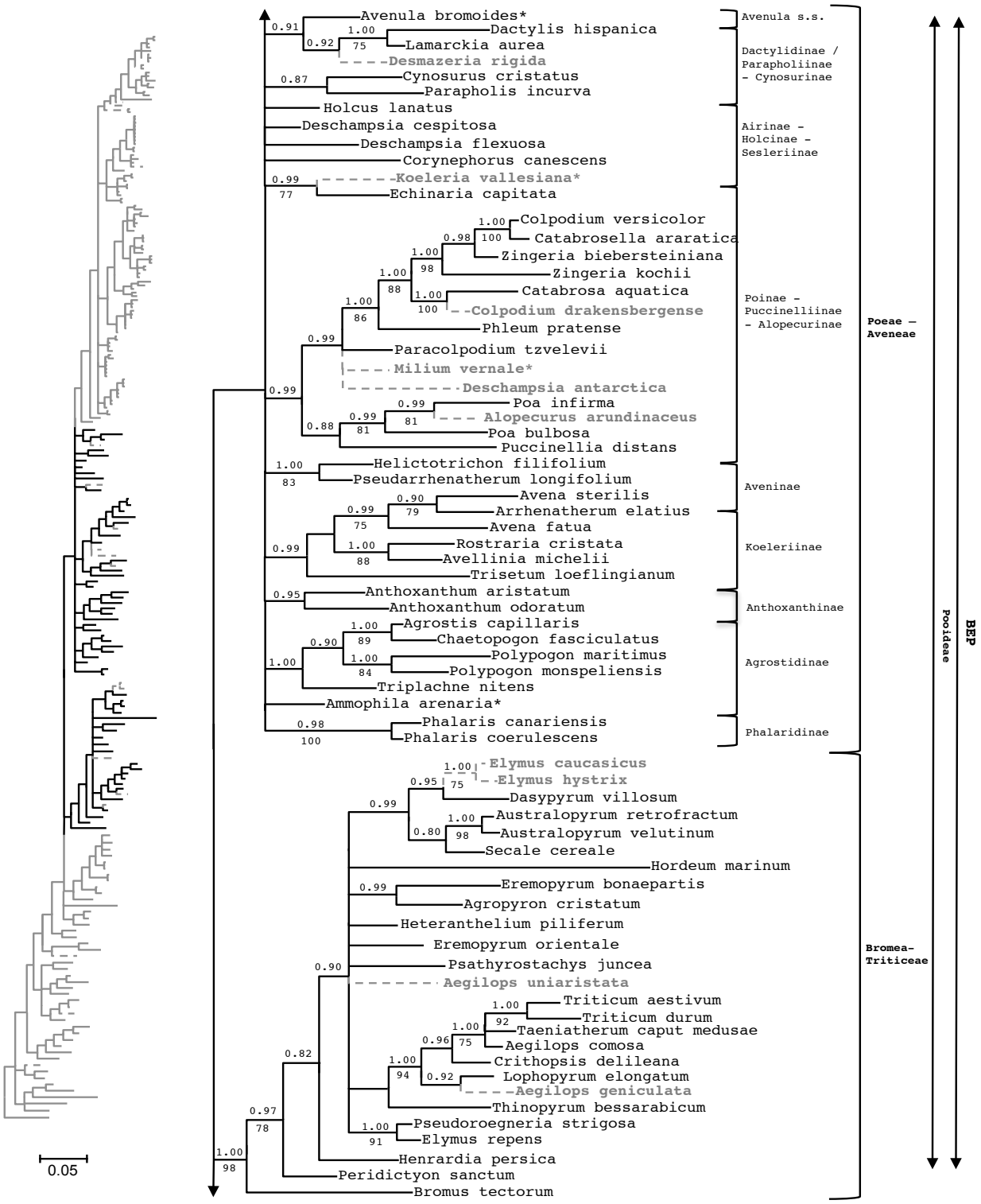


Figure S1. *(Continued)*

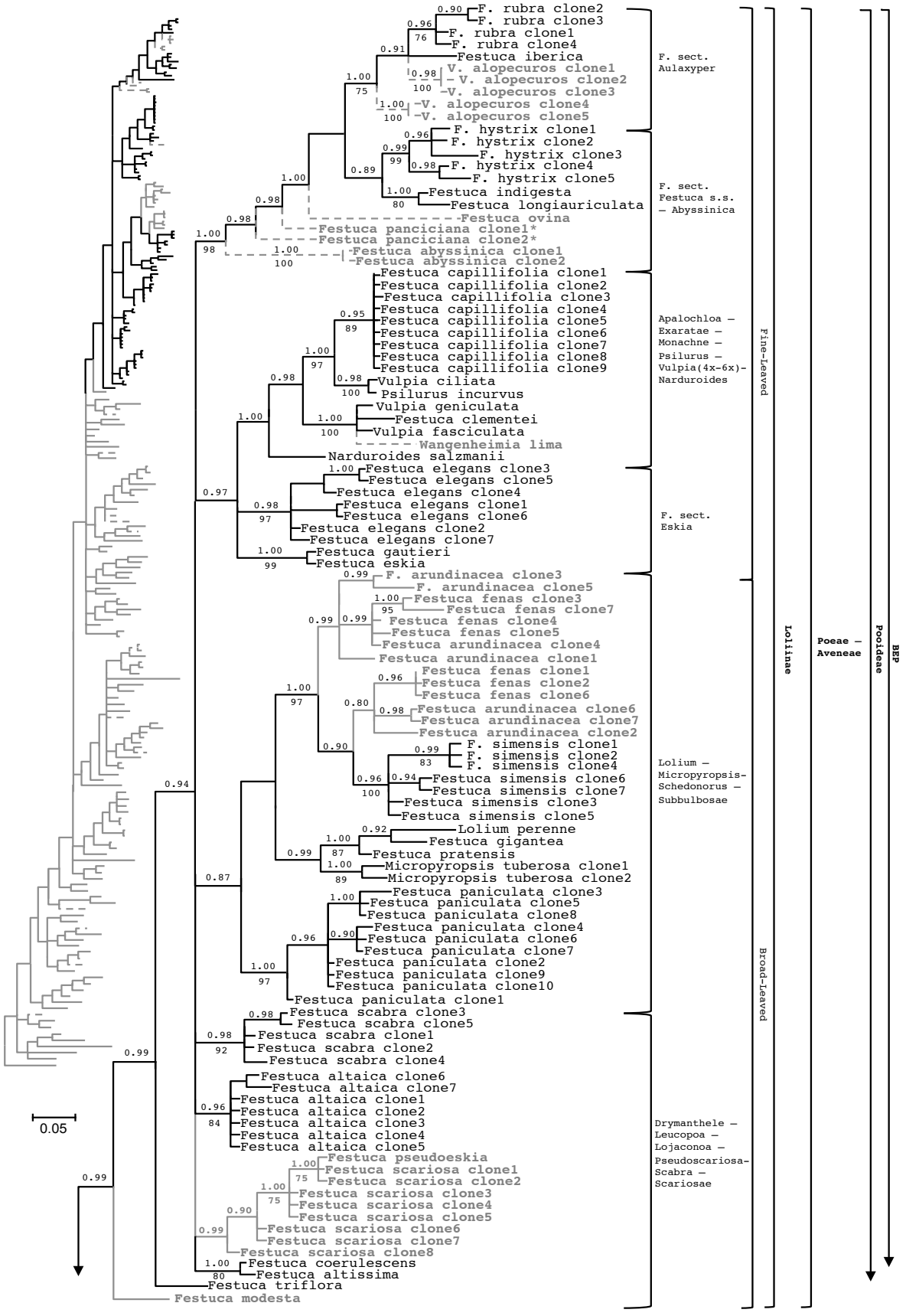


Figure S1. *(Continued)*

Figure S1. Bayesian phylogenetic inference of the Poaceae family based on the *ß-amylase* exon data set (Table 1, *ß-amylase* data set 1; 588 nucleotide positions, 142 species, 223 sequences). Bayesian posterior probabilities ≥ 90% and bootstrap support ≥ 75% are shown above and below the branches, respectively. Species assignment to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure. Dashed boxes and bold gray species indicate recombination events detected in the exon-only *ß-amylase* data set (Table 1, *ß-amylase* data set 2; 588 nucleotide positions, 142 species, 146 sequences) by two or more methods implemented in the RDP4 program Rec. n: indicates the number of recombination events (Table 3). Bold gray species and gray branches indicate that these accessions were detected with a high proportion of neutral, positive selected sites, and a complete absence of purifying selective positions, based on BS-REL method (Table 4; Fig. 2). Bold gray species and gray dotted branches indicate that these accessions are misplaced when assessed in comparison with their placement in the Poaceae organismal tree (Fig. 1). Asterisks (*) indicate placements different in the phylogeny based on the *ß*-amylase exon+intron data set (Table 1, *ß-amylase* data set 4; 850 nucleotide positions, 142 species, 146 sequences; Fig. 2) versus the exon-only data set (Table 1, *ß-amylase* data set 1; 588 nucleotide positions, 142 species, 223 sequences; this figure).

CHAPTER II

DISTRIBUTION AND EVOLUTIONARY DYNAMICS OF *STOWAWAY* MINIATURE INVERTED REPEAT TRANSPOSABLE ELEMENTS (MITES) IN GRASSES

1. Abstract

- *Stowaway* Miniature Inverted repeat Transposable Elements (*Stowaway* MITEs) are a heterogeneous group of small, non-autonomous mobile elements derived from class II DNA transposons. Their occurrence and their potential footprints in the grasses was assessed within an explicit phylogenetic framework.
- An organismal tree was used to analyze the distribution and evolutionary dynamics of *Stowaway* MITEs and their potential excision footprints in the fourth intron of the *ß-amylase* gene and in other introns of several nuclear genes across the Poaceae. Megablast and discontiguous megablast searches in the Entrez nucleotide database were performed for the *ß-amylase*, *blz-1, dmc1, nuc,* and *xly* genes MITEs.
- *Stowaway* MITEs and their potential footprints were distributed in introns and intergenic spacers of many other nuclear genes throughout the BEP lineages; however, they were absent in the studied PACCMAD lineages.
- A plausible underlying dynamic of successive acquisitions and deletions of *ß-amylase Stowaway* MITEs in the temperate grasses could be explained by three alternative hypotheses: (i) a single early acquisition of a palindrome element, similar to *Tc1-Mariner*, in the fourth intron of the *ß-amylase* gene in the ancestor of the Pooideae, followed by multiple independent losses; (ii) multiple independent acquisitions of MITEs in non-related pooid lineages; or (iii) different waves of acquisition of MITEs, followed by multiple losses and horizontal transfers in the temperate grasses. This last hypothesis seems to fit best with the evidence found to date.

2. Introduction

Miniature Inverted repeat Transposable Elements (MITEs) are a heterogeneous group of small, non-autonomous mobile elements. They have between a few dozen and a few hundred base pairs, and are found in both prokaryotic and eukaryotic genomes. Numerically, they are the most abundant of all types of transposable elements in plants, with copy numbers in the

thousands (FESCHOTTE *ET AL.*, 2002A, 2003; JIANG *ET AL.*, 2004). MITEs are closely associated with euchromatic single-copy genes (LU *ET AL.*, 2012, and references therein). The close physical association between MITEs and plant genes has caused some authors to hypothesize that MITEs might play a role in gene regulation (BUREAU AND WESSLER, 1994; BUREAU *ET AL.*, 1996; MAO *ET AL.*, 2000; ZHANG *ET AL.*, 2004; WICKER *ET AL.*, 2007). Recently, KUANG *ET AL.*, (2009) revealed a novel mechanism by which MITEs could affect gene regulation via a small RNA pathway. They also identified several MITEs that have the capacity to alter the structure of the genes.

MITEs likely originated from defective remnants of a subset of much longer autonomous class II DNA transposons (FESCHOTTE AND WESSLER, 2002; FESCHOTTE *ET AL.*, 2002A,B, 2003, 2005; TURCOTTE AND BUREAU, 2002; ZHANG *ET AL.*, 2004; MACAS *ET AL.*, 2005; MENZEL *ET AL.*, 2006) that lost their transposase-encoding ORFs due to internal deletions (FLAVELL *ET AL.*, 1994). MITEs have no coding capacity, and thus their mobility depends on the activity *in trans* of transposases encoded by cognate full-length autonomous class II transposons (JIANG *ET AL.*, 2003, 2004; FESCHOTTE *ET AL.*, 2005) that recognize a common terminal inverted repeated (TIR) sequence. These specific transposases catalyze both the insertion and the excision of their associated transposons (FESCHOTTE *ET AL.*, 2002A,B; JIANG *ET AL.*, 2003; KIKUCHI *ET AL.*, 2003; NAKAZAKI *ET AL.*, 2003; YANG *ET AL.*, 2007). Thus, MITEs should be expected to excise with a certain frequency, and—as is the case in class II transposons—leave footprints upon excision. Neither substitutions nor deletions lead to compensatory changes; hence, the highly stable secondary structure of the elements may gradually be reduced (PETERSEN AND SEBERG, 2000; FESCHOTTE *ET AL.*, 2002B). Consequently, the dynamics underlying the acquisition, loss, and high copy number of MITEs are still obscure (MASON-GAMER, 2007).

Plant *Stowaway* elements of the *Mariner* family are short non-autonomous sequences derived from class II DNA transposons (WICKER *ET AL.*, 2007). These elements vary considerably in length, from 59 to 323 bp (BUREAU AND WESSLER, 1994; TURCOTTE AND BUREAU, 2002; FESCHOTTE *ET AL.*, 2002B, 2003, 2005; MACAS *ET AL.*, 2005), and can account for up to 2% of the entire genome (e.g., *Oryza*; MAO *ET AL.*, 2000). *Stowaway* MITEs are flanked by TIRs with the almost invariant 5´-CTCCTCCC…GGGAGGAG-3´pattern and, like all *Tc1-Mariner* elements, always produce a TA target site duplication (TSD) upon insertion (WICKER *ET AL.*, 2007). Class II transposable elements can increase their numbers by transposing during chromosome replication (GREENBLATT AND BRINK, 1962). Alternatively, they can exploit gap repair following excision to create an extra copy at the donor site (NASSIF

ET AL., 1994). Transposition is mediated by a transposase that recognizes the TIRs and cuts both strands at each end (WICKER *ET AL.*, 2007). An interaction between *Mariner*-like transposases and *Stowaway* MITEs' TIRs was detected in rice (FESCHOTTE *ET AL.*, 2005), giving insight into the mechanism that is likely responsible for the evolutionary dynamics of these mobile elements. Recently, YANG *ET AL.* (2009) demonstrated that the high transposition activity achieved by rice MITEs was caused by their competitive use of the transposases encoded by related autonomous class II transposons. Comparative genomic analysis of *Oryza sativa* genomes (subsp. *japonica* and *indica*) have shown evidence of several amplification bursts of MITEs at different evolutionary times, as well as their important regulatory role in gene expression and their potential influence in species diversification (LU *ET AL.*, 2012).

The distribution and long-term evolution of *Stowaway* MITEs can be traced by examining their sequences in the context of a well-known phylogeny, such as the Triticeae (MASON-GAMER, 2007; PETERSEN AND SEBERG, 2000, 2009). MASON-GAMER (2007) found up to seven non-related *Stowaway* MITEs in different Triticeae genera at the fourth intron of the *ß-amylase* gene. Four of them showed greater similarity to elements in other genes or in other Triticeae genomes than to each other. The polyphyletic distribution of the element was also confirmed by the detection of five distinct putative excision footprints and two distinct deletions of excision-flanking regions in both related and unrelated species, revealing a complex history of gains and losses of *Stowaway ß-amylase* MITEs in the wheat tribe (MASON-GAMER, 2007).

In order to further investigate the distribution and evolutionary dynamics of the *ß-amylase Stowaway* MITEs, an explicit phylogenetic framework was used to analyze the evolutionary patterns of sequence diversity and presence vs. absence of *Stowaway* MITEs—and their potential footprints—at the fourth intron of the *ß-amylase* gene in a broader representation of grass lineages, with the temperate pooids sampled more exhaustively. For this analysis, an organismal tree constructed from nuclear (*ß-amylase,* ITS) and plastid (*mat*K*, ndh*F, *trn*TL, *trn*LF) sequences, was used as reference. The *ß-amylase* marker was used after excluding sequences with likely pseudogenization, paralogy, homeology, recombination and misplacements (cf. chapter 1). Our study aimed to test whether these mobile elements and their footprints reflect the organismal phylogeny, or could have been acquired from other non-related lineages through horizontal gene transfer (HGT) events or from intragenomic mobility. Additionally, we assessed the presence versus absence of other *Stowaway* MITEs in the *blz*1, *dmc*1, *nuc* and *xyl* genes of the Triticeae (PETERSEN AND SEBERG, 2000, 2009) in the current

genomic data repositories (e. g. Entrez nucleotide database). Our goal was to compare the patterns of distribution of different MITEs in the major BEP and PACCMAD lineages of Poaceae. Particularly, we wanted to test whether the presence of *ß-amylase* MITEs in the BEP clade and their absence in the PACCMAD clade (see Results) would be congruent wit similar distributions of different MITEs from other nuclear genes, and whether this would add support to the existence of an 'evolutionary barrier' for the transference of nuclear MITEs between representatives of these two main grass lineages.

3. Material and Methods

3.1 Taxon sampling

The sample includes a broad representation of Pooideae taxa, especially of the more recently evolved Aveneae / Poeae and Bromeae / Triticeae core-pooid lineages, which show high levels of divergence and evolutionary rates (GPWG, 2001, 2012; MASON-GAMER, 2005; CATALAN, 2006; CATALAN *ET AL.,* 2006; QUINTANAR *ET AL.,* 2007). Several representatives of the Ehrhartoideae and of the PACCMAD clade were also included in the study, to determine whether the *ß-amylase Stowaway* MITEs are found in more distantly related grass lineages.

The sampling covered in 117 species representing 82 genera and 6 subfamilies (GPWG, 2001, 2012; SORENG *ET AL.,* 2007, and http://www.tropicos.org): Centothecoideae: 1 species, Panicoideae: 6, Danthonioideae: 4, Chloridoideae: 5, Ehrhartoideae: 2, and Pooideae: 100. Within Pooideae the sampling included the following tribes: Lygeeae / Nardeae: 2 species, Stipeae: 3, Diarrheneae: 1, Brachypodieae: 3, Bromeae / Triticeae: 22, and Aveneae / Poeae: 69. Amplified sequences from 16 samples were cloned (not shown in this chapter): *Brachypodium distachyon* (9 clones), *Festuca abyssinica* (2), *F. altaica* (7), *F. arundinacea* (7), *F. capillifolia* (9), *F. elegans* (7), *F. fenas* (7), *F. hystrix* (5), *F. panciciana* (2), *F. paniculata* (10), *F. rubra* (5), *F. scabra* (5), *F. scariosa* (8), *F. simensis* (7), *Micropyropsis tuberosa* (2), *Vulpia alopecuros* (5). The origins of the plants studied, the locations of voucher specimens, and the GenBank sequence accession numbers are listed in the appendix of chapter II; Table S1 (page 131).

3.2 DNA sequencing and alignment

DOYLE AND DOYLE'S (1987) CTAB method was used to isolate DNA from silica-gel-dried leaves and from fresh materials for most of the studied samples. For herbarium samples, DNA extracts were obtained using the DNAeasy® Plant Mini Kit (QIAGEN Ltd., West

Crawley, UK) procedure. A 2370-nucleotide region (see Fig. S1 in the appendix of chapter II; page 130), extending from exons 2 to 5 of the low-copy nuclear *ß-amylase* gene was amplified and sequenced using the primers 2a-for, 2b-for, 2c-for, 2d-for, 2g-for, 3a-for, 3a-bac, 4a-bac, 4b-for, 5a-bac and 5b-bac, following the procedures indicated in MASON-GAMER (2005). Procedures for DNA amplification, cloning and sequencing are indicated in the appendix of chapter II (page 126).

Alignment of the *ß-amylase* sequences was done with Protein Multiple Sequence Alignment (MUSCLE) Software version 3.7 (EDGAR, 2004). The alignments were manually adjusted using the program MacClade 4.08 OS X (MADDISON AND MADDISON, 2008). Amino acid translations were used to guide the nucleotide alignments. Alignments of the fourth *ß-amylase* intron from the newly generated grass sequences were inspected visually and compared with some previously published *ß-amylase* sequences in *Triticeae,* in which a collection of *Stowaway* MITEs had been found (MASON-GAMER, 2007). Following alignment, a concatenated data matrix of sequences from the six genes yielded 5,485 nucleotide positions after removal of the highly variable or homoplasious positions (including the *ß-amylase Stowaway* elements).

3.3 Organismal tree reconstruction

Phylogenetic distribution of the presence and absence of *Stowaway ß-amylase* transposable elements and their target site footprints was assessed using a reference multigenic organismal tree of the Poaceae. The Poaceae organismal tree was constructed using the nuclear single-copy *ß-amylase* gene, the multicopy ribosomal *ITS*1-5.8S-*ITS*2 region, and the plastid *mat*K, *ndh*F, *trn*TL and *trn*LF genes from a collection of 116 selected grass taxa plus the close relative *Joinvillea* (Joinvillaceae) included as an out-group (see the appendix of chapter II; Table S1 (page 131). The ITS, *trn*TL and *trn*LF regions were amplified and sequenced using the primers and procedures from CATALAN *ET AL*. (2004) and QUINTANAR *ET AL*. (2007), the *ndh*F gene followed CATALAN *ET AL*. (1997), and the *mat*K DÖRING *ET AL*. (2007). The potential combinability of the data sets was assessed using the Partition Homogeneity (PH) test (Incongruence Length Difference of FARRIS *ET AL*., 1994), implemented in PAUP* 4.0 beta 10 (SWOFFORD, 2002). The PH test was conducted through heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON to estimate if the separate and combined nuclear and plastid data sets were significantly different from random partitions of the same sizes. Bayesian and Maximum parsimony (MP) based searches were performed for both the

independent and the combined data sets using, respectively, MrBayes 3.2.2 (ALTEKAR *ET AL.*, 2004; RONQUIST *ET AL.*, 2011) and PAUP*. All gaps were treated as missing data. Procedures for the Bayesian and parsimony-based searches are indicated in the appendix of chapter II (page 126).

3.4 Distribution and dynamics of the *Stowaway* MITEs

Megablast and discontiguous Megablast searches were performed in the Entrez nucleotide database using as queries the *Stowaway*-like sequences found in the *ß-amylase* gene (MASON-GAMER, 2005, 2007, and current study). The searches aimed to target other potentially homologous *Stowaway* sequences present in other grass representatives or in other genes, and to retrieve information on their respective genomic locations. Similar searches were performed using as query sequences the *Stowaway*-like elements described by PETERSEN AND SEBERG (2009) in the *blz1, dmc1, nuc*, and *xyl* genes of the Triticeae in order to determine whether other described TEs show similar inter-tribal distribution patterns in the grass family as those found in the fourth intron of the *ß-amylase* gene (see results). In all cases we used a query coverage > 80%, threshold E-values > 1e-08, and sequence similarity values > 69% to compare their potential homologies. Searches were performed both including and excluding the TSD and the conserved TIR. The specific *Stowaway*-bearing genes were examined visually at each specific locus. The secondary structure of these elements was inspected through their minimum-energy folding values using MFOLD (ZUKER, 2003) and their free energy values following SANTALUCIA (1998).

4. Results

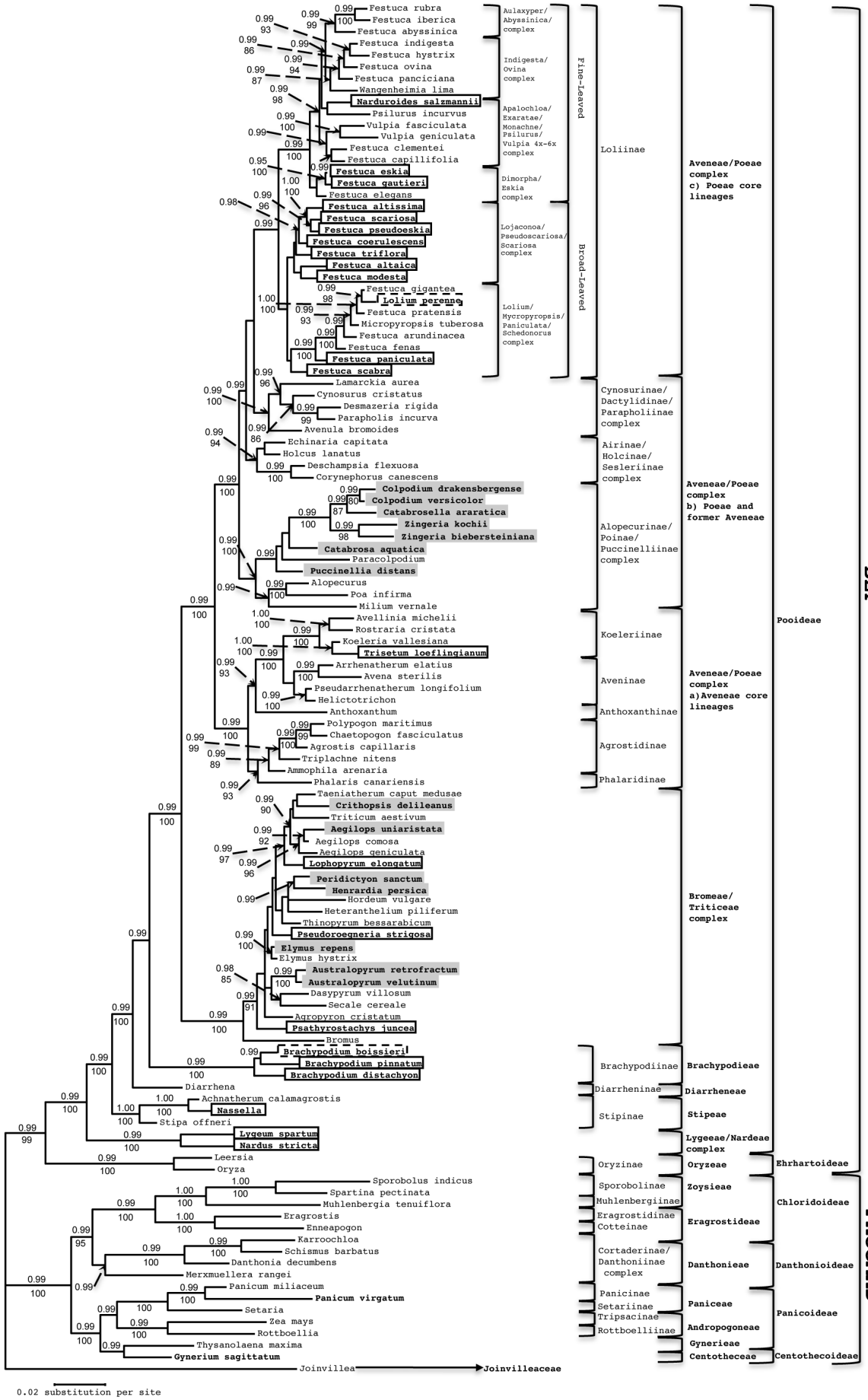
4.1 The organismal tree of grasses

Evolutionary dynamics of the mobile *Stowaway* MITEs could only be confidently inferred within a solid phylogenetic framework and from their remaining conserved termini (MASON-GAMER, 2007; PETERSEN AND SEBERG, 2009). The Poaceae organismal tree (Fig. 1) was constructed by combining sequence data from the nuclear *ß-amylase* (117 taxa / 1143 aligned nucleotides) and ITS (112 / 609) genes, and the plastid *mat*K (99/ 1259), *ndh*F (102 / 710), *trn*TL (98 / 789) and *trn*LF (102 / 975) genes.

The independent phylogenetic searches conducted with the six separate *ß-amylase*, ITS, *mat*K, *ndh*F, *trn*TL and *trn*LF data sets yielded highly congruent topologies (results not shown except the *ß-amylase* tree; Fig. 2). The PH test detected incongruence among the cpDNA data

sets, except for *mat*K – *trn*TL (P= 0.4) and *ndh*F – *trn*TL (P= 0.3). Partition incongruence was also found between the nuclear *ß-amylase* and ITS data (P= 0.01), and between the plastid and nuclear data (P= 0.01). Despite this, the separate topologies (e.g. *ß-amylase* tree; Fig. 2) were congruent with the tree based on the six concatenated nuclear and plastid genes (Fig. 1), with incongruence observed mostly among some terminal tips. Furthermore, those topological conflicts were not well supported. Since the incongruence did not affect the main results and conclusions of this research, we chose to conduct phylogenetic analyses on the combined concatenated data set of the six sequenced genes, which had a total length of 5571 characters, of which 2274 (40.81%) were parsimony informative. The BI tree (Fig. 1) and MP tree (not shown) from the combined data set are highly congruent with one another and with those obtained for the family by previous authors.

Figure 1. Phylogenetic distribution of the *ß-amylase Stowaway* MITEs and their potential footprints in the Poaceae organismal tree. Bayesian Inference estimation was based on nuclear (ITS, *ß-amylase*) and plastid (*trn*TL, *trn*LF, *ndh*F, *mat*K) gene sequences. Bayesian posterior probabilities of ≥ 95% are shown above the nodes, and MP bootstrap support of ≥ 75% is shown below the nodes. Solid line boxes indicate species with putative degraded excision footprints within the two conserved TIR sequences of 10-bp. Grey boxes indicate species with *Stowaway* MITE footprints among the TA---GTA motifs. Dotted line boxes indicate species lacking the TA---GTA motifs along with portions of the flanking intron. Black boldfaced species in the PACCMAD clade have different kinds of insertions in the same position.

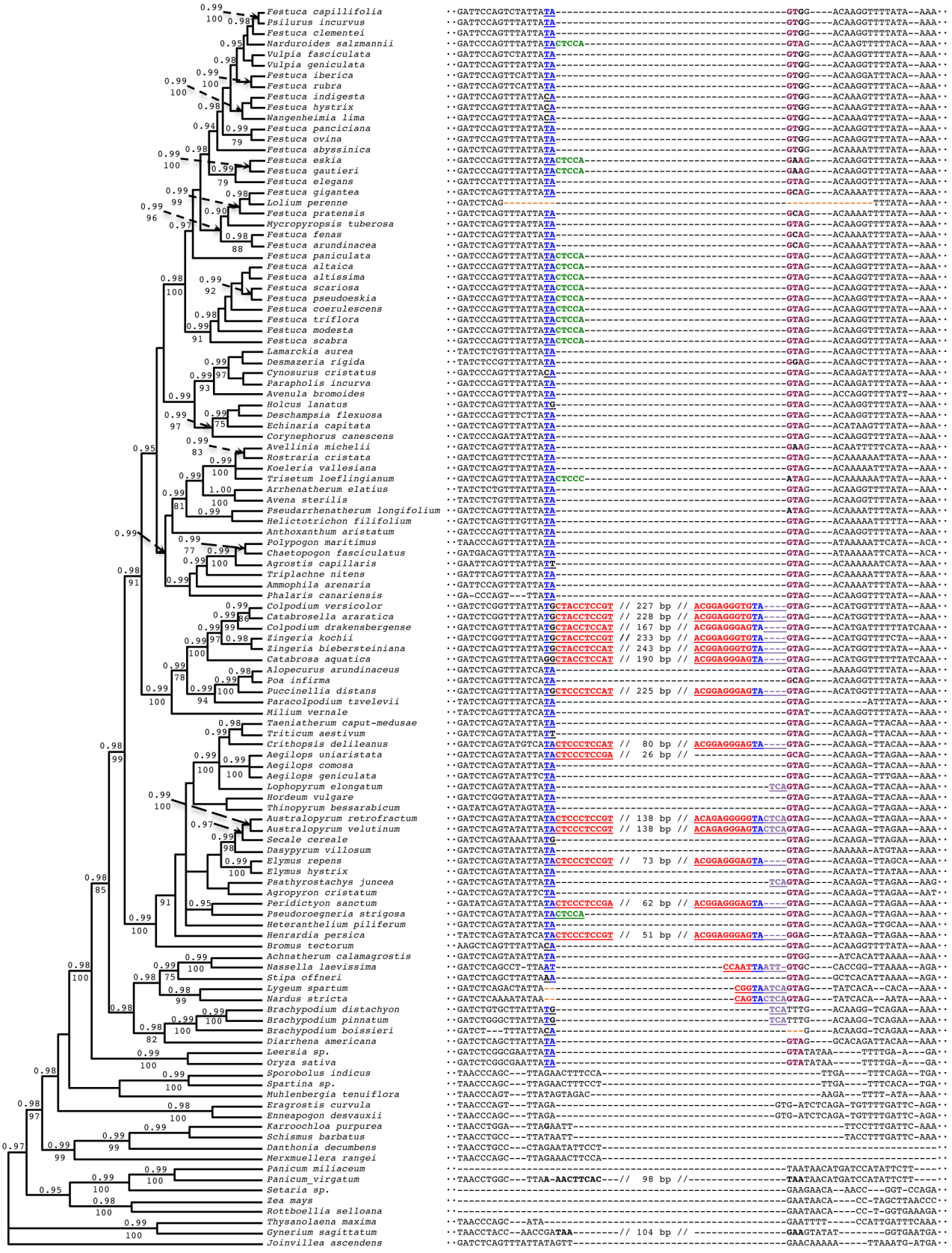

Festuca rubra
Festuca iberica
Festuca abyssinica
Festuca indigesta
Festuca hystrix
Festuca ovina
Festuca panciciana
Wangenheimia lima
Narduroides salzmannii
Psilurus incurvus
Vulpia fasciculata
Vulpia geniculata
Festuca clementei
Festuca capillifolia
Festuca eskia
Festuca gautieri
Festuca elegans
Festuca altissima
Festuca scariosa
Festuca pseudoeskia
Festuca coerulescens
Festuca triflora
Festuca altaica
Festuca modesta
Festuca gigantea
Lolium perenne
Festuca pratensis
Micropyropsis tuberosa
Festuca arundinacea
Festuca fenas
Festuca paniculata
Festuca scabra
Lamarckia aurea
Cynosurus cristatus
Desmazeria rigida
Parapholis incurva
Avenula bromoides
Echinaria capitata
Holcus lanatus
Deschampsia flexuosa
Corynephorus canescens
Colpodium drakensbergense
Colpodium versicolor
Catabrosella araratica
Zingeria kochii
Zingeria biebersteiniana
Catabrosa aquatica
Paracolpodium
Puccinellia distans
Alopecurus
Poa infirma
Milium vernale
Avellinia michelii
Rostraria cristata
Koeleria vallesiana
Trisetum loeflingianum
Arrhenatherum elatius
Avena sterilis
Pseudarrhenatherum longifolium
Helictotrichon
Anthoxanthum
Polypogon maritimus
Chaetopogon fasciculatus
Agrostis capillaris
Triplachne nitens
Ammophila arenaria
Phalaris canariensis
Taeniatherum caput medusae
Crithopsis delileanus
Triticum aestivum
Aegilops uniaristata
Aegilops comosa
Aegilops geniculata
Lophopyrum elongatum
Peridictyon sanctum
Henrardia persica
Hordeum vulgare
Heteranthelium piliferum
Thinopyrum bessarabicum
Pseudoroegneria strigosa
Elymus repens
Elymus hystrix
Australopyrum retrofractum
Australopyrum velutinum
Dasypyrum villosum
Secale cereale
Agropyron cristatum
Psathyrostachys juncea
Bromus
Brachypodium boissieri
Brachypodium pinnatum
Brachypodium distachyon
Diarrhena
Achnatherum calamagrostis
Nassella
Stipa offneri
Lygeum spartum
Nardus stricta
Leersia
Oryza
Sporobolus indicus
Spartina pectinata
Muhlenbergia tenuiflora
Eragrostis
Enneapogon
Karroochloa
Schismus barbatus
Danthonia decumbens
Merxmuellera rangei
Panicum miliaceum
Panicum virgatum
Setaria
Zea mays
Rottboellia
Thysanolaena maxima
Gynerium sagittatum
Joinvillea
Joinvilleaceae
0.02 substitution per site
Aulaxyper/ Abyssinica/ complex
Indigesta/ Ovina complex
Apalochloa/ Exaratae/ Monachne/ Psilurus/ Vulpia 4x-6x complex
Dimorpha/ Eskia complex
Lojaconoa/ Pseudoscariosa/ Scariosa complex
Lolium/ Mycropyropsis/ Paniculata/ Schedonorus complex
Fine-Leaved
Broad-Leaved
Loliinae
Cynosurinae/ Dactylidinae/ Parapholiinae complex
Airinae/ Holcinae/ Sesleriinae complex
Alopecurinae/ Poinae/ Puccinelliinae complex
Koeleriinae
Aveninae
Anthoxanthinae
Agrostidinae
Phalaridinae
Brachypodiinae
Diarrheninae
Stipinae
Oryzinae
Sporobolinae
Muhlenbergiinae
Eragrostidinae
Cotteinae
Cortaderinae/ Danthoniinae complex
Panicinae
Setariinae
Tripsacinae
Rottboelliinae
Aveneae/Poeae complex c) Poeae core lineages
Aveneae/Poeae complex b) Poeae and former Aveneae
Aveneae/Poeae complex a)Aveneae core lineages
Bromeae/ Triticeae complex
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae complex
Oryzeae
Zoysieae
Eragrostideae
Danthonieae
Paniceae
Andropogoneae
Gynerieae
Centotheceae
Pooideae
Ehrhartoideae
Chloridoideae
Danthonioideae
Panicoideae
Centothecoideae
BEP
PACCMAD

4.2 *ß-amylase Stowaway* MITEs in the Poaceae

According to WICKER'S *ET AL*. (2007) classification, the insertions found at the fourth intron of the *ß-amylase* gene in the BEP clade of the Poaceae were classified in the class II (DNA transposon) - subclass 1 -, TIR order, *Tc1-Mariner* superfamily as non-autonomous *Stowaway* MITEs (but see also SEBERG AND PETERSEN (2009) and WICKER *ET AL*. (2009) for a discussion on the potential classification system of transposable elements) based on: i) their two conserved TIR sequences of 10-bp (5´-CTMCCTCCRW…ACRGAGGGWG-3´), and the characteristic 2-bp TSD "TA" (Fig. 2); ii) their high similarity (69) 80 - 100%) to other MITEs deposited in the Entrez nucleotide database (overall > 80% Blast matches) (Table 1); and iii) their hairpin-like secondary structure (Fig. 3). The target site 5'-TA-3' and the 5'-GTA-3' motif were found in all the 100 species sampled within the BEP clade. *Stowaway* MITEs and other potential excision footprints within the 3´-TA---GTA-5 region were found in 37 of the 100 species sampled within the BEP clade (Fig. 1, 2; Table 1). Within the PACCMAD clade, 2 out of the 16 surveyed species showed a different kind of insertion in the same position where the BEP-clade *Stowaway* MITEs resided (e.g. *Gynerium sagittatum* and *Panicum virgatum*, insertions of 104 bp and 98 bp, respectively; Fig. 2). However, the absence of typical *Stowaway* features such as the flanked 2-bp target site preference (TA), and similar TIR sequences, ruled out the possibility that these insertions could be the remnants of *Stowaway* MITE footprints.

In a previous study, MASON-GAMER (2007) described a collection of three main *Stowaway* MITE types that were found within the fourth intron of the *ß-amylase* gene in several Triticeae species (Fig. 1, 2; Table 1). Two *Australopyrum* species, *A. retrofractum* AY821692, and *A. velutinum* AY821693.1, showed the same 138 bp MITE. A second type was shared by the partially degraded *A. uniaristata* AY821691.1; 26 bp (94.7% identity in the 5'-end region) and *Peridictyon sanctum* AY821714.1, 62 bp (89% identity with the former). The third type was shared by *Crithopsis delileana* AY821694.1, 80 bp, and *Henrardia persica* AY821703.1, 51 bp (94.4% identity with the former).

Figure 2. *ß-amylase* tree and alignment of the fourth intron region of the gene containing *Stowaway* elements and their footprints in the studied Poaceae species. Halfcompatible bayesian consensus heuristic tree based on a reduced unambiguous *ß-amylase* data set (see text for details of the analyses); values above and below branches represent PPS and BS values, respectively. The data set shows: i) the 2-bp target site (TA) underlined in blue boldface; ii) the 10 bp TIR sequence underlined in red boldface; iii) the likely degraded excision footprints adjacent to the 5' flanking regions of the TIR sequence underlined in green boldface; iv) the likely degraded excision footprints adjacent to the 3' flanking regions of the TIR sequence underlined in purple boldface; v) deletions including the target site in red boldface; vi) the 3-bp (GTA) motif downstream from the *Stowaway* elements (common in the BEP clade) in brown boldface, and vii) the lengths of the internal portions of the inserted elements inside double slashes.


Festuca capillifolia
Psilurus incurvus
Festuca clementei
Narduroides salzmannii
Vulpia fasciculata
Vulpia geniculata
Festuca iberica
Festuca rubra
Festuca indigesta
Festuca hystrix
Wangenheimia lima
Festuca panciciana
Festuca ovina
Festuca abyssinica
Festuca eskia
Festuca gautieri
Festuca elegans
Festuca gigantea
Lolium perenne
Festuca pratensis
Mycropyropsis tuberosa
Festuca fenas
Festuca arundinacea
Festuca paniculata
Festuca altaica
Festuca altissima
Festuca scariosa
Festuca pseudoeskia
Festuca coerulescens
Festuca triflora
Festuca modesta
Festuca scabra
Lamarckia aurea
Desmazeria rigida
Cynosurus cristatus
Parapholis incurva
Avenula bromoides
Holcus lanatus
Deschampsia flexuosa
Echinaria capitata
Corynephorus canescens
Avellinia michelii
Rostraria cristata
Koeleria vallesiana
Trisetum loeflingianum
Arrhenatherum elatius
Avena sterilis
Pseudarrhenatherum longifolium
Helictotrichon filifolium
Anthoxanthum aristatum
Polypogon maritimus
Chaetopogon fasciculatus
Agrostis capillaris
Triplachne nitens
Ammophila arenaria
Phalaris canariensis
Colpodium versicolor
Catabrosella araratica
Colpodium drakensbergense
Zingeria kochii
Zingeria biebersteiniana
Catabrosa aquatica
Alopecurus arundinaceus
Poa infirma
Puccinellia distans
Paracolpodium tzvelevii
Milium vernale
Taeniatherum caput-medusae
Triticum aestivum
Crithopsis delileanus
Aegilops uniaristata
Aegilops comosa
Aegilops geniculata
Lophopyrum elongatum
Hordeum vulgare
Thinopyrum bessarabicum
Australopyrum retrofractum
Australopyrum velutinum
Secale cereale
Dasypyrum villosum
Elymus repens
Elymus hystrix
Psathyrostachys juncea
Agropyron cristatum
Peridictyon sanctum
Pseudoroegneria strigosa
Heteranthelium piliferum
Henrardia persica
Bromus tectorum
Achnatherum calamagrostis
Nassella laevissima
Stipa offneri
Lygeum spartum
Nardus stricta
Brachypodium distachyon
Brachypodium pinnatum
Brachypodium boissieri
Diarrhena americana
Leersia sp.
Oryza sativa
Sporobolus indicus
Spartina sp.
Muhlenbergia tenuiflora
Eragrostis curvula
Enneapogon desvauxii
Karroochloa purpurea
Schismus barbatus
Danthonia decumbens
Merxmuellera rangei
Panicum miliaceum
Panicum_virgatum
Setaria sp.
Zea mays
Rottboellia selloana
Thysanolaena maxima
Gynerium sagittatum
Joinvillea ascendens

Most of the *ß-amylase Stowaway* MITEs found by MASON-GAMER (2007) formed almost perfect hairpin-like secondary structures that had very low free energies (Fig. 3; type A). Our analyses indicated that the different MITEs could be classified into two main types of *ß-amylase* MITEs within the Triticeae. The Blast searches performed in the Entrez nucleotide database showed that the identical insertion of *Australopyrum retrofractum* and *A. velutinum* (138 bp; dG=-67.44) was also present in other genes of other Triticeae species (e.g. 1st intron of the *HvPKABA1* gene, 6th intron of the *p5cdh* gene and an intergenic spacer in *Hordeum vulgare*; 3rd intron of the *vp1D* gene and an intergenic spacer in *Triticum aestivum*; 1st intron the *phosphate synthase II* gene in *Triticum urartu*; intergenic spacers of *Aegilops tauschii, Triticum monococcum* and *T. turgidum*; Table 1). Apart from the Triticeae, similar elements were also found in other genes of Poeae s.l. taxa (e.g. 2nd intron of the *fl* gene in *Avena agadiriana, A. hirtula, A. insularis,* and *A. lusitanica*). However, the Blast searches did not find similarities among these MITEs and those found at the same *ß-amylase* site in other Triticeae taxa. Conversely, the insertions found at the fourth intron of the *ß-amylase* gene in *Aegilops uniaristata* (26 bp; dG=-4.53), *Crithopsis delileanus* (80 bp; dG=-36.78), *Elymus repens* (73 bp; dG=-30.31), *Henrardia persica* (51 bp; dG=-40.30) and *Peridictyon sanctum* (62 bp; dG= -44.90) were similar to each other (Table 1) but not to the *Australopyrum* ones. They were respectively similar to sequences found at other loci in other Triticeae species (e.g., 4th intron of the *sey1* gene in *Secale cereale*; intergenic spacer sequence in *Hordeum vulgare* and *Triticum aestivum*) and Brachypodieae (e.g., 2nd intron of the ribosomal subunit 8E gene in *Brachypodium sylvatium*; intergenic spacer in *B. distachyon*) taxa.

Table 1. Megablast searches for Stowaway-like *ß-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

***ß-amylase Stowaway* MITEs found in representatives of Alopecurinae/Poinae/Puccinelliinae**

Catabrosa aquatica (*ß-amylase* intron; 4, 212 bp; range 1645-1818; HE565906)
Catabrosella araratica (*ß-amylase* intron; 4, 249 bp; range 1532-1822; HE565912)
Colpodium drakensbergense (*ß-amylase* intron; 4, 436 bp; range 1639-1820; HE565915)
Colpodium versicolor (*ß-amylase* intron; 4, 249 bp; range 1532-1822; HE565911)
Puccinellia distans (*ß-amylase* intron; 4, 245 bp; range 1647-2120; JX536564)
Zingeria biebersteiniana (*ß-amylase* intron; 4, 265 bp; range 1532-1839; HE565930)
Zingeria kochii (*ß-amylase* intron; 4, 263 bp; range 1531-1836; HE565927)

Genome	Location, intron	Gen Bank Accession, Range	Query coverage >80%	E-value	Sequence similarity
Ehrhartoideae: Oryzeae					
Oryza glaberrima	*Ogl12g0016G16_1* gene, 5	FQ377947.1, 220949-221207	- / 86% / - / 86% / - / - / -	- / 2e-12 / - / 2e-12 / - / - / -	- / 69% / - / 69% / - / - / -
Oryza minuta	Intergenic spacer	AC232150.1, 41198-41464	- / 87% / - / 87% / - / - / -	- / 4e-20 / - / 4e-20 / - / - / -	- / 71% / - / 71% / - / - / -
Oryza sativa	Intergenic spacer	AC123528.2, 78005-78266	- / 86% / - / 86% / - / - / 83%	- / 4e-14 / - / 4e-14 / - / - / 2e-18	- / 69% / - / 69% / - / - / 70%

Table 1. Megablast searches for Stowaway-like *ß-amylase, blz-1*, *dmc1*, *nuc,* and *xly* elements

***ß-amylase Stowaway* MITEs found in representatives of Triticeae**

Aegilops comosa (*ß-amylase* intron; 4, 84 bp; range 1205-1334; AY821690.1)
Aegilops uniaristata (*ß-amylase* intron; 4, 35 bp; range 1197-1274; AY821691.1)
Australopyrum retrofractum (*ß-amylase* intron; 4, 165 bp; range 1207-1412; AY821692)
Australopyrum velutinum (*ß-amylase* intron; 4, 165 bp; range 1647-1895; AY821693.1)
Crithopsis delileana (*ß-amylase* intron; 4, 105 bp; range 1647-1777; AY821694.1)
Elymus repens (*ß-amylase* intron; 4, 95 bp; range 1188-1324; EU282245.1)
Henrardia persica (*ß-amylase* intron; 4, 74 bp; range 1647-1777; AY821703.1)
Lophopyrum elongatum (*ß-amylase* intron; 4, 47 bp; range 1647-1728; AY821731.1)
Peridictyon sanctum (*ß-amylase* intron; 4, 84 bp; range 1647-1777; AY821714.1)

Genome	Location, intron	Gen Bank Accession, Range	Query coverage >80%	E-value	Sequence similarity
Aveneae					
Avena agadiriana	*Floricaula/leafy* gene (FL), 2	GQ979788.1, 882-718	- / - / 80% / 84% / - / - / - / - / - / - / -	/ - / - / 6e-47 / 2e-48 / - / - / - / - / - / - / -	- / - / 87% / 88% / - / - / - / - / - / - / -
Avena hirtula	*Floricaula/leafy* gene (FL), 2	GQ979745.1, 715-879	- / - / 80% / 84% / - / - / - / - / - / - / -	- / - / 2e-48 / 1e-49 / - / - / - / - / - / - / -	- / - / 88% / 89% / - / - / - / - / - / - / -
Avena insularis	*Floricaula/leafy* gene (FL), 2	GQ979809.1, 715-879	- / - / 80% / 84% / - / - / - / - / - / - / -	- / - / 6e-47 / 2e-48 / - / - / - / - / - / - / -	- / - / 87% / 88% / - / - / - / - / - / - / -
Avena lusitanica	*Floricaula/leafy* gene (FL), 2	GQ979764.1, 709-873	- / - / 80% / 84% / - / - / - / - / - / - / -	- / - / 2e-48 / 1e-49 / - / - / - / - / - / - / -	- / - / 88% / 89% / - / - / - / - / - / - / -
Triticeae					
Aegilops comosa	*ß-amylase* gene, 4	AY821690, 1209-1334	- / 98% / - / - / - / - / - / 100% / 99% / -	- / 9e-13 / - / - / - / - / - / 2e-55 / 2e-12 / -	- / 96% / - / - / - / - / - / 100% / 75% / -
Aegilops tauschii	Intergenic spacer	AY534123, 69953-70071	- / - / 81% / 81% / - / - / - / - / - / - / -	- / - / 1e-49 / 8e-46 / - / - / - / - / - / - / -	- / - / 89% / 87% / - / - / - / - / - / - / -
Hordeum vulgare	*Protein kinase HvPKABA1* gene, 1	AB058924.1, 7401-7573	- / - / 83% / 83% / - / - / - / - / - / - / -	- / - / 8e-46 / 1e-49 / - / - / - / - / - / - / -	- / - / 85% / 87% / - / - / - / - / - / - / -
Hordeum vulgare	Intergenic spacer	AY853252, 121911-122077	- / - / 85% / 92% / - / - / 86% / - / - / 86%	- / - / 3e-44 / 2e-48 / - / - / 2e-26 / - / - / 1e-08	- / - / 96% / 88% / - / - / 89% / - / - / 77%
Hordeum vulgare	*Delta-1-pyrroline-5carboxylate dehydrogenase* (p5cdh) gene, 6	DQ154923, 7412-7576	- / - / 81% / 81% / - / - / - / - / - / - / - /	- / - / 2e-52 / 5e-55 / - / - / - / - / - / - /	- / - / 88% / 89% / - / - / - / - / - / - / -
Secale cereale	*Protein SEY1* gene, 4	FJ374546.1, 463-559	85% / - / - / - / - / - / - / 92% / - / -	0.040 / - / - / - / - / - / - / 2e-06 / - / -	71% / - / - / - / - / - / - / 77% / - / -
Triticum aestivum	*Viviparous1* gene (vp1D), 3	AJ400714.1, 2782-2967	- / - / 90% / 90% / - / - / - / - / - / - / -	- / - / 6e-47 / 3e-51 / - / - / - / - / - / - / -	- / - / 83% / 85% / - / - / - / - / - / - / -
Triticum aestivum	Intergenic spacer	FN645450, 1073786-1073951	86% / - / 91% / 91% / - / - / 89% / - / - / -	0.001 / - / 2e-48 / 8e-52 / - / - 5e-33 / - / - / -	84% / - / 97% / 97% / - / - / 100% / - / - / -
Triticum monococcum	Intergenic spacer	AY951945, 23112-23276	- / - / 85% / 85% / - / - / - / - / - / - / -	- / - / 2e-48 / 3e-51 / - / - / - / - / - / - /	- / - / 87% / 89% / - / - / - / - / - / - / -
Triticum turgidum	Intergenic spacer	DQ871219.1, 173020-172854	- / - / 85% / 83% / - / - / - / - / - / - / -	- / - / 6e-47 / 1e-49 / - / - / - / - / - / - / -	- / - / 87% / 94% / - / - / - / - / - / - / -

Table 1. Megablast searches for Stowaway-like *ß-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

***ß-amylase Stowaway* MITEs found in representatives of Triticeae**

Aegilops comosa (*ß-amylase* intron; 4, 84 bp; range 1205-1334; AY821690.1)
Aegilops uniaristata (*ß-amylase* intron; 4, 35 bp; range 1197-1274; AY821691.1)
Australopyrum retrofractum (*ß-amylase* intron; 4, 165 bp; range 1207-1412; AY821692)
Australopyrum velutinum (*ß-amylase* intron; 4, 165 bp; range 1647-1895; AY821693.1)
Crithopsis delileana (*ß-amylase* intron; 4, 105 bp; range 1647-1777; AY821694.1)
Elymus repens (*ß-amylase* intron; 4, 95 bp; range 1188-1324; EU282245.1)
Henrardia persica (*ß-amylase* intron; 4, 74 bp; range 1647-1777; AY821703.1)
Lophopyrum elongatum (*ß-amylase* intron; 4, 47 bp; range 1647-1728; AY821731.1)
Peridictyon sanctum (*ß-amylase* intron; 4, 84 bp; range 1647-1777; AY821714.1)

Genome	Location, intron	Gen Bank Accession, Range	Query coverage >80%	E-value	Sequence similarity
Triticeae					
Triticum urartu	*Phosphate synthase II* gene, 1	GU797179.1, 495-663	- / - / 84% / 84% / - / - / - / - / - / - / -	- / - / 3e-57 / 1e-61 / - / - / - / - / - / - / -	- / - / 90% / 91% / - / - / - / - / - / - / -
Brachypodieae					
Brachypodium distachyon	*Intergenic spacer*	EU730895.1, 40543-40628	81% / - / - / - / - / - / - / - / - / - / 86%	8e-05 / - / - / - / - / - / - / - / - / - / 7e-12	79% / - / - / - / - / - / - / - / - / - / 80%
Brachypodium sylvaticum	*Ribosomal subunit 8E* gene, 2	EF059989.1, 239308-239420	- / - / - / - / - / - / - / - / - / - / 86%	- / - / - / - / - / - / - / - /- / - / 5e-07	- / - / - / - / - / - / - / - / - / - / 87%

Table 1. Megablast searches for Stowaway-like *ß-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

***Xylose isomerase Stowaway* MITE found in representatives of Triticeae and other Pooideae**		
Triticeae:		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Hordeum bogdanii	*Xylose isomerase* (xly) gene, 1	EU862645.1, 174-262
Hordeum brachyantherum	*Xylose isomerase* (xly) gene, 1	EU862641.1, 171-259
Hordeum brevisubulatum	*Xylose isomerase* (xly) gene, 1	EU862625.1, 171-259
Hordeum bulbosum	*Xylose isomerase* (xly) gene, 1	EU862644.1, 164-249
Hordeum chilense	*Xylose isomerase* (xly) gene, 1	EU862639.1, 172-260
Hordeum comosum	*Xylose isomerase* (xly) gene, 1	EU862631.1, 174-262
Hordeum cordobense	*Xylose isomerase* (xly) gene, 1	EU862647.1, 175-244
Hordeum erectifolium	*Xylose isomerase* (xly) gene, 1	EU862630.1, 174-250
Hordeum euclaston	*Xylose isomerase* (xly) gene, 1	EU862632.1, 174-262
Hordeum flexuosum	*Xylose isomerase* (xly) gene, 1	EU862629.1, 174-262
Hordeum intercedens	*Xylose isomerase* (xly) gene, 1	EU862640.1, 167-255
Hordeum marinum	*Xylose isomerase* (xly) gene, 1	EU862624, 171-259
Hordeum murinum	*Xylose isomerase* (xly) gene, 1	EU862627.1, 174-262
Hordeum muticum	*Xylose isomerase* (xly) gene, 1	EU862628.1, 171-259
Hordeum patagonicum	*Xylose isomerase* (xly) gene, 1	EU862646.1, 164-251
Hordeum pubiflorum	*Xylose isomerase* (xly) gene, 1	HQ619240.1, 174-261
Hordeum pusillum	*Xylose isomerase* (xly) gene, 1	EU862642.1, 172-260
Hordeum roshevitzii	*Xylose isomerase* (xly) gene, 1	EU862648.1, 163-251
Hordeum stenostachys	*Xylose isomerase* (xly) gene, 1	EU862638.1, 174-262
Hordeum vulgare	*Xylose isomerase* (xly) gene, 1	EU862643.1, 162-250
Aegilops speltoides	*ADP-glucose pyrophosphorylase small subunit* gene, 2	DQ145552.1, 348-488
Aegilops speltoides	*Pina gene for puroindoline*	AJ302096.1, 1389-1476
Aegilops speltoides	*Phytoene synthase* 1 (psy-B1) gene, 3	FJ393537, 2257-2220
Aegilops tauschii	*Calcium-dependent protein kinase* (cpk9) gene, 1	EU183492, 3393-3481
Aegilops tauschii	*Phosphomannomutase D1* (pmm-d1) gene, 9	GQ412271, 2077-2240
Aegilops tauschii	*Pyruvate dehydrogenase E1* (pdha1) gene, 1	GU211253, 274-353
Agropyron cristatum	*3-phosphoglycerate kinase* (pgk1) gene, 4	JF965633, 1218-1304
Agropyron mongolicum	*3-phosphoglycerate kinase* (pgk1) gene, 4	JF965627.1, 1210-1296
Dasypyrum villosum	*Serine/threonine protein kinase* (stpk-v) gene, 1	HQ864471, 724-761
Heteranthelium piliferum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	AF277239, 1065-1115

Table 1. Megablast searches for Stowaway-like *ß-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

***Xylose isomerase Stowaway* MITE found in representatives of Triticeae and other Pooideae**		
Triticeae:		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Hordeum bogdanii	*DP62 hypothetical protein* gene, 2	FJ898044.1, 619-776
Hordeum vulgare	*Putative ankyrin repeat protein* gene, 4	DQ900685.1, 40922-41189
Hordeum vulgare	*BKIN12 gene protein kinase*, 2	X65606.1, 2256-2346
Hordeum vulgare	Intergenic spacer	AJ507214.1, 4295-4376
Kengyilia grandiflumis	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965613.1, 1212-1298
Kengyilia hirsuta	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965609.1, 1242-1328
Kengyilia kokonorica	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965615.1, 1212-1296
Kengyilia longiglumis	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965611.1, 1210-1296
Kengyilia melanthera	*3-phosphoglycerate kinase* (pgk1) gene, 4	JF965614, 1211-1297
Kengyilia mutica	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965610.1, 1210-1296
Kengyilia rigidula	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965616.1, 1214-1298
Kengyilia stenachyra	*3-phosphoglycerate kinase* (Pgk1) gene, 4	JF965612.1, 1242-1328
Secale cereale	*Putative expressed protein* gene, 2	FJ374603, 1394-1479
Secale strictum	*Pyruvate dehydrogenase E1* (*pdha1*) gene, 1	GU211256.1, 1091-1249
Secale sylvestre	*Pyruvate dehydrogenase E1* (*pdha1*) gene, 1	GU211264, 1091-1249
Triticum aestivum	*KN1 homeobox protein* (*Wknox1b*) gene, 3	AB182944.1, 7704-7848
Triticum aestivum	*Putative ankyrin repeat protein* gene, 4	AB298185.1, 18766-19031
Triticum aestivum	*Acetyl-coenzyme A carboxylase* (Acc-1) gene, 2	AF029897.1, 6382-6467
Triticum aestivum	*Phytoene synthase 1* (Psy-B1) gene, 2	FJ393540.1, 1065-1155
Triticum aestivum	*Viviparous 1* (VP-1B) gene, 3	FJ643535.1, 2731-2818
Triticum aestivum	Intergenic spacer	FN564429, 184303-184447
Triticum aestivum	Intergenic spacer	AF325196, 21653-21737
Triticum aestivum	Intergnic spacer	EU660901, 97519-97677
Triticum aestivum	Intergenic spacer	HQ391120.1
Triticum aestivum	Intergenic spacer	HQ390612.1, 283-347
Triticum monococcum	Intergenic spacer	AY485644.1, 355527-355671
Triticum monococcum	*G2528 tousled-like kinase* (mtk4) gene, 1	AY244513, 778-857
Triticum monococcum	*Mitochondrial putative neurolysin precursor*	FJ394375.1, 393-452
Triticum turgidum	Intergenic spacer	DQ871219.1, 172873-173018
Triticum turgidum	*Phytoene synthase 1* (Psy-B1) gene, 2	FJ393539.1, 1065-1155

Table 1. Megablast searches for Stowaway-like *ß-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

***Xylose isomerase Stowaway* MITE found in representatives of Triticeae and other Pooideae**		
Triticeae:		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Triticum turgidum	*Acetyl-CoA carboxylase* (Acc-2) gene, 1	EU660894.1, 12431-12525
Triticum turgidum	*Phytoene synthase 1* (Psy-B1) gene, 2	FJ393539.1, 2053-2211
Triticum urartu	*Sucrose phosphate synthase II* gene, 1	GU797179.1, 502-660
Triticum urartu	*RGA2 resistance protein* (RGA2) gene, 1	EU675952.1, 1103-1189
Triticum urartu	Intergenic spacer	FJ184992.1
Triticum urartu	*psi-acc-2* gene, 1	EU626553, 30113-30167
Brachypodieae:		
Brachypodium distachyon	*FHA domain protein* gene, 13	EU730897.1, 66190-89409
Brachypodium sylvaticum	*Methyl sulfoxide reductase* gene, 9	EF059989, 117355-117400
Poeae: Loliinae:		
Lolium perenne	Intergenic spacer	JN051258, 351-389
Barley leucine zipper Stowaway* MITEs found in representatives of *Hordeum		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Hordeum bulbosum	*Barley leucine zipper* gene (blz-1), 3	EU862588.1, 611-677
Hordeum murinum	*Barley leucine zipper* gene (blz-1), 3	EU862571.1, 596-662
Hordeum vulgare	*Barley leucine zipper* gene (blz-1), 3	X80068, 2806-2906
Nucellin Stowaway* MITEs found in representatives of *Hordeum		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Hordeum vulgare	*Nucellin* (nuc) gene, 5	AY485643, 93732-93894
Short *disrupted meiotic cDNA Stowaway* MITEs found in representatives of Triticeae		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Australopyrum velutinum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	AF277253, 807-881
Hordeum fuegianum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	GU734666.1, 842-938
Hordeum jubatum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	GU734673.1, 842-938
Hordeum tetraploidum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	GU734680, 841-937
Taeniatherum caput-medusae	*Disrupted meiotic cDNA 1 (dmc1) gene, 5*	AF277249, 797-893

Table 1. Megablast searches for Stowaway-like *β-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* elements

Long *Disrupted meiotic cDNA Stowaway* MITEs found in representatives of Triticeae		
Genome	**Location, intron**	**Gen Bank Accession, Range**
Heteranthelium piliferum	*Disrupted meiotic cDNA* 1 (dmc1) gene, 5	AF277239, 874-1183

In this study a new *Stowaway* MITE was found at the 4th intron of the *ß-amylase* gene in *Puccinellia distans* (Genbank JX536564, 225 bp), a member of the Alopecurinae / Poinae / Puccinelliinae clade that is nested within the large Poeae s.l. clade (Figs. 2, 3; Table 1). This insertion is highly similar to other *ß-amylase* 4th intron MITE sequences in Alopecurinae / Poinae / Puccinelliinae in Genbank, such as *Catabrosa aquatica* (HE565906, 190 bp), *Catabrosella araratica* (HE565912, 228 bp), *Colpodium drakensbergense* (HE565915, 167 bp), *C. versicolor* (HE565911, 227 bp), *Zingeria biebersteiniana* (HE565930, 243 bp) and *Z. kochii* (HE565927, 233 bp). Most of these Alopecurinae / Poinae / Puccinelliinae *Stowaway* MITEs formed stem-and-loop secondary structures with small additional insertions (Fig. 3; type B). Although they showed low free energies, the additional insertions disturbed, at least in part, their hairpin-like structures. Consequently, almost all of these TEs were longer and less stable than those detected in the Triticeae (MASON-GAMER, 2007). The exception was the *Zingeria biebersteiniana* insertion (243 bp; dG=-37.13), which showed a stem-and-loop secondary structure similar to the *Australopyrum* element (Fig. 3A). The remaining Alopecurinae / Poinae / Puccinelliinae MITEs showed identities >80% (Fig. 3; type B). This group included *P. distans* (225 bp; dG=-37.82), *C. aquatica* (190 bp; dG=-26.34), *C. araratica* (228 bp; dG=-44.16), *C. versicolor* (227 bp; dG=-44.16), and *Z. kochii* (233 bp; dG=-43.93) (Table 1). The *Stowaway* MITEs of *C. araratica*, *C. versicolor* and *Z. kochii* were also highly similar (>80% identity) to those found within the Ehrhartoideae / Oryzeae clade (e.g. 5th intron of the *Ogl12g0016G16_1* gene in *Oryza glaberrima*; intergenic spacer in *O. sativa* and *O. minuta*). The *C. aquatica* insertion had the highest free energy (dG=-26.34) of all insertions found within the Alopecurinae / Poinae / Puccinelliinae group. The insertion found in *C. drakensbergense* (167 bp; dG=-14.02) did not form a stem-and-loop secondary structure as almost one third of the sequence has been lost (see *ß-amylase* alignment). Additionally, this sequence showed a contiguous 232 bp insertion located 36 bp downstream of the 3'-end TA target duplication site. This insertion, flanked by a duplicated 5'-AGAAGTTGCATCT-3' sequence on its 5' and 3' ends could correspond to a piece of another mobile element, though it does not fold into a hairpin-like structure.

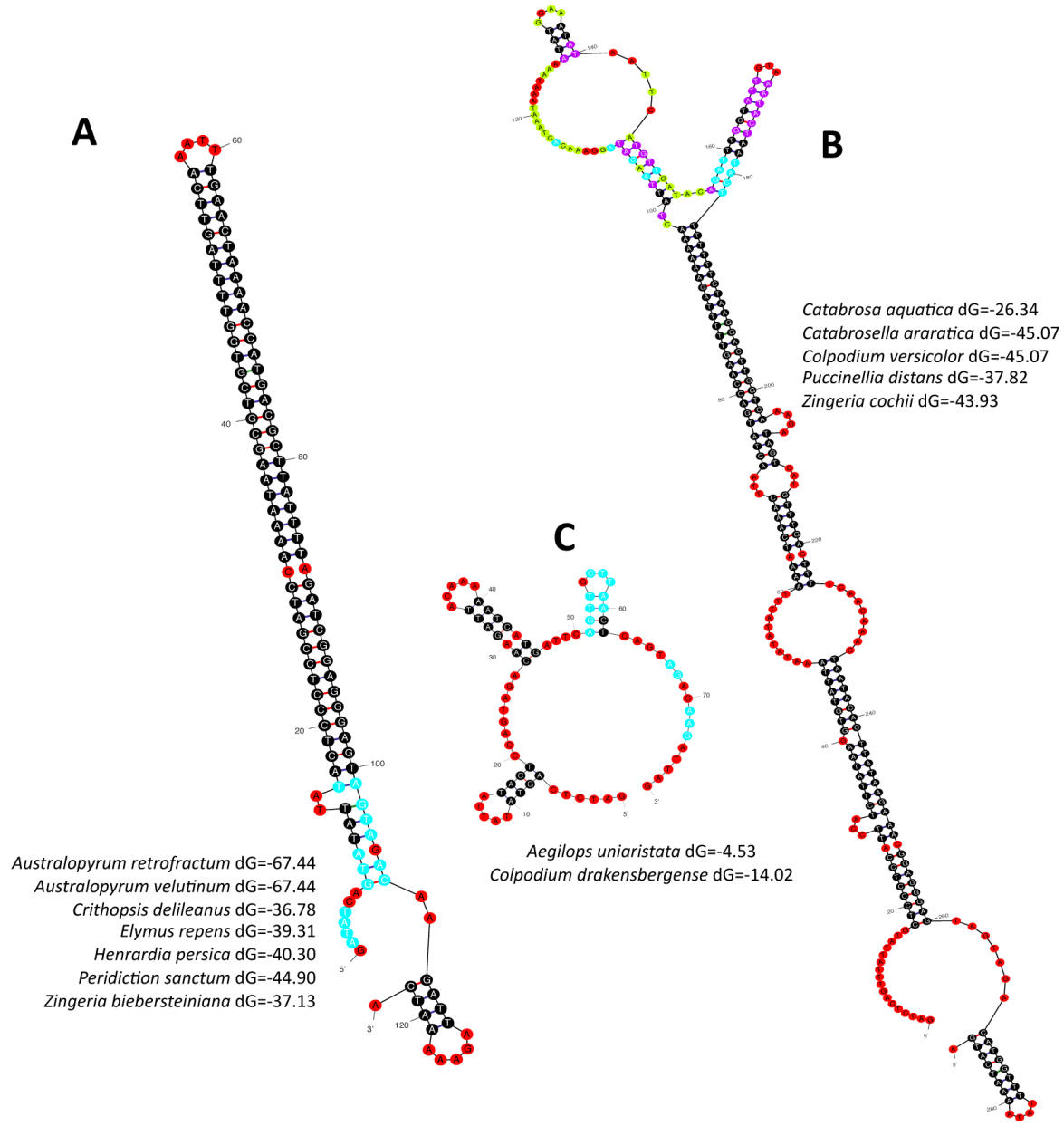


Figure 3. The secondary structures of the three types of *Stowaway* MITEs found at the fourth intron of the *ß-amylase* gene in grasses. Type A: Stem-and-loop secondary structure; Type B: Stem-and-loop secondary structure with small additional insertions; Type C: Additional small insertions that do not form a stem-and-loop secondary structure. The free energy is listed after the name of each species. The color of the nucleotide represents the probability that the base is single-stranded in the computed holdings. Red: 0.999%, green: 0.90, blue: 0.50, purple: 0.022, and black: 0.001.

Potential *Stowaway* MITE footprints have also been observed at the fourth intron of the *ß-amylase* gene in BEP clade grasses. These footprints consist of short insertions and/or deletions that could indicate the potential presence of a TE in the common ancestor. This is consistent with the proposed mechanism of excision of the *Tc1/mariner* element in rice (YANG *ET AL.*, 2006), and could help to explain the evolutionary dynamics of the *Stowaway* MITEs.

The *Stowaway* MITE footprints fell into six categories (Fig. 2): i) the target site 5'-TA-3' and the 5'-GTA-3' motif were present in nearly all representatives of the Pooideae. Mutations at these target sites, which could alter excision-insertion events (Fig. 2; underlined blue boldface), have been also occasionally found in Triticeae (MASON-GAMER, 2007). Several mutations in the preferred target sites were observed in distinct pooid lineages (e.g. *Bromus tectorum* '**C**A····GTA'; *Secale cereale* 'T**G**····GTA'; *Triticum aestivum* 'T**T**····GTA'; *Agrostis capillaris* 'T**T**····GTA'; *Pseudarrhenatherum longifolium* 'TA····**A**TA'; *Avellinia michelli* 'TA····G**A**A'; *Poa infirma* 'TA····G**C**A'; *Desmazeria rigida* 'TA····G**G**A'; *Stipa tenacissima* and *S. offneri* '**A**A····GTA'; *Brachypodium distachyon* and *B. pinnatum* T**G**; *Lolium* / *Micropyropsis* / *Festuca* subgen. *Schedonorus* (except *Micropyropsis tuberosa*) 'TA····G**C**A'; *Festuca eskia* / *F. gautieri* 'TA····G**A**A'; all fine-leaved *Festuca* (except the *F.* sects. *Dimorpha* / *Eskia* group and *Narduroides salzmannii*) share the GT**G** extreme; all the *Festuca* sect. *Festuca* taxa, except *F. panciciana* and *F. ovina,* share the **C**A extreme); ii) Small insertions of 2-5 bp adjacent to the 5'-TA duplication site have been detected in 14 different pooid species. They correspond to short-TIR remnants, with sequences varying from the 5'-CTCCA-3' (in *Pseudoroegneria strigosa*, *Festuca scabra*, *F. modesta, F. triflora*, *F. coerulescens*, *F. pseudeskia*, *F. scariosa, F. altisima*, *F. altaica, F. paniculata*, *F. gautieri*, *F. eskia* and *Narduroides salzmani*) to 5'-CTCCC-3' (in *Trisetum loeflingianum*); iii) Small insertions of 2-9 bases adjacent to the 3'-GTA motif were detected in 7 species. They likely correspond to flanking regions of TIR and short-TIR remnants with sequences varying from: 5´-TCA-3 (in *Brachypodium pinnatum, B. distachyon, Psathyrostachys juncea* and *Lophopyrum elongatum*), to 5'-CAGTACTCA-3' (*Nardus stricta*), 5'-CGGTAATCA-3' (*Lygeum spartum*), 5'-CCAATTAATT-3' (*Nassella laevissima*); iv) A 5'-GTA-3' deletion in *Brachypodium boissieri*; v) Two 5'-TA-3' deletions in *Nardus stricta* and *Lygeum spartum*; and vi) Two small deletions in, respectively, the 5' (9 bp) and 3' (15 bp) ends of the intron sequence flanking the empty element site were found in *Lolium perenne*. It is not clear whether all these deletions (iv, v, vi) are associated with element loss.

4.3 Triticeae *Stowaway blz1, dmc1, nuc*, and *xyl* MITEs found in other genes and in other pooid lineages

The *Stowaway*-like elements described in the *blz1, dmc1, nuc*, and *xyl* genes of Triticeae (PETERSEN AND SEBERG, 2000, 2009) were occasionally present in the same loci though they were mostly distributed in intergenic spacers and in the introns of other nuclear genes in the

newly searched species (Table 1). These insertions were almost exclusively found in representatives of the Triticeae; however one of them (*xyl* MITE) was also detected in representatives of other Pooideae lineages.

The 68-73 bp TE detected in the first intron of the *xyl* gene was only found in this locus in the 26 originally studied diploid *Hordeum* species (PETERSEN AND SEBERG, 2009; Table 1). Similar sequences were found, however, in other loci of species outside of *Hordeum*. These sequences were mainly observed within representatives of Triticeae (e.g., introns of the *pdha1* (*Aegilops*), *stpk-v* (*Dasypyrum*), and *mtk4* (*Triticum*) genes, and in intergenic spacers (*Triticum*)), but they were also found in representatives of Brachypodieae (e.g., introns of *fha domain protein* and *methyl sulfoxide reductase* genes (*Brachypodium distachyon, B. sylvaticum*) and of Loliinae (e.g., intergenic spacer, *Lolium perenne*) (Table 1)). All the sequences showed a similar hairpin-like secondary structure (cf. PETERSEN AND SEBERG, 2009 and results not shown).

The 96-104 bp TE detected in the third intron of the *blz1* gene was found only in this locus in three species of *Hordeum* (PETERSEN AND SEBERG, 2009; Table 1). Similar sequences were found at different nuclear loci and at one putative mitochondrial locus in other representatives of Triticeae (e.g., introns of the *psy-B1* (*Aegilops, Triticum*), *psi-acc-2* (*Triticum*), and *dmc1* (*Heteranthelium*) and mitochondrial *putative neurolysin precursor* (*Triticum*) genes, and intergenic spacers in *Triticum*). All these sequences showed a hairpin-like secondary structure similar to that described by PETERSEN AND SEBERG (2009) for their long *blz1-A* MITE. No matches were recovered for their short *blz1-B* MITE (47–49 bp), which was a defective TE, having one conserved TIR sequence and lacking a TSD (PETERSEN AND SEBERG, 2009).

The 164 bp TE detected in the fifth intron of the *nuc* gene in *Hordeum vulgare* (AY485643, Table 1) was only found in this locus in this species. Similar sequences were found in other loci in other species of Triticeae (e.g., non-coding sequences of *ß-amylase* (*Australopyrum*), and introns of *dp62 hypothetical protein* (*Hordeum*), *pdha1* (*Secale*), *pmm-d1* (*Aegilops*), and *sucrose phosphate synthase II* (*Triticum*) genes, and in intergenic spacers in *Triticum*). These sequences also showed an almost perfect hairpin-like secondary structure (cf. PETERSEN AND SEBERG, 2009, and results not shown).

PETERSEN AND SEBERG (2000, 2009) found two types of MITEs in the fifth intron of the *dmc1* gene: a short insertion of 59-76 bp, and a long insertion of 283 bp. The short element was found in representatives of *Australopyrum* and *Taeniatherum* (PETERSEN AND SEBERG, 2000)

and has been also described in other representatives of *Hordeum* (Table 1). We found similar sequences of this element in other loci of other Triticeae representatives and of *Hordeum* (e.g., introns of the *acc-1* and *acc-2* (*Triticum*), *bkini2* (*Hordeum*), *cpk9* (*Aegilops*), *pgk1* (*Agropyron, Kengyilia*), *pina* (*Aegilops*), *psy-b1* (*Triticum*), *putative expressed protein* (*Secale*), *rga2* (*Triticum*), and *vp-1b* (*Triticum*) genes and in intergenic spacers in *Hordeum*. The long *dmc1* element was recorded in *Heteranthelium piliferum*, with potential footprints detected in *Dasypyrum* (PETERSEN AND SEBERG, 2000). We found similar sequences in other loci of other Triticeae representatives (e.g., introns of the *adp-glucose pyrophosphorylase small subunit* (*Aegilops*), *putative ankyrin repeat protein* (*Hordeum, Triticum*), and *wknox1b* (*Triticum*) genes and in intergenic spacers in *Triticum*. Those sequences also showed a hairpin-like secondary structure (results not shown), similar to that of *H. piliferum* (PETERSEN AND SEBERG, 2000).

5. Discussion

5.1 Evolutionary dynamics of the *Stowaway* MITEs in grasses

The occurrence of *Stowaway* MITEs and their potential footprints in the grasses was assessed within a robust Poaceae phylogeny, under the assumption that the tree obtained from the combined analysis of the nuclear and plastid genes (Fig. 1) represented a good estimate of the true evolutionary history of the family.

The 116 Poaceae species sampled fell into the two traditionally described clades PACCMAD and BEP (e.g. GPWG, 2001, 2012; SÁNCHEZ-KEN *ET AL.,* 2007; BOUCHENAK-KHELLADI *ET AL.,* 2008) (Fig. 1). The monophyletic PACCMAD clade was split into two strongly supported sister lineages, the Centothecoideae-Panicoideae and the Danthonioideae-Chloridoideae. The successive divergences within each of these clades were in agreement with previously published phylogenies (e.g. BARKER *ET AL.,* 1995; DUVALL AND MORTON, 1996; BARKER, 1997; SORENG AND DAVIS, 1998; MATHEWS *ET AL.,* 2000; SÁNCHEZ-KEN *ET AL.,* 2007; BOUCHENAK-KHELLADI *ET AL.,* 2008). The BEP clade included members of the Ehrhartoideae and Pooideae. The topological resolution within the Ehrhartoideae and the Pooideae groups was in agreement with previous studies (e.g. SORENG AND DAVIS, 2000; GPWG, 2001, 2012; SORENG *ET AL.,* 2003; DAVIS AND SORENG, 2007; BOUCHENAK-KHELLADI *ET AL.,* 2008; SCHNEIDER *ET AL.,* 2009; TRIPLETT AND CLARK, 2010). The Pooideae clade showed the successive divergences of the more ancestral Lygeeae / Nardeae, Stipeae, Diarrheneae and Brachypodieae lineages, and the separation of the more recently evolved Bromeae / Triticeae

and Aveneae / Poeae 'core pooid' lineages. The divergences within the Aveneae / Poeae lineage agreed with those proposed by DÖRING *ET AL.* (2007), QUINTANAR *ET AL.* (2007), and SCHNEIDER *ET AL.* (2009). QUINTANAR *ET AL.* (2007) used an intertribal grouping system to reflect the morphological and evolutionary heterogeneity of the large Poeae s.l. clade that consisted of two strongly supported sister groups, the "core-Aveneae" lineage and the "core-Poeae + Poeae *pro parte* + former Aveneae" lineage. The results of the successively enlarged phylogenies of Loliinae obtained by TORRECILLA AND CATALAN (2002), TORRECILLA *ET AL.* (2004), CATALAN *ET AL.* (2004, 2006), and INDA *ET AL.* (2008) were confirmed in our analyses; the Loliinae were split into two main groups, the more ancestral broad-leaved Loliinae and the more recently evolved fine-leaved Loliinae (Fig. 1).

Seven types of *Stowaway* MITEs were found within the Pooideae, confirming the polyphyletic origins of these elements in some cases. The apparent homogeneity observed in the TEs located in the *blz1, dmc1, nuc*, and *xyl* genes of the Triticeae (PETERSEN AND SEBERG, 2000, 2009), and occasionally in other loci of other pooids ('*xyl*' element in *Brachypodium* and *Lolium*), contrasts with the heterogeneity among the elements in the fourth intron of the *ß-amylase* gene in different BEP lineages and in other loci of the temperate grasses. Three different types of *Stowaway* MITEs have been detected in the *ß-amylase* gene, two of them in the Triticeae, and one in the Alopecurinae / Poinae / Puccinelliinae. It was not possible to align their disparate sequences, and megablast and discontiguous megablast searches in the Entrez nucleotide database showed their presence in a variety of genes throughout the BEP clade (Table 1), thus confirming their polyphyletic nature.

Although the heterogeneous and uneven distributions of four of the seven types of *Stowaway* MITEs would preclude any stem-based vertically-inherited transmission, most of these elements were found within the same phylogenetic lineage (e.g., *bzl*1, *dmc*1, and *nuc* MITEs in Triticeae; *ß-amylase Puccinellia*-type MITEs in Alopecuriinae / Poiinae / Puccinelliinae; *ß-amylase Australopyrum*-type and *Aegilops-Peridictyon*-type MITEs in Triticeae; Table 1), with a few cases of apparent transmissions from/to less-related grass groups (e.g. *xyl*-type Triticeae MITEs in Brachypodieae and Poeae (*Lolium*); *ß-amylase Aegilops-Peridictyon*-type MITEs in Brachypodieae; *ß-amylase Australopyrum*-type MITEs in Poeae (Avena), and *ß-amylase* Alopecuriinae / Poiinae / Puccinelliinae MITEs in *Oryza* (Table 1). This suggests a high intragenomic mobility and a more restricted interspecific and intergeneric transference of these elements. Remarkably, none of the seven BEP MITEs has been found in any loci of any representative of the PACCMAD grasses. This is consistent with the strong

evolutionary divergence and reproductive and genomic isolation of these two main grass lineages, which separated between 50-70 Mya (SALSE *ET AL.*, 2008; VICENTINI *ET AL.*, 2008).

5.2 Phylogenetic distributions of the *ß-amylase Stowaway* MITEs in the BEP clade

The reconstruction of the evolutionary history of the two types of *ß-amylase* 4th intron Triticeae MITEs is difficult to decipher. One potential hypothesis would begin with the transference of an 84bp TE from a more basal Brachypodieae species to some members of the more recently evolved Triticeae (e.g. *P. sanctum*). From these species the MITE could have been transferred to other Triticeae (*C. delileanus*, *E. repens*, and *H. persica*), and to other loci of other representatives of the wheat tribe (*H. vulgare*, *S. cereale* and *T. aestivum*). These transfers would have involved, or been followed by, the putative acquisition of new nucleotides in some cases or the loss of others (e.g. *A. uniaristata*), including the loss of their stem-and-loop secondary structures. By contrast, another hypothesis must be invoked to explain the distribution of the *Australopyrum*-type insertion. This element has not been found in the basal Pooideae and, therefore, could have been transferred from the Triticeae clade to the Poeae (*Avena*) clade or vice versa.

The evolution of the Alopecurinae / Poinae / Puccinelliinae *Stowaway* MITE could be explained by alternative hypotheses. This large insertion could have been transferred from the more basal Ehrhartoideae clade (*Oryza* spp) to the Alopecurinae / Poinae clade (*C. araratica*, *C. versicolor*, *Z. kochii*), after which it could have been transferred within this group (to *Z. biebersteiniana*) and to the close Puccinelliinae (*P. distans*) group. The TE found in *Z. biebersteiniana* is the only one found within the Poeae s.l. clade that shows a perfect stem-and-loop secondary structure. The insertions found in *C. aquatica* (190 bp) and *C. drakensbergense* (167 bp) (sequence similarity to *C. aquatica*: 86%; query coverage: 100%; e-value: 5e-81) might have resulted from additional transferences from *C. araratica*, *C. versicolor* and *Z. kochii*. During these processes, the *C. aquatica* and *C. drakensbergense* TEs could have experienced deletions, thus reducing their similarities to other Alopecurinae / Poinae / Puccinelliinae *Stowaway* MITEs.

All the potential intertribal transferences invoked to explain the disparate distributions of the three types of *ß-amylase Stowaway* MITEs involve largely isolated, divergent lineages which show no evidence of current crossing. The oldest split corresponds to that of Ehrhartoideae (Oryza) and Pooideae, dated at 50-47 Mya, followed by those of Brachypodieae and the 'core-pooid' clade, 30-25 Mya, and of Triticeae and Poeae s. l. (including Aveneae), c.

25 Mya (VICENTINI *ET AL.,* 2008). The sexual isolation of these lineages would preclude any recent intertribal transfer, and their respective long evolutionary distances would require complex evolutionary scenarios to explain the observed distribution of their MITEs (see comments below).

5.3 Acquisition of the *ß-amylase Stowaway* MITEs in the Poaceae

The *ß-amylase* elements had to be mobilized from ancestral sequences in accordance with the mechanisms proposed to explain the potential independent losses of MITEs. IZSVÁK *ET AL.* (1998) proposed a model in which the stable secondary structures of MITEs (e.g., Fig. 3 type A and B) could dislocate from the replication complex. This model does not, however, exclude the possibility that MITEs could also be mobilized by trans-complementary transposases in a cross-mobilization process, but these elements would not be able to copy or excise if they have any mutation or deletion in the TIR region, or if they lose their secondary structure (SANTIAGO *ET AL.,* 2002; PETERSEN AND SEBERG, 2000, 2009). In *Oryza sativa*, the transpositionally active MITEs contain internal sequences that enhance transposition (YANG *ET AL.,* 2009), and the structure of these mobile elements shows integrity for their TIR and TSD regions (LU *ET AL.,* 2012).

Because the *ß-amylase* insertions are in an intron, it is unlikely that they have been subjected to selection pressure. Most of the *ß-amylase Stowaway* MITEs found by MASON-GAMER (2007) formed almost perfect hairpin-like secondary structures that had very low free energies (Fig. 3; type A). Accumulation of mutations, insertions, and deletions in the original element would result in the observed diversity detected among the present *Stowaway* MITEs. These mechanisms could be invoked to explain the insertions without stem-and-loop secondary structures (Fig. 3; type C) in, for example, *Aegilops uniaristata* and *Colpodium drakensbergense*. Despite the large number of *Stowaway* MITEs and other types of MITEs studied, MITE mobility is an open question, because their transposase binding sites and transposition enhancer regions are almost always degraded. As a consequence, estimation of their absolute or relative age is difficult, because the dynamics underlying their gain, loss, and high copy number are not yet well understood (FESCHOTTE *ET AL.,* 2002B; PETERSEN AND SEBERG, 2000, 2009; LU *ET AL.,* 2012).

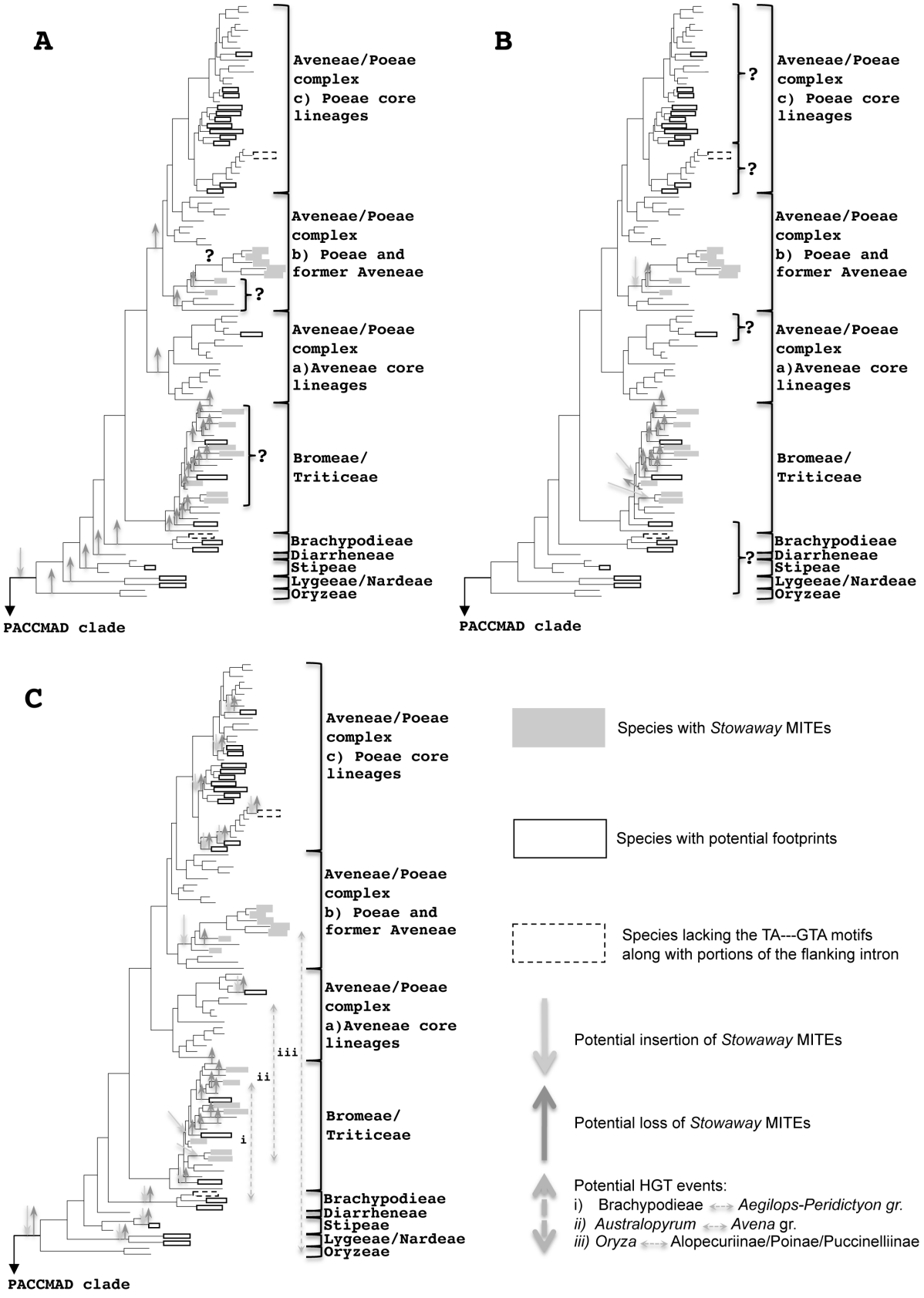

A
Aveneae/Poeae complex c) Poeae core lineages
Aveneae/Poeae complex b) Poeae and former Aveneae
Aveneae/Poeae complex a)Aveneae core lineages
Bromeae/ Triticeae
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae
Oryzeae
PACCMAD clade
B
Aveneae/Poeae complex c) Poeae core lineages
Aveneae/Poeae complex b) Poeae and former Aveneae
Aveneae/Poeae complex a)Aveneae core lineages
Bromeae/ Triticeae
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae
Oryzeae
PACCMAD clade
C
Aveneae/Poeae complex c) Poeae core lineages
Aveneae/Poeae complex b) Poeae and former Aveneae
Aveneae/Poeae complex a)Aveneae core lineages
Bromeae/ Triticeae
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae
Oryzeae
PACCMAD clade
i
ii
iii
Species with Stowaway MITEs
Species with potential footprints
Species lacking the TA---GTA motifs along with portions of the flanking intron
Potential insertion of Stowaway MITEs
Potential loss of Stowaway MITEs
Potential HGT events:
i) Brachypodieae ⟷ Aegilops-Peridictyon gr.
ii) Australopyrum ⟷ Avena gr.
iii) Oryza ⟷ Alopecuriinae/Poinae/Puccinelliinae

Figure 4. Alternative hypotheses on the potential evolutionary dynamics of gains, losses and transferences of *ß-amylase Stowaway* MITEs across the studied Poaceae lineages. a) Single early acquisition hypothesis; b) three independent acquisitions hypothesis; c) different acquisitions followed by multiple losses and different horizontal transfers hypothesis (see text for more detailed explanations on each hypothesis). Light and dark grey arrows represent potential gains and losses of MITEs, respectively. Question marks indicate uncertain events. Grey dashed arrows represent potential HGT events for the *Aegilops-Peridictyon*-type (i), *Australopyrum*-type (ii) and Alopecuriinae / Poinae / Puccinelliinae-type (iii) MITEs. A double arrow indicates uncertainty in the direction of the transference.

All the compiled information provides evidence that both gains and losses of *Stowaway* MITEs at the *ß-amylase* locus may have occurred repeatedly in the BEP clade of Poaceae. The plausible underlying dynamic of acquisitions and deletions of these elements and their footprints could be explained by three, non-mutually-exclusive, alternative hypotheses (Fig. 4).

i) A single early acquisition of a ß-amylase TE in the Pooideae followed by multiple independent losses

This first hypothesis favors a single early insertion of a palindrome element, similar to *Tc1-Mariner*, in the fourth intron of the *ß-amylase* gene in the ancestor of the BEP clade, followed by multiple independent losses (Fig. 4A). This potentially autonomous element was flanked by its TDS (5'-TA-3') and the 5'-GTA-3' motif, explaining the wide distribution of the 5'-TA---GTA-3'footprints across the pooids. This early insertion could have been vertically transmitted to successive generations. This hypothesis would explain the broad phylogenetic distribution of the 5'-TA---GTA-3' region (and derived mutated sequences) in almost all the analyzed BEP members (Fig. 1, 2). By contrast, the lack of target sites, *Stowaway* MITEs, and TE footprints in the PACCMAD members sequenced to date suggest that these insertions were never present at the same locus in that clade. However, early acquisition by the BEP ancestor does not explain the observed dissimilarities between the three types of *ß-amylase Stowaway* MITEs detected in the Triticeae and Alopecurinae / Poinae / Puccinelliinae. These MITEs are more similar to elements found in other parts of the same genomes, or in other species' genomes, than they are to each other (Table 1), raising doubts about their potential monophyletic origin.

Additionally, the single early acquisition scenario would require at least 23 independent partial losses to explain the absence of *ß-amylase Stowaway* insertions in many of the surveyed species, and the presence of the different types of footprints found in the BEP clade (Fig. 2). Losses of footprint-like sites and the adjacent intron regions in *Brachypodium boissieri* and *Lolium perenne* might additionally be interpreted as secondary parallel deletions. Shared

insertions adjacent to the TA target site (CTCCM) in 14 pooid taxa (Fig. 2) and adjacent to the GTA motif in 7 pooid taxa could be explained by an acquisition followed by later independent losses, by multiple independent acquisitions, or by a mixture of both processes.

ii) Independent insertions of three different Stowaway MITEs at the same locus in the ß-amylase gene in the Triticeae and in the Alopecurinae / Poinae / Puccinelliinae lineages

The second hypothesis favors multiple independent insertions of different *Stowaway* elements at the same *ß-amylase* locus in non-related Pooideae lineages (Fig. 4B). This scenario would require at least three distinct insertions, which correspond to the three types of *Stowaway* MITEs analyzed in this study (Table 1). This hypothesis would explain the polyphyletic distribution of the homoplasious insertions, their differences in sequence diversity and size, and their marked similarities to elements found in non-homologous sites in other genomes. This hypothesis is also consistent with the high number of MITE copies found in the grass genomes, where they are clearly abundant (JIANG *ET AL.*, 2004). This theory of multiple insertions at homologous sites has traditionally been viewed as the result of an active history of amplification and insertion of MITEs (FESCHOTTE *ET AL.*, 2002A,B; WESSLER *ET AL.*, 1995; MASON-GAMER, 2007) and other types of transposable elements (SAVILLE *ET AL.*, 1999; WALKER *ET AL.*, 1997; BRINGAUD *ET AL.*, 2002; PENTON *ET AL.*, 2002). However, this hypothesis does not explain the existence of the TDS (5'-TA-3') and the 5'-GTA-3' motif and derived sequences found in almost all members of the Pooideae.

Although the MITEs' very short target site (TA) is extremely frequent, empirical observations suggest that it is the preferred insertion site for this type of transposable element (BUREAU *ET AL.*, 1996; WESSLER, 2006; MASON-GAMER, 2007; PETERSON AND SEBERG, 2009). Since all grass MITEs found in the fourth intron of *ß-amylase* to date are in the Pooideae, and the 5'-TA---GTA-3' region is widespread only in the BEP clade, it is plausible to hypothesize that this motif is important in making the fourth intron of *ß-amylase* a MITE hotspot.

iii) A single early acquisition and multiple losses, followed by three independent insertions, additional losses, and horizontal transfer

The third hypothesis depicts a more complex scenario of a single early acquisition of a *ß-amylase Stowaway* MITE in the ancestor of the Pooideae followed by one or a few early losses. This would explain the wide distribution of 5'-TA---GTA-3' footprints. A series of more recent independent insertions, losses, and potential horizontal transfers (Fig. 4C) is required to explain

their current distribution. *Stowaway* MITEs are a highly dynamic group of rapidly evolving sequences, and their history appears to have involved several temporally separated waves of amplification (FESCHOTTE *ET AL.*, 2003; LU *ET AL.*, 2012). Because of this, and since neither of the previous two hypotheses completely explained the evolutionary dynamic of the *ß-amylase Stowaway* MITEs found in the grasses, a third hypothesis which draws from both of them, is proposed. According to this hypothesis, two different waves of element acquisitions, accompanied by multiple losses and horizontal gene transfers, might have taken place in the history of the Pooideae. This scenario could be explained by five consecutive steps:

1) An element with a palindromic organization and a hairpin-like secondary structure (FESCHOTTE AND WESSLER, 2002) inserted into the fourth intron of the *ß-amylase* gene in the ancestor of the BEP group. This insertion probably took place c. 55 Mya, at the time of the divergence between the BEP and PACCMAD lineages (VICENTINI *ET AL.*, 2008). This element likely inserted with a mechanism similar to the *Tc1-Mariner* superfamily, with analogous TIR and TSD sequences.

2) The initial transposable element underwent one or more early excisions using a trans-system transposase that recognized the TIRs, leaving 5'-TA---GTA-3' footprints, which are currently widespread throughout the Ehrhartoideae / Pooideae clade (there is no data on Bambusoideae).

3) At least three independent insertions of orthologous *Stowaway* MITEs may have occurred in members of the Triticeae (e.g., in the *Aegilops-Peridictyon* group and in the *Australopyrum* group), and in members of the Alopecurinae / Poinae / Puccinelliinae lineage, within the previously created flanking site 5'-TA---GTA-3' (Fig. 2). It is also possible that some of the present-day elements are direct descendents of the initial insertion, and have been degraded by the accumulation of mutations, insertions, and deletions, with some even losing their stem-and-loop secondary structure, as in the case of *C. drakensbergense* (Fig. 3; type C).

4) Subsequent deletions are needed to explain such observations as partial elements (e.g., *Ae. uniaristata*) and the existence and distribution of apparent atypical footprints within the widespread 5'-TA---GTA-3' motif (e.g., the green boldface sequences in Fig. 2). These events probably followed the same mechanism of cross-mobilization described above. Little is known about the methods of transposition *in trans* and their requirements, because the vast majority of MITEs identified so far are probably remnants of ancient insertion events. However, it is plausible that some of those hypothetical deletions or transpositions left the collection of potential *Stowaway* footprints found today (Fig. 2).

5) Hybridization, polyploidization, and lineage sorting may have contributed to the distribution of *Stowaway* MITEs and their footprints throughout the Pooideae. These processes have often been used to explain the incongruence between the phylogeny of the nuclear and plastid genes in the grasses (KELLOGG *ET AL.*, 1996; GPWG, 2001, 2012; CATALAN *ET AL.*, 2004, 2006; CATALAN, 2006; SORENG *ET AL.*, 2007). Horizontal transfer of mobile elements has been documented in grasses (DIAO *ET AL.*, 2006), as has the non-sexual HGT of a functional copy of the *pgiC* gene between two non-related and sexually isolated Poeae lineages (GHATNEKAR *ET AL.*, 2006; VALLENBACK *ET AL.*, 2008, 2010). These findings demonstrate that both non-expressed and expressed sequences can be transferred among non-related grasses. Horizontal transfer could explain the unquestionable disconnect found between the evolutionary relationships of the inserted *Stowaway ß-amylase* elements and the relationships among the species and genera that harbor them (Fig. 2). Nonetheless, this could also be explained by recurrent transfer of ancestrally variable elements across the genome. If a genome has a wide collection of elements inherited from an early ancestral population, with some or all types shared across some or all taxa, then moving them around within genomes could also cause an apparent disconnect between element types and the organismal phylogeny. Ongoing wide genome sequencing research in several model pooid species and in other grass representatives will likely help to decipher the potential intergenomic vs. intragenomic mobilities of the transposable elements and the evolutionary dynamics of the *ß-amylase* MITEs.

6. Literature cited

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC bioinformatics* 20, 407-415.

Barker, N.P., Linder, H.P., Harley, E.H., 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbcL* sequence data. *Systematic Biology* 20, 423-435.

Barker, N.P., 1997. The relationships of *Amphipogon*, *Elytrophorus*, and *Cyperochloa* (Poaceae) as suggested by *rbcL* sequence data. *Telopea* 7, 205-213.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Bringaud, F., Biteau, N., Melville, S.E., Hez, S., El-Sayed, N.M., Leech, V., Berriman, M., Hall, N., Donelson, J.E., Baltz, T., 2002. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. *Eukaryotic Cell* 1, 137-151.

Bureau, T.E., Wessler, S.R., 1994. Stowaway: a new family of inverted repeat elements associated with genes of both monocotyledonous and dicotyledonous plants. *The Plant Cell* 6, 907-916.

Bureau, T.E., Ronald, P.C., Wessler, S.R., 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. *Proceedings of the National Academy of Sciences, U.S.A.* 93, 8524-8529.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and trnL-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., 2006. Phylogeny and evolution of *Festuca* L. and related genera of subtribe *Loliinae* (Poeae, Poaceae). In: A.K. Sharma, A. Sharma (Eds.), *Plant genome. Biodiversity and evolution*. Science Publishers, Enfield (New Hampshire), pp. 255-303.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British and European flora. Botanical Society of the British Isles*, London, pp. 45-70.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristic and parallel computing. *Nature Methods* 9, 772.

Davis, J.I., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in GBSSI. *Aliso* 23, 335-348.

DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology* 49, 171-179.

Diao, X., Freeling, M., Lisch, D.R., 2006. Horizontal transfer of a plant transposon. *PLoS Biology* 4, 119–128.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex Pooideae, Poaceae. *Kew Bulletin* 62, 407-424.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Duvall, M.R., Morton, B.R., 1996. Molecular phylogenetics of Poaceae: an expanded analysis of rbcL sequence data. *Molecular Phylogenetics and Evolution* 5, 352-358.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315-319.

Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783-791.

Feschotte, C., Wessler, S.R., 2002. Mariner-like transposases are widespread and diverse in flowering plants. *Proceedings of the National Academy of Sciences, U.S.A.* 99, 280-285.

Feschotte, C., Jiang, N., Wessler, S.R., 2002a. Plant transposable elements: where genetics meets genomics. *Nature Reviews Genetics* 3, 329-341.

Feschotte, C., Zhang, X., Wessler, S.R., 2002b. Miniature inverted-repeat transposable elements MITEs and their relationship with established DNA transposons. In: N.L. Craig, R. Craigie, M. Gellert, A.M. Lambovitz (Eds.), *Mobile DNA II*. American Society of Microbiology Press, Washington DC, pp. 1147-1158.

Feschotte, C., Lakshmi, S., Wessler, S.R., 2003. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway miniature inverted repeat transposable elements MITEs. *Genetics* 165, 747-758.

Feschotte, C., Osterlund, M.T., Peeler, R., Wessler, S.R. 2005. DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. *Nucleic Acids Research* 33, 2153-2165.

Flavell, A.J., Pearce, S.R., Kumar, A., 1994. Plant transposable elements and the genome. *Current Opinion in Genetics and Development* 4, 838-844.

Ghatnekar, L., Jaarola, M., Bengtsson, B.O., 2006. The introgression of a functional nuclear gene from *Poa* to *Festuca ovina*. *Procceedings of the Royal Society B* 273, 395-399.

Guindon, S., Gascuel, O., 2003. A simple, fasta and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52, 696-704.

G.P.W.G., The Grass Phylogeny Working Group., 2001. Phylogeny and subfamilial classification of the grasses. *Annals of the Missouri Botanical Garden* 88, 373-457.

G.P.W.G., The Grass Phylogeny Working Group., 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Greenblatt, I.M., Brink, R.A., 1962. Twin mutations in medium variegated pericarp maize. *Genetics* 47, 489-501.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae Poaceae, Pooideae grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Izsvák, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., Hackett, P.B., 1998. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. *Journal of Molecular Evolution* 48, 13-21.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., Wessler, S.R., 2003. An active DNA transposon family in rice. *Nature* 421, 163-167.

Jiang, N., Feschotte, C., Zhang, X., Wessler, S.R., 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements MITEs. *Current Opinion in Plant Biology* 7, 115-119.

Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). *Systematic Botany* 21, 321-347.

Kikuchi, K., Terauchi, K., Wada, M., Hirano, H.Y., 2003. The plant MITE mPing is mobilized in anther culture. *Nature* 421, 167-170.

Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang, J., Buell, C.R., Baker, B., 2009. Identification of miniature inverted-repeat transposable elements MITEs and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. *Genome Research* 19, 42-56.

Lu, C., Chen J., Zhang, Y., Hu, Q., Su, W., Kuang, H., 2012. Miniature Inverted-Repeat Transposable Elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Molecular Biology and Evolution* 29, 1005-1017.

Macas, J., Koblízková, A., Neumann, P., 2005. Characterization of Stowaway MITEs in pea *Pisum sativum* L. and identification of their potential master elements. *Genome* 48, 831-839.

Maddison, D.R., Maddison, W.P., 2008. *MacClade 4.08 version for OS X*. Sinauer Associates Inc., Sunderland (MA), USA

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, S.A., Wing, R.A., 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. *Genome Research* 10, 982-990.

Mason-Gamer, R.J., 2005. The β-amylase of grasses and phylogenetic analysis of the Triticeae (Poaceae). *American Journal of Botany* 92, 1045-1058.

Mason-Gamer, R.J., 2007. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. *Evolutionary Biology* 7, 92-102.

Mason-Gamer, R.J., Kellogg, E.A., 1997. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45, 524-545.

Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. *American Journal of Botany* 87, 96-107.

Menzel, G., Dechyeva, D., Keller, H., Lange, C., Himmelbauer, H., Schmidt, T., 2006. Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. *Cromosome research* 14, 831-844.

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., Tanisaka, T., 2003. Mobilization of transposons in the rice genome. *Nature* 421, 170-172.

Nassif, N., Penney, J., Pal, S., Engels, W., Gloor, G., 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-indiced gap repair. *Molecular and Cellular Biology* 14, 1613-1625

Penton, E.H., Sullender, B.W., Crease, T.J., 2002. Pokey, a new DNA transposons in *Daphnia Cladocera*: Crustacea. *Journal of Molecular Evolution* 55, 664-673.

Petersen, G., Seberg, O., 2000. Phylogenetic evidence for excision of Stowaway miniature inverted-repeat transposable elements in Triticeae (Poaceae). *Molecular Biology and Evolution* 17, 1589-1596.

Petersen, G., Seberg, O., 2009. Stowaway MITEs in *Hordeum* (Poaceae): evolutionary history, ancestral elements and classification. *Cladistics* 25, 198-208.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid trnT-F and nuclear ITS sequences. *American Journal of Botany* 94, 1554-1596.

Ronquist, F., Huelsenbeck, J.P., Teslenko M., 2011. *MrBayes version 3.2 manual: tutorials and model summaries* (http://mrbayes.sourceforge.net/mb3.2_manual.pdf).

Salse, J., Bolot, S., Throude, M.I., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuilleta, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20, 11-24.

Sánchez-Ken, J.G., Clark, L.G., Kellogg, E.A., Kay, E.E., 2007. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Systematic Botany* 32, 71-80.

SantaLucia, J.R., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences, U.S.A.* 95, 1460-1465.

Santiago, N., Herráiz, C., Goñi, J.R., Messeguer, X., Casacuberta, J.M., 2002. Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis thaliana*. *Molecular Biology and Evolution* 19, 2285-2293.

Saville, K.J., Warren, W.D., Atkinson, P.W., O'Brochta, D.A., 1999. Integration specificity of the hobo element of *Drosophila melanogaster* in dependent on sequences flanking the integration site. *Genetica* 105, 133-147.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid matK gene-3'trnK exon and nuclear ITS sequences. *Taxon* 58, 405-424.

Seberg, O., Petersen, G. 2009. Correspondence: A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 10: 276, doi:10.1038/nrg2165-c3.

Soreng, R.J., Davis, J.I., 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction character sets. *Botanical Review* 64, 1-85.

Soreng, R.J., Davis, J.I., 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus

reticulation. In: S.W.L. Jacobs, J. Everett (Eds.), *Grasses: Systematics and Evolution*. CSIRO, Melbourne, pp. 61-74.

Soreng, R.J., Peterson, P.M., Davidse, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. Catalogue of New World grasses Poaceae: IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Swofford, D.L., 2002. PAUP*: *Phylogenetic analysis using parsimony and other methods. version 4b10*. Sunderland (MA). Sinauer Associates.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear ITS sequences. *Systematic Botany* 27, 241-251.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and trnL-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Triplett, J.K., Clark, L.G., 2010. Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on *Arundinaria* and allies. *Systematic Botany* 35, 102-120.

Turcotte, K., Bureau, T., 2002. Phylogenetic analysis reveals Stowaway-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily. *Genome* 45, 82-90.

Vallenback, P., Jaarola, M., Ghatnekar, L., Bengtsson, B., 2008. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. *Molecular Phylogenetics and Evolution* 46, 890-896.

Vallenback, P., Ghatnekar, L., Bengtsson, B., 2010. Structure of the Natural Transgene PgiC2 in the Common Grass *Festuca ovina*. *PlosOne*. e13529. doi:10.1371/ journal.pone. 0013529.

Vicentini, A., Barber, J.C., Aliscioni, A.A., Giussani, L.M., Kellogg, E.M., 2008. The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology* 14, 2693-2977.

Walker, E.L., Eggleston, W.B., Demopoulos, D., Kermicle, J., Dellaporta, S.L., 1997. Insertions of a novel class of transposable elements with a strong target site preference for the r locus of maize. *Genetics* 146, 681-693.

Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. *Proceedings of the National Academy of Sciences, U.S.A.* 21, 17600-17601.

Wessler, S.R., Bureau, T.E., White, S.E., 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. *Current Opinion in Genetics and Development* 5, 814-821.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, B., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8, 973-982.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, B., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2009. Correspondence: Reply: A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 10, 276.

Yang, G., Weil, C.F., Wessler, S.R., 2006. A rice Tc1/Mariner-like element transposes in yeast. *The Plant Cell* 18, 2469-2478.

Yang, G., Zhang, F., Hancock, C.N., Wessler, S.R., 2007. Transposable of the rice miniature inverted repeat transposable element mPing in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, U.S.A.* 104, 10962-10967.

Yang, G., Nagel, D. H., Feschotte, C., Hancock, C. N., Wessler, S.R., 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. *Science* 325, 1391-1394.

Zhang, X., Jiang, N., Feschotte, C., Wessler, S.R., 2004. PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. *Genetics* 166, 971-986.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31, 3406-3415.

7. Appendix

7.1 Procedures for DNA isolation, amplification, cloning and sequencing

Amplification reactions were performed using a reaction mix containing 0.5 μl of forward and reverse primer (10mM), 1 μl of 10X PCR buffer, 3μl of $MgCl_2$ (50 mM), 0.3 μl of *Taq* DNA polymerase (5U/μl), 1 μl of deoxynucleoside triphosphate set PCR grade (10mM), between 0.5 and 1 μl of template DNA, and an aliquot of sterile ultrapure water (MilliQ) added

to achieve a final volume of 10 μl. The amplification parameters consisted of an initial denaturing step of 1 min at 94 ºC, followed by 4 cycles of 45 sec denaturing at 94 ºC, 2 min annealing at 65 ºC, and 1 min extension at 72 ºC, followed by 29 cycles of 30 sec denaturing at 94 ºC, 40 sec annealing at 65 ºC, and 40 sec extension at 72 ºC, with a final extension step of 7 min at 72 ºC. All PCR products were cleaned using ExoSAP-IT® following the protocol indicated by the manufacturer. PCR products of 17 species were cloned into pGEM-T Easy vectors (Promega), and transformed into *E. coli* JM109 competent cells (Promega) following the manufacturer's protocol, except that all reactions were halved. Between 2 and 10 cloned fragments were amplified directly from white colonies using the same primers and recipe as those used for the original PCR.

Beta amylase products were employed as templates for the cycle sequencing reaction using the ABI Big-Dye Terminator Cycle Sequencing Kit® (Applied Biosystems). The external and internal primers listed above were used in this step. The sequencing reactions were prepared with 2 μl of Premix Ready Reaction, 2 μl of primer (5mM), between 1 and 3 μl of purified DNA, and an aliquot of sterile ultrapure water (MilliQ) to achieve a final volume of 10 μl. PCR was performed with one denaturing cycle of 1 min at 96 ºC, 24 cycles of 10 sec denaturing at 96 ºC, 5 sec annealing at 50 ºC, and 4 min extension at 60 ºC, then a termination step of 4 min at 60 ºC. Forward and reverse fragments were generated for each sample and contigs were assembled and manually adjusted using the program Sequencher® 4.2.2 (Genes Codes Corporation, Ann Arbor, Michigan, U.S.A).

7.2 Procedures for the Bayesian and parsimony-based searches

Bayesian and Maximum parsimony (MP) based searches were performed using, respectively, MrBayes 3.1.2 (ALTEKAR *ET AL.*, 2004; RONQUIST AND HUELSENBECK, 2003) and PAUP*. All gaps were treated as missing data. Bayesian analysis was conducted according to the GTR+I+ gamma model, which was selected as the optimal model for each partition using the tests of goodness of fit for alternative nucleotide substitution models performed through the Hierarchical Likelihood Ratio Test (hLRTs), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in jModelTest 2 (GUINDON AND GASCUEL, 2003; DARRIBA *ET AL.*, 2012). Two independent analyses were run with one million generations each using the Markov Chain of Monte Carlo (MCMC) algorithm. Trees were sampled every 1,000 generations, and 25% of the generations were discarded as burn-in once stability in the likelihood values was attained. A half-compatible consensus Bayesian tree was computed from

the 750 posterior probability saved trees. MP analysis was based on heuristic searches of 10,000 random-order-entry trees, with TBR branch swapping and saving no more than 10 trees of length equal to or shorter than 10 per replicate. The most parsimonious trees were used to compute the respective strict consensus trees. Branch support was estimated through 1,000 bootstrap replicates (FELSENSTEIN, 1985) using the TBR-M (Tree Bisection Reconstruction swapping, MULPARS OFF) strategy of DEBRY AND OLMSTEAD (2000) as a method to reduce computational time. Following the criteria of MASON-GAMER AND KELLOGG (1997) and INDA *ET AL.* (2008), clades with bootstrap support values (BS) of 75-100% (PAUP 4b.10-86-macosx) or Posterior Probability support values (PPS) of 90-100% (MrBayes 3.1.2) were considered moderately to strongly supported.

7.3 Supplementary literature

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC Bioinformatics* 20, 407-415.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British and European flora. Botanical Society of the British Isles*, London, pp. 45-70.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest2: more models, new heuristic and parallel computing. *Nature methods* 9, 772.

DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology* 49, 171-179.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.

G.P.W.G., The Grass Phylogeny Working Group., 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large

phylogenies by maximum-likelihood. *Systematic Biology* 52, 696-704.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Inda, L.A., Sanmartín, I., Buerki, S., Catalan, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (Festuca subgenus Schedonorus and Lolium). *Journal of Biogeography* 41, 600-614.

Mason-Gamer, R.J., Kellogg, E.A., 1997. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Systematic Biology* 45, 524-545.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid trnT-F and nuclear ITS sequences. *American Journal of Botany* 94, 1554-1596.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *BMC Bioinformatics* 19, 1572-1574.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *mat*K gene-3' *trn*K exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

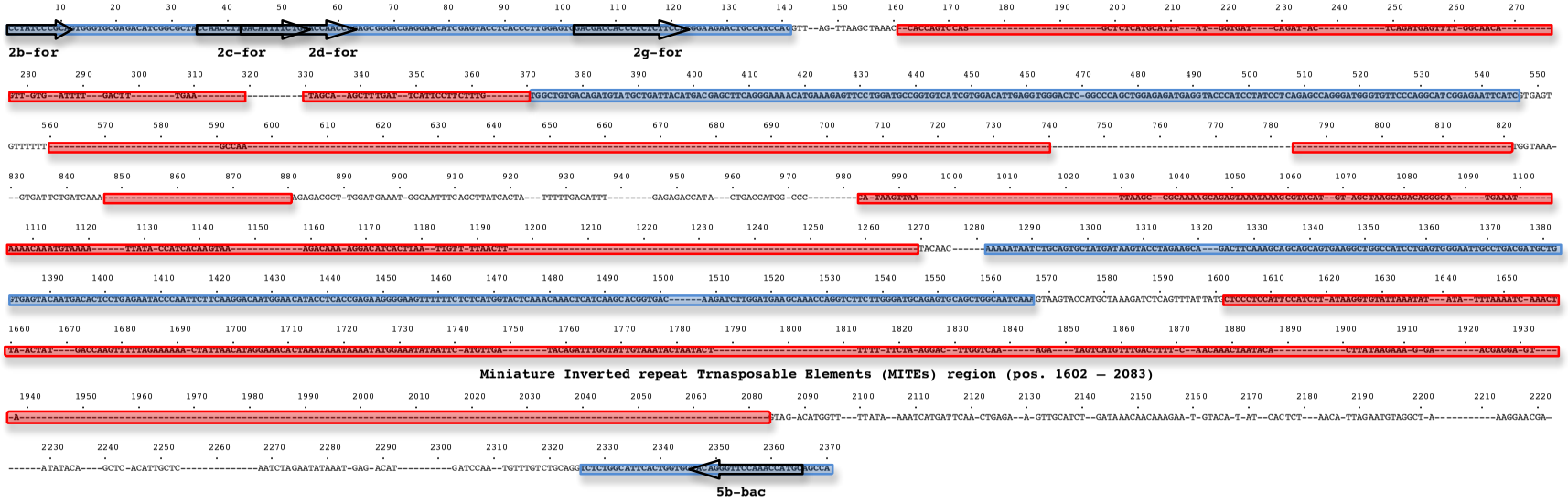
Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., (Eds.), 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

7.4. Supplementary figures



Primer	Sequence
2b-for	GTCAACATCCCNATCCCRCA
2c-for	SCGAKCCNGACATTTTCTRC
2d-for	GACATTTTCTRCACCAACCG
2g-for	GAYGACSAGCCTCTCTTCMR
5b-bac	TACAGGGTTCCAAACCATGC

Figure S1. Partial sequence of the *ß-amylase* gene from the Poaceae *Puccinellia distans* (Poinae – Puccinelliinae – Alopecurinae group). Gaps were introduced after aligning this sequence with the remaining grass sequences analyzed in this research. Nucleotide position is indicated above the sequence. Blue rectangles indicate the locations of the exon regions (from 2 to 5). Red rectangles indicate the location of the intron regions that were ambiguous in terms of alignment and excluded from recombination, evolutionary and phylogenetic analyses. The arrows show the location of the primers designed for this study. The sequences of these new primers are shown in a table below the figure (see Materials and Methods for more details).

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Joinvilleaceae Tomi. and A.C.Sm.							
Joinvillea Gaudich. ex Brongn. & Gris spp.	E. A. Kellogg collection, GenBank	JX536496	AF019784.1	JQ972979	JQ972955	U21973.1	AF164380.1
Poaceae Bamhart							
BEP							
Ehrhartoideae Link							
Oryzeae Dumort.							
Oryzinae Griseb.							
Leersia Sw. spp.	USA. Indiana: Bloomington, GenBank	JX536520	AF019793.1	JQ972993	JQ972960	FJ908350.1	AF148677.1
Oryza L. spp.	GenBank	L10346.1	EF221613.1	DQ415935.1	EF137577.1	FJ908349.1	AF148650.1
Bambusoideae Luerss.							
Olyreae Kunth ex Spenn.							
Parianinae Hack.							
Pariana radiciflora Sagot ex Döll (unknown)	USA. Cultivated in Chicago Botanic Garden, GenBank	JX536521	-	FJ644168.1	-	KC020504.1	AF164387.1
Pooideae Benth.							
Lygeeae J. Presl / Nardeae W.D.J. Koch							
Lygeum spartum L. (2x)	Spain. Zaragoza: Vedado de Peñaflor, UZ 136.07, GenBank	JX536522	FM179417.1	-	EU434098.1	JX438110	JX438063
Nardus stricta L. (2x)	Spain. Granada: Sierra Nevada. UZ 229.07, GenBank	JX536523	FM179420.1	JQ972994	EU434097.1	JX438109	JX438062
Stipeae Dumort.							
Achnatherum calamagrostis (L.) P. Beauv. (2x)	Spain. Huesca: Sierra Guara. UZ 282.07, GenBank	JX536525	EU194678.1	JQ972996	EU200806.1	JX438111	JX438064
Nassella (Trin.) E. Desv. spp.	Chile. Cultivated in IBB, GenBank	JX536524	JQ972931	JQ972995	DQ887454.1	AF251450.1	AF164406.1
Stipa offneri Breistr. (2x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 21.2000, GenBank	JX536527	FN434568.1	JQ972997	JQ972961	JX438112	AF164407.1
Diarrheneae C.S. Campb.							
Diarrhena P. Beauv. spp.	USA. Indiana: Belmont. UZ 116.08, GenBank	JX536528	FM179400.1	JQ972998	JQ972962	DQ786833.1	FM253123.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Brachypodieae Harz							
Brachypodium boissieri Nyman (2x)	Spain, Granada: Sierra Nevada: Dornajo. UZ 211.07, GenBank	JX536538	JN187607.1	JQ973900	JQ972964	JX438113	JX438065
Brachypodium distachyon (L.) P.Beauv. (2x) clone 1 – 9	Spain. Caceres: Malpartida. UZ 28.07	JX536529-JX536537	JN187616.1	JQ972999	JQ972963	U71043.1	AM234568.1
Brachypodium pinnatum (L.) P.Beauv. (2x)	South Africa. Western Cape: Cape Town. Gardens, GenBank	JX536539	AF019782.1	-	JN187671.1	AY622312.1	DQ786891.1
Bromeae Dumort. - Triticeae Dumort.							
Bromeae Dumort.							
Bromus L. spp. *	GenBank	AY821734.1	U83354.1	EU036162.1	EU036187.1	DQ786822.1	DQ786894.1
Triticeae Dumort.							
Aegilops comosa Sibth. & Sm. (2x)	GenBank	AY821696.1	AF149201.1	EU013754.1	EU013515.1	DQ247879.1	HM770818.1
Aegilops geniculata Roth (2x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 138.07, GenBank	JX536544	AF156998.1	EU013843.1	EU013604.1	JX438116	JX438067
Aegilops uniaristata Vis. (2x)	GenBank	AY821691.1	AF149200.1	EU013922.1	EU013684.1	DQ247910.1	EU013209.1
Agropyron cristatum (L.) Gaertn. (2x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 134.07, GenBank	JX536542	JQ972932	AF519115.1	AY740792.1	JX438114	JX438066
Australopyrum retrofractum (Vickery) Á. Löve (2x)	GenBank	AY821692.1	EU617249.1	-	EU617319.1	AF267662.1	HM540008.1
Australopyrum velutinum (Nees) B. K. Simon (2x)	GenBank	AY821693.1	-	AF519119.1	EU617319.1	DQ247878.1	HM770817.1
Crithopsis delileana (Schult.) Roshev. (unknown)	GenBank	AY821694.1	L36487.1			DQ247895.1	HM770829.1
Dasypyrum villosum (L.) P.Candargy (2x)	GenBank	AY821699	JQ972933	-	JQ972965	JX438115	HM770820.1
Elymus caucasicus (K. Koch) Tzvelev (2x)	GenBank	GQ847690.1	AY740809.1	DQ159289.1	DQ159289.1	AF267667.1	HQ652701.1
Elymus hystrix L. (2x)	USA. Indiana: Belmont. UZ 113.08, GenBank	JX536543	EF396973.1	-	EF396985.1	-	HM352813.1
Elymus repens (L.) Gould (6x)	GenBank	EU282245.1	GQ373267.1	DQ914534.1	DQ912406	AF267682.1	HM850576.1
Eremopyrum bonaepartis (Spreng.) Nevski (2x)	GenBank	AY821700	-	-	-	-	-
Eremopyrum triticeum (Gaertn.) Nevski (2x)	GenBank	-	-	-	-	DQ247883.1	HM770822.1
Henrardia persica (Boiss.) C.E.Hubb. (unknown)	GenBank	AY821703.1	L36491.1	-	-	DQ247885.1	HM770824.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Triticeae Dumort.							
Heteranthelium piliferum Hochst. (2x)	GenBank	AY821704.1	L36492.1	AF519153	-	DQ247886.1	HM770825.1
Hordeum marinum Huds. (2x)	GenBank	EU282257.1	KC193784.1	KF600707.1	KF600707.1	HQ619401.1	HM850542.1
Lophopyrum Á. Löve spp. (2x)	GenBank	AY821731.1	EF014249.1	-	EU139481.1	DQ247892.1	HM770826.1
Peridictyon sanctum (Janka) Seberg, Fred. & Baden (2x)	GenBank	AY821714.1	L36497.1	AF519154.1	AF519154.1	DQ247894.1	HM770828.1
Psathyrostachys juncea (Fisch.) Nevski (2x)	GenBank	AY821716.1	EF581984.1	-	EF581911.1	AF267707.1	AB078140.1
Pseudoroegneria strigosa (Bieb.) Löve (2x)	GenBank	EU282267.1	AY740795.1	EU036137.1	EU617284.1	AF267712.1	HQ652698.1
Secale cereale L. (2x)	GenBank	AY821723	AF303400.1	DQ336856.1	AF478501.1	U71022.1	DQ420046.1
Taeniatherum caput-medusae (L.) Nevski (2x)	GenBank	AY821726	L36505.1	-	-	DQ247902.1	EF137512.1
Thinopyrum bessarabicum (Săvul. & Rayss) Á. Löve (2x)	GenBank	AY821730.1	L36506.1	AF519165.1	AF519165.1	DQ247903.1	HM770837.1
Triticum aestivum L. (2x)	Spain, Huesca. cultivated from seeds, GenBank	JX536540	FM998928.1	JQ973001	AF148757.1	DQ247921.1	DQ420028.1
Triticum durum Desf. (2x)	Spain, Huesca. cultivated from seeds, , GenBank	JX536541	HQ285920.1	-	AB732942	DQ247920	-
Poeae R.Br – Aveneae Dumort.							
Aveneae Dumort.							
Phalaridinae Fr.							
Phalaris canariensis L. (2x)	Spain. Valencia: Albal. UZ 338.07, GenBank	JX536545	FJ178782.1	DQ631509.1	DQ631443.1	JX438117	JX438068
Phalaris coerulescens Desf. (2x)	Spain. Badajoz: La Albuera. UZ 50.07GenBank	JX536546	JF951081.1	DQ631510.1	DQ631444.1	HE575745.1	FN908066.1
Agrostidinae Fr.							
Agrostis capillaris L. (2x)	Spain. Lugo: Rábade. UZ 299.07, GenBank	JX536549	FM179384.1	AY450936.1	JQ972966	JX438119	JX438069
Ammophila arenaria (L.) Link (2x)	Spain. Huelva: Mazagón. UZ 75.07, GenBank	JX536547	DQ539590.1	DQ631522.1	DQ631456.1	JX438118	AM234561.1
Chaetopogon fasciculatus (Link) Hayek (2x)	Spain. Huelva: El Rocío. UZ 77.07, GenBank	JX536550	DQ539593.1	DQ631523.1	DQ631457.1	JX438121	-
Polypogon monspeliensis (L.) Desf. (unknown)	Spain. Badajoz: Entrin Alto-Entrin Bajo. UZ 52.07	JX536552	-	-	-	-	-
Polypogon maritimus Willd. (unknown)	Spain. Badajoz: Entrin Alto-Entrin Bajo. UZ 52.07, GenBank	JX536551	DQ336818.1	DQ336863.1	DQ336838.1	JX438122	JX438071
Triplachne nitens Link (2x)	Spain. Almeria: Cabo de Gata. UZ 365.07, GenBank	JX536548	DQ336816.1	DQ336861.1	DQ336836.1	JX438120	JX438070

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Anthoxanthinae A. Gray							
Anthoxanthum aristatum Boiss. (2x)	Spain. Caceres. Ctra.Navalmoral. UZ 24.07, GenBank	JX536553	JQ972934.1	DQ631515.1	DQ631449.1	JX438123.1	-
Anthoxanthum odoratum L. (2x)	Spain. León: Ancares,Cuiña. UZ 287.07, GenBank	JX536554	FJ010630.1	-	EF137590.1	DQ786812.1	HM850562.1
Aveninae J. Presls							
Arrhenatherum elatius (L.) P. Beauv. ex J. Presl & C. Presl (2x)	Spain. Badajoz: Embalse de la Tentudis. UZ 55.07, GenBank	JX536558	FM179388.1	DQ336866.1	EF137591.1	JX438126	JX438074
Avena fatua L. (unknown)	Masson-Gamer, R., GenBank	JX536556	FJ794719.1	JF904749.1	JF904749.1	U71018.1	HQ593189.1
Avena sterilis L. (2x)	Spain.Caceres: Navalmoral. UZ 19.07, GenBank	JX536555	EU833817.1	JQ973002	EU833883.1	JX438124	JX438073
Helictotrichon Besser spp.	Spain. Valencia: Requena. UZ 346.07, GenBank	JX536557	DQ336819.1	DQ336864.1	DQ336839.1	JX438125	DQ786920.1
Pseudarrhenatherum longifolium (Thore) Rouy (2x)	Spain. Coruña: Castro Elviña. UZ 298.07, GenBank	JX536559	FM179430.1	JQ973003	JQ972967	JX438127	JX438075
Koeleriinae Asch. & Graebn.							
Avellinia michelii (Savi) Parl. (2x)	Spain. Cádiz: Sanlucar de Barrameda. UZ 98.07, GenBank	JX536563	JQ972935	DQ631531.1	DQ631465.1	JX438131	JX438079
Koeleria vallesiana (Honck.) Gaudin (2x)	Spain. Madrid: Torrelaguna-El Verrueco. UZ 4.07, GenBank	JX536561	DQ539601.1	DQ631536.1	DQ631468.1	JX438129	JX438077
Rostraria cristata (L.) Tzvelev (2x)	Spain. Toledo: Ontígola. UZ 9.07, GenBank	JX536562	DQ336833.1	DQ336879.1	DQ336853.1	JX438130	JX438078
Trisetum loeflingianum (L.) P. Beauv. (unknown)	Spain. Toledo: Ontígola. UZ 10.07, GenBank	JX536560	DQ539608.1	DQ631539.1	DQ631473.1	JX438128	JX438076
Poinae Dumort. – Puccinelliinae Soreng & J.I. Davis – Alopecurinae Dumort.							
Alopecurus L. spp.	Spain. Badajoz: Entrin Alto-Entrin Bajo. UZ 53.07, GenBank	JX536567	EU792345.1	JQ973004	JQ972968	JX438133	JX438080
Catabrosa aquatica (L.) P. Beauv. (2x)	Spain. Huesca: Formigal. UZ 55.07, GenBank	HE565906.1	DQ539565.1	JQ973005	JQ972969	JX438135	JX438061
Catabrosella araratica (Lipsky) Tzvelev (unknown)	GenBank	HE565912.1	AJ196300.1	-	-	-	-
Colpodium drakensbergense Hedberg & I. Hedberg (unknown)	GenBank	HE565915.1	-	-	-	-	-
Colpodium versicolor (Steven) Schmalh. (unknown)	GenBank	HE565911.2	AJ867445.1	-	-	-	FM253122.1
Milium vernale M. Bieb. (2x)	GenBank	HE565910.1	EU792340.1	DQ353963.1	DQ353963.1	DQ786856.1	DQ786928.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Poinae Dumort. – Puccinelliinae Soreng & J.I. Davis – Alopecurinae Dumort.							
Paracolpodium (Tzvelev) Tzvelev spp.	GenBank	HE565909.1	EF432735.1	-	-	-	-
Phleum pratense L. (2x)	GenBank	HE565899.1	HQ600524.1	DQ353964.1	DQ353964.1	DQ786860.1	HQ593382.1
Poa bulbosa L. (2x)	Spain. Alicante: Alcoy, Font Roja. UZ 334.07 GenBank	JX536565	GQ324492.1	DQ354038	DQ354039.1	AY589098.1	HE966970.1
Poa infirma Kunth (2x)	Spain. Caceres: Tornavacas. UZ 37.07, GenBank	JX536566	AF393012.1	DQ367407.1	AF488773.1	JX438132	JX438058
Puccinellia distans (Jacq.) Parl. (2x)	Spain. Barcelona: Prat de Llobregat. UZ 156.07, GenBank	JX536564	AF532934.1	DQ336859.1	AF533024.1	JX438134	JX438081
Zingeria biebersteiniana (Claus) P.A. Smirn. (unknown)	GenBank	HE565930.1	FJ196301.1	-	-	-	-
Zingeria trichopoda (Boiss.) P.A. Smirn. (unknown)	GenBank	-	FM179441.1	-	-	-	AM234551.1
Airinae Fr. - Holcinae Dumort. - Sesleriinae Parl.							
Corynephorus canescens (L.) P. Beauv. (2x)	Spain. Avila: Puerto de Tornavacas. UZ 35.07, GenBank	JX536568	DQ539578.1	DQ631506.1	EF137598.1	JX438136	JX438082
Deschampsia antarctica É.Desv. (unknown)	GenBank	HE565901.1	JF904807.1	EU792463.1	EU792463.1	-	-
Deschampsia cespitosa (L.) P. Beauv. (2x)	Spain. Lugo: Curtis, Sobrado dos Montes. UZ 305.07. GenBank	JX536569	HQ114385.1	DQ631507.1	DQ631441.1	U71012.1	KC474527.1
Deschampsia flexuosa (L.) Trin. (2x)	Spain. Lugo: Xistral-Pena Vella. UZ 313.07, GenBank	JX536570	JQ972936	DQ631505.1	AY237913.1	JX438137	JX438083
Echinaria capitata (L.) Desf. (2x)	Spain. Toledo: Ontígola. UZ 13.07, GenBank	JX536572	JQ972938	DQ631519.1	DQ631453.1	JX438138	AM234599.1
Holcus lanatus L. (2x)	Spain. Leon: Sierra. Ancares. UZ 291.07, GenBank	JX536571	JQ972937	DQ631503.1	EF137606.1	DQ786849.1	JX438084
Avenula *sensu stricto*							
Avenula bromoides (Gouan) H. Scholz (unknown)	Spain. Zaragoza: Vedado de Peñaflor. UZ 133.07, GenBank	JX536578	JQ972940	DQ631525.1	DQ631459.1	JX438142	JX438087
Dactylidinae Stapf / Parapholiinae Caro – Cynosurinae Fr.							
Cynosurus cristatus L. (2x)	Spain. Leon: Sierra Ancares. UZ 285.07, GenBank	JX536573	JQ972939	JQ973006	EF137599	JX438139	DQ786901.1
Dactylis hispanica Roth (2x)	Spain. Zaragoza: Peñaflor. UZ 116.07	JX536575	AF393014.1	DQ353961.1	DQ353961.1	-	-
Desmazeria rigida (L.) Tutin (2x)	Spain. Zaragoza: La Alfranca. UZ 124.07, GenBank	JX536577	FM179399.1	JQ973007	JQ972970	JX438141	JX438086

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Dactylidinae Stapf / Parapholiinae Caro – Cynosurinae Fr.							
Lamarckia aurea (L.) Moench (2x)	Spain. Caceres: Tornavacas. UZ 36.07, GenBank	JX536576	AF532936.1	DQ631490.1	AF533029.1	JX438140	JX438085
Parapholis incurva (L.) C. E. Hubb. (4x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 127.07, GenBank	JX536574	FM179422.1	DQ631491.1	AF533036.1	DQ786859.1	EF137508.1
Loliinae Dumort.							
Broad-leaved Loliinae							
***Festuca* subgen. *Drymanthele* – *F.* subgen. *Leucopoa* – *F.* sect. *Lojaconoa* Gand. – *F.* sect. *Pseudoscariosa* Krivot. – *F. scabra* Vahl - *F.* sect. *Scariosae* Hack.**							
Festuca altaica Trin. (4x) clone 1 – 7	Canada. Yukon Territory: Teslin Lake. GenBank	JX536665-JX536671	JQ972953	EF585002.1	JQ972977	-	FN668464.1
Festuca altissima All. (2x)	Norway. Akershus: Frogn parish, GenBank	JX536583	FM179404.1	EF585003.1	AF478505.1	JX438147	AM234585.1
Festuca coerulescens Desf. (2x)	Spain. Cádiz: Jerez de la Frontera. UZ19.08, GenBank	JX536581	JQ972941	EF585027.1	JQ972971	JX438145	JX438089
Festuca modesta Nees ex Steud. (2x)	China. Yunnan. Fugong. GenBank	JX536579	EF584953.1	EF585068.1	EF592985.1	JX438143	-
Festuca pseudoeskia Boiss. (2x)	Spain. Granada: Sierra Nevada. UZ 216.07, GenBank	JX536582	AF519979.1	EF585084.1	AY099000.1	JX438146	JX438090
Festuca scabra Vahl (unknown) clone 1 – 5	Lesotho. Sehlabathebe Nat'l Park. PRE 522214.	JX536672-JX536676	JQ972954	JQ973017	JQ972978	-	JX438108
Festuca scariosa (Lag.) Asch. & Graebn. (2x) clone 1 – 8	Spain. Granada: Sierra Nevada. UZ 219.07, GenBank	JX536584-JX536591	AF519978.1	EF585100.1	AY098999.1	JX438148	-
Festuca triflora J.F. Gmel. (2x)	Spain. Ciudad Real: Viso del Marqués. UZ 166.07, GenBank	JX536580	AF538362.1	DQ631483.1	AF533052.1	JX438144	JX438088
***Lolium* L. – *Micropyropsis* Romero Zarco & Cabezudo – *Festuca* subgen. *Schedonorus* (P. Beauv.) Peterm. – *F.* sect. *Subbulbosae* Nyman ex Hack.**							
Festuca arundinacea Schreb. (6x) clone 1 – 7	Spain. Lugo: Láncara. GenBank	JX536609-JX536615	AJ240154.1	DQ367405.1	AF533042.1	JX438151	JX438093
Festuca fenas Lag. (4x) clone 1 – 7	France: ABY BN354. UZ 142.07	JX536602-JX536608	JQ972943	JQ973009	JQ972973	JX438150	JX438092
Festuca gigantea (L.) Vill. (6x)	Norway, GenBank	JX536619	AF303416.1	JQ973012	AF533043.1	JX438154	JX438096
Festuca paniculata subsp. *spadicea* (L.) Litard. (6x) clone 1 – 10	Spain. León: Sierra Ancares. UZ 290.07	JX536592-JX536601	JQ972942	JQ973008	JQ972972	JX438149	JX438091
Festuca pratensis Huds. (2x)	Kazasthan. Cultivated seeds W6-13201. UZ 141.07, GenBank	JX536618	AF303421.1	JQ973011	AF478503.1	JX438153	JX438095
Festuca simensis (4x) Hochst. ex A. Rich. clone	Uganda. Kabale, Kisoro border, Namaganda	JX536677-	-	KJ529291	GU573750	KJ529485	-

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
***Lolium* L. – *Micropyropsis* Romero Zarco & Cabezudo – *Festuca* subgen. *Schedonorus* (P. Beauv.) Peterm. – *F.* sect. *Subbulbosae* Nyman ex Hack.**							
Lolium perenne L. (2x)	UK. Wales. Cultivated seeds PI 619001, GenBank	JX536620	AJ240138.1	DQ367404	AF478504.1	JX438155	JX438097
Micropyropsis tuberosa Romero-Zarco & Cabezudo (unknown) clone 1 – 2	Spain. Huelva: Almonte. UZ89.07, GenBank	JX536616-JX536617	JQ972944	JQ973010	AF533037.1	JX438152	JX438094
Fine-leaved Loliinae							
***Festuca* sect. *Eskia* Willk.**							
Festuca elegans Boiss. (2x, 4x) clone 1 – 7	Spain. Granada: Sierra. Nevada. UZ 231.07, GenBank	JX536621-JX536627	AF303406.1	EF585038.1	AF478509.1	JX438156	JX438098
Festuca eskia Ramond ex DC. (2x)	Spain. Huesca: Benasque. GenBank	JX536629	AF303412.1	EF585040.1	AF478508.1	JX438158	JX438100
Festuca gautieri (Hack.) K.Richt. (2x)	Spain. Huesca: Panticosa. UZ 152.07, GenBank	JX536628	AF303414.1	EF585044.1	AF478507.1	JX438157	JX438099
***Festuca* subsect. *Exaratae* St-Yves – *Narduroides* Rouy – *Psilurus* Trin. – *Vulpia* sect. *Apalochloa* (Dumort.) Stace – *V.* sect. *Monachne* Dumort. – *V.* sect. *Vulpia* (4x-6x)**							
Festuca capillifolia Dufour ex Roem. & Schult. (2x) clone 1 – 9	Spain. Almeria: Sierra. Gador. UZ 179.07, GenBank	JX536633-JX536641	AF303419.1	EF585022.1	AF478511.1	JX438161	JX438102
Festuca clementei Boiss. (2x)	Spain. Granada: Sierra Nevada. UZ 223.07, GenBank	JX536642	AF478482.1	EF585025.1	AF478524.1	JX438162	JX438060
Narduroides salzmannii (Boiss.) Rouy (2x)	Spain. Madrid: Arganda. UZ 111.07, GenBank	JX536632	JQ972946	JQ973014	AF478535.1	JX438160	JX438101
Psilurus incurvus (Gouan) Schinz & Thell. (4x)	Spain. Caceres: Malpartida de Plasencia. UZ 31.07, GenBank	JX536631	JQ972945	JQ973013	AF478533.1	JX438159	JX438059
Vulpia alopecuros (Schousb.) Dumort. (2x) clone 1 – 5	Portugal. Algave: Meia Praia.	JX536653-JX536657	AF478491.1	EF585117.1	AF487617.1	-	-
Vulpia ciliata Dumort. (4x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 112.07	JX536630	AF478486.1	EF585120.1	AY118105.1	-	JN894723.1
Vulpia fasciculata (Forssk.) Samp. (4x)	Spain. Barcelona: Vilanova. SP 15.2000	JX536643	JQ972948	EF585121.1	AF478528.1	JX438163	JN894722.1
Vulpia geniculata (L.) Link (2x)	Spain. Sevilla: Constantina. JACA J29397, GenBank	JX536644	JQ972947	EF585123.1	AF478531.1	JX438164	-
***Festuca* sect. *Aulaxyper* Dumort. s.l. – *F. abyssinica* Hochst. ex A. Rich.**							
Festuca abyssinica Hochst. ex A. Rich. (4x) clone 1 – 2	Uganda. Elgon Mnt. MHU1566	JX536663-JX536664	JQ972952	JQ973016	JQ972976	JX438171	JX438107
Festuca iberica (Hack.) K. Richt. (6x)	Spain. Granada: Sierra Nevada. UZ 218.07, GenBank	JX536662	AY118087.1	EF585052.1	AF478516.1	JX438170	-
Festuca rubra L. (6x, 8x) clone 1 – 5	UK: Scotland. Cultivated in Huesca. UZ 150.07,	JX536658-	AJ240158.1	EF585097.1	AY118099.1	JX438169	JX438106

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008), SCHNEIDER *ET AL.* (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL.* (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Festuca* sect. *Festuca s.s.							
Festuca hystrix Boiss. (2x) clone 1 – 5	Spain.Almeria.UZ185.07 GenBank	JX536646-JX536650	AF478480.1	EF585051.1	AF478520.1	JX438168	-
Festuca indigesta Boiss. (6x)	Spain. Granada: Sierra Nevada. UZ 212.07, GenBank	JX536651	AF303426.1	EF585054.1	AF478519.1	JX438167	JX438105
Festuca longiauriculata Fuente, Ortúñez & Ferrero (2x)	Spain. Almería: Sra. Filabres. UZ- 59.2000 y UZ192.07	JX536652	AF478479.1	EF585062.1	AF478518.1	-	-
Festuca ovina L. (2x)	Germany. Thüringen: Saale-Holzland-Kreis. Müller 6879, GenBank	JX536684	JQ972950	EF585076.1	JQ972975	JX438166	JX438104
Festuca panciciana K. Richt. (unknown) clone 1 – 2	Bosnia-Hercegovina. Troglav: Sajkovacko zdrlo.	JX536685-JX536686	JQ972951	JQ973015	JQ972974	-	-
Wangenheimia lima (L.) Trin. (2x)	Spain. Zaragoza: Vedado de Peñaflor. UZ 113.07, GenBank	JX536645	JQ972949	EF585131.1	AF478536.1	JX438165	JX438103
PACCMAD							
Centothecoideae Soderstr.							
Centotheceae Ridl.							
Thysanolaena maxima (Roxb.) Kuntze (unknown)	UK. Cultivated in Kew Garden. 1979-3225, GenBank	JX536497	AF019854.1	JQ972980	EF137520.1	TMU21984	EF137433.1
Panicoideae Link							
Gynerieae Sanchez-Ken & L.G. Clark							
Gynerium sagittatum (Aubl.) P. Beauv. (unknown)	Peru. Cultivated in Kew Garden 1991-1276, GenBank	JX536498	AF019858.1	JQ972981	JQ972956	AY847120.1	EF137431.1
Paniceae R.Br.							
Panicinae Fr.							
Panicum miliaceum L. (unknown)	GenBank	JX536499	FJ606748.1	JQ972982	JQ972957	AY188472.1	FR667662.1
Panicum virgatum L. (unknown)	E. A. Kellogg collection, GenBank	JX536500	DQ005062	-	-	U21986.1	EU434294.1
Setariinae Dumort.							
Setaria P. Beauv. spp.	USA. Indiana: Bloomington. , GenBank	JX536501	GQ870175.1	JQ972983	GU561492.1	AY029678.1	EF137472.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL*. (2003, 2007), TORRECILLA *ET AL*. (2003, 2004), CATALAN *ET AL*. (2006), QUINTANAR *ET AL*. (2007), BOUCHENAK-KHELLADI *ET AL*. (2008), INDA *ET AL*. (2008), SCHNEIDER *ET AL*. (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL*. (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Andropogoneae Dumort.							
Tripsacinae Dumort.							
Zea mays L. (2x)	GenBank	AF068119.1	ZMU46616	EF541347.1	EF541269.1	U21985.1	X86563
Rottboelliinae J. Presl							
Rottboellia L. f. spp.	E. A. Kellogg collection, GenBank	JX536502	DQ005063.1	JQ972984	JQ972958	-	-
Danthonioideae H.P. Linder & N.P. Barker							
Danthonieae Zotov							
Cortaderiinae Conert / Danthoniinae Fr.							
Danthonia decumbens (L.) DC. (2x)	Spain. Cádiz: Jerez de la Frontera. UZ 31.08, GenBank	JX536507	EU401308.1	JQ972986	EU401189.1	EU400785.1	JX438057
Danthonia spicata (L.) P. Beauv. ex Roem. & Schult. (2x)	USA. Indiana: Belmont, Yellowwood State Forest. UZA 112.08	JX536506	DQ887163.1	-	EU401198.1	EU400789.1	HQ593264.1
Karroochloa Conert & Türpe spp.	P. Linder 5360 (BOL), GenBank	JX536509	DQ655793.1	JQ972988	DQ655852.1	AF251458.1	EU400735.1
Merxmuellera macowanii (Stapf) Conert (2x)	USA. Cultivated in Chicago Botanic Garden	JX536505	AF019863.1	-	EU401213.1	AM849184.1	EU400745.1
Merxmuellera rangei (Pilg.) Conert (2x)	N. Barker 960 (GRA), GenBank	JX536504	AF019862.1	JQ972985	EU401215.1	EU400803.1	EU400747
Schismus barbatus (L.) Thell. (unknown)	Spain. Zaragoza: Vedado de Peñaflor, UZ 139.07, GenBank	JX536508	JQ972930	JQ972987	JQ972959	GQ471663.1	EU400767.1
Stipagrostis zeyheri (Nees) De Winter (unknown)	N. Barker 1133 (BOL)	JX536503	AF019845.1	-	DQ172301.1	GU222711.1	HE573958.1
Chloridoideae Kuth ex Beilschm.							
Eragrostideae Stapf							
Eragrostidinae J. Presl							
Eragrostis curvula (Schrad.) Nees (4x)	L. Clark 1303 (ISC)	JX536511	AF498394.1	JQ972990.1	DQ655870.1	U21989.1	HE586097.1
Eragrostis spectabilis (Pursh) Steud. (2x)	USA. Cultivated in Chicago Botanic Garden	JX536513	-	-	-	HE577868.1	HE577884.1
Eragrostis Wolf spp.	Masson-Gamer, R., GenBank	JX536512	DQ655822.1	JQ972990	DQ655877.1	U21989.1	AF312341.1
Cotteinae Reeder							
Enneapogon P. Beauv. spp.	Columbus 3133, GenBank	JX536510	EF153043.1	JQ972989	EF156692.1	GU359699.1	AF312360.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL*. (2003, 2007), TORRECILLA *ET AL*. (2003, 2004), CATALAN *ET AL*. (2006), QUINTANAR *ET AL*. (2007), BOUCHENAK-KHELLADI *ET AL*. (2008), INDA *ET AL*. (2008), SCHNEIDER *ET AL*. (2009), and GPGW (2012). spp. indicates that sequences from more than one species have been collected for that genus. Information on ploidy levels of Loliinae taxa is taken from INDA *ET AL*. (2008, 2014) and the Index to Plant Chromosome number (http://www.tropicos.org/Project/IPCN).

Taxa	Source	GenBank accession					
		ß-amylase	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Zoysieae Benth.							
Muhlenbergiinae Pilg.							
Muhlenbergia montana (Nutt.) Hitchc. (4x)	Columbus 3375, GenBank	JX536519	EF153067.1	JQ972991	EF156716.1	HM143488.1	AF144600.1
Muhlenbergia tenuifolia (Kunth) Kunth (6x)	Columbus 3269	JX536518	GQ397908.1	JQ972991.1	GQ397962.1	HM143519.1	-
Sporobolinae Benth.							
Spartina Schreb. spp.	E. A. Kellogg collection, GenBank	JX536517	EF153082.1	AF372625.1	EF137568.1	AF251465.1	AF312353.1
Sporobolus heterolepis (A. Gray) A. Gray (unknown)	USA. Cultivated in Chicago Botanic Garden	JX536516	-	-	-	-	HQ713377.1
Sporobolus indicus (L.) R.Br. (unknown)	L. Clark 1293 (ISC), GenBank	JX536514	EU646098.1	JQ972992	EF137569.1	U21983.1	AF144601
Sporobolus wrightii Munro ex Scribn. (unknown)	Columbus 2507	JX536515	EF153085.1	-	EF156734.1	GU359668.1	-

CHAPTER III

HORIZONTAL GENE TRANSFER AND PSEUDOGENEIZATION OF THE MITOCHONDRIAL *RPS3* GENE IN POACEAE

1. Abstract

- Angiosperms' mitochondrial DNA is the most complex and the least understood of the three plant cell genomes. The evolution of plant mtDNA is characterized by complex structural rearrangements and recombinations, which eventually lead to chimeric molecules.
- In this chapter, we analyze the molecular evolution of the mitochondrial *rps3* gene throughout a wide selection of Poaceae species, evaluating the potential existence of Horizontal Gene Transfer (HGT) and other evolutionary events through visual inspection of the sequences, Recombination Detection (RDP4) approaches, topological (Shimodaira-Hasegawa (SH) tests, Premature Terminal Codon (PTC) detection and Branch-Specific Branch-Site REL (Random Effects Likelihood) modeling.
- The highly chimeric nature of the mitochondrial *rps3* gene in the grasses was confirmed by 12 cases of microconversions (short gene conversions, 1 – 140 nucleotide length *rps3* sequences, shared by phylogenetically unrelated accessions), which involved 123 grass taxa (87.8% of the tested sequences), and three long recombinations, which involved 20 grass taxa (16.13%). The comparison of the *rps*3 phylogeny with a Poaceae organismal tree based on nuclear and plastid sequences revealed the existence of 38 misplacements (26.4%) within the mitochondrial reconstruction, which was corroborated by the topological incongruence detected among these phylogenies. Thirty-three *rps*3 sequences (48.5%) showed high proportions of sites with neutral or slightly positive selection and no purifying selection, suggesting the existence of pseudogenic copies or relaxed selection pressure.
- The potential occurrence of different HGT events in the evolutionary history of Poaceae is the most likely hypothesis to explain the microconversions, recombinations and misplacements observed within the mitochondrial *rps3* gene of the grasses. Pseudogeneization, or the relaxation of evolutionary pressure within the mitochondrial *rps*3 copy would explain the absence of purifying selection detected in half of grass sequences tested.

2. Introduction

Five hundred million years of land plant evolution resulted in relatively few alterations to the chloroplast genome but considerable changes in the mitochondrial genome (mtDNA) (KNOOP, 2004). The angiosperms' mitochondrial genome is among the least understood, in part because it is the largest and most complex of the three plant cell genomes (WOLFE *ET AL.*, 1987; EYRE-WALKER AND GAUT, 1997). However, its gene content is much less than that of a eubacterial genome (a possible ancestral symbiont), suggesting the occurrence of intracellular gene transfers (IGT) during the course of cell evolution (GRAY *ET AL.*, 1992; MARIENFELD *ET AL.* 1999; and KOULINTCHENKO *ET AL.*, 2003). Non-coding regions, intergenic spacer regions, pieces of foreign DNA of unknown origin, many small repeated sequences, and regulatory elements such as promoter and terminator elements, occupy a majority of the plant mitochondrial genome (ADAMS AND PALMER, 2003; KUBO AND MIKAMI, 2007; KUBO AND NEWTON, 2008). The evolution of the plant mitochondrial genome is characterized by a very low substitution rate, complex structural rearrangements, and Horizontal Gene Transfer (HGT), which eventually lead to gene conversion in foreign and native sequences (PALMER AND HERBON, 1988; ALBERT *ET AL.*, 1998; BERGTHORSSON *ET AL.*, 2003; KUBO AND MIKAMI, 2007; KUBO AND NEWTON, 2008). These events result in the generation of chimeric molecules and pseudogenes and the acquisition of sequences from non-related taxa (cf. NEWTON *ET AL.*, 2004; BARKMAN *ET AL.*, 2007; HAO AND PALMER, 2009; WEILONG AND PALMER, 2009).

The *rps3* (*ribosomal protein S3*) gene is widespread in the mitochondrial genome of protists, fungi and plants (GRAY *ET AL.*, 1998; LANG *ET AL.* 1999), though it has been lost from the mtDNA of a variety of fungal and angiosperm lineages (cf. ADMS ET AL., 2002; ADAMS AND PALMER, 2003). The *rps3* gene is also present in the chloroplast genome of higher plants (LAROCHE AND BOUSQUET, 1999), but with the characteristic absence of a 1649-bp intron included in the mitochondrial *rps3* gene of flowering plants (BOCK *ET AL.*, 1994; LAROCHE AND BOUSQUET, 1999). The singular features of the mitochondrial genome, and specifically the *rps3* gene, drove our interest in analyzing their molecular evolution, evaluating the potential existence of HGT and other evolutionary events throughout a wide selection of Poaceae species.

The evolutionary fixation of foreign DNA which replaces or overlaps native sequences through recombination has been extensively described by BERGTHORSSON *ET AL.* (2003, 2004), WON AND RENNER (2003), NICKRENT *ET AL.* (2004), WOLOSZYNSKA *ET AL.* (2004), RICHARDSON AND PALMER (2007), HAO AND PALMER (2009), and HAO *ET AL.* (2010), among others.

Specifically, recombination via short sequences is a key evolutionary process, likely the most common and least recognized type of recombination, and plays a central role in the structural evolution of angiosperm mitochondrial genomes (ANDRE *ET AL.*, 1992; CONKLIN AND HANSON, 1994; HAO AND PALMER, 2009). Many of the reported plant mitochondrial pseudogenes are thought to have arisen via short sequence motifs (c.f. HANSON AND FOLKERTS, 1992; CONKLIN AND HANSON, 1994; WOLSTENHOLME AND FAURON, 1995).

We also investigated the potential deviation from the expected purifying selection of the *rps3* exonic sequences within the studied Poaceae as a method of detecting evidence of pseudogenization. This method is based on a site-by-site approach, using the generalized branch-site methods of KOSAKOVSKY-POND *ET AL.* (2011) and fitting a distribution of substitution rates across sites (random effect model = REL; NIELSEN AND YANG, 1998), thus inferring the rate at which individual sites evolve given this distribution (KOSAKOVSKY-POND ET AL., 2005). Theory predicts that the number of substitutions leading to adaptive amino acid changes (non-synonymous mutations, d_N) should be significantly larger than the number of substitutions at synonymous sites (d_S) (MIYATA AND YASUNAGA, 1980; LI, 1997). Conversely, if selection consistently weeds out amino acid changes, then d_N should be significantly smaller than d_S. Thus, whereas an estimated d_N/d_S value near 1 suggests neutrality, a d_N/d_S greater than or less than 1 indicates positive or negative directional selection, respectively (reviewed by YANG AND BIELAWSKI, 2000; PRESTON AND KELLOGG, 2006; BRUNNER *ET AL.*, 2009).

The grass family (Poaceae) is one of the largest and most diverse families of flowering plants. It contains approximately 10,000 species and 600-700 genera (CLAYTON AND RENVOIZE, 1986; WATSON AND DALLWITZ, 1992). Previous evolutionary studies of the Poaceae (e. g. BOUCHENAK-KHELLADI *ET AL.*, 2008; SCHNEIDER *ET AL.*, 2009; MINAYA *ET AL.*, 2013, and references therein) have provided valuable insights into the phylogeny of the family and the structure of the nuclear and plastid grass genomes. However, the mitochondrial genome belongs to a separate group with its own singular background. The main goal of our study was to determine whether the grass *rps3* sequences might have experienced HGT and pseudogeneization events, similarly to other mitochondrial genes, and to identify them. For this, the specific aims of our study were: (i) to detect whether recombination is present in the studied *rps3* grass sequences; (ii) to compare the phylogeny of a multigenic Poaceae organismal tree with the mitochondrial *rps3* reconstruction based on a broad representation of the main Pooideae lineages (with special emphasis on the Poeae-Aveneae), and to detect the potential misplacements in the *rps*3 tree; and (iii) to explore the extent of sequences with a high

proportion of sites under purifying selection in the studied *rps3* grass sequences as an indicator of relaxation of selection pressure, which would reflect the likely existence of pseudogenization.

3. Material and methods

3.1 Taxon sampling and DNA analyses

Taxon sampling was designed to include a broad representation of the Poaceae, with special attention to the Poodae tribe (Poeae – Aveneae) and the more recently evolved Loliinae subtribe (see in the appendix of chapter III; Table S1, page 183). Sampling included 141 grass species classified in 97 genera and 10 subfamilies (GPWG, 2001, 2012; SORENG *ET AL.*, 2007, and http://www.tropicos.org/): Anomochlooideae: 2 species, Centothecoideae: 2, Panicoideae: 11, Micrairoideae: 3, Danthonioideae: 5, Aristidoideae: 2, Chloridoideae: 7, Ehrhartoideae: 2, Bambusoideae: 7, and Pooideae: 100 (Brachyelytreae: 1, Lygeeae/ Nardeae: 2, Stipeae: 9, Diarrheneae: 1, Brachypodieae: 1, Bromeae – Triticeae: 6, Poeae – Aveneae: 80). 136 new grass *rps*3 sequences were generated for this study. The origins of the plants studied, the locations of voucher specimens, and the GenBank sequence accession numbers are listed in the appendix of chapter III (Table S1, page 183). *Joinvillea ascendens* (Joinvilleaceae) was selected to root the phylogenetic inferences because of its close relationship to the grasses (CAMPBELL AND KELLOGG, 1987; LINDER AND RUDALL, 1993).

DOYLE AND DOYLE'S (1987) CTAB method was used to isolate DNA from silica-gel-dried leaves and from fresh materials for most of the studied samples. For herbarium samples, DNA extracts were obtained using the DNAeasy® Plant Mini Kit (QIAGEN Ltd., West Crawley, UK) procedure. A 1693-nucleotide region, extending from the second to the third exon of the mitochondrial *ribosomal protein S3* (*rps*3) gene, was amplified and sequenced using the primers 2a-for (external): gttcgatacgtccacctac, 2b-for (internal): caaagaagtcccgctctgac, 2c-bac (internal): aagcttcctcccttcgctcc, and 2d-bac (external): gtacgtttcggatatagcac (appendix of chapter III; Fig. S1; page 182). To avoid DNA contamination, sample mix-up, mislabeled samples or PCR artifacts, multiple and independent DNA extractions and PCR amplification and sequencing were conducted using independently prepared samples and varying the primer combinations. Procedures for DNA amplification and sequencing are indicated in the appendix of chapter III (page 178).

Table 1. Detection of microconversions and other analyses performed in the analysis of the mitochondrial *rps*3 gene in grasses. Tree statistics in Maximum Parsimony (MP) analyses of the datasets used. CI: consistency index (KLUGE AND FARRIS, 1969), excluding uninformative characters; RI: retention index (FARRIS, 1989).

Analysis performed	Data set used	Number of accessions	Number of nucleotide positions	MP tree length [steps]	Parsimony-informative characters	CI	RI
Poaceae organismal tree (Fig. 1)	ITS, *ß-amylase*, *mat*K, *ndh*F, *trn*TL, *trn*LF	138	5192	11540	2267 (43.68%)	0.4578	0.7499
Poaceae organismal tree (Fig. 2A inset)	ITS, *ß-amylase*, *mat*K, *ndh*F, *trn*TL, *trn*LF	64	5192	11540	2267 (43.68%)	0.4578	0.7499
Poaceae organismal tree (Fig. 2B inset)	ITS, *ß-amylase*, *mat*K, *ndh*F, *trn*TL, *trn*LF	66	5192	11540	2267 (43.68%)	0.4578	0.7499
Poaceae organismal tree (Fig. 2C inset)	ITS, *ß-amylase*, *mat*K, *ndh*F, *trn*TL, *trn*LF	71	5192	11540	2267 (43.68%)	0.4578	0.7499
Visual detection of microconversions (Figs. 2A-C)	*rps3* sequences (from exon 2 to exon 3), excluding ambiguous intron positions	141	1693	-	-	-	-
***rps3* Majority rule ML consensus tree (Fig. 3)**	*rps3* sequences (from exon 2 to exon 3, partial sequences), excluding ambiguous intron positions, including 3 outgroups	144	1693	2681	427 (25.22%)	0.4659	0.7041
Recombination Detection Program (RDP4) (Table 2)	*rps3* sequences (from exon 2 to exon 3, partial sequences), excluding ambiguous intron positions and accessions with missing data	124	1693	-	-	-	-
Existence of Premature Terminal Codons	*rps3* sequences (exons 2 and 3, partial sequences)	144	1533	-	-	-	-
Branch-Site REL (Table 3)	*rps3* sequences (exons 2 and 3, partial sequences), including accessions with likely recombination and misplacement	68	1533	-	-	-	-

Sequence contigs were assembled and manually inspected with the program Sequencher® 4.2.2 (Gene Codes Corporation, Ann Arbor, Michigan, U.S.A) to check the overlapping chromatogram peaks. Alignment of the *rps3* sequences was done with Protein Multiple Sequence Alignment Software version 3.7 (MUSCLE) (EDGAR, 2004). The alignments were manually adjusted using the program MacClade 4.08 OS X (MADDISON AND MADDISON, 2008). Improvement of the nucleotide alignments was performed using amino acid translations and by comparing our sequences with the previously published cDNA of the mitochondrial *rps*3 of *Zea perennis* (DQ645538), *Tripsacum dactyloides* (DQ984517), *Sorghum bicolor* (DQ984518), and *Bambusa oldhamii* (EU365401). These comparisons were also used to check the exon-intron-exon boundaries in the *rps3* gene. While alignment was generally straightforward in the coding regions, there were some intron positions (pos. 560 – 741) where patterns of length variation were complex enough to make homology assessment difficult. Regions that were ambiguous in terms of alignment were excluded from phylogenetic, recombination and substitution-rate analyses (Table 1).

3.2 Phylogenetic inference

3.2.1 Organismal tree reconstruction

A multigenic organismal tree of the Poaceae was used as a reference to assess the existence of phylogenetic bias (e. g. potential gene conversion and HGT) within the *rps3* reconstructions. The Poaceae organismal tree was constructed using the multicopy nuclear ribosomal *ITS*1-5.8S-*ITS*2 region, the low-copy nuclear *ß-amylase* marker and the plastid *mat*K, *ndh*F, *trn*TL and *trn*LF regions from a collection of 137 selected grass taxa, plus the close relative *Joinvillea* (Joinvillaceae) included as an outgroup (Table 1). Three additional organismal trees were reconstructed, but with reduced numbers of accessions (71, 66 and 64), to comparatively illustrate the presence of short *rps3* sequences shared by phylogenetically unrelated accessions (microconversions). The *ß-amylase* marker was used after excluding sequences with likely pseudogenization, paralogy, homeology, recombination and misplacements (see chapter I). The ITS and *trn*TL-*trn*LF regions were amplified and sequenced using the primers and procedures from CATALAN *ET AL*. (2004) and QUINTANAR *ET AL*. (2007); the *ß-amylase* gene followed the procedures indicated in MASON-GAMER (2005); the *ndh*F gene followed CATALAN *ET AL*. (1997); and the *mat*K gene was amplified and sequenced following the criteria of DÖRING *ET AL*. (2007). The potential combinability of the data sets was assessed using the Partition Homogeneity (PH) test (Incongruence Length Difference of FARRIS *ET AL*.,

1994), implemented in PAUP* 4.0 beta 10 (SWOFFORD, 2002). The PH test was conducted through heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees ON to estimate if the separate and combined nuclear and plastid data sets were significantly different from random partitions of the same sizes. Bayesian Inference (BI) and Maximum parsimony (MP) based searches were performed for both the independent and the combined data sets using, respectively, MrBayes 3.2.2 (ALTEKAR *ET AL.,* 2004; RONQUIST *ET AL.*, 2011) and PAUP*v4.10b (SWOFFORD, 2002). All gaps were treated as missing data. Procedures for the Bayesian and parsimony-based searches are indicated in the appendix of chapter III (page 178).

3.2.2 Reconstruction of the rps3 phylogeny

Maximum likelihood (ML) and maximum parsimony (MP) analyses were used to estimate the phylogeny of the *rps3* data set, which included 1693 nucleotide positions and 144 grass accessions, plus *Joinvillea ascendens, Nicotiana tabacum* and *Cycas revoluta* as out-groups (Table 1). All gaps were treated as missing data. Maximum likelihood analyses (FELSENSTEIN, 1973) were carried out in the program GARLI 2.01 (Genetic Algorithm for Rapid Likelihood inference; ZWICKL, 2006) following the procedures used by CERROS-TLATILPA *ET AL*. (2011). We performed five independent ML runs in GARLI using an initial random tree and the automated stopping criterion (70.000 generations without improving topology, lnL increase for significantly better topology = 0.01, and score improvement threshold = 0.05). The five independent ML searches were conducted according to the GTR+I+gamma model, which was selected as the optimal model using the test of goodness of fit for alternative nucleotide substitution models performed through the Hierarchical Likelihood Ratio Test (hLRTs), the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and the Decision Theory Performance-based Selection (DT) in jModelTest 2 (GUINDON AND GASCUEL, 2003; DARRIBA *ET AL.*, 2012). The *rps3* ML Majority-Rule tree from the five independent reconstructions was computed in PAUP*v4.10b (SWOFFORD, 2002). Maximum likelihood bootstrap values were obtained from 100 replicates using GARLI and the automated stopping criterion. Procedures for the MP-based searches are indicated in the appendix of chapter III (page 178).

3.3 Detection of potential recombinants, misplacements and pseudogenic sequences within the mitochondrial *rps3* gene

Two complementary methods were used to infer the occurrence of recombination events. Firstly, we conducted a detailed visual examination of the *rps3* matrix with 141 grass taxa and 1693 positions between the second and third exons of the *rps3* gene, after excluding the ambiguous intron positions (Table 1) and ordering the accessions according to their positions in the organismal tree. This search allowed us to identify the existence of recombinations involving short DNA regions. These regions could eventually appear punctually or discontinuously throughout the sequences sampled (as gene conversions or microconversions, respectively). The size of these microconversions was limited to the span of the shared nucleotides, which was calculated based on the distance to the closest apomorphic site. Secondly, the exchange of DNA sequences was evaluated using the Recombination Detection Program (RDP4) version 4.16 (MARTIN *ET AL.*, 2010). This program distinguishes continuous patterns of recombination signals that include a high number of nucleotides, which are difficult to differentiate visually. The RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, LARD, and 3SEQ methods implemented in the RDP4 program were used to identify recombinant and parental sequences and to calculate probability scores for potential recombination events. We used the default settings in all cases, except that we activated the circular genome option. Only potential recombinant events detected by at least two methods were considered significant. The Bonferroni multiple comparison correction test was performed to diminish the expected number of false positive results. Masking of similar sequences was allowed in order to increase the power of the recombination detection methods. For these analyses we used a data set with 1693 positions after excluding the ambiguous intron positions (pos. 560 – 741). This matrix included 124 accessions and also contained *Arabidopsis thaliana, Beta vulgaris*, *Cycas revoluta,* and *Nicotiana tabacum* outgroups, in order to detect evidence of recombination from non-related grass lineages, but excluded sequences with missing data (Table 1).

The visual and topological phylogenetic reconciliations between the *rps3* gene phylogeny and the multigenic Poaceae organismal tree were done following the criterion of LAFAY *ET AL.* (1999) and the Shimodaira–Hasegawa (SH) nonparametric bootstrap test (SHIMODAIRA AND HASEGAWA, 1999), respectively. Specifically, the SH test determined whether the topology of the ML tree derived from the *rps3* gene based on 138 grass accessions was significantly different from the Poaceae Bayesian tree derived from the nuclear ITS and *ß*-

amylase markers and the plastid *mat*K, *ndh*F, *trn*TL and *trn*LF regions. This test was done using PAUP*v4.10b with RELL (resampling estimated log-likelihood) optimization and 1000 bootstrap replicates.

Two complementary methods were used to infer the occurrence of pseudogenic copies among the *rps3* sequences sampled. First, we looked for the existence of Premature Terminal Codons (PTCs) among the coding positions of the 144 sampled *rps3* sequences (exons 2 and 3: 1533 positions; Fig. S1, page 182). According to AKHUNOV *ET AL*. (2013), some potential pseudogenic copies could accumulate mutations and eventually produce PTCs at the early stages of evolution.

Secondly, we considered that most of the coding positions in a functional protein are constrained by purifying selection, rather than positive (non-purifying) selection or neutral variation (KIMURA, 1968). Selection was investigated through the analysis of distribution of substitution rates across sites and branches (REL model; NIELSEN AND YANG, 1998; KOSAKOVSKY-POND ET AL., 2005), as detailed in chapter I. REL models assume that a continuous-time stationary Markov process characterizes substitutions along a branch of a phylogenetic tree, where a transition rate matrix gives the likelihood of change along the branch. Considering that selective pressures may vary over both sites and time, branch-site REL methods are able to estimate the proportion of sites with different ω values for all lineages of the tree (KOSAKOVSKI-POND *ET AL*., 2011). In this case, each branch-site combination has instantaneous rate and transition matrices. We restricted our analysis to terminal lineages to evaluate recent pressures associated with a representative subset of 68 BEP sequences with 1533 exon positions (Table 1), to detect sequences with high proportions of sites under neutral or positive selection, indicating potential relaxed selection pressure (represented by $\omega = dN/dS \geq 1$) and evidence of pseudogenization. We used the Branch-Specific Branch-Site REL (BS-REL; REL = Random Effect Likelihood) model as suggested in KOSAKOVSKY-POND *ET AL*. (2008) to estimate the *rps3* lineage-specific proportion of sites (Pr1, Pr2 and Pr3) showing rates circumscribed to three discrete classes: $\omega1 \leq \omega2 \leq 1 \leq \omega3$ We considered that classes $\omega1$ and $\omega2$ can be less than or equal to 1, and are associated with the neutral model specifying stronger and weaker purifying selection, respectively, whereas $\omega3$ (unconstrained class) is related to adaptive selection if it is greater than one. According to the estimated $\omega1$, $\omega2$, and $\omega3$ values, we assigned each of these classes and its proportion to one out of five selection scenarios: (I) $\omega = 0 – 0.6$ (strong negative selection), (II) $\omega = 0.61 – 0.9$ (weak negative selection), (III) $\omega = 0.91 – 1.5$ (neutral selection), (IV) $\omega = 1.51 – 5$ (weak positive selection), and (V) $\omega > 5$

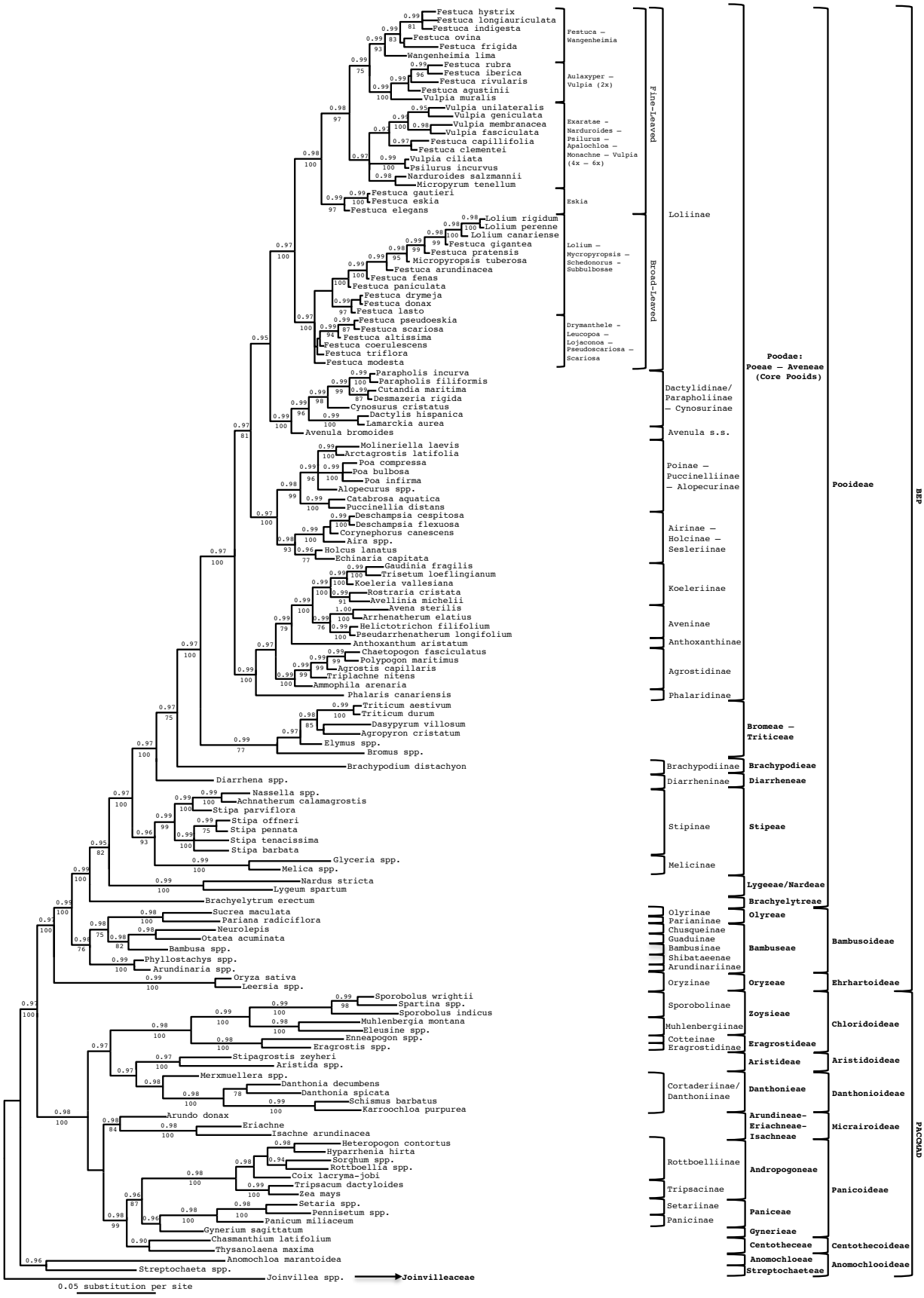
(strong positive selection). BS-REL was performed using the maximum likelihood-based tools implemented in the WEB interface DataMonkey (POND AND FROST, 2005).

4. Results

4.1 The organismal tree of the grasses

The multigenic organismal tree of the Poaceae (Fig. 1) provided a framework for assessing the evolutionary and phylogenetic patterns observed in the *rps3* tree. The multigenic data set consisted of nuclear ITS (131 sequences/ 609 nucleotides), *ß-amylase* (76/850) and plastid *mat*K (120/1259), *ndh*F (138/710), *trn*TL (133/789) and *trn*LF (138/ 975) sequences. GenBank accession numbers for all sequences are indicated in the appendix of chapter III (Table S1, page 183). The independent phylogenetic searches conducted with the six separate ITS, *ß*-amylase, *mat*K, *ndh*F, *trn*TL and *trn*LF data sets yielded highly congruent topologies. The PH test detected incongruence among the plastid (cpDNA) data sets (P< 0.01), except for *mat*K – *trn*TL (P= 0.4) and *ndh*F – *trn*TL (P= 0.3). Partition incongruence was also found between the plastid and nuclear data (P= 0.01). Despite this, the separate topologies were similar to the tree based on the combined nuclear and plastid datasets, with incongruence observed mostly among some terminal tips. Furthermore, those topological conflicts were not well supported. Since the incongruence did not affect the main results and conclusions of this research, we chose to conduct phylogenetic analyses on the combined dataset of the five sequenced genes. The BI tree (Fig. 1) and MP tree (Fig. 1 Bootstrap support values) from the combined data set were highly congruent to those previously obtained for the grasses (MINAYA *ET AL.*, 2013, chapter I and II). The MP statistics of the Poaceae organismal tree are shown in Table 1.

Figure 1. Poaceae organismal tree. Bayesian Inference estimation based on nuclear (ITS, *ß-amylase*) and plastid (*trn*TL, *trn*LF, *ndh*F, *mat*K) gene sequences (see Table S1; page 183). Bayesian posterior probabilities of ≥ 90% and bootstrap support values of ≥ 75% are shown above and below the nodes, respectively. Species assignment to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure following SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.,* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). Identical criteria were followed in the remaining phylogenetic inferences presented in this paper.


Festuca hystrix
Festuca longiauriculata
Festuca indigesta
Festuca ovina
Festuca frigida
Wangenheimia lima
Festuca rubra
Festuca iberica
Festuca rivularis
Festuca agustinii
Vulpia muralis
Vulpia unilateralis
Vulpia geniculata
Vulpia membranacea
Vulpia fasciculata
Festuca capillifolia
Festuca clementei
Vulpia ciliata
Psilurus incurvus
Narduroides salzmannii
Micropyrum tenellum
Festuca gautieri
Festuca eskia
Festuca elegans
Lolium rigidum
Lolium perenne
Lolium canariense
Festuca gigantea
Festuca pratensis
Micropyropsis tuberosa
Festuca arundinacea
Festuca fenas
Festuca paniculata
Festuca drymeja
Festuca donax
Festuca lasto
Festuca pseudoeskia
Festuca scariosa
Festuca altissima
Festuca coerulescens
Festuca triflora
Festuca modesta
Parapholis incurva
Parapholis filiformis
Cutandia maritima
Desmazeria rigida
Cynosurus cristatus
Dactylis hispanica
Lamarckia aurea
Avenula bromoides
Molineriella laevis
Arctagrostis latifolia
Poa compressa
Poa bulbosa
Poa infirma
Alopecurus spp.
Catabrosa aquatica
Puccinellia distans
Deschampsia cespitosa
Deschampsia flexuosa
Corynephorus canescens
Aira spp.
Holcus lanatus
Echinaria capitata
Gaudinia fragilis
Trisetum loeflingianum
Koeleria vallesiana
Rostraria cristata
Avellinia michelii
Avena sterilis
Arrhenatherum elatius
Helictotrichon filifolium
Pseudarrhenatherum longifolium
Anthoxanthum aristatum
Chaetopogon fasciculatus
Polypogon maritimus
Agrostis capillaris
Triplachne nitens
Ammophila arenaria
Phalaris canariensis
Triticum aestivum
Triticum durum
Dasypyrum villosum
Agropyron cristatum
Elymus spp.
Bromus spp.
Brachypodium distachyon
Diarrhena spp.
Nassella spp.
Achnatherum calamagrostis
Stipa parviflora
Stipa offneri
Stipa pennata
Stipa tenacissima
Stipa barbata
Glyceria spp.
Melica spp.
Nardus stricta
Lygeum spartum
Brachyelytrum erectum
Sucrea maculata
Pariana radiciflora
Neurolepis
Otatea acuminata
Bambusa spp.
Phyllostachys spp.
Arundinaria spp.
Oryza sativa
Leersia spp.
Sporobolus wrightii
Spartina spp.
Sporobolus indicus
Muhlenbergia montana
Eleusine spp.
Enneapogon spp.
Eragrostis spp.
Stipagrostis zeyheri
Aristida spp.
Merxmuellera spp.
Danthonia decumbens
Danthonia spicata
Schismus barbatus
Karroochloa purpurea
Arundo donax
Eriachne
Isachne arundinacea
Heteropogon contortus
Hyparrhenia hirta
Sorghum spp.
Rottboellia spp.
Coix lacryma-jobi
Tripsacum dactyloides
Zea mays
Setaria spp.
Pennisetum spp.
Panicum miliaceum
Gynerium sagittatum
Chasmanthium latifolium
Thysanolaena maxima
Anomochloa marantoidea
Streptochaeta spp.
Joinvillea spp.
0.05 substitution per site
Joinvilleaceae
Festuca – Wangenheimia
Aulaxyper – Vulpia (2x)
Exaratae – Narduroides – Psilurus – Apalochloa – Monachne – Vulpia (4x – 6x)
Eskia
Lolium – Mycropyropsis – Schedonorus – Subbulbosae
Drymanthele – Leucopoa – Lojaconoa – Pseudoscariosa – Scariosa
Fine-Leaved
Broad-Leaved
Loliinae
Dactylidinae/ Parapholiinae – Cynosurinae
Avenula s.s.
Poinae – Puccinelliinae – Alopecurinae
Airinae – Holcinae – Sesleriinae
Koeleriinae
Aveninae
Anthoxanthinae
Agrostidinae
Phalaridinae
Bromeae – Triticeae
Brachypodiinae
Diarrheninae
Stipinae
Melicinae
Olyrinae
Parianinae
Chusqueinae
Guaduinae
Bambusinae
Shibataeenae
Arundinariinae
Oryzinae
Sporobolinae
Muhlenbergiinae
Cotteinae
Eragrostidinae
Cortaderiinae/ Danthoniinae
Rottboelliinae
Tripsacinae
Setariinae
Panicinae
Poodae: Poeae – Aveneae (Core Pooids)
Pooideae
Brachypodieae
Diarrheneae
Stipeae
Lygeeae/Nardeae
Brachyelytreae
Olyreae
Bambuseae
Bambusoideae
Oryzeae
Ehrhartoideae
Zoysieae
Chloridoideae
Eragrostideae
Aristideae
Aristidoideae
Danthonieae
Danthonioideae
Arundineae-Eriachneae-Isachneae
Micrairoideae
Andropogoneae
Panicoideae
Paniceae
Gynerieae
Centotheceae
Centothecoideae
Anomochloeae
Streptochaeteae
Anomochlooideae
BEP
PACCMAD

4.2 Estimates of recombination within the mitochondrial *rps3* gene in grasses

The detailed visual examination of the *rps3* sequences throughout the 141 grass accessions (Fig. 2), ordered according to the Poaceae organismal tree (Fig. 1), allowed us to find 12 cases of continuous and discontinuous short (1-140 nucleotides) sequences shared by phylogenetically unrelated accessions, involving 123 grass taxa (87.8% of the tested sequences). These events likely represent the existence of gene conversion or microconversion within the *rps3* gene throughout the Poaceae (Fig. 2A-C). Six cases were observed between distantly related species, each involving species from the PACCMAD and species from the BEP clade (Fig. 2A): (1) Positions 43 – 45 (labeled with dark red in Figure 2A): *Hyparrhenia hirta* (Andropogoneae), *Muhlenbergia montana* (Zoysieae) and *Melica minuta* (Stipeae) share this region with accessions nested within the core pooids clade; (2) pos. 49 – 492 (dark blue): *Stipa tenacissima* (Stipeae) and *Schismus barbatus* (Danthonieae) share this region with the core pooids; (3) pos. 422 – 436 (blue) *Stipagrostis zeyheri* (Aristideae) and *Sporobolus indicus* (Zoysieae) share these positions with most of the sampled taxa within the Poodae (Poeae – Aveneae) clade; (4) pos. 497 – 513 (brown): *Sporobolus indicus* (Zoysieae) shares these positions with the Dactylidinae/ Parapholiinae group, Puccinelliinae, Holcinae and Aveninae; (5) pos. 422 – 559 (purple): *Muhlenbergia montana* (Zoysieae), Triticodae (Bromeae – Triticeae), *Agrostis* (Agrostidinae) and *Poa* (Poinae) share this region; (6) pos. 49 – 1300 (red): *Catabrosa aquatica* (Puccinelliinae) and *Pseudarrhenatherum longifolium* (Aveninae) shared a collection of discontinuous symplesiomorphies with most members of the PACCMAD clade, Ehrhartoideae, Bambusoideae and the basal pooids.

Three cases were observed between distantly related members of the Pooideae (Fig. 2B): (1) pos. 43 – 1692 (dark blue): the Poeae – Aveneae clade shared a collection of discontinuous positions with some basal Pooideae lineages such as *Lygeum spartum*, *Nardus stricta*, *Melica minuta*, *Glyceria declinata*, *Stipa tenacissima*, *Stipa offneri* and *Brachypodium distachyon*; (2) pos. 1198 – 1227 (brown): unrelated *Dasypyrum villosum* (Triticeae), *Agrostis capillaris* (Agrostidinae), *Trisetum loeflingianum* (Koeleriinae), *Holcus lanatus* (Holcinae) and *Festuca clementei* (fine-leaved Loliinae) share these identical positions; (3) pos. 1277 – 1288 (pink): the Festuca and Aulaxyper + Vulpia (2x) clades share an identical insertion with unrelated *Holcus lanatus* (Holcinae), *Koeleria vallesiana, Rostraria cristata* (Koeleriinae), *Agrostis capillaris* (Agrostidinae) and *Dasypyrum villosum* (Triticeae).

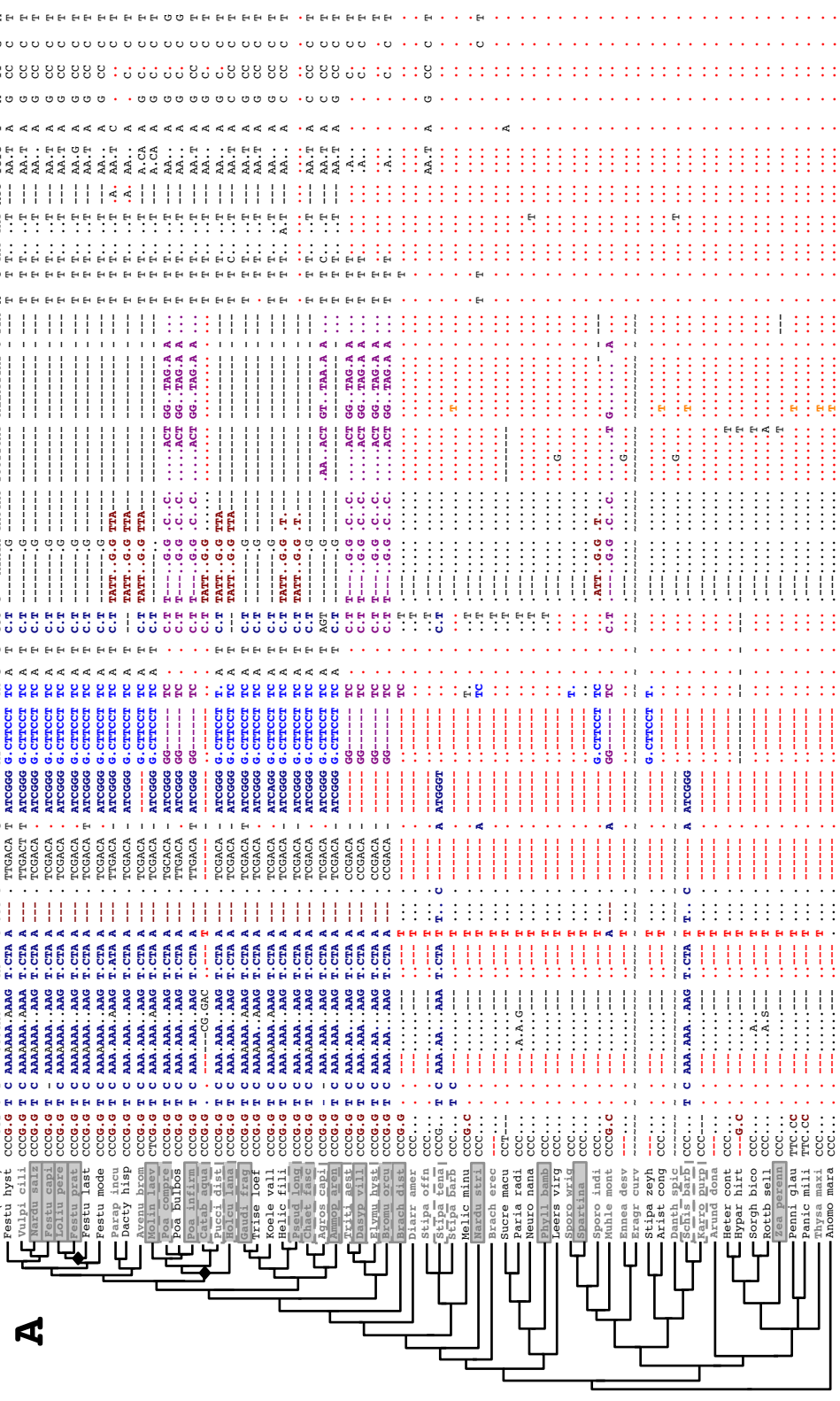


Figure 2. *(Continued)*

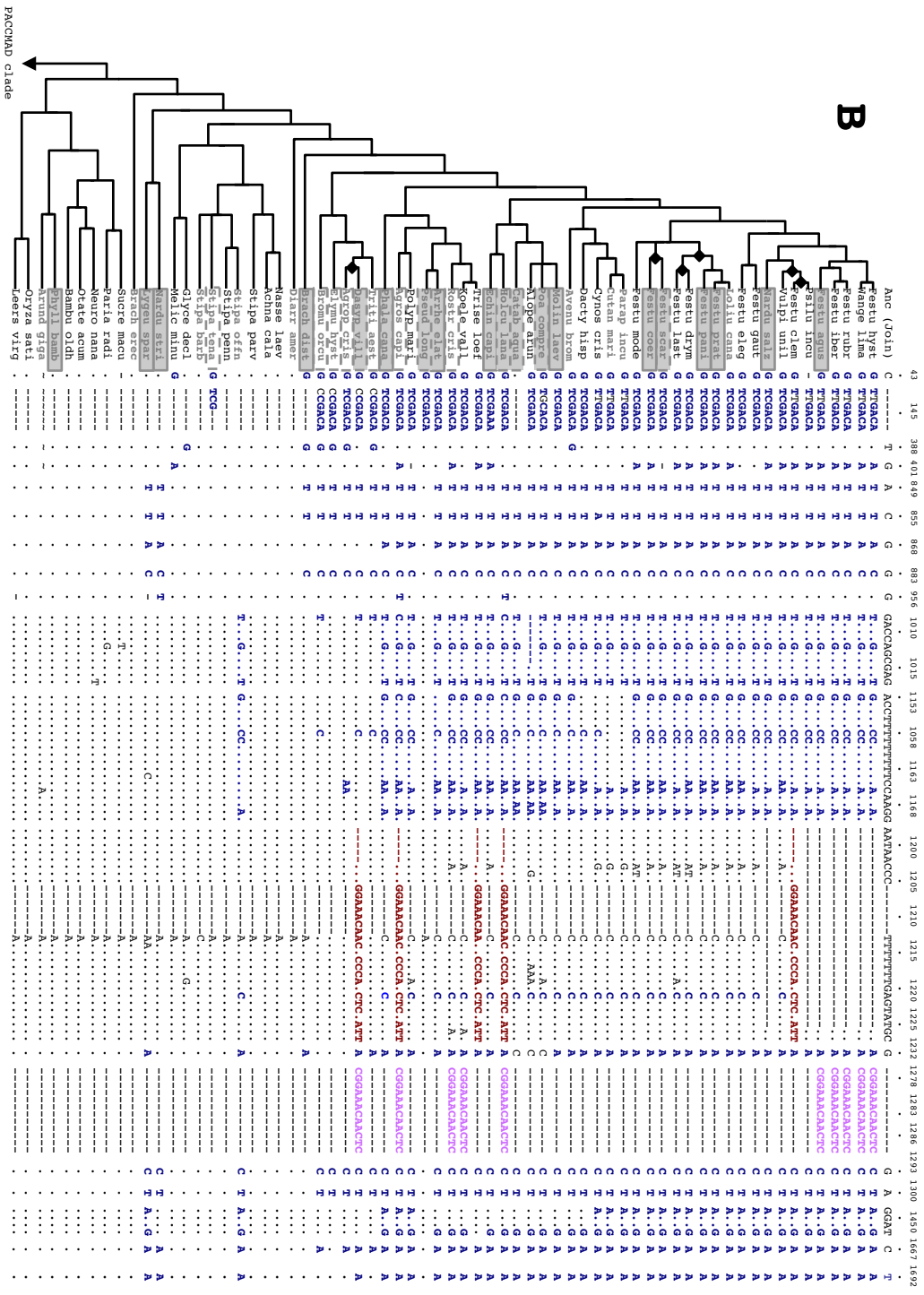


Figure 2. *(Continued)*

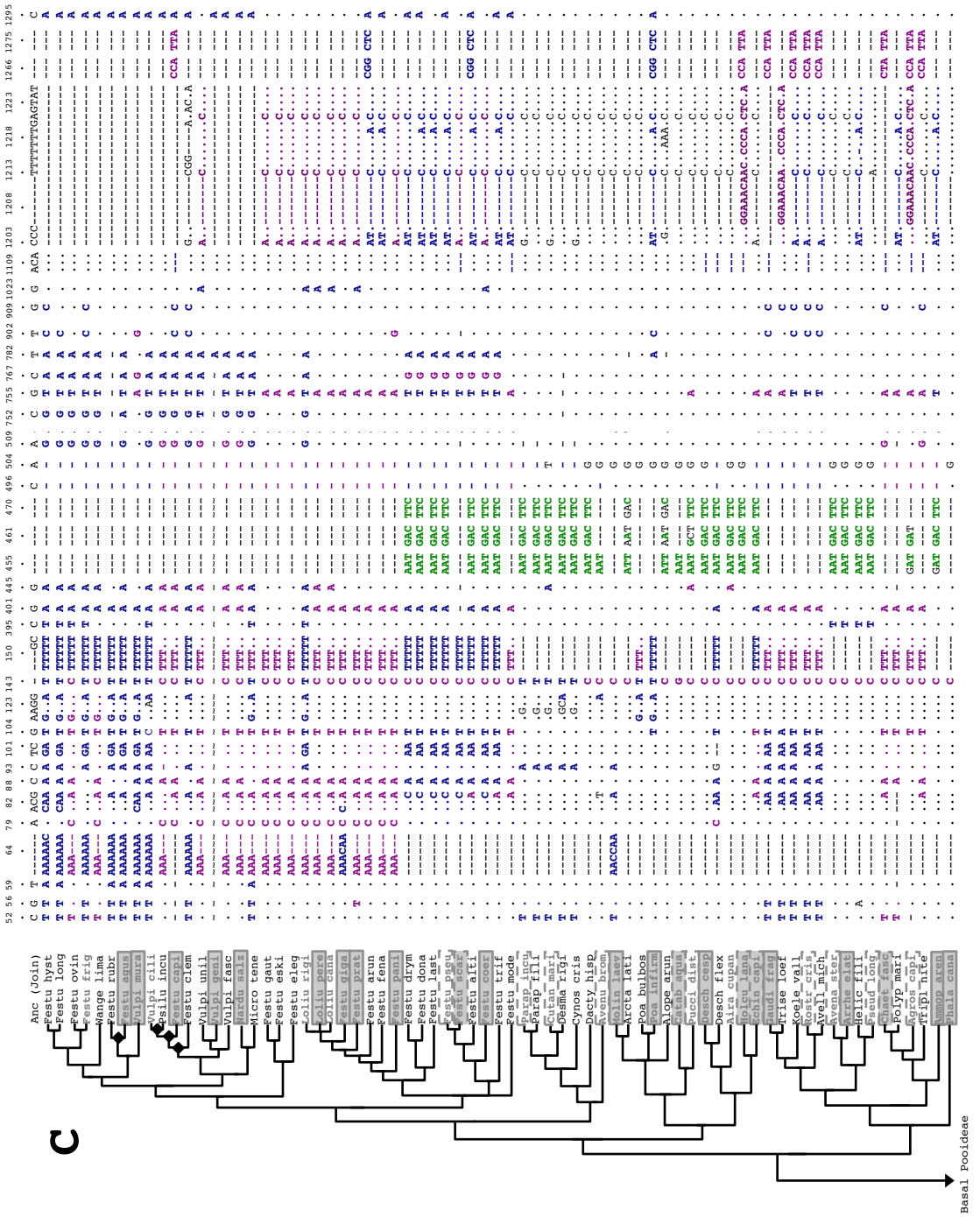


Figure 2. *(Continued)*

Figure 2. Representation of the twelve detected cases of microconversions within the mitochondrial *rps3* gene. Short *rps3* sequences (1 – 140 nt) shared by phylogenetically unrelated grass accessions were identified through visual examination of the 1693 positions sequenced between the second and third exons of the mitochondrial *rps3* gene throughout 141 grass taxa ordered according to the organismal tree (left side of the figure, Table 1). The *rps3* sequence of *Joinvillea ascendens* was used to root the tree [ancestral taxa; Anc (Join)]. Figure 2A represents the 6 microconversions, distinguished by colors, which were found between members of the distantly related PACCMAD and BEP clades (dark red, dark blue, blue, brown, purple, and red; see text). Figure 2B shows the 3 regions shared between non-monophyletic members of the Pooideae (dark blue, brown, and pink). Figure 2C indicates the 3 regions shared between non-related members of the Poodae (Poeae – Aveneae) (dark blue, purple and green; see text). The three Bayesian organismal trees presented on the left side of this figure (A to C, insets) were based on BI from ITS, *ß-amylase*, *mat*K, *ndh*F, *trn*TL and *trn*LF sequences, following the procedures indicated in material and methods but with a reduced and representative number of accessions (Table 1). Branch diamonds indicate clades with low bootstrap (<75%) or posterior probability (<90%) support. Dashed boxes and bold gray species indicate that these accessions presented a recombination event (Table 2). Grey boxes and bold gray species indicate that the *rps3* sequences of these accessions presented a high proportion of neutral or positively selected sites and a complete absence of purifying selective positions (Table 3). Bold gray species indicate that these accessions were misplaced when comparing the *rps3* tree (Fig. 3) with the Poaceae organismal tree (Fig. 1). The combination of grey boxes and bold gray labels also shows accessions with an absence of purifying selective positions and misplacements. Abbreviations of species are presented in brackets in table S1 (page 183). The symbol (~) indicates the occurrence of missing positions.

Three cases were observed between species classified within the Poodae (Fig. 2C): (1) pos. 52 – 1295 (dark blue) and (2) pos. 52 – 1275 (purple): these two events appeared discontinuously spread among most of the Aveneae/ Poeae taxa sampled. Species such as *F. eskia*, *Micropyrum tenellum* (fine-leaved Loliinae), *Lolium* rigidum (broad-leaved Loliinae), and *K. vallesiana* (Koeleriinae) include a blended sequence with a mix of both events; (3) pos. 454 – 471 (green): this insertion is widespread among members of the Poodae (Poeae – Aveneae) clade, with the exception of *Phalaris* (Phalaridinae), *Triplachne*, *Polypogon*, *Chaetopogon* (Agrostidinae), Koeleriinae, *Poa bulbosa*, *P. infirma*, and *Molineriella laevis* (Poinae). Within Loliinae, this insertion was only observed within the broad-leaved Drymanthele – Leucopoa – Lojaconoa – Pseudoscariosa – Scariosa clade, with the exception of *F. modesta* and *F. scariosa*.

Three recombination events (Table 2, Fig. 3) were detected by two or more algorithms ($p<0.05$) implemented in the program RDP4 within an *rps3* data set that included 124 accessions and 1693 positions (Table 1), once the accessions with missing data and the ambiguous intron positions (pos. 560 – 741) were discarded. The first recombination involved the distantly related *Schismus, Karroochloa* (Danthonieae), *Stipa tenacissima* (Stipeae), all sampled representatives of Triticodae (Bromeae – Triticeae), *Pseudarrhenatherum* and *Avena* (Aveninae) lineages. Putative major (*Holcus lanatus*, Holcinae, 99.2%) and minor (*Achnatherum calamagrostis*, Stipeae, 99.9%) parental sequences were also distantly related to each other and to the recombinant lineages. The second recombination was detected in the non-monophyletic *Holcus, Agrostis* and *Chaetopogon* (Agrostidinae), *F. scariosa* and *F.*

pseudoeskia (broad-leaved Loliinae) lineages. The third recombination was observed in the unrelated *Rostraria* (Koeleriinae), *Catabrosa, Poa* and *Puccinellia* (Poinae - Puccinelliinae) lineages. A total of twenty accessions (16.13% of the tested sequences) were involved in any of the recombinations detected by the RDP4 program; 2 from taxa nested within the PACCMAD clade, and the rest from taxa nested within BEP; (Stipeae (1), Triticodae (Bromeae – Triticeae) (6) and Poodae (Poeae – Aveneae) (11). Interestingly, the second and third recombination events presented the largely divergent dicot *Nicotiana tabacum* (Solanaceae), which was included as outgroup, as the most likely minor parental sequence, and the non-related *Pariana radicifolia* (Bambusoideae, rec. event 2) and *Achnatherum calamagrostis* (Stipeae, rec. even 3) as the respective potential major parental sequences (Table 2, Fig. 3).

Table 2. Summary of recombination events identified by two or more methods implemented in the program RDP4 in the studied sequences of the mitochondrial *rps*3 gene. Events detected in 124 accessions /1693 positions, after excluding ambiguous intron positions and species with missing data. Number of Rec. Event: Events are numbered according to their statistical importance in the RDP4 analysis. Rec. sp.: Species under recombination; Pos.: Beginning and Ending breakpoint position in alignment.

Nº Rec. Event.	Rec. sp	Pos.	Parental sequence(s) (% similarity)		Probability of significant tests for different detection methods in RDP4					
			Major	Minor	GENECONV	BootScan	MaxChi	Chimaera	SiScan	3Seq
1.-	*Agropyron cristatum* *Avena sterilis* *Bromus orcuttianus* *Dasypyrum villosum* *Elymus hystrix* *Karroochloa purpurea* *Pseudarrhenatherum longifolium* *Schismus barbatus* *Stipa tenacissima* *Triticum aestivum* *Triticum durum*	442 – 1533	Holcus lanatus (99.2%)	Achnatherum calamagrostis (99.9%)	3.04E-4	2.65E-4	1.37E-6	2.98E-4	<0.0001	0.018
2.-	*Agrostis capillaris* *Chaetopogon fasciculatus* *Festuca pseudoeskia* *Festuca scariosa* *Holcus lanatus*	852 – 1139	*Pariana radiciflora* (96.2%)	Unknown (*Nicotiana tabacum*)	-	-	0.0017	9.31E-4	1.21E-5	-
3.-	*Catabrosa aquatica* *Poa compressa* *Puccinellia distans* *Rostraria cristata*	841 – 1150	*Achnatherum calamagrostis* (95.9%)	Unknown (*Nicotiana tabacum*)	-	-	0.016	0.015	2.39E-4	0.023

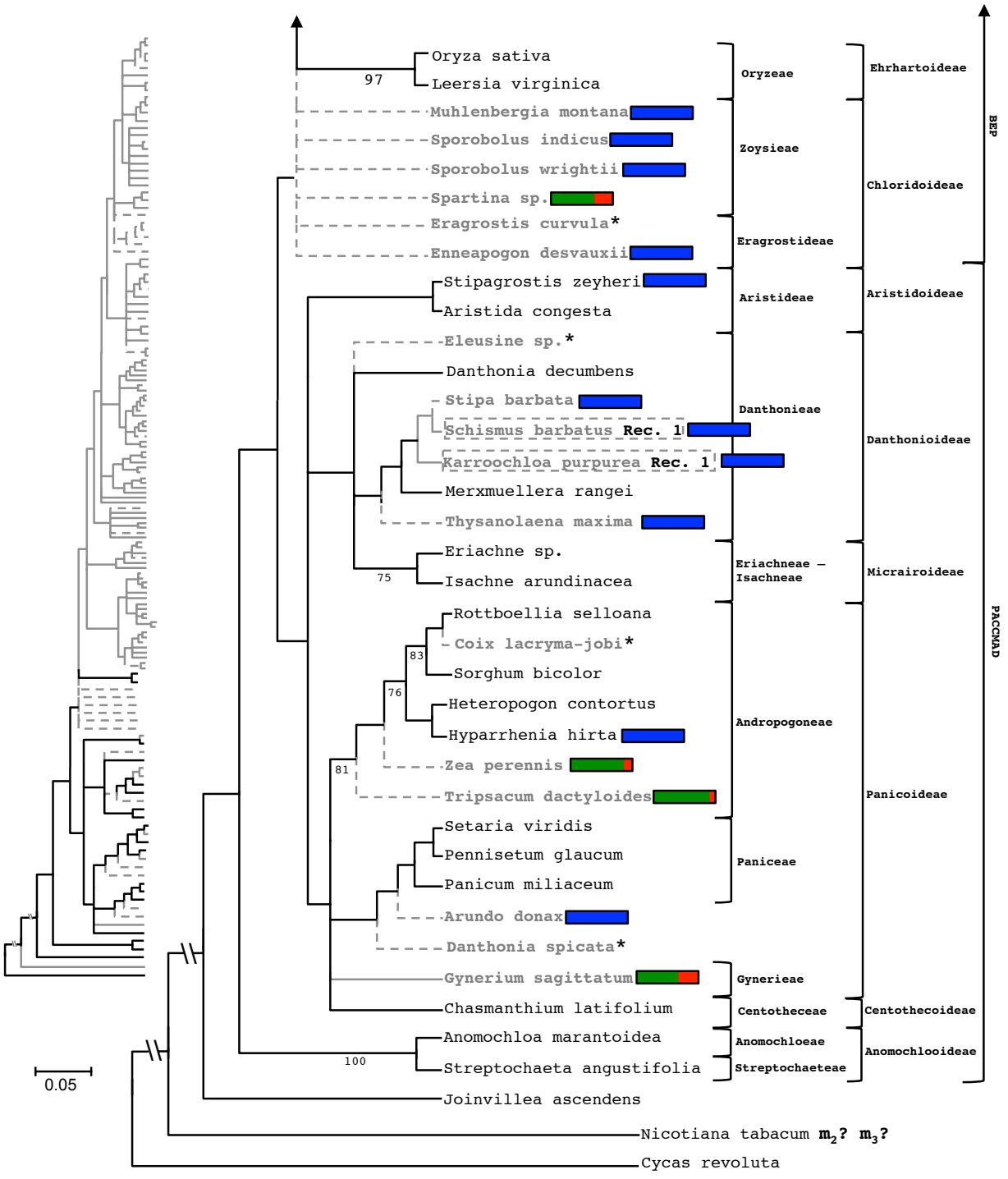


Figure 3 *(Continued)*

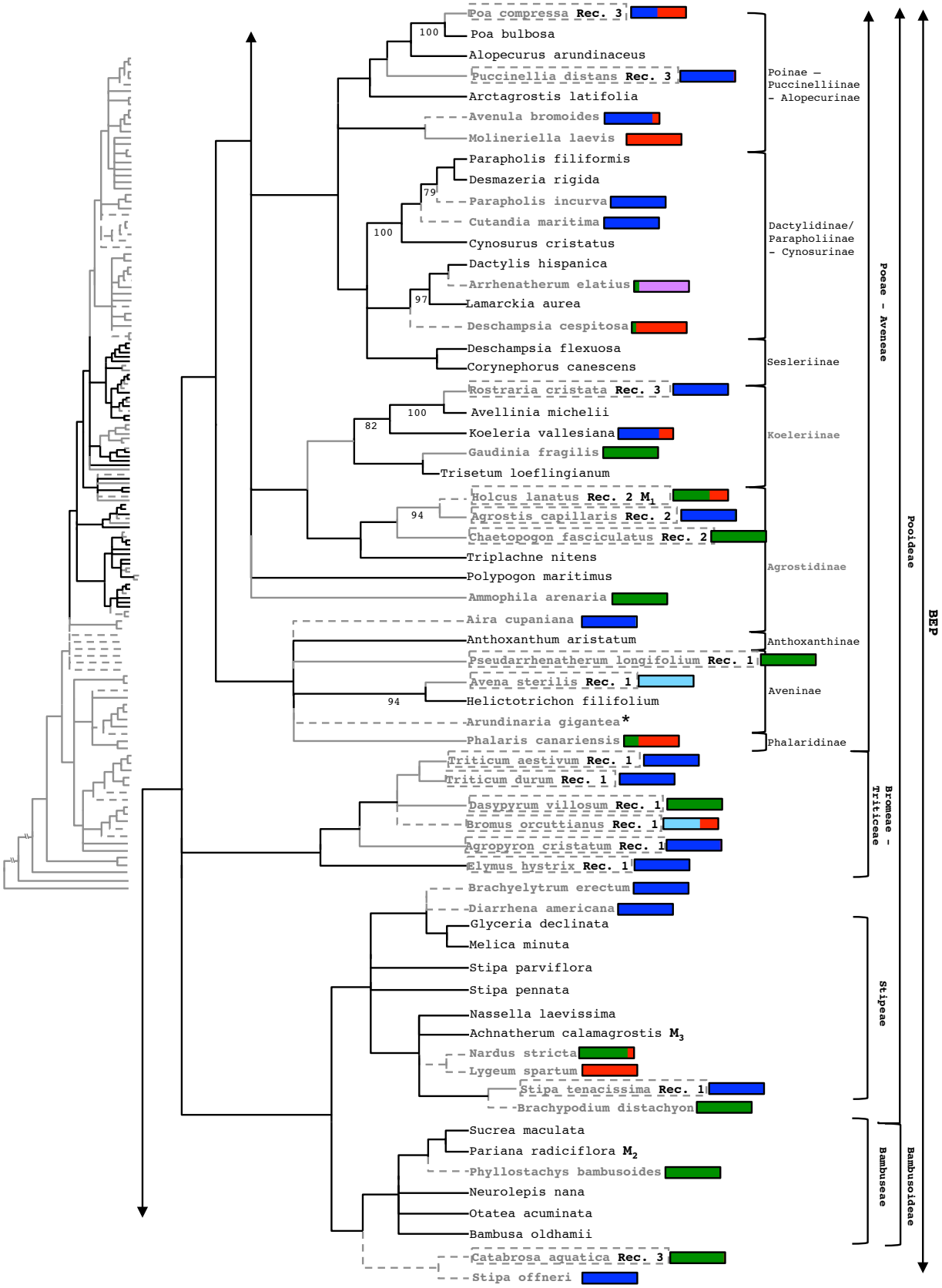


Figure 3 *(Continued)*

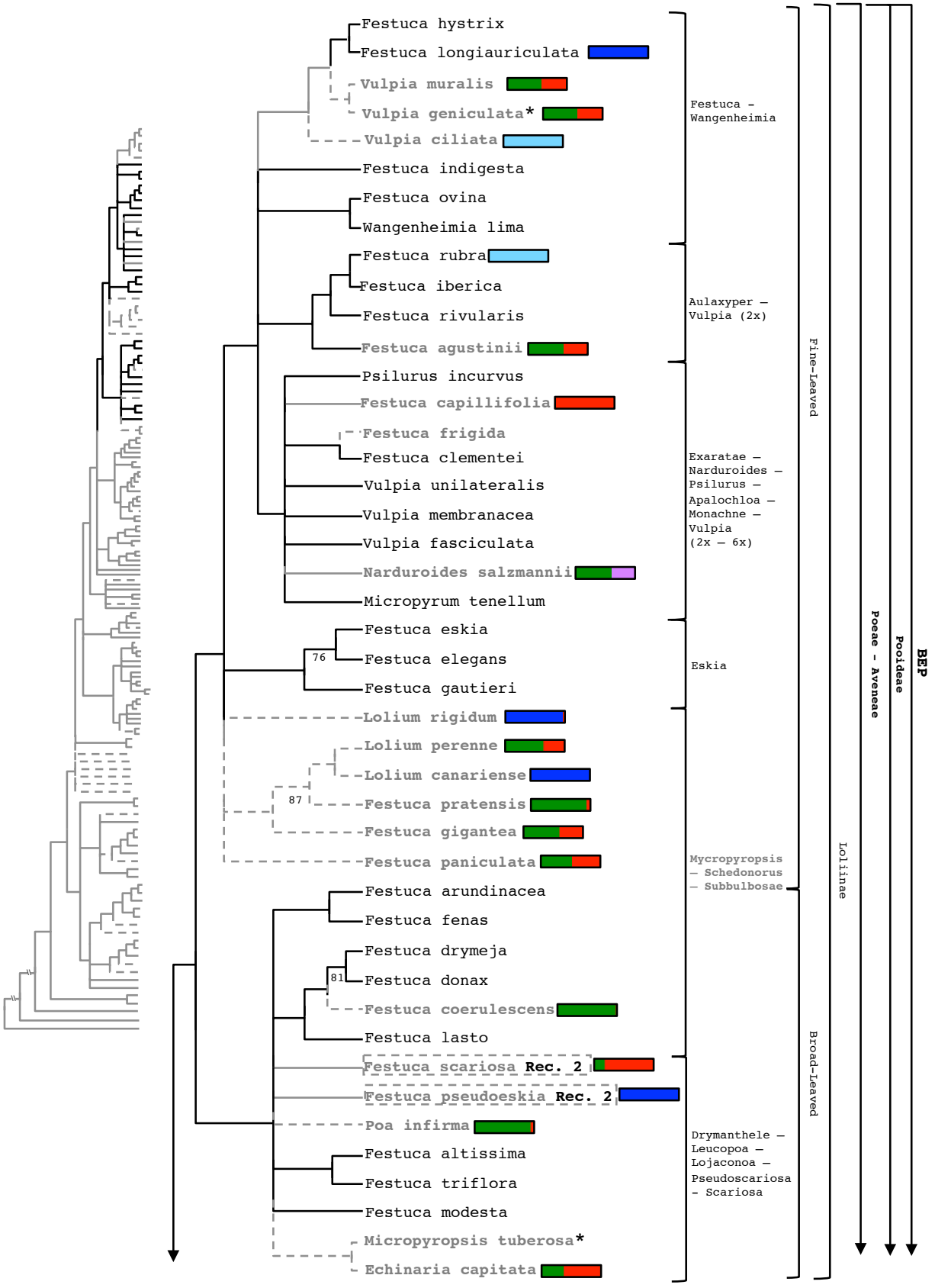


Figure 3 *(Continued)*

Figure 3. Majority-Rule ML consensus tree of Poaceae based on analysis of the mitochondrial *rps3* gene. This inference was computed from five independent Maximum Likelihood analyses based on the *rps3* exons + unambiguously aligned introns data set (Table 1). Parsimony bootstrap values > 75% are indicated below the branches. Species assignments to sections, subgenera, tribes, subfamilies and evolutionary clades are represented on the right side of the figure. Gray branches, dashed boxes and bold gray species indicate recombination events detected by two or more methods implemented in the RDP4 program. Rec. n: indicates the number of each recombination event (Table 2). M_n: major parents in the recombination n. m_n: minor parents in the recombination n. m_n?: most likely parents in the recombination n. Colored proportions of the painted horizontal bars mean strong negative (dark blue), negative (light blue), neutral (green), positive (pink), and strong positive (red) selection based on the BS-REL method (Table 3). Gray dotted branches and bold gray species indicate that these accessions are misplaced when assessed in comparison with their placement in the Poaceae organismal tree (Fig. 1). Asterisks (*) indicate that these species presented a high number of missing positions and were excluded from the recombination analyses.

4.3 Pseudogenization and misplacement within the mitochondrial *rps3* gene

The criterion used to detect potential PTCs (AKHUNOV *ET AL.*, 2013) in the *rps3* coding sequences (exons 2 and 3; 1533 positions; Table 1; Fig. S1; page 182) did not retrieve any evidence of existence of pseudogenic sequences among the studied Poaceae accessions. However, the Branch-Site REL (BS-REL) analysis detected different sequence positions under neutral, slightly positive and strongly positive selection within the second and third *rps3* exons in the studied samples (68 accessions, exons 2 and 3: 1533 positions, Tables 1, 3; Fig. 3). Strong negative selection ($0 \leq \omega \leq 0.9$, range I in Table 3, dark blue bar in Fig. 3) was predominant across sequences and positions (46.76% average value), followed by neutral selection ($0.91 \leq \omega \leq 1.5$; III; green bar; 29.31%), and then by low percentages of strongly positive selection ($\omega > 5$; V; red bar in; 17.50%), slightly negative selection ($0.9 \leq \omega \leq 0.61$; II; dark blue bar; 4.42%) and slightly positive selection ($0.5 \leq \omega \leq 1.51$; IV; pink bar; 2%). Twenty-five out of 68 sequences (36.8%) tested with the BS-REL method showed 100% of their sites under negative selection (I and II in Table 3), and 4 more sequences (5.8%) presented less than 10% of their sites under neutral or positive selection. Thirty-three sequences (48.5%) showed a high proportion of neutral or positively selected sites and a complete absence of purifying selective positions (Table 3, Fig. 3), suggesting the existence of pseudogenic copies or other relaxed selection pressure events [*Ammophyla arenaria* (99% neutral sites / 1% positively selected sites), *Arrhenatherum elatius* (7% / 93%), *Brachypodium distachyon* (100% / -), *Bromus orcuttianus* (65% / 35%), *Catabrosa aquatica* (97% / 3%), *Chaetopogon fasciculatus* (99% / 1%), *Dasypyrum villosum* (99% / 1%), *Deschampsia cespitosa* (5% / 95%), *Echinaria capitata* (36% / 64%), *Festuca agustinii* (58% / 42%), *F. capillifolia* (- / 100%), *F. coerulescens* (99% / 1%), *F. gigantea* (60% / 40%), *F. paniculata* (51% / 49%), *F. pratensis* (93% / 7%), *F. scariosa* (16% / 84%), *Gaudinia fragilis* (96% / 4%), *Gynerium sagittatum* (67% / 33%), *Holcus lanatus* (64% /36%), *Lolium perenne* (60% / 40%), *Lygeum spartum* (- /

100%), *Molineriella laevis* (- / 100%), *Narduroides salzmannii* (57% / 43%), *Nardus stricta* (83% / 17%), *Phalaris canariensis* (22% / 78%), *Phyllostachys bambusoides* (98% / 2%), *Poa infirma* (92% / 8%), *Pseudarrhenatherum longifolium* (98% / 2%), *Spartina* (69% / 31%), *Tripsacum dactyloides* (61% / 39%), *Vulpia geniculata* (60% / 40%), *V. muralis* (56% / 44%), *Zea perennis* (86% / 14%)].

Table 3. Summary of the Branch-Site REL (BS-REL) results of the studied grass mitochondrial rps3 sequences analyzed in the web interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at each class along each branch. $\omega = d_N/d_S$, where d*N* and d*S* are the numbers of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) $\omega = 0 - 0.6$, Strongly Negative Selection (dark blue bar in Fig. 3); (II) $\omega = 0.61 - 0.9$, Weak Negative Selection (light blue bar); (III) $\omega = 0.91 - 1.5$, Neutral Selection (green bar); (IV) $\omega = 1.51 - 5$, Weak Positive Selection (pink bar); and (V) $\omega > 5$, Strongly Positive Selection (red bar).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
Agropyron cristatum	0.00	0.97	0.00	0.01	38.95	0.02	98%	0	0	0	2%
Agrostis capillaris	0.00	0.96	0.00	0.03	89.04	0.01	99%	0	0	0	1%
Aira cupaniana	0.42	0.97	1.00	0.02	10000	0.01	97%	2%	0	0	1%
Ammophila arenaria	1.00	0.95	1.00	0.04	10000.00	0.01	0	0	99%	0	1%
Arrhenatherum elatius	0.99	0.00	1.00	0.07	2.46	0.93	0	0	7%	93%	0
Arundo donax	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Avena sterilis	0.89	1.00	0.77	0.00	461.86	0.00	0	100%	0	0	0
Avenula bromoides	0.00	0.83	<0.0001	0.00	6.56	0.17	83%	0	0	0	17%
Brachyelytrum erectum	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Brachypodium distachyon	1.00	1.00	1.00	0.00	852.19	0.00	0	0	100%	0	0
Bromus orcuttianus	1.00	0.65	1.00	<0.0001	324.89	0.35	0	0	65%	0	35%
Catabrosa aquatica	1.00	0.76	1.00	0.21	202.95	0.03	0	0	97%	0	3%
Chaetopogon fasciculatus	1.00	0.95	1.00	0.04	10000	0.01	0	0	99%	0	1%
Cutandia maritima	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Dasypyrum villosum	1.00	0.99	1.00	0.01	743.67	0.01	0	0	99%	0	1%
Deschampsia cespitosa	1.00	0.05	1.00	<0.0001	2.11	0.95	0	0	5%	0	95%
Diarrhena americana	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Echinaria capitata	1.00	0.36	0.76	0.00	1605.95	0.64	0	0	36%	0	64%
Elymus hystrix	0.22	1.00	0.17	0.00	1.10	0.00	100%	0	0	0	0
Enneapogon desvauxii	0.17	1.00	0.17	0.00	0.99	0.00	100%	0	0	0	0
Festuca agustinii	1.00	0.58	0.95	0.00	1683.23	0.42	0	0	58%	0	42%
Festuca capillifolia	0.97	0.00	0.87	0.00	7819.28	1.00	0	0	0	0	100%
Festuca coerulescens	1.00	0.94	1.00	0.05	261.95	0.01	0	0	99%	0	1%
Festuca gigantea	1.00	0.60	0.87	<0.0001	580.75	0.40	0	0	60%	0	40%
Festuca longiauriculata	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Festuca paniculata subsp. *spadicea*	1.00	0.51	0.06	<0.0001	3087.65	0.49	0	0	51%	0	49%
Festuca pratensis	1.00	0.93	0.98	0.00	2908.27	0.07	0	0	93%	0	7%
Festuca rubra	0.12	0.99	<0.0001	0.00	10000	0.01	0	99%	0	0	1%

Table 3. Summary of the Branch-Site REL (BS-REL) results of the studied grass mitochondrial rps3 sequences analyzed in the web interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at each class along each branch. $\omega = d_N/d_S$, where d*N* and d*S* are the numbers of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) $\omega = 0 - 0.6$, Strongly Negative Selection (dark blue bar in Fig. 3); (II) $\omega = 0.61 - 0.9$, Weak Negative Selection (light blue bar); (III) $\omega = 0.91 - 1.5$, Neutral Selection (green bar); (IV) $\omega = 1.51 - 5$, Weak Positive Selection (pink bar); and (V) $\omega > 5$, Strongly Positive Selection (red bar).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
Festuca pseudoeskia	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Festuca scariosa	1.00	0.16	0.94	0.00	2.32	0.84	0	0	16%	0	84%
Gaudinia fragilis	1.00	0.93	1.00	0.03	10000	0.04	0	0	96%	0	4%
Gynerium sagittatum	1.00	0.67	0.99	0.00	9999.98	0.33	0	0	67%	0	33%
Holcus lanatus	1.00	0.61	1.00	0.03	1.70	0.36	0	0	64%	0	36%
Hyparrhenia hirta	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Karroochloa purpurea	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Koeleria vallesiana	0.00	0.70	0.00	0.01	11.92	0.29	71%	0	0	0	29%
Lolium canariense	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Lolium perenne	1.00	0.60	0.99	<0.0001	124.80	0.40	0	0	60%	0	40%
Lolium rigidum	0.00	0.95	<0.0001	0.00	108.45	0.05	95%	0	0	0	5%
Lygeum spartum	0.00	0.00	0.86	0.00	11.65	1.00	0	0	0	0	100%
Molineriella laevis	0.87	0.00	0.78	0.00	30.90	1.00	0	0	0	0	100%
Muhlenbergia montana	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Narduroides salzmannii	1.00	0.57	0.96	0.00	2.37	0.43	0	0	57%	43%	0
Nardus stricta	0.95	0.83	<0.0001	0.00	6.56	0.17	0	0	83%	0	17%
Parapholis incurva	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Phalaris canariensis	1.00	0.22	0.86	0.00	1312.15	0.78	0	0	22%	0	78%
Phyllostachys bambusoides	1.00	0.94	1.00	0.04	1.63	0.02	0	0	98%	0	2%
Poa compressa	0.00	0.43	<0.0001	0.03	16.26	0.54	46%	0	0	0	54%
Poa infirma	0.94	0.92	0.95	0.00	30.71	0.08	0	0	92%	0	8%
Pseudarrhenatherum longifolium	1.00	0.95	1.00	0.03	780.08	0.02	0	0	98%	0	2%
Puccinellia distans	0.45	0.95	0.53	0.03	298.52	0.02	98%	0	0	0	2%
Rostraria cristata	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Schismus barbatus	0.34	0.90	0.34	0.10	1.36	0.00	100%	0	0	0	0
Spartina spp.	1.00	0.69	0.86	<0.0001	920.86	0.31	0	0	69%	0	31%
Sporobolus indicus	0.00	0.00	0.00	0.00	0.0	1.00	100%	0	0	0	0
Sporobolus wrightii	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0

Table 3. Summary of the Branch-Site REL (BS-REL) results of the studied grass mitochondrial rps3 sequences analyzed in the web interface DataMonkey. Maximum likelihood estimates of rate classes with $\omega_1 \leq 1$, $\omega_2 \leq 1$ and ω_3 (unconstrained) and proportion of sites (Pr1, Pr2 and Pr3, respectively) evolving at each class along each branch. $\omega = d_N/d_S$, where d*N* and d*S* are the numbers of non-synonymous and synonymous mutations, respectively. Roman numerals (I to IV) correspond to the total percentage of sites (across ω_1, ω_2 and ω_3) evolving according to distinct selection scenarios: (I) $\omega = 0 - 0.6$, Strongly Negative Selection (dark blue bar in Fig. 3); (II) $\omega = 0.61 - 0.9$, Weak Negative Selection (light blue bar); (III) $\omega = 0.91 - 1.5$, Neutral Selection (green bar); (IV) $\omega = 1.51 - 5$, Weak Positive Selection (pink bar); and (V) $\omega > 5$, Strongly Positive Selection (red bar).

Species	ω_1	Pr1	ω_2	Pr2	ω_3	Pr3	I	II	III	IV	V
Stipa barbata	0.38	0.37	0.42	0.63	1.18	0.00	100%	0	0	0	0
Stipa offneri	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Stipa tenacissima	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Stipagrostis zeyheri	0.00	0.95	0.00	0.01	77.59	0.05	95%	0	0	0	5%
Thysanolaena maxima	0.34	0.96	0.35	0.04	1.15	0.00	100%	0	0	0	0
Tripsacum dactyloides	1.00	0.61	0.97	0.00	977.06	0.39	0	0	61%	0	39%
Triticum aestivum	0.00	0.00	0.00	0.00	0.00	1.00	100%	0	0	0	0
Triticum durum	0.33	1.00	0.00	0.00	0.00	0.56	100%	0	0	0	0
Vulpia ciliata	0.72	0.99	0.73	0.01	672.76	0.00	0	100%	0	0	0
Vulpia geniculata	1.00	0.60	0.87	0.00	96.06	0.40	0	0	60%	0	40%
Vulpia muralis	1.00	0.51	1.00	0.05	3.69	0.44	0	0	56%	0	44%
Zea perennis	1.00	0.86	0.92	0.00	728.00	0.14	0	0	86%	0	14%
Average percentage of the mitochondrial *rps3* marker (this Chapter)							46.76%	4.42%	29.31%	2%	17.5%
Average percentage of the low copy nuclear gene *ß-amylase* (Chapter I; page 63)							85.42%	1.7%	9.72%	0.38%	2.78%

The visual comparison between the Poaceae organismal tree (Fig. 1) and the *rps3* majority rule ML consensus inference (Fig. 3) showed 45 misplaced species. Although most of these misplacements are likely due to the chimeric nature observed within the mitochondrial *rps3* gene, twenty-one of the 45 misplaced accessions were also associated with the existence of neutral and/or positive selection at most of their sites (Table 3, Fig. 3: *Arrhenatherum elatius, Brachypodium distachyon, Catabrosa aquatica, Deschampsia cespitosa, Echinaria capitata, Festuca coerulescens, F. paniculata, F. gigantea, F. pratensis, Lolium perenne, Holcus lanatus, L. rigidum, Lygeum spartum, Nardus stricta, Phyllostachys bambusoides, Poa infirma, Spartina, Tripsacum dactyloides*, *Vulpia geniculata, V. muralis, Zea perennis*). Seven of the 45 misplacements were likely caused by extensive missing positions (Fig. 3; *Arundinaria gigantea* (780 missing positions), *Coix lacryma-jobi* (780), *Danthonia spicata* (270), *Eleusine* (780), *Eragrostis* curvula (780), *Micropyropsis tuberosa* (441), *Vulpia geniculata* (200). The SH test detected incongruence (P=0.54) between the topology of the Poaceae organismal tree (Fig. 1) and the *rps3* majority rule ML consensus tree based on the exons + unambiguously aligned intron positions based on 138 grass accessions.

5. Discussion

5.1 Evidence of HGT within the mitochondrial *rps3* gene in the grasses

We have detected 12 cases of microconversion (1 – 140 nt) (Fig. 2A-C), involving 123 grass taxa (87.8% of the tested sequences). These microconversions add significance to the number of known cases of mitochondrial gene conversion between distantly related sequences (e.g. PALMER *ET AL.*, 2003; KNOOP, 2004; HAO AND PALMER, 2009; HAO *ET AL.*, 2010; and ONG AND PALMER, PERSONAL COMMUNICATION). We have also described the existence of three longer recombinations (Table 2, Fig. 3), involving 20 grass taxa (16.13% of the tested sequences). A number of evolutionary processes could be invoked to explain the presence of the recombinations described. The existence of random mutation in conjunction with homoplasy, where independently derived mutational changes occur in parallel in two or more taxa, could explain several of the observed cases of microconversion; however, this explanation is highly improbable, mainly due to the high number of symmetric mutations in non-related taxa that would be required. Alternatively, the recombinations observed could be explained by nucleotide inversions caused by stem-and-loop formations (KELCHNER AND WENDEL, 1996; MINAYA *ET AL.*, 2013; ONG AND PALMER, PERSONAL COMMUNICATION). However, these structures have not been detected in the studied *rps*3 sequences. Instead, we must consider the

possibility that the nucleotides involved in the observed recombinations are of foreign origin, supporting the existence of HGT within the mitochondrial *rps3* marker. This hypothesis is not without precedent, since the grasses have also provided the first testable case of HGT in a functional nuclear gene (pgiC; GHATNEKAR *ET AL.*, 2006; VALLENBACK *ET AL.*, 2008, 2010).

The most likely hypothesis would be that the foreign fragments originated as the result of a hybridization or other HGT process, since hybridization and HGT occur frequently among members of the grass family due to their allogamous pollination syndrome (cf. JACOBS AND EVERETT, 2000; HEGDE AND WAINES, 2004) and the existence of common potential transmission vectors, such as the ubiquitous systemic and non-systemic fungal endophytes of grasses (SCHARDL *ET AL.*, 2004; SÁNCHEZ MÁRQUEZ *ET AL.*, 2012). The foreign fragments would then be integrated into the resident mtDNA by subsequent rearrangements (cf. BERGTHORSSON *ET AL.*, 2003; PETERSEN *ET AL.*, 2006). Thus, the recipient genome would theoretically be more receptive to foreign DNA that came from closely related species due to the similarities between the donor and recipient sequences (cf. GRAY *ET AL.*, 1992; HANSON AND FLOKERTS, 1992). Our results partially corroborate this assessment, since the major parental sequences hypothesized by the RDP4 program are relatively phylogenetically close to the recombinant species (Table 2). Nevertheless, we have observed more microconversions between species classified in the PACCMAD and BEP clades (6 cases: Fig. 2A), than between members of the Pooideae subfamily (3 cases: Fig. 2B) or between species classified within the Poodae (Poeae – Aveneae complex) (3 cases: Fig. 2C). Since the plant mitochondrial genome is characterized by a low substitution rate (WOLFE *ET AL.*, 1987; PALMER AND HERBON, 1988; LAROCHE *ET AL.*, 1997; MUSE, 2000), the most likely explanation for this unexpected result would be that the foreign sequences were inserted frequently into closely related taxa, but in most cases they are not divergent enough to stand out within the native genome, so they are lost in a sea of apparent irresolution (RICHARDSON AND PALMER, 2007). This limitation may present a significant barrier to obtaining a comprehensive view of the speed and pattern of HGT events in mtDNA, especially if any of the dominant modes of transfer involve mechanisms, such as hybridization, that theorically favor closely related donors and recipients. Alternatively, the presence of other unknown biological (e. g. ubiquitous fungal endophytes) or evolutionary processes (cf. LIAN *ET AL.*, 2013) cannot be ruled out, specifically for explaining the presence of *Nicotiana tabacum* as the most likely minor parental sequence in the second and third recombinations (Table 2).

The chimeric nature observed within the *rps3* gene contributes to explaining the existence of the 38 misplacements (26.4% of the sampled taxa) seen when comparing the

phylogeny based on the *rps3* gene (Fig. 3) with the Poaceae organismal tree (Fig. 1). Recombination events likely were improperly overweighed in the *rps3* phylogenetic reconstructions that, in combination with the very low substitution rates that characterize the mtDNA, produced the phylogenetic misplacements observed, as previously hypothesized by WEILONG AND PALMER (2009). Further, the results of the topological SH test (P=0.54) clearly indicate a significant topological incongruence between these phylogenies. Consequently, the putative HGT events hypothesized support the exclusion of the mitochondrial *rps3* gene as a source of phylogenetic information in the grasses.

5.2 Evolutionary dynamics of the mitochondrial *rps*3 gene and the potential transfer of a functional copy to the plastid genome

Our study also shows additional evidence for the complex evolutionary dynamics of the mitochondrial *rps3* gene in the grasses. Of the two complementary methods used to detect evidence of pseudogenization within the 144 studied *rps3* sequences, the existence of PTCs (AKHUNOV *ET AL.,* 2013) failed to detect potential pseudogenic copies that might have accumulated mutations at the early stages of evolution. The absence of PTCs throughout the sequences studied was corroborated using the amino acid translations to guide the alignment of these coding sequences. However, since exon 1 and parts of exons 2 and 3 of the mitochondrial *rps3* gene were not sequenced in this research (appendix of chapter III; Fig. S1, page 182), this criterion alone was not conclusive to rule out the presence of pseudogenes. Thus, as a complementary method to detect pseudogenization, we evaluated the existence of sequences with a large proportion of sites under neutral or positive selection, as indicative of potential relaxation of selection pressure (cf. KOSAKOVSKY-POND *ET AL.*, 2008, 2011). The 33 sequences (48.52% of the analyzed accessions) found in the mitochondrial *rps3* gene of the grasses with a high proportion of neutral sites, slightly positively selected sites or with a complete absence of purifying selection suggest that the existence of pseudogenization and/or different mechanisms of relaxation of selection pressure could be extensive throughout this gene in the Poaceae. Furthermore, the comparison of the BS-REL results between the mitochondrial *rps3* gene and the low-copy nuclear *ß-amylase* gene in the grasses (chapter I, Table 3) demonstrates that the mitochondrial marker has three times the average percentage of neutral positions, and six times the average percentage of sites under strongly positive selection.

The existence of HGT within the mitochondrial *rps3* gene of the grasses is the most likely hypothesis to explain the recombinations and misplacements observed. Given the more

ancestral split of the BEP and PACCMAD clades, the six microconversions observed between these Poaceae clades (Fig. 2A) were likely mainly resposible for the putative loss of function of the mitochondrial *rps3* gene in the grasses. The subsequent relaxation of evolutionary pressure within the mitochondrial copy would explain the absence of purifying selection detected in 48.52% of the tested grass sequences.

The recombinations, misplacements and pseudogenizations described within the *rps3* gene seem to indicate that it has lost functionality in the mitochondrial genome of the grasses. This assumption supports the hypothesis that the *rps3* gene has been incorporated into the plastid genome, as previously suggested by ADAMS *ET AL*. (2002), and ADAMS AND PALMER (2003). The Intracellular gene transfer (IGT) hypothesis, proposed for diverse plant mitochondrial genes (MARIENFELD *ET AL*., 1999; ADAMS *ET AL*., 2002; ADAMS AND PALMER, 2003; KOULINTCHENKO *ET AL*., 2003; RICHARDSON AND PALMER, 2007; and HAO AND PALMER, 2009) could also be invoked for explaining the observed evolutionary pattern of the mitochondrial rps3 copy in the grasses.

6. Literature cited

Adams, K.L., Qiu, Y.L., Stoutemyer, M., Palmer, J.D., 2002. Punctuated evolution of mitochonrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences, U.S.A.* 99, 9905-9912.

Adams, K.L., Palmer, J.D., 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29, 380-395.

Akhunov, E.D., Sehgal, S., Liang, H., Wang, S., Akhunova, A.R., Kaur, G., Li, W., Forrest, K.L., See, D., Simková, H., Ma, Y., Hayden, M.J., Luo, M., Faris, J.D., Dolezel, J., Gill, B.S., 2013. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. *Plant Physiology* 161, 252-265.

Albert, B., Godelle, B., Gouyon, P.H., 1998. Evolution of the plant mitochondrial genome: dynamics of duplication and deletion of sequences. *Journal of Molecular Evolution* 46, 155-158.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel metropolis coupled markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC Bioinformatics* 20, 407-415.

Andre, C., Levy, A., Walbot, V., 1992. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends in Genetics* 8, 128-132.

Barkman, T.J., McNeal, J.R., Lim, S., Coat, G., Croom, H.B., Young, N.D., dePaphilis, C.W., 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evolutionary Biology* 7, 248-263.

Bergthorsson, U., Adams, K.L., Thomason, B., Palmer, J.D., 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424, 197-201.

Bergthorsson, U., Richardson, A.O., Young, G.J., Goertzen, L.R., Palmer, J.D., 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences, U.S.A.* 51, 17747-17752.

Bock, H., Brennicke, A., Schuster, W., 1994. *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondrial: GTG as a potential translation initiation codon in plant mitochondria?. *Plant Molecular Biology* 24, 811-818.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Brunner, P.C., Keller, N., McDonald, A., 2009. Wheat domestication accelerated evolution and triggered positive selection in the ß-Xylosidase enzyme of *Mycosphaerella graminicola*. *PLoS ONE* 11, *e7884*.

Campbell, C.S., Kellogg, E.A., 1987. Sister group relationships of the Poaceae. In: T.R. Soderstrom, K.W. Hilu, C.S. Campbell, M.E. Barkworth (Eds.), *Grass Systematics and Evolution*. Smithsonian Institution Press, Washington DC, pp. 217-224.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related

grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British and European flora.* Botanical Society of the British Isles, London, pp. 45-70.

Cerros-Tlatilpa, R., Travis Columbus, J., Barker, N.P., 2011. Phylogenetic relationships of *Aristida* and relatives (Poaceae, Aristidoideae) based on noncoding chloroplast (*trnL-F, rpl16*) and nuclear (ITS) DNA sequences. *American Journal of Botany* 98, 1868-1886.

Clayton, W.D., Renvoize, S.A., 1986. *Genera Gramineum: Grasses of the World*. *Kew Bulletin*. Series XIII, Royal Botanic Gardens, Kew.

Conklin, P.L., Hanson, M.R., 1994. Recombination of plant mitochondrial genomes. In: J. Paszkowski (Ed.), *Homologous Recombination and Gene Silencing in Plants*. Kluwer Academic Publishers, Netherlands, pp. 61-81.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristic and parallel computing. *Nature Methods* 9, 772.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex (Pooideae, Poaceae). *Kew Bulletin* 62, 407-424.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procesure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19, 11-15.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.

Eyre-Walker, A., Gaut, B.S., 1997. Correlated Rates of synonymous site evolution Across plant genomes. *Molecular Phylogenetics and Evolution* 14, 455-460.

Farris, J.S., 1989. The retention index and the rescaled consistency index. *Cladistics* 5, 417-419.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. *Cladistics* 10, 315-319.

Felsenstein, J., 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* 22, 240-249.

Ghatnekar, L., Jaarola, M., Bengtsson, B.O., 2006. The introgression of a functional nuclear gene from Poa to Festuca ovina. *Proceedings of the Royal society B* 273, 395–399.

G.P.W.G., The Grass Phylogeny Working Group., 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88, 373-457.

G.P.W.G., The Grass Phylogeny Working Group., 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Gray, M.W., Hanic-Joyce, P.J., Covello, P.S., 1992. Transcription, processing and editing in plant mitochondria. *Annual Review of Plant Physiology and Plant Molecular Biology* 43, 145-175.

Gray, M.W., Lang, B.F., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T.G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., Burger, G., 1998. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Research* 26, 865-878.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52, 696-704.

Hanson, M.R., Folkerts, O., 1992. Structure and function of the higher plant mitochondrial genome. *International Review of Cytology* 141, 129-165.

Hao, W., Palmer, J.D., 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proceedings of the National Academy of Sciences, U.S.A.* 106, 16728-16733.

Hao, W., Richardson, A.O., Zheng, Y., Palmer, J.D., 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. *Proceedings of the National Academy of Sciences, U.S.A.* 107, 21576-21581.

Hegde, S.G., Waines, J.G., 2004. Hybridization and introgression between bread wheat and wild and weedy relatives in North America. *Crop Science Society of America* 44, 1145-1155.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Jacobs, S.W.L., Everett, J., 2000. *Grasses: systematics and evolution*. CSIRO publishing.

Kelchner, S.A., Wendel, J.F., 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. *Current Genetics* 30, 259-262.

Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624-626.

Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18, 1-32.

Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics* 46, 123-139.

Kosakovsky-Pond, S.L., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. *BMC Bioinformatics* 21, 676-679.

Kosakovsky-Pond, S.L., Poon, A.F.Y., Leigh Brown, A.J., Frost, S.D.W., 2008. A Maximum Likelihood Method for Detecting Directional Evolution in Protein Sequences and its Application to Influenza A Virus. *Molecular Biology and Evolution* 25, 1809-1824.

Kosakovsky-Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* 28, 3033-3043.

Koulintchenko, M., Konstantinov, Y., Dietrich, A., 2003. Plant mitochondria actively import DNA via the permeability transition pore complex. *European Molecular Biology Organization* 22, 1245-1254.

Kubo, T., Mikami, T., 2007. Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum* 129, 6-13.

Kubo, T., Newton, K.J., 2008. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8, 5-14.

Lafay, B., Lloyd, A.T., McLean, M.J., Devine, K.M., Sharp, P.M., Wolfe, K.H., 1999. Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Research* 27, 1642-1649.

Lang, F.F., Gray, M.W., Burger, G., 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* 33, 351-397.

Laroche, J., Li, P., Maggia, L., Bousquet, J., 1997. Molecular evolution of angiosperm mitochondrial introns and exons. *Proceedings of the National Academy of Sciences, U.S.A.* 94, 5722-5727.

Laroche, J., Bousquet, J., 1999. Evolution of the mitochondrial rps3 intron in perennial and annual angiosperms and homology to nad5 intron1. *Molecular Biology and Evolution* 16, 441-452.

Li, W.-H., 1997. *Molecular Evolution*. Sinauer, Sunderland, MA.

Lian, S., Lee, J., Cho, W.K., Yu, J., Kim, J. Choi, H., Kim, K., 2013. Phylogenetics and recombination analyses of tomato spotted wilt virus. *PLoS ONE* 8, e63380.

Linder, H.P., Rudall, P.J., 1993. The megagametophyte in *Anarthria* (Anarthriaceae, Poales) and its implications for the phylogeny of the Poales. *American Journal of Botany* 80, 1455-1464.

Maddison, D.R., Maddison, W.P., 2008. *MacClade 4.08 version for OS X*. Sinauer Associates, Inc.

Marienfeld, J., Unseld, M., Brandt, P., Brennicke, A., 1999. The mitochondrial genome of the flowering plant *Arabidopsis thaliana* is a composition of native and immigrant information. *Trends in Plant Science* 4, 495-502.

Martin D.P, Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. *BMC bioinformatics* 26, 2462-2463.

Mason-Gamer R.J., 2005. The *b-amylase* genes of grasses and a phylogenetic analysis of the Triticeae (Poaceae). *American Journal of Botany* 92, 1045-1058.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

Miyata, T., Yasunaga,T., 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* 16, 23–36.

Muse, S.V., 2000. Examining rates and patterns of nucleotide substitution in plants. *Plant Molecular Biology* 42, 25-43.

Newton, K.J., Gabay-Laughnan, S., dePaepe, R., 2004. Mitochondrial mutations in plants. In: D.A. Day, A.H. Millar, J. Whelan (Eds.), *Plant Mitochondria: From Genome to Function*. Kluwer, Dordrecht, The Netherlands, pp. 121-142.

Nickrent, D.L., Blarer, A., Qiu, Y-L., Vidal-Russell, R., Anderson, F.E., 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. *Evolutionary Biology* 4, 40.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positive selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929-936.

Palmer, J.D., Herbon, L.A., 1988. Plant mitochondrial-DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28, 87-97.

Palmer, S., Schildkraut, E., Lazarin, R., Nguyen, J., Nickoloff, J.A., 2003. Gene conversion tracts in *Saccharomyces cerevisiae* can be extremely short and highly directional. *Nucleic Acids Research* 31, 1164-1173.

Petersen, G., Seberg, O., Davis, J.T., Goldman, D.H., Stevenson, W., Campbell, L.M., Michelangeli, F.A., Specht, C.D., Chase, M.W. Fay, M. F., Pires, J.C., Freudenstein, J.V., Hardy, C.R., Simmons, M.P., 2006. Mitochondrial data in monocot phylogenetics. *Aliso* 22, 52-62.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *BMC Bioinformatics* 21, 2531-2533.

Preston, J.C., Kellogg, E.A., 2006. Reconstructing the evolutionary history of paralogous APETALA1/ FRUITFULL-Like genes in grasses (Poaceae). *Genetics* 174, 421-437.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from platid *trnT-F* and nuclear *ITS* sequences. *American Journal of Botany* 94, 1554-1596.

Richardson A.O., Palmer, J.D., 2007. Horizontal gene transfer in plants. *Journal of Experimental Botany* 58, 1-9.

Ronquist, F., Huelsenbeck, J.P., Teslenko M., 2011. MrBayes version 3.2 manual: tutorials and model summaries (http://mrbayes.sourceforge.net/mb3.2_manual.pdf).

Sánchez Márquez, S., Bills, G.F., Herrero, N., Zabalgogeazcoa, I., 2012. Non systemic fungal endophytes of grasses. *Fungal Ecology* 5, 289-297.

Schardl, C.L., Leuchtmann, A., Spiering, M.J., 2004. Symbioses of grasses with seedborne fungal endophytes. *Annual Review of Plant Biology* 55, 315-340

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *matK* gene-3'*trnK* exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16, 1114-1116.

Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three

plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Swofford, D.L., 2002. *PAUP*: Phylogenetic analysis using parsimony and other methods. Version 4b10*. Sunderland (MA): Sinauer Associates.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear *ITS* sequences. *Systematic Botany* 27, 241-251.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Vallenback, P., Jaarola, M., Ghatnekar, L., Bengtsson, B., 2008. Origin and timing of the horizontal transfer of a PgiC gene from Poa to Festuca ovina. *Molecular Phylogenetics and Evolution* 46, 890-896.

Vallenback, P., Ghatnekar, L., Bengtsson, B., 2010. Structure of the Natural Transgene PgiC2 in the Commo Grass Festuca ovina. *PLoS ONE*, e13529.

Watson, L., Dallwitz, M.J., 1992 onwards. The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. Version: 12th August 2014. http://delta-intkey.com

Weilong, H., Palmer, J., 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitocondrial genes. *Proceedings of the National Academy of Sciences, U.S.A.* 29, 16728-16733.

Wolfe, K.H., Li, W., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, U.S.A.* 84, 9054-9058.

Woloszynska, M., Bocer, T., Mackiewicz, P., Janska, H., 2004. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of *Phaseolus*. *Plant Molecular Biology* 56, 811-820.

Wolstenholme, D.R., Fauron, C.M., 1995. Mitochondrial genome organization. In: C.S. Levings, I.K. Vasil (Eds.), *The molecular biology of plant mitochondria*. Kluwer Academic, Amsterdam, pp. 1-59.

Won, H., Renner, S.S., 2003. Horizontal gene transfer from flowering plants to *Gnetum*. *Proceedings of the National Academy of Sciences, U.S.A*. 100, 10824-10829.

Yang, Z., Bielawski, J.P., 2000. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution* 19, 908-917.

Zwickl , D.J., 2006. *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence data sets under the maximum likelihood criterion*. Ph.D. dissertation, University of Texas, Austin, Texas, USA.

7. Appendix

7.1 Procedures for DNA isolation, amplification, and sequencing

Amplification reactions were performed using a reaction mix containing 0.8 μl of forward and reverse primer (20mM), 2 μl of 10X PCR buffer, 3μl of $MgCl_2$ (50 mM), 0.2 μl of *Taq* DNA polymerase (5U/μl), 0.5 μl of Deoxynucleoside Triphosphate Set PCR Grade (10mM), between 1 and 1.5 μl of template DNA, and an aliquot of sterile ultrapure water (MilliQ) added to achieve a final volume of 20 μl. The amplification parameters consisted of an initial denaturing step of 3 min at 94 ºC, followed by 35 cycles of 2 min denaturing at 92 ºC, 1 min annealing at 55 ºC, and 2 min extension at 72 ºC, followed by a final extension step of 7 min at 72 ºC. All PCR products were cleaned up using ExoSAP-IT® following the protocol indicated by the manufacturer.

rps3 products were used as templates for the cycle sequencing reaction using the ABI Big-Dye Terminator Cycle Sequencing Kit® (Applied Biosystems). The external and internal primers listed above were used in this step. The sequencing reactions were prepared with 3 μl of $MgCl_2$ (50mM), 1 μl of primer, 1 μl of Premix Ready Reaction and between 1 and 3 μl of purified DNA. An aliquot of sterile ultrapure water (MilliQ) was added to achieve a final volume of 10 μl. PCR was performed with one denaturing cycle of 30 sec at 96 ºC, 24 cycles of 10 sec denaturing at 96 ºC, 5 sec annealing at 50 ºC, and 4 sec extension at 60 ºC; then a termination step of 4 min at 60 ºC. Finally, the region was sequenced at the Genome Sequencing Service of the Center for Genomics and Bioinformatics, Indiana University.

7.2 Procedures for the Bayesian and parsimony-based searches

Bayesian analysis of both the independent (ITS, *ß-amylase, mat*K, *ndh*F, *trn*TL and *trn*LF) and the combined data sets used to infer the organismal tree was conducted according to the GTR+I+ gamma model. Two independent analyses were run with one million generations each using the Markov Chain of Monte Carlo (MCMC) algorithm. Trees were sampled every 1,000 generations, and 25% of the generations were discarded as burn-in once stability in the likelihood values was attained. A half-compatible consensus Bayesian tree was computed from the 750 posterior-probability saved trees.

Maximum Parsimony analyses conducted to infer the organimal and mitochondrial *rps3* phylogenies were based on heuristic searches of 10,000 random-order-entry trees, with TBR branch swapping and saving no more than 10 trees of length equal to or shorter than 10 per replicate. The most parsimonious trees were used to compute the respective strict and 50% Majority-Rule consensus trees. Branch support was estimated through 1,000 bootstrap replicates (FELSENSTEIN, 1985) using the TBR-M (Tree Bisection Reconstruction swapping, MULPARS OFF) strategy of DEBRY AND OLMSTEAD (2000) as a method to reduce computational time. Following the criteria of MINAYA *ET AL*. (2013) and DÍAZ-PÉREZ *ET AL*. (2014), clades with bootstrap support values (BS) of 75-100% (PAUP*) or Posterior Probability support values (PPS) of 90-100% (MrBayes 3.2.2) were considered moderately to strongly supported.

7.3 Supplementary Literature

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca, Lolium, Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: Bailey, J., Ellis, R.G. (Eds.), *Current taxonomic research on the British & European flora*. Botanical Society of the British Isles, pp. 45-70.

DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology* 49, 171-179.

Díaz-Pérez, A., Sharifi-Tehrani, M., Inda, L.A., Catalan, P., 2014. Polyphyly, gene-duplication and extensive allopolyploidy framed the evolution of the ephemeral *Vulpia* grasses and other fine-leaved Loliinae (Poaceae). *Molecular Phylogenetics and Evolution* 79, 92-105.

Felsenstein J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783-791.

G.P.W.G., The Grass Phylogeny Working Group, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from platid *trnT-F* and nuclear *ITS* sequences. *American Journal of Botany* 94, 1554-1596.Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *matK* gene-3'*trnK* exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid *matK* gene-3'*trnK* exon and nuclear *ITS* sequences. *Taxon* 58, 405-424.

Soreng, R.J., Peterson, P.M., Davidse, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved Festuca lineages (Poaceae) based on nuclear *ITS* sequences. *Systematic Botany* 27, 241-251.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trnL-F* sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

7.4 Supplementary figure

Primer	Sequence
2a-for	GTTCGATACGTCCACCTAC
2b-for	CAAAGAAGTCCCGCTCTGAC
2c-bac	AAGCTTCCTCCCTTCGCTCC
2d-bac	GTACGTTTCGGATATAGCAC

Figure S1. Schematic drawing of the *rps3* gene. Rectangles and connecting line show the relative locations of exons and introns respectively. The arrows show the amplification and sequencing primers used in this study. The sequences of these primers are shown in a table below the figure (see Materials and Methods for more details).

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Cycadaceae Pers.								
Cycas revoluta Thunb.	GenBank	AY345867	-	-	-	-	-	-
Brassicaceae Burnett								
Arabidopsis thaliana (L.) Heynh.	GenBank	KF923189.1	-	-	-	-	-	-
Solanaceae Juss.								
Nicotiana tabacum L.	GenBank	BA000042	-	-	-	-	-	-
Amaranthaceae Juss.								
Beta vulgaris L.	GenBank	BA000009	-	-	-	-	-	-
Joinvilleaceae Tomi. and A.C.Sm.								
Joinvillea ascendens Gaudich. ex Brongn. and Gris (Anc [Join])	E. A. Kellogg collection, GenBank	xxxx	JX536496	-	JQ972979	JQ972955	U21973.1	AF164380.1
Joinvillea plicata (Hook. f.) Newell & B.C. Stone	,GenBank	-	-	AF019784.1	-	-	-	-
Poaceae Bamhart								
BEP								
Ehrhartoideae Link								
Oryzeae Dumort.								
Oryzinae Griseb.								
Leersia virginica Willd. (Leers virg)	E. A. Kellogg collection	xxxx	-	-	-	-	-	-
Leersia Sw. spp.	GenBank	-	JX536520	AF019793.1	JQ972993	JQ972960	FJ908350.1	AF148677.1
Oryza sativa L. (Oryza sati)	GenBank	AP011077	L10346.1	EF221613.1	DQ415935.1	EF137577.1	FJ908349.1	AF148650.1
Bambusoideae Luerss.								
Bambuseae Kunth ex Dumort.								
Arundinariinae Nees ex Lindl.								
Arundinaria gigantea (Walter) Muhl. (Arund giga)	J. Triplett 197, GenBank	xxxx	-	AY004759.1	GQ468376.1	EF137522	FJ643707.1	-
Arundinaria tecta (Walter) Muhl.	GenBank	-	-	-	-	-	-	EF125165.1
Shibataeeae Nakai								
Phyllostachys bambusoides Sielbold and Zucc. (Phyll bamb)	J. Triplett 121, GenBank	xxxx	-	-	FJ644171.1	-	-	AB088805.1
Phyllostachys Siebold & Zucc. spp.	GenBank	-	-	DQ131503.1	-	EF137544.1	U21970.1	-

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*β-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Bambusinae J. Presl								
Bambusa vulgaris Schrad. ex J.C. Wendl.	GenBank	-	-	-	FJ644133.1	-	FJ643709	-
Bambusa oldhamii Munro (Bambu oldh)	GenBank	xxxx	-	AY839707.1		EU434054.1	.-	EU434246.1
Guaduinae Soderstr. & R.P. Ellis								
Otatea acuminata (Munro) C.E. Calderón & Soderstr. (Otate acum)	L. Clark 1312 (ISC), GenBank	xxxx	-	GQ384316.1	JQ408612	EF137542.1	AF182350.1	EF137436.1
Chusqueinae Soderstr. & R.P. Ellis								
Neurolepis nana L.G. Clark (Neuro nana)	L. Clark & P. Asimbaya 1453 (ISC), GenBank	xxxx	-	-	xxxx	xxxx	FJ751653.1	-
Olyreae Kunth ex Spenn.								
Parianinae Hack.								
Pariana radiciflora Sagot ex Döll (Paria radi)	L. Clark & W. Zhang 1344 (ISC), GenBank	xxxx	-	-	FJ644168.1	-	AF182354.1	AF164387.1
Olyrinae Kromb.								
Sucrea maculata Soderstr. (Sucre macu)	L. Clark & W. Zhang 1345 (ISC), GenBank	xxxx	-	-	FJ644212.1	xxxx	AF182343.1	-
Pooideae Benth.								
Brachyelytreae Ohwi								
Brachyelytrum erectum (Schreb.) P. Beauv. (Brach erec)	Indiana: Salt Creek. UZ 111.08, GenBank	xxxx	-	EU489105.1	FJ644136.1	EU489326.1	U22005.1	AF164384.1
Lygeeae J. Presl / Nardeae W.D.J. Koch								
Lygeum spartum L. (Lygeu spar)	Spain. Zaragoza: Vedado de Peñaflor. UZ 136.07, GenBank	xxxx	JX536522	FM179417.1	-	EU434098.1	JX438110	JX438063
Nardus stricta L. (Nardu stri)	Spain. Granada: Sierra Nevada. UZ 229.07, GenBank	xxxx	JX536523	FM179420.1	JQ972994	EU434097.1	JX438109	JX438062
Stipeae Dumort.								
Melicinae Fr.								
Glyceria declinata Bréb. (Glyce decl)	Spain. Caceres: Navalmoral. UZ 25.07, GenBank	xxxx	-	EF674560.1	KJ529258	EU223364.1	KJ529425	-
Glyceria nemoralis (R. Uechtr.) R. Uechtr. & Körn.	GenBank	-	-	-	-	-	-	AM234578.1
Melica minuta L. (Melic minu)	Spain: Córdoba: Cabra. UZ 102.07, GenBank	xxxx	-	-	KJ529259	KJ529403	KJ529426	KJ529313
Melica picta K. Koch	GenBank	-	-	FM179418.1	-	-	-	-
Stipinae Griseb.								
Achnatherum calamagrostis (L.) P. Beauv. (Achna cala)	Spain. Huesca: Sierra de Guara. UZ282.07, GenBank	xxxx	JX536525	EU194678.1	JQ972996	EU200806.1	JX438111	JX438064
Nassella laevissima (Phil.) Barkworth (Nasse laev)	Chile. Cultivated in IBB	xxxx	JX536524	JQ972931	JQ972995	-	-	-
Nassella (Trin.) E. Desv.spp.	GenBank	-	-	-	-	DQ887454.1	AF251450.1	AF164406.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Stipinae Griseb.								
Stipa barbata Desf. (Stipa barb)	Spain. León. Santa Lucia. LEB. GenBank	xxxx	-	-	-	-	xxxx	xxxx
Stipa offneri Breistr. (Stipa off)	Spain. Zaragoza: Vedado de Peñaflor. UZ 21.2000, GenBank	xxxx	JX536527	FN434568.1	JQ972997	JQ972961	JX438112	AF164407.1
Stipa parviflora Desf. (Stipa parv)	Spain. Zaragoza: Vedado de Peñaflor. UZ 21.2000, GenBank	xxxx	-	EU194790.1	xxxx	EU200930.1	U71045.1	GU254719.1
Stipa pennata L. (Stipa penn)	Spain. Huesca: Casbas. UZ 10.2001, GenBank	xxxx	-	FM179438.1	xxxx	xxxx	GU254759.1	GU254718.1
Stipa tenacissima L. (Stipa tena)	Spain. Almeria: Gador. UZ 180.07, GenBank	xxxx	JX536526	xxxx	KJ529260	KJ529404	KJ529427	KJ529316
Diarrheneae C.S. Campb.								
Diarrheninae Ohwi								
Diarrhena americana P. Beauv. (Diarr amer)	USA. Indiana: Belmont. UZ 116.08, GenBank	xxxx	-	FM179400.1	JQ972998	JQ972962	-	FM253123.1
Diarrhena obovata (Gleason) Brandenburg	GenBank	-	-	-	-	-	DQ786833.1	-
Brachypodieae Harz								
Brachypodiinae Hack.								
Brachypodium distachyon (L.) P. Beauv. (Brach dist)	Spain. Caceres: Malpartida. UZ 28.07, GenBank	xxxx	-	JN187616.1	JQ972999	JQ972963	U71043.1	AM234568.1
Bromeae Dumort. - Triticeae Dumort.								
Bromus orcuttianus Vasey (Bromu orcu)	Peterson, Saarela & Sears 19787, GenBank	xxxx	-	-	EU036161.1	EU036186.1	-	-
Bromus L. spp.	,GenBank	-	HE565905	FM179394.1	-	-	DQ786821.1	AM234570.1
Agropyron cristatum (L.) Gaertn. (Agrop cris)	Spain. Zaragoza: Vedado de Peñaflor. UZ 134.07, GenBank	xxxx	JX536542	JQ972932	AF519115.1	AY740792.1	JX438114	JX438066
Dasypyrum villosum (L.) P. Candargy (Dasyp vill)	Italy. Sardinia: Cagliari. UZ 158.07, GenBank	xxxx	AY821699	JQ972933	-	JQ972965	JX438115	HM770820.1
Elymus hystrix L. (Elymu hyst)	USA. Indiana: Belmont. UZ 113.08	xxxx	-	-	-	-	-	-
Elymus repens (L.) Gould	GenBank	-	EU282245	GQ373267.1	DQ914534.1	DQ912406	AF267682.1	HM850576.1
Triticum aestivum L. (Triti aest)	Spain. Huesca, cultivated from seeds, GenBank	xxxx	JX536540	FM998928.1	JQ973001	AF148757.1	DQ247921.1	DQ420028.1
Triticum durum Desf.	Spain. Huesca, cultivated from seeds, GenBank	xxxx	JX536541	HQ285920.1	DQ419995.1	EU013671.1	DQ247920.1	DQ420050.1
Poeae R.Br – Aveneae Dumort.								
Phalaridinae Fr.								
Phalaris canariensis L. (Phala cana)	Spain. Valencia: Albal. UZ 338.07, GenBank	xxxx	JX536545	FJ178782.1	DQ631509.1	DQ631443.1	JX438117	JX438068
Agrostidinae Fr.								
Agrostis capillaris L. (Agros capi)	Spain. Lugo: Rábade. UZ 299.07, GenBank	xxxx	JX536549	FM179384.1	AY450936.1	JQ972966	JX438119	JX438069
Ammophila arenaria (L.) Link (Ammo aren)	Spain. Huelva: Mazagón. UZ 75.07, GenBank	xxxx	-	DQ539590.1	DQ631522.1	DQ631456.1	JX438118	AM234561.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Agrostidinae Fr.								
Chaetopogon fasciculatus (Link) Hayek (Chaet fasc)	Spain. Huelva: El Rocío. UZ 77.07, GenBank	xxxx	JX536550	DQ539593.1	DQ631523.1	DQ631457.1	JX438121	-
Polypogon maritimus Willd. (Polyp mari)	Spain. Badajoz: Entrin Alto-Entrin Bajo. UZ 52.07, GenBank	xxxx	JX536551	DQ336818.1	DQ336863.1	DQ336838.1	JX438122	JX438071
Triplachne nitens Link (Tripl nite)	Spain. Almeria: Cabo de Gata. UZ 365.07, GenBank	xxxx	JX536548	DQ336816.1	DQ336861.1	DQ336836.1	JX438120	JX438070
Anthoxanthinae A. Gray								
Anthoxanthum aristatum Boiss.	Spain. Caceres. Crta. Navalmoral. UZ 24.07, GenBank	xxxx	JX536553	JQ972934	DQ631515.1	DQ631449.1	JX438123	-
Aveninae J. Presls								
Arrhenatherum elatius (L.) P. Beauv. ex J. Presl & C. Presl (Arrhe elat)	Spain. Badajoz: Tentudis. UZ55.07, GenBank	xxxx	JX536558	FM179388.1	DQ336866.1	EF137591.1	JX438126	JX438074
Avena sterilis L. (Avena ster)	Spain. Caceres: Navalmoral. UZ19.07, GenBank	xxxx	JX536555	EU833817.1	JQ973002	EU833883.1	JX438124	JX438073
Helictotrichon filifolium (Lag.) Henrard (Helic fili)	Spain. Valencia: Requena. UZ346.07, GenBank	xxxx	JX536557	DQ336819.1	DQ336864.1	DQ336839.1	JX438125	-
Pseudarrhenatherum longifolium (Thore) Rouy (Pseud long)	Spain. Coruña: Castro Elvira. UZ 298.07, GenBank	xxxx	JX536559	FM179430.1	JQ973003	JQ972967	JX438127	JX438075
Koeleriinae Asch. & Graebn.								
Avellinia michelii (Savi) Parl. (Avell mich)	Spain. Cádiz: Sanlucar de Barrameda. UZ 98.07, GenBank	xxxx	JX536563	JQ972935	DQ631531.1	DQ631465.1	JX438131	JX438079
Gaudinia fragilis (L.) P. Beauv. (Gaudi frag)	Spain. Caceres: Malpartida. UZ 30.07, GenBank	xxxx	-	DQ539600.1	DQ631545.1	DQ631478.1	KJ529444	EF137499.1
Koeleria vallesiana (Honck.) Gaudin (Koele vall)	Spain. Madrid: Torrelaguna. UZ 4.07, GenBank	xxxx	-	DQ539601.1	DQ631536.1	DQ631468.1	JX438129	JX438077
Rostraria cristata (L.) Tzvelev (Rostr cris)	Spain, Toledo: Ontígola. UZ 9.07, GenBank	xxxx	JX536562	DQ336833.1	DQ336879.1	DQ336853.1	JX438130	JX438078
Trisetum loeflingianum (L.) P. Beauv. (Trise loef)	Spain. Toledo: Ontígola. UZ 10.07, GenBank	xxxx	JX536560	DQ539608.1	DQ631539.1	DQ631473.1	JX438128	JX438076
Poinae Dumort. – Puccinelliinae Soreng & J.I. Davis – Alopecurinae Dumort.								
Alopecurus arundinaceus Poir. (Alope arun)	Spain. Badajoz: Entrin Alto-Entrín Bajo. UZ 53.07, GenBank	xxxx	-	-	JQ973004	JQ972968	JX438133	JX438080
Alopecurus borealis Trin.	GenBank	-	-	EU792345.1	-	-	-	-
Arctagrostis latifolia (R. Br.) Griseb. (Arcta lati)	Gillespie collection, GenBank	xxxx	-	EU921830.1	DQ353969.1	AY237904.1	DQ786813.1	DQ786885.1
Catabrosa aquatica (L.) P. Beauv. (Catab aqua)	Spain. Huesca: Formigal. UZ.55.07, Gen Bank	xxxx	HE565906	DQ539565.1	JQ973005	JQ972969	JX438135	JX438061
Molineriella laevis (Brot.) Rouy (Molin laev)	Spain. Toledo: Alcañizo. UZ 17.07	xxxx	-	xxxx	KJ529275	KJ529413	KJ529449	DQ786929
Poa bulbosa L. (Poa bulbos)	Spain. Alicante: Alcoy. UZ 334.07, GenBank	xxxx	JX536565	EU792388.1	EU792472.1	DQ354035.1	AY589098.1	HE966970.1
Poa compressa L. (Poa compre)	Spain. Huesca: Abiego. UZ 51.08	xxxx	-	EU792395.1	DQ354003.1	AY504649.	AY589115.1	KJ529347

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Poinae Dumort. – Puccinelliinae Soreng & J.I. Davis – Alopecurinae Dumort.								
Poa infirma Kunth (Poa infi)	Spain. Caceres: Tornavacas. UZ 37.07, GenBank	xxxx	JX536566	AF393012.1	DQ367407.1	AF488773.1	JX438132	JX438058
Puccinellia distans (Jacq.) Parl. (Pucci dist)	Spain. Barcelona: Prat. UZ 156.07, GenBank	xxxx	JX536564	AF532934.1	DQ336859.1	AF533024.1	JX438134	JX438081
Airinae Fr. - Holcinae Dumort. - Sesleriinae Parl.								
Aira cupaniana Guss. (Aira cupan)	Spain. Caceres: Navalmoral. UZ 21.07, GenBank	xxxx	-	-	DQ631508.1	DQ631442.1	KJ529458	DQ786879.1
Aira caryophyllea L.	GenBank	-	-	AM049252	-	-	-	-
Corynephorus canescens (L.) P. Beauv.	Spain. Avila: Tornavacas. UZ 35.07, GenBank	xxxx	JX536568	DQ539578.1	DQ631506.1	EF137598.1	JX438136	JX438082
Deschampsia cespitosa (L.) P. Beauv. (Desch cesp)	Spain. Lugo: Curtis. UZ 305.07, GenBank	xxxx	JX536569	HQ114385.1	EF584984.1	EF137601.1	U71012.1	KC474527.1
Deschampsia flexuosa (L.) Trin. (Desch flex)	Spain. Lugo: Xistral-Peña Vella. UZ 313.07, GenBank	xxxx	JX536570	JQ972936	DQ631505.1	AY237913.1	JX438137	JX438083
Echinaria capitata (L.) Desf. (Echin capi)	Spain. Toledo: Ontígola. UZ 13.07, GenBank	xxxx	JX536572	JQ972938	DQ631519.1	DQ631453.1	JX438138	AM234599.1
Holcus lanatus L. (Holcu lana)	Spain. Leon: Sierra de Ancares. UZ 291.07, GenBank	xxxx	JX536571	JQ972937	DQ631503.1	EF137606.1	DQ786849.1	JX438084
Avenula *sensu stricto*								
Avenula bromoides (Gouan) H. Scholz (Avenu brom)	Spain. Zaragoza: Vedado de Peñaflor. UZ 133.07, GenBank	xxxx	-	JQ972940	DQ631525.1	DQ631459.1	JX438142	JX438087
Dactylidinae Stapf / Parapholiinae Caro – Cynosurinae Fr.								
Cutandia maritima (L.) Barbey (Cutan mari)	Spain. Valencia: El Puig. UZ 324.07. GenBank	xxxx	-	EF584915.1	EF584987.1	AF487618.1	KJ529472	KJ529370
Cynosurus cristatus L. (Cynos cris)	Spain. Leon: Sierra de Ancares. UZ 285.07	xxxx	JX536573	JQ972939	JQ973006	EF137599	JX438139	DQ786901.1
Dactylis hispanica Roth (Dacty hisp)	Spain. Zaragoza: Vedado de Peñaflor. UZ 116.07, GenBank	xxxx	JX536575	AF393014.1	DQ353961.1	AF533027.1	-	-
Desmazeria rigida (L.) Tutin (Desma rigi)	Spain. Zaragoza: La Alfranca. UZ 124.07, GenBank	xxxx	-	FM179399.1	JQ973007	JQ972970	JX438141	JX438086
Lamarckia aurea (L.) Moench	Spain. Caceres: Tornavacas. UZ 36.07, GenBank	xxxx	JX536576	AF532936.1	DQ631490.1	AF533029.1	JX438140	JX438085
Parapholis filiformis (Roth) C.E. Hubb. (Parap fili)	Spain. Castellón: Sagunto. UZ 326.07	xxxx	-	xxxx	KJ529282	KJ529415	KJ529467	KJ529365
Parapholis incurva (L.) C.E. Hubb. (Parap incu)	Spain. Zaragoza: Vedado de Peñaflor. UZ 127.07, GenBank	xxxx	JX536574	FM179422.1	DQ631491.1	AF533036.1	DQ786859.1	EF137508.1
Loliinae Dumort.								
Broad-leaved Loliinae								
***Festuca* subgen. *Drymanthele* – *F.* subgen. *Leucopoa* – *F.* sect. *Lojaconoa* Gand. – *F.* sect. *Pseudoscariosa* Krivot. – *F.* sect. *Scariosae* Hack.**								
Festuca altissima All. (Festu alti)	Norway. Akershus: Frogn parish, GenBank	xxxx	JX536583	FM179404.1	EF585003.1	AF478505.1	JX438147	AM234585.1
Festuca coerulescens Desf. (Festu coer)	Spain. Cádiz: Jerez de la Frontera. UZ 19.08, GenBank	xxxx	JX536581	JQ972941	EF585027.1	JQ972971	JX438145	JX438089
Festuca modesta Nees ex Steud. (Festu mode)	China. Yunnan. Fugong, GenBank	xxxx	-	EF584953.1	EF585068.1	EF592985.1	JX438143	-

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
***Festuca* subgen. *Drymanthele* – *F.* subgen. *Leucopoa* – *F.* sect. *Lojaconoa* Gand. – *F.* sect. *Pseudoscariosa* Krivot. – *F.* sect. *Scariosae* Hack.**								
Festuca pseudoeskia Boiss. (Festu pseu)	Spain. Granada: Sierra Nevada. UZ 216.07	xxxx	JX536582	AF519979.1	EF585084.1	AY099000.1	JX438146	JX438090
Festuca scariosa Lag. ex Willk. (Festu scar)	Spain. Granada: Sierra Nevada. UZ 219.07	xxxx	-	AF519978.1	EF585100.1	AY098999.1	JX438148	-
Festuca triflora J.F. Gmel. (Festu trif)	Spain. Ciudad Real. UZ 166.07, GenBank	xxxx	JX536580	AF538362.1	DQ631483.1	AF533052.1	JX438144	JX438088
***Lolium* L. – *Micropyropsis* Romero Zarco & Cabezudo – *Festuca* subgen. *Schedonorus* (P. Beauv.) Peterm. – *F.* sect. *Subbulbosae* Nyman ex Hack.**								
Festuca arundinacea Schreb. (Festu arun)	Spain. Lugo: Láncara: Santa Bárbara, GenBank	xxxx	-	AJ240154.1	DQ367405.1	AF533042.1	JX438151	JX438093
Festuca donax Lowe (Festu dona)	Portugal. Madeira: Porto Moniz, GenBank	xxxx	-	EF584935.1	EF585033.1	EF592968.1	KJ529475	KJ529376
Festuca drymeja Mert. & W.D.J. Koch (Festu drym)	Spain. Huesca, cultivated from seeds, GenBank	xxxx	-	HM453181	EF585034.1	AY098997.1	KJ529474	KJ529375
Festuca fenas Lag. (Festu fena)	France. ABY BN354. UZ 142.07	xxxx	-	JQ972943	JQ973009	JQ972973	JX438150	JX438092
Festuca gigantea (L.) Vill. (Festu giga)	Norway, GenBank	xxxx	JX536619	AF303416.1	JQ973012	AF533043.1	JX438154	JX438096
Festuca lasto Boiss. (Festu last)	Spain. Cádiz: Sierra Bermeja, GenBank	xxxx	-	AF303418.1	EF585060.1	AY098998.1	KJ529473	KJ529374
Festuca paniculata subsp. *spadicea* (L.) Litard. (Festu pani)	Spain. León. UZ 290.07, GenBank	xxxx	JX536592	JQ972942	JQ973008	JQ972972	JX438149	JX438091
Festuca pratensis Huds. (Festu prat)	Kazasthan. Cultivated from seeds. UZ141.07, GenBank	xxxx	JX536618	AF303421.1	JQ973011	AF478503.1	JX438153	JX438095
Lolium canariense Steud. (Loliu cana)	Spain. Canary isles. Tenerife. GenBank	xxxx	-	EF379062	EF379014	AY228162.1	KJ529477	KJ529380
Lolium perenne L. (Loliu pere)	UK. Wales. Cultivada. PI619001. GenBank	xxxx	JX536620	AJ240138.1	DQ367404	AF478504.1	JX438155	JX438097
Lolium rigidum Gaudin (Loliu rigi)	Spain. Zaragoza: Vedado de Peñaflor, GenBank	xxxx	-	EU814830.1	EF379034	AF533040.1	KJ529478	KJ529381
Micropyropsis tuberosa Romero Zarco & Cabezudo	Spain. Huelva: Almonte. UZ 89.07, GenBank	xxxx	JX536616	JQ972944	JQ973010	AF533037.1	JX438152	JX438094
Fine-Leaved Loliinae								
***Festuca* sect. *Eskia* Willk.**								
Festuca elegans Boiss. (Festu eleg)	Spain. Granada: Sierra Nevada. UZ 231.07, GenBank	xxxx	JX536621	AF303406.1	EF585038.1	AF478509.1	JX438156	JX438098
Festuca eskia Ramond ex DC. (Festu eski)	Spain. Huesca: Benasque. UZ 262.07 GenBank	xxxx	JX536629	AF303412.1	EF585040.1	AF478508.1	JX438158	JX438100
Festuca gautieri (Hack.) K.Richt. (Festu gaut)	Spain. Huesca: Panticosa. UZ 152.07. GenBank	xxxx	JX536628	AF303414.1	EF585044.1	AF478507.1	JX438157	JX438099
***Festuca* subsect. *Exaratae* St-Yves – *Narduroides* Rouy – *Psilurus* Trin. – *Vulpia* sect. *Apalochloa* (Dumort.) Stace – *V.* sect. *Monachne* Dumort. – *V.* sect. *Vulpia* (4x-6x)**								
Festuca capillifolia Dufour ex Roem. & Schult. (Festu capi)	Spain. Almeria: Sierra Gador. UZ 179.07, GenBank	xxxx	JX536633	AF303419.1	EF585022.1	AF478511.1	JX438161	JX438102
Festuca clementei Boiss. (Festu clem)	Spain. Granada: Sierra Nevada. UZ 223.07, GenBank	xxxx	JX536642	-	EF585025.1	AF478524.1	JX438162	JX438060
Micropyrum tenellum (L.) Link (Micro tene)	Spain. Caceres: Puerto Castaños. UZ 41.07, GenBank	xxxx	-	-	EF585116.1	AF478534.1	KJ529488	KJ529387
Narduroides salzmannii (Boiss.) Rouy	Spain. Madrid: Arganda del Rey. UZ 111.07, GenBank	xxxx	JX536632	JQ972946	JQ973014	AF478535.1	JX438160	JX438101

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
***Festuca* subsect. *Exaratae* St-Yves – *Narduroides* Rouy – *Psilurus* Trin. – *Vulpia* sect. *Apalochloa* (Dumort.) Stace – *V.* sect. *Monachne* Dumort. – *V.* sect. *Vulpia* (4x-6x)**								
Psilurus incurvus (Gouan) Schinz & Thell. (Psilu incu)	Spain. Caceres: Malpartida. UZ31.07, GenBank	xxxx	JX536631	JQ972945	JQ973013	AF478533.1	JX438159	JX438059
Vulpia ciliata Dumort. (Vulpi cili)	Spain. Zaragoza: Vedado de Peñaflor UZ 112.07, GenBank	xxxx	JX536630	AF478486.1	EF585120.1	AF478527.1	-	JN894723.1
Vulpia fasciculata (Forssk.) Samp. (Vulpi fasc)	Spain. Barcelona: Vilanova. SP15.2000, GenBank	xxxx	JX536643	JQ972948	EF585121.1	AF478528.1	JX438163	JN894722.1
Vulpia geniculata (L.) Link (Vulpi geni)	Spain. Sevilla: Constantina. J29397, GenBank	xxxx	JX536644	JQ972947	EF585123.1	AF478531.1	JX438164	-
Vulpia membranacea (L.) Dumort.	Spain. Cádiz: Sanlucar de Barrameda, UZ 8.2002 GenBank	xxxx	-	xxxx	EF585124.1	AY118101.1	KJ529492	-
Vulpia unilateralis (L.) Stace (Vulpi unil)	Spain. Zaragoza Vedado de Peñaflor. UZ 114.07, GenBank	xxxx	-	xxxx	EF585130.1	AY118107.1	KJ529491	KJ529388
***Festuca* sect. *Aulaxyper* Dumort. s.l. – *Vulpia* sect. *Vulpia* (2x)**								
Festuca agustinii Linding. (Festu agus)	Spain. Canary isles. La Palma, UZ2000 GenBank	xxxx	-	AY099005.1	EF585000.1	EF592949.1	KJ529504	KJ529396
Festuca iberica (Hack.) K. Richt. (Festu iber)	Spain. Granada: Sierra Nevada. UZ218.07, GenBank	xxxx	JX536662	AY118087.1	EF585052.1	AF478516.1	JX438170	-
Festuca rivularis Boiss.	Spain. Granada: Sierra Nevada, UZ78.2000, GenBank	xxxx	-	AF478475.1	EF585093.1	EF593000.1	KJ529502	KJ529395
Festuca rubra L. (Festu rubr)	UK. Scotland: Lamrkshire. UZ 150.07, GenBank	xxxx	JX536658	AJ240158.1	EF585097.1	AY118099.1	JX438169	JX438106
Vulpia muralis (Kunth) Nees (Vulpi mura)	Spain. Sevilla: Sierra Morena: Palancar, UZ1.2002, GenBank	xxxx	-	xxxx	EF585126.1	AY118102.1	KJ529496	KJ529391
***Festuca* sect. *Festuca s.s.* – *Wangenheimia* Moench**								
Festuca frigida Grossh. (Festu frig)	Spain. Granada: Sierra Nevada. UZ 226.07, GenBank	xxxx	-	AF478481.1	DQ631485.1	AF478521.1	KJ529494	-
Festuca hystrix Boiss. (Festu hyst)	Spain. Almeria: Sierra Gador. UZ185.07, GenBank	xxxx	JX536646	AF478480.1	EF585051.1	AF478520.1	JX438168	-
Festuca indigesta Boiss.	Spain. Granada: Sierra Nevada. UZ 212.07, GenBank	xxxx	JX536651	AF303426.1	EF585054.1	AF478519.1	JX438167	JX438105
Festuca longiauriculata Fuente, Ortúñez & Ferrero (Festu long)	Spain. Almería: Sierra Filabres. UZ 192.07, GenBank	xxxx	JX536652	AF478479.1	EF585062.1	AF478518.1	-	-
Festuca ovina L. (Festu ovin)	Germany. Thüringen: Saale-Holzland, GenBank	xxxx	-	JQ972950	EF585076.1	JQ972975	JX438166	JX438104
Wangenheimia lima (L.) Trin. (Wange lima)	Spain. Zaragoza: Vedado de Peñaflor. UZ 113.07, GenBank	xxxx	-	JQ972949	EF585131.1	AF478536.1	JX438165	JX438103
PACCMAD								
Anomochlooideae Pilg. ex Potztal								
Streptochaeteae C.E. Hubb.								
Streptochaeta angustifolia Soderstr.	L. Clark 1304 (ISC), GenBank	xxxx	-	-	-	xxxx	EF422918.1	-
Streptochaeta spicata Schrad. ex Nees	GenBank	-	-	-	-	-	-	AF164383.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		rps3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Anomochloeae C.E. Hubb.								
Anomochloa marantoidea Brongn. (Anomo mara)	L. Clark 1302 (ISC), GenBank	xxxx	-	-	-	-	EF422917.1	AF164381.1
Centothecoideae Soderstr.								
Centotheceae Ridl.								
Chasmanthium latifolium (Michx.) H.O. Yates	E. A. Kellogg collection, GenBank	xxxx	-	DQ172079.1	KJ529257	EF137558.1	GU359720	KJ529308
Thysanolaena maxima (Roxb.) Kuntze (Thysa maxi)	UK. Cultivated in Kew Garden. 1979-3225, GenBank	xxxx	JX536497	AF019854.1	JQ972980	EF137520.1	TMU21984	EF137433.1
Panicoideae Link								
Gynerieae Sanchez-Ken & L.G. Clark								
Gynerium sagittatum (Aubl.) P. Beauv.	Peru. Cultivated in Kew Garden. 1991-1276, GenBank	xxxx	-	AF019858.1	JQ972981	JQ972956	AY847120.1	EF137431.1
Paniceae R. Br.								
Panicinae Fr.								
Panicum miliaceum L. (Panic mili)	E. A. Kellogg collection, GenBank	xxxx	-	FJ606748.1	JQ972982	JQ972957	AY188472.1	FR667662.1
Setariinae Dumort.								
Pennisetum glaucum (L.) R. Br. (Penni glau)	E. A. Kellogg collection, GenBank	xxxx	-	FJ766182.1	xxxx	xxxx	AF499149.1	-
Pennisetum macrourum Trin.	GenBank	-	-	-	-	-	-	EF137467.1
Setaria vididis (L.) P. Beauv.	USA. Indiana: Bloomington. UZ102.08, GenBank	xxxx	-	-	-	EF137585.1	U21976.1	-
Setaria P. Beauv. spp.	GenBank	-	-	AF019831.1	JQ972983	-	-	EF137472.1
Andropogoneae Dumort.								
Tripsacinae Dumort.								
Tripsacum dactyloides (L.) L.	E. A. Kellogg collection, GenBank	DQ984517	-	DQ005086.1	xxxx	DQ005007.1	AF117433.1	-
Zea perennis (Hitchc.) Reeves & Mangelsd. (Zea perenn)	GenBank	DQ645538.1	-	-	-	-	-	-
Zea mays L.	GenBank	-	AF068119	ZMU46616	EF541347.1	EF541269.1	U21985.1	X86563
Rottboelliinae J. Presl								
Coix lacryma-jobi L.	E. A. Kellogg collection, GenBank	xxxx	-	DQ005034.1	xxxx	EF137580.1	AF117403.1	EF137458.1
Heteropogon contortus (L.) P. Beauv. ex Roem. & Schult. (Heter cont)	E. A. Kellogg collection, GenBank	xxxx	-	EU646132.1	xxxx	DQ004984.1	AF117411.1	-
Hyparrhenia hirta (L.) Stapf (Hypar hirt)	E. A. Kellogg collection, GenBank	xxxx	-	DQ005056.1	xxxx	DQ004986.1	AM849124	AF164417.1
Rottboellia selloana Hack. (Rottb sell)	E. A. Kellogg collection, GenBank	xxxx	JX536502	-	JQ972984	JQ972958	-	-

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Rottboelliinae J. Presl								
Rottboellia aurita Steud.	GenBank	-	-	DQ005063.1	-	-	-	-
Sorghum bicolor (L.) Moench (Sorgh bico)	USA. Indiana: Bloomington. UZ107.08, GenBank	DQ984518	-	DQ190420.1	xxxx	xxxx	U21981.1	AF164418.1
Micrairoideae Pilg.								
Arundineae Dumort – Eriachneae Eck-Boorsb. – Isachneae Benth.								
Arundo donax L. (Arund dona)	Spain. Zaragoza: Daroca, GenBank	xxxx	-	DQ172077.1	KF169820	GQ869907	U21998.1	AF164408.1
Eriachne R.Br. spp.	E. A. Kellogg collection, GenBank	xxxx	-	AF019811.1	-	xxxx	AM849155	-
Isachne arundinacea (Sw.) Griseb.	Panama. Sanchez-Ken 680 (ISC), GenBank	xxxx	-	-	xxxx	xxxx	AY847119.1	-
Danthonioideae H.P. Linder & N.P. Barker								
Danthonieae Zotov								
Cortaderiinae Conert / Danthoniinae Fr.								
Danthonia decumbens (L.) DC.	Spain. Cádiz: Jerez. UZ31.08, GenBank	xxxx	-	EU401308.1	JQ972986	EU401189.1	EU400785.1	JX438057
Danthonia spicata (L.) P. Beauv. ex Roem. & Schult. (Danth spic)	USA. Indiana: Belmont. UZA112.08, GenBank	xxxx	JX536506	DQ887163.1	-	EU401198.1	EU400789.1	AF164409.1
Karroochloa purpurea (L. f.) Conert & Türpe (Karro purp)	P. Linder 5360 (BOL), GenBank	xxxx	JX536509	DQ655793.1	JQ972988	DQ655852.1	AF251458.1	-
Merxmuellera Conert spp.	N. Barker 960 (GRA), GenBank	xxxx	JX536504	AF019862.1	JQ972985	EU401215.1	EU400803.1	EU400747
Schismus barbatus (L.) Thell. (Schis barb)	Spain. Zaragoza: Vedado de Peñaflor. UZ139.07, GenBank	xxxx	-	JQ972930	JQ972987	JQ972959	GQ471663.1	EU400767.1
Aristidoideae Caro								
Aristideae C.E. Hubb.								
Aristida congesta Roem. & Schult. (Arist cong)	N. Barker 1130 (BOL) GenBank	xxxx	-	DQ171986.1	xxxx	DQ172211.1	-	-
Aristida adscensionis L.	GenBank	-	-	-	-	-	AM849125	AF164412.1
Stipagrostis zeyheri (Nees) De Winter (Stipa zeyh)	N. Barker 1133 (BOL), GenBank	xxxx	-	DQ172076.1	-	DQ172301.1	AF251455.1	-
Chloridoideae Kunth ex Beilschm.								
Eragrostideae Stapf								
Eragrostidinae J. Presl								
Eragrostis curvula (Schrad.) Nees (Eragr curv)	L. Clark 1303 (ISC), GenBank	xxxx	-	DQ655822.1	JQ972990	DQ655877.1	U21989.1	-
Eragrostis Wolf spp.	GenBank	-	JX536512	-	-	-	-	AF312341.1

Table S1. List of taxa included in this study. Source and GenBank accession numbers. Systematic assignment follows SORENG *ET AL.* (2003, 2007); TORRECILLA *ET AL.* (2002, 2003, 2004); CATALAN *ET AL.* (2004, 2006); QUINTANAR *ET AL.* (2007); BOUCHENAK-KHELLADI *ET AL.* (2008); INDA *ET AL.* (2008); SCHNEIDER *ET AL.* (2009); GPGW (2012); and MINAYA *ET AL.* (2013). xxxx indicates sequences to be submitted to GenBank. Spp. indicates that sequences from more than one species have been collected for that genus. Abbreviations of species used in the Figure 2 are included in brackets.

Taxa	Source	GenBank accession						
		*rps*3	*ß-amylase*	ITS	*trn*TL	*trn*LF	*ndh*F	*mat*K
Cotteinae Reeder								
Enneagopon desvauxii P. Beauv. (Ennea desv)	Columbus 3133, GenBank	xxxx	-	EF153043.1	JQ972989	EF156692.1	GU359699.1	-
Enneapogon glaber N.T. Burb.	GenBank	-	-	-	-	-	-	AF312360.1
Zoysieae Benth.								
Muhlenbergiinae Pilg.								
Eleusine Gaertn. spp.	E. A. Kellogg collection, GenBank	xxxx	xxxx	EF153042.1	AY523497.1	EF137563.1	AM849151	AF144580.1
Muhlenbergia montana (Nutt.) Hitchc. (Muhle mont)	Columbus 3375, GenBank	xxxx	-	EF153067.1	JQ972991	EF156716.1	HM143488	AF144600.1
Sporobolinae Benth.								
Spartina Schreb. spp. (Spartina)	E. A. Kellogg collection, GenBank	xxxx	JX536517	EF153082.1	AF372625.1	EF137568.1	AF251465.1	AF312353.1
Sporobolus indicus (L.) R.Br. (Sporo indi)	L. Clark 1293 (ISC), GenBank	xxxx	JX536514	EU646098.1	JQ972992	EF137569.1	U21983.1	AF144601
Sporobolus wrightii Munro ex Scribn. (Sporo wrig)	Columbus 2507, GenBank	xxxx	JX536515	EF153085.1	-	EF156734.1	GU359668.1	-

CHAPTER IV

DIVERSIFICATION RATES AND CHROMOSOME EVOLUTION IN THE TEMPERATE GRASSES (POOIDEAE) ARE ASSOCIATED WITH MAJOR ENVIRONMENTAL CHANGES IN THE OLIGOCENE-MIOCENE

1. Abstract

- Pooideae is one of the most species-rich subfamilies in the Poaceae, and includes some of its most economically important representatives, such as wheat, barley and oats. However, little is known about the chromosomal evolution, the tempo of divergence and the diversification rates of its lineages. This chapter explores the phylogeny of the Pooideae, including a temporal framework and diversification rates. We also look at the evolution of Pooideae, chromosomal changes and the shifts in its diversification rates, addressing the debated question of the importance of polyploidization in diversification.
- We sequenced five plastid DNA regions, two coding (*ndhF*, *mat*K) and three non-coding (*trn*H-*psb*A, *trn*TL and *trn*LF), in 170 Poaceae taxa, representing all subfamilies of the grasses and all but three Pooideae tribes. Parsimony and Bayesian phylogenetic analyses were conducted, and divergence times were inferred in BEAST using a relaxed molecular clock method. Chromosome evolution was analyzed using the chromEvol software, and diversification rates were assessed using the MEDUSA approach.
- Our phylogenetic tree agrees with those previously proposed for the main subfamily clades. Diversification of the Pooideae started in the Middle Eocene with the divergence of Brachyelytreae, and was especially intense during the Oligocene-Miocene. By the Middle Miocene most subtribal lineages within the Pooideae had diverged. The background diversification rate increased significantly at the time of the origin of the Poodae + Triticodae clade. Three decreases, one moderately supported and two marginally significant, were inferred for the subtribe Parapholiinae-Cynosuriinae-Dactylidiinae and the tribes Nardeae. *s.l.* and Brachyelytreae, respectively. Analysis of chromosome number evolution showed that the base haploid chromosome number (n = 7) has remained stable throughout the phylogenetic history

of the core pooids. Most chromosome number changes detected occurred at shallow levels of the phylogeny, but not all chromosome number mutations behaved in the same way.

- Shifts in diversification rates were registered during the Oligocene-Miocene for the most species-rich lineages of the C3 temperate Pooideae, paralleling similar major cases observed in the tropical C4 grasses. These shifts in diversification occurred in a context of falling temperatures that might have increased ecological opportunities for all grasses around the world. We found no link between chromosome transitions and major diversification events in the Pooideae. Polyploidy was favored over disploidy in the core pooids. This contradicts some previous findings in the angiosperms, though all those cases corresponded to relatively recent Oligocene-Neogene neopolyploids.

2. Introduction

Recently developed evolutionary models and statistical tools to test them allow researchers to include different sorts of data when calculating phylogenetic inferences (DRUMMOND AND RAMBAUT, 2007). Information on chromosome numbers and diversity of taxonomic groups is commonly used in phylogenetic analyses in order to investigate aspects such as the *tempo* and *mode* of diversification in a lineage and the impact of chromosome changes on diversification (e.g. ESCUDERO *ET AL.*, 2012). The phylogenetic structure of the grass subfamily Pooideae has been thoroughly studied in recent decades (e.g. GPWG 2001, BOUCHENAK-KHELLADI *ET AL.,* 2008, and references therein), but recent revisions on this topic have called for larger datasets to increase the robustness of the results (GPWG, 2012). Combining the phylogenetic information on this group with chromosomal and species diversity data would open an insight into the evolution of these economically and ecologically important taxa.

Molecular phylogenetic studies support the monophyly of the grasses (MICHELANGELI *ET AL.*, 2003, DAVIS AND SORENG, 2007; BLANER *ET AL.*, 2014). The Pooideae subfamily is also reconstructed as monophyletic and sister to the Bambusoideae within the BEP clade in the most recent phylogenetic studies (BOUCHENAK-KHELLADI *ET AL.,* 2008; GPWG, 2012). The systematic positions of the different grass tribes and subtribes, including those in the Pooideae, are currently under discussion, and their evolutionary relationships are not totally resolved (e.g. BOUCHENAK-KHELLADI *ET AL.,* 2008; PIRIE *ET AL.*, 2008; SANCHEZ-KEN AND

CLARKE, 2010; CAHOON *ET AL.,* 2010; GPWG, 2012).

The Pooideae include about one third of the species richness of the grasses (approximately 150 genera and 3560 species). They have high ecological (covering 20% of the terrestrial surface) and economic importance, and include some of the most prominent crops such as wheat, rye, oats and barley (CLAYTON AND RENVOIZE, 1986, GPWG, 2001). The tribal arrangement of the Pooideae is controversial and has varied widely over the last century (e.g. ASCHERSON AND GRAEBNER, 1898-1902, TZVELEV, 1976, CLAYTON AND RENVOIZE, 1986, WATSON AND DALLWITZ, 1992). Combined molecular and morphological approaches (SORENG AND DAVIS, 2000; HSIAO *ET AL.,* 1995; CATALAN *ET AL.,* 1997, GWPG, 2001) have allowed for the establishment of 13 tribes in the Pooideae. However, recent works with broader sampling, either at the subfamily level or of particular lineages, have uncovered inconsistencies in this taxonomic arrangement (GILLESPIE *ET AL.,* 2007, QUINTANAR *ET AL.,* 2007, CATALAN *ET AL.,* 2007, DORING *ET AL.,* 2007, SORENG *ET AL.,* 2007, SAARELA *ET AL.,* 2010). Several of the defined tribes are not monophyletic, and a wide range of changes to the taxonomy of the Pooideae subfamily has been suggested. The Stipeae tribe was recently disaggregated into three smaller monophyletic tribes (SORENG *ET AL.,* 2007). The closely related tribes Nardeae and Lygeae were united in one larger tribe, Nardeae *s.l.* (SCHNEIDER *ET AL.,* 2009). The Triticeae and Bromeae tribes were both included as subtribes in tribe Triticeae *s.l.* (SCHNEIDER *ET AL.*, 2009), while the Littledalea subtribe was separated from Bromeae (SORENG *ET AL.,* 2007, SCHNEIDER *ET AL.,* 2009), and Hordeinae was created as a new subtribe within Triticeae *s. l.* (SCHNEIDER *ET AL.*, 2009). The Aveneae and Poeae tribes, traditionally separated based on morphological characters (CLAYTON AND RENVOIZE, 1986, WATSON AND DALLWITZ, 1992), were merged into a controversial supertribe, Poodae (DAVIS AND SORENG, 2007) or Poeae *s.l.* (SORENG AND DAVIS, 2000; SORENG *ET AL.,* 2003, 2007; QUINTANAR *ET AL.*, 2007) based on molecular phylogenetic evidence. This supertribe includes most of the species described within Pooideae (2260, 135 genera) and includes not only the former Poeae *s.str.* and Aveneae *s.str.* tribes, but also the Hainardeae, Phleeae and Seslerieae tribes (QUINTANAR *ET AL.,* 2007; DAVIS AND SORENG, 2007). Poodae has been split into 22 subtribes, arranged into two groups based on plastid DNA (see the appendix of chapter IV; Fig. S1, page 230). Sixteen subtribes belong to the Poeae-type plastid DNA clade and 6 to the Aveneae-type plastid DNA clade (SORENG *ET AL.,* 2007). Different studies focusing on particular subtribes such as the Airinae (CHIAPELLA, 2007), Loliinae (TORRECILLA AND CATALAN, 2002;

TORRECILLA *ET AL.*, 2004; CATALAN, 2006; CATALAN *ET AL.*, 2004, 2006, 2007; HUNTER *ET AL.*, 2004), Poinae (GILLESPIE AND SORENG, 2005; GILLESPIE *ET AL.*, 2007, 2008, SORENG *ET AL.*, 2007) and Aveninae (QUINTANAR *ET AL.*, 2007, SAARELA *ET AL.*, 2010) have suggested that further changes to the taxonomy of the supertribe Poodae may be necessary.

We have analyzed five plastid DNA regions in order to build a robust phylogeny of the Pooideae. We have also increased the number of Pooideae tribes and genera analyzed with respect to previous molecular studies (SORENG *ET AL.*, 2007). Overall, 10 of the GPWG's (2001) 13 tribes are represented in our survey (about 80%; see the appendix of chapter IV; Table S1; page 232). We combined our molecular data with cytogenetic information and diversity data taken from our own records and the literature (e.g. tropicos.org, Missouri Botanical Garden; WATSON AND DALLWITZ, 1992). The combined analysis of molecular and cytogenetic data allowed us to infer trends of chromosome number evolution in this group of grasses. Grass genome evolution models (e.g. SALSE *ET AL.*, 2008) have inferred a chromosome base number of x = 5 for the ancestor of the grasses (c. 90 My), followed by paleoduplication (paleopoliploidization) to x = 10 and additional breakage plus fusion to x = 12 (c. 70-50 My). From this putative diploidized paleopolyploid ancestor, distinctly inferred genomic arrangements led to some current genomes that are relatively conserved (e.g. *Oryza*, x = 12), others that are more severely reduced (e.g. *Triticum*, x = 7), and alternative rearrangements that resulted in different genomic reductions (e.g. *Sorghum*, x = 10; *Zea*, x = 10) (cf. GAUT, 2002; SALSE *ET AL.*, 2008). In the highly diverse Pooideae, x = 7 is considered to be the most common chromosome base number in several groups (e.g. Triticeae, SALSE *ET AL.*, 2008; Poodae, HSIAO *ET AL.*, 1995). A decreasing aneuploid series was proposed in the Pooideae (CATALAN *ET AL.*, 1997), ranging from x = 13, 10 in the more ancestral Lygeae (and Brachyelytreae, x = 11; SAARELA *ET AL.*, 2003), through x = 12, 11, 10, 9, 8 in the successively diverged Stipeae, Meliceae and Brachypodieae, to x = 7 in the more recently evolved core pooids (Poodae + Triticeae s. l., including the former tribes Poeae and Aveneae and Triticeae and Bromeae). A reduction to very small chromosome base numbers has also been documented in some pooid lineages [e.g. x = 5 Brachypodieae (CATALAN AND OLMSTEAD, 2000), x = 5 Anthoxanthiinae (PIMENTEL *ET AL.*, 2013)], including one of the lowest numbers in the angiosperms [x = 2 *Zingeria, Colpodium* (KIM *ET AL.*, 2009)].

Diversification rates in angiosperm clades have been measured using very different methods, and the obtained results have reflected a great disparity (e.g., MAGALLÓN AND

SANDERSON, 2001; MAGALLÓN AND CASTILLO, 2009). Speciation rates in the Poales have been estimated to be between r = 0.0611 spp. My^{-1} and r = 1013 spp. My^{-1}, depending on the parameters of the analysis (MAGALLÓN AND SANDERSON, 2001; MAGALLÓN AND CASTILLO, 2009), but no data exists for the Poaceae or any of its lineages. A temporal framework for the phylogeny is necessary to estimate speciation rates (MAGALLÓN AND SANDERSON, 2001). The origin of the Poaceae family, together with the main diversification events in the grasses (including the BEP clade; WU AND GE, 2012), have been dated using a variety calibration points (e.g. VICENTINI *ET AL.*, 2008; CHRISTIN *ET AL.*, 2014). The divergence times of the main lineages within the Pooideae are not as well understood.

The specific aims of this study were: i) to analyze the diversification dynamics of the Pooideae, including the timing of diversification and diversification rates of its main lineages; ii) to investigate the patterns of chromosome number evolution operating in the Pooideae and assess their impact on the diversification of the different groups; and iii) to compare the times of divergence, diversification rates and patterns of chromosome evolution of the main Pooideae lineages with the major climatic and biome changes that occurred on Earth throughout the Cenozoic in order to assess their possible correlation.

3. Materials and Methods

3.1 Sampling, DNA sequencing

A total of 170 species representing nine subfamilies within the Poaceae (GPWG, 2001) were included in this study. Sampling focused mainly on the temperate Pooideae; all traditionally accepted tribes except Ampelodesmae, Phaenospermae and Brylkinia (realigned within Meliceae; SCHNEIDER, 2013) were included in the survey. Representatives from other BEP groups and from several PACCMAD lineages were added to increase the robustness of the Pooideae tree (see the appendix of chapter IV; Table S1, page 232). DNA sequences from five chloroplast DNA (cpDNA) regions, two coding (*mat*K, *ndh*F) and three non-coding (*trn*TL, *trn*LF, *trn*H-*psb*A) regions, were used in the phylogenetic analysis. Procedures for DNA isolation, amplification and sequencing and for sequence alignment are indicated in the appendix of chapter IV (page 227).

3.2 Phylogenetic analyses

Phylogenetic analyses were conducted using independent and concatenated cpDNA data. Three different data sets were used: (i) only coding regions; (ii) only non-coding regions and (iii) both coding and non-coding regions. The analysis of the data set including only coding regions was two-fold: (a) all sites were considered equally and (b) non-synonymous and synonymous sites were analyzed independently. Missing sequences caused by PCR and/or sequencing failures were coded as missing data in the concatenated data sets, making a final data set with at least three sequenced genes per sample (five genes, 70.6%; four genes, 25.3%, three genes 4.1%; Tables 1). Independent data sets were combined when no conflicting node was supported by more than 0.95 Bayesian posterior probability (PP) or 90 bootstrap support (BS) (PIRIE *ET AL.*, 2009). Overall, 798 sequences were used in this study (154 *trn*H-*psb*A; 166 *trn*LF; 161 *trn*TL; 149 *mat*K and 168 *ndh*F). The combined matrices including only coding and only non-coding regions were composed of 481 and 317 sequences, respectively.

Phylogenetic analyses were performed using maximum parsimony (MP) and Bayesian inference (BI) methods using the programs PAUP* 4.0 beta 10 (SWOFFORD, 2002) and MrBayes 3.2.2 (ALTEKAR *ET AL.*, 2004; RONQUIST *ET AL.*, 2011), respectively. Independent analyses were conducted on each data matrix using *Joinvillea ascendens* Gaudich ex Brongn. and Gris (Joinvillaceae) to root the tree. The MP analysis was conducted following CATALAN *ET AL.* (2012), with branch support computed through 10,000 bootstrap replicates.

Tests of goodness of fit for alternative nucleotide substitution models were performed through the Hierarchical Likelihood Ratio Test (hLRTs), AIC and BIC tests in jModelTest 2 (GUINDON AND GASCUEL, 2003; DARRIBA *ET AL.*, 2012). The estimated GTR+I+gamma model was fixed for all Bayesian inferences performed in this study. Gaps were coded in the matrix as presence/absence following the simple method proposed by SIMMONS AND OCHOTERENA (2000), as implemented in the software SeqState (MÜLLER, 2005). All Bayesian analyses conducted with MrBayes were carried out with and without considering the gaps following RONQUIST *ET AL.* (2005) and DWIVEDI AND GADAGKAR (2009). However, as the gap characters did not improve the support of the clades (see Results), they were excluded from all final analyses.

The Bayesian analysis of each data set was conducted with 50 000 000 generations and using a random starting tree. Sampled trees were saved every 2000 generations, and the

program was allowed to estimate the likelihood parameters required. Results collected prior to stationarity (25-35%) were discarded as burn-in. Convergence was evaluated using the "compare" function in the online application AWTY (NYLANDER *ET AL.*, 2008) and using TRACER v1.6 (RAMBAUT AND DRUMMOND, 2007). Fifty-percent Majority Rule consensus trees of all the saved posterior probability trees were computed for each analysis.

3.3 Divergence time estimation

Bayesian divergence time analysis was conducted on the complete cpDNA data set, including coding and non-coding regions using BEAST v.2.1.2. (BOUCKAERT *ET AL.*, 2014). The obtained topology was congruent with the ones recovered using partial data sets (Results not shown).

The GTR + I + G model, a lognormal uncorrelated relaxed clock model (WERTHEHEIM *ET AL.*, 2010; BROWN AND YANG, 2011) and a Yule tree prior were imposed in all the analyses. Two fossils were used for the calibration of the tree: the phytholiths described by ZUCOL *ET AL.* (2010) from the Middle Eocene and attributed to the Pooideae (ZUCOL *ET AL.*, 2010; STRÖMBERG, 2011); and the macrofossil remains of a stipoid grass from the Late Eocene (MANCHESTER, 2001; STRÖMBERG, 2011). Accordingly, the ancestral node of the Pooideae was constrained to have a mean age of 44 ± 4 My using a normal distribution (Mean = 44; σ = 1.95). The ancestral node of the Stipeae tribe was assigned a mean age of 36 ± 3 My using a normal distribution (Mean = 36; σ = 1.5). Although normal distributions are not generally considered suitable for primary calibrations, we consider that the fossils used fulfill the conditions indicated by HO AND PHILLIPS (2009) that make use of a normal prior advisable. The selected dates roughly agree with the divergence time estimates for the main Poaceae splits proposed by CHRISTIN *ET AL.* (2014) when Cretaceous grass phytolith fossils are excluded from the analysis. Substitution rates in all analyses were established following WOLFE *ET AL.* (1987). Wide ranges and a uniform distribution were imposed for all substitution rates in order to cover most biologically realistic values.

Four independent Markov Chains were run for a total of 600 000 000 generations. The impact of the different prior sets on posterior values was assessed by running an extra chain for 50 000 000 generations without data, sampling only the prior following POPP *ET AL.* (2011). We used TRACER v1.6 to analyse the log files and to assess convergence through the effective sample size (ESS ≥ 200; DRUMMOND AND HO, 2007). Resulting trees from the four

searches were combined using LogCombiner v.1.7.2 (DRUMMOND AND RAMBAUT, 2007) with a burn-in of 25%. A maximum credibility tree was constructed using TreeAnnotator v.1.7.2 (DRUMMOND AND RAMBAUT, 2007).

3.4 Shifts in diversification in the Pooideae

We used MEDUSA (ALFARO *ET AL.*, 2009) as implemented in the R package Geiger (HARMON *ET AL.*, 2008; R DEVELOPMENT CORE TEAM, 2011) in order to estimate diversification rates and test for changes in speciation and relative extinction rates in the Pooideae clade. This approach is based on the equations from the birth-death likelihood model described by RABOSKY (2006) and RABOSKY *ET AL.* (2007). MEDUSA uses the Akaike Information Criterion (AIC) to compare different models, each model including a particular combination of phylogenetic relationships (a time-calibrated tree topology with branch lengths) and taxonomic data (species richness of extant groups) with particular values of diversification and relative extinction (ALFARO *ET AL.*, 2009). We selected the MEDUSA approach to avoid the problem of unresolved, incompletely and/or non-randomly sampled lineages in the Pooideae (ESCUDERO AND HIPP, 2013). We used a skeleton-phylogenetic tree with one accession per main lineage, which was constructed by pruning the original time-calibrated tree with the drop.tip option in the R package APE (PARADIS *ET AL.*, 2004; R DEVELOPMENT CORE TEAM, 2011). In the analysis, we considered ten out of the thirteen Pooideae tribes [Brachyelytreae, Meliceae, Lygeae, Nardeae, Stipeae, Diarrheneae, Brachypodieae, Bromeae, Triticeae, and the supertribe Poodae., sensu SORENG *ET AL.* (2007)] that were sampled and reconstructed as monophyletic groups in our plastid tree (see Results). One sample per tribe was included in the skeleton tree, except for the supertribe Poodae, for which a larger representation of 15 out of 22 subtribes (~ 98% of Poodae diversity; cf. CLAYTON *ET AL.*, 2006; WATSON AND DALLWITZ, 1992; SAARELA *ET AL.*, 2010) was included. Species diversity in each tribal and subtribal lineage was extracted from the GRIN taxonomy database (www.ars-grin.gov). All MEDUSA analyses were conducted with the options "estimateExtinction" and "cutAtStem" set to TRUE, with the maximum number of models set to 51. This number represented the sum of the tips and nodes of the trees after pruning to major clades (ESCUDERO AND HIPP, 2013). In order to include phylogenetic uncertainty in the analyses (DRUMMOND *ET AL.*, 2012), we estimated ΔAICs, net diversification rates and relative extinction rates (epsilon, ε = extinction rate / speciation rate) from the respective

consensus trees and also from 500 trees randomly sampled from the BEAST MCMC analysis. Results from the 500 randomly chosen trees were subsequently summarized following ESCUDERO AND HIPP (2013).

3.4 Chromosome evolution modeling

The chromosomal evolution of the Pooideae was modeled on the Bayesian dated phylogeny produced with BEAST using the software CHROMEVOL v. 2.0 (MAYROSE *ET AL.*, 2010). This program implements a likelihood-based method for tracking the pattern of haploid chromosome number changes along a phylogeny (MAYROSE, 2014). Haploid chromosome numbers for the sampled species were taken from our own records and from contrasted public databases and literature (e.g. GOLDBLATT AND JOHNSON, 1979; WATSON AND DALLWITZ, 1992; www.tropicos.org). The input tree was pruned to eliminate all terminals for which chromosome numbers were unknown using the package APE (126 terminals were included in the analysis). Ten models of chromosome evolution in CHROMEVOL v. 2.0 are based on different combinations of three to six of the following eight parameters of rates of chromosome transition: 1) whole genome duplication; 2) half-duplication or demipolyploidy; 3) gain or 4) loss of a single chromosome; 5) gain or 6) loss of a single chromosome proportional to the current chromosome number; 7) a specified chromosome number which characterizes a phylogenetic group and 8) rate for transitions by base number (MAYROSE *ET AL.*, 2010; MAYROSE, 2014). Demiploidy is interpreted by the program as a 'half-duplication' of the chromosome number. The fit of each of the ten available models was investigated using the AIC (BOURNHAM AND ANDERSON, 2002). Our analysis focused on the better-sampled core pooids (including the supertribes Triticodae and Poodae). All taxa from this group for which chromosome numbers were found were included. For each extant taxon the number of chromosome transitions was inferred from the root to the leaf representing that taxon (MAYROSE *ET AL.*, 2010). All inferred chromosome transitions were subsequently dated and classified based on whether they were followed by cladogenesis or not (*i.e.*, whether each was mapped on a node or on a terminal tip).

4. Results

4.1 Phylogenetic reconstruction and times of diversification

The number of sequences used in the analysis, the number of base pairs per gene and the percentage of parsimony informative characters in each data set are summarized in Table 1. Interestingly, we have observed a higher average number of parsimony informative characters in coding than in non-coding regions (35.2% *vs* 23.9%), although the difference was not significant. The topologies recovered in the Bayesian and parsimony trees (see the appendix of chapter IV; Figs. S1, S2, page 230) were also congruent, and only the first will be discussed. Bayesian topologies recovered with MrBayes and BEAST did not differ significantly, and only the latter will be discussed.

Table 1. Number of samples sequenced per region (one sample per species); size of the alignment; percentage of parsimony informative characters (p.i.); consistency index (CI) and retention index (RI).

DNA regions	No. of seq. per DNA region	Size (bp)	Percentage of p. i. characters	CI/RI
*ndh*F	168	713	39.4	0.375/0.823
*mat*K	149	990	31	0.434/0.832
*trn*H-*psb*A	154	599	19.7	0.420/0.818
*trn*LF	166	2033	18.7	0.486/0.810
*trn*TL	161	1253	33.4	0.451/0.760
TOTAL	798	5588	26 (1505)	

Phylogenetic trees from different cpDNA data sets largely coincided (results not shown), although the levels of support for the main clades varied. Coding (*mat*K, *ndh*F) and non-coding (*trn*LF, *trn*TL, *trn*H-*psb*A) Bayesian trees recovered similar topologies, with the former showing higher posterior probability values. Topologies based only on coding regions showed no differences regardless of whether all bases were considered equally or non-synonymous vs. synonymous positions were independently analysed.

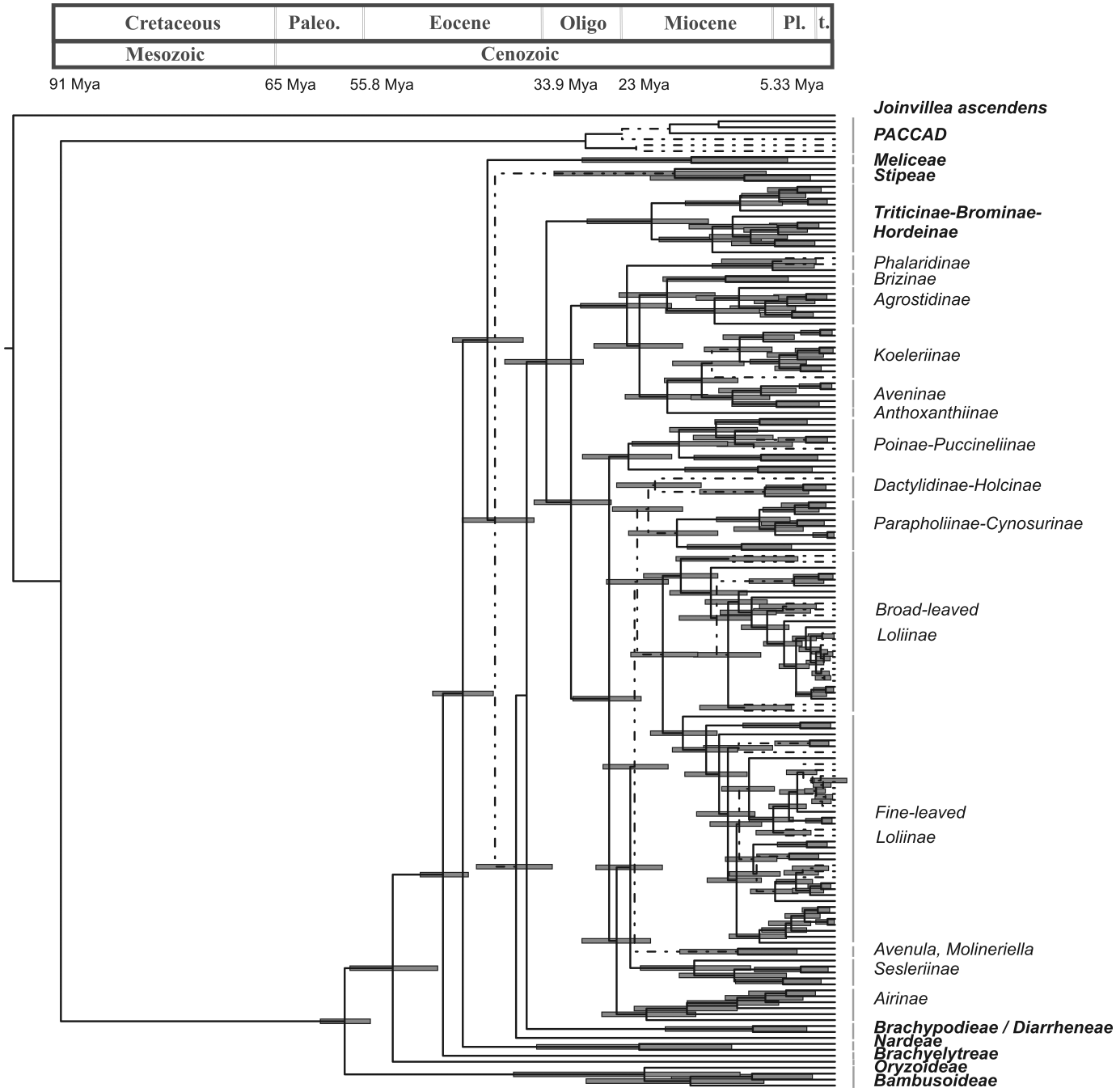


Figure 1.- Maximum clade credibility tree from the Bayesian analysis of plastid DNA sequences (*trn*H-*psb*A, *trn*TL, *trn*LF, *ndh*F and *mat*K) from 163 samples representing 152 Pooideae, 6 PACCAD (PACCMAD in text), 3 Bambusoideae and 1 Erhartoideae grass species and *Joinvillea ascendens* as outgroup, constructed with BEAST using a Yule prior. Divergence times were inferred using a relaxed molecular clock. Dashed lines represent branches with PP < 0.8. // represents shortened branches. Pl., Pliocene; t., Pleistocene. The grey bars indicate 95% highest posterior density (HPD) intervals for the divergence time estimates.

The concatenated data set built using the five plastid regions included 170 taxa (see the appendix of chapter IV; Table S1; page 232) and 5588 characters, 26% of which (1505) were parsimony informative. The Bayesian trees (MCC Beast tree, Fig. 1; 50% MR MrBayes consensus tree, Fig. S1; page 230) showed a topology largely consistent with previously published phylogenies based on combined plastid and nuclear data (GPWG, 2001; SCHNEIDER *ET AL.*, 2009). The recovered phylogenies (Figs. 1, S1; page 230) were congruent and supported the lineage arrangements proposed for the Poaceae by the GPWG (2001). The dates recovered in our analysis for the onset of the diversification of Pooideae, as well as for the divergences of its main lineages, were roughly consistent with previous results by other authors (e.g. VICENTINI *ET AL.*, 2008). Within the BEP clade, the Erhartoideae were sister to a clade including the Pooideae and the Bambusoideae with moderate support (Posterior probability -PP- 0.92). In the PACCMAD clade there was a split of two highly supported lineages (PP 1.0), one including Danthonioideae and Chloridoideae and another comprising Centothecoideae, Panicoideae and Arundinoideae.

Within the Pooideae, all tribes defined by the GPWG (2001) were recovered as monophyletic with strong support, except for Stipeae (Figs. 1, S1, page 230). Diversification of the pooids took place from the Middle Eocene onwards. The early diverging lineage leading to the Brachyelytreae (PS 1.0) originated around 45.5 My (42.5-48.1 My). The consecutive divergences of Nardeae *s.l.* (*sensu* SCHNEIDER *ET AL.*, 2009, including Lygeae), Meliceae and Stipeae were dated to the Middle to Late Eocene [Nardeae: 43.2 My (HPD 39.7-45.6 My); Meliceae 40.3.1 My (HPD 38.2-44.3 My); Stipeae 39.4 My (HPD 34.9-43.2 My, but with low support)].

Diversification increased during the Oligocene and the Miocene, when most of the Pooideae lineages branched off (Figs. 1, S1). Diarrheneae and Brachypodieae diverged in the Late Eocene-Early Oligocene [37 My (HPD 32.8-41.6 My) and 35.7 My (HPD 31.1-40 My), respectively] (Fig. 1). Within the core pooids, the supertribes Triticodae and Poodae diverged in the Early Oligocene [33.5 My (HPD 29.2-38.3 My)]. The splits of the three main Triticodae lineages occurred in the Early to Middle Miocene [Brominae 21.3 My (HPD 14.6-28.7 My); Triticinae and Hordeinae 14.2 My; (HPD 8.5-20.4 My)]. Within the Poodae, the sister "Aveneae-type cpDNA" and "Poeae-type cpDNA" lineages (sensu SORENG *ET AL.*, 2007) split in the Early Oligocene [(30.6 My (HPD 25.9-34.9 My)]. The former clade showed the early divergence of the Phalaridinae lineage in the Late Oligocene [24.1 My (HPD 18.9-

29.5 My)], then the split of Agrostidinae and Brizinae (including *Airopsis*), and then the more recent divergence of Koelerinae and Aveninae in the Middle Miocene [15.4 My (HPD 11.2-19.8 My)]. The second clade showed the successive splits of the Poinae / Puccineliinae and the Airinae lineages in the Late Oligocene [(26.2 My (HPD 19.6-27.5 My) and [25.3 My (HPD 21.4-29.3 My)], respectively, and the Sesleriinae (including *Mibora*) and the Loliinae in the Late Oligocene-Early Miocene [23.7 My (HPD 20-27.7 My) and 22.9 My (HPD 19.3-26.5 My), respectively].

Our results supported an Early Miocene [20 My (HPD 15.9-23.7 Mya)] divergence for the well supported broad-leaved and fine-leaved lineages of Loliinae (CATALAN *ET AL.*, 2004) (Fig. 1). Diversification within these clades took place mostly between the Late Miocene and the Pleistocene. Within the broad-leaved clade, the reconstruction of the Asian-American, *Lojaconoa*, *Drymanthele*, *Scariosae*, *Pseudoscariosae*, *Leucopoa*, *Subbulbosae*, and *Schedonorus-Micropyropsis-Lolium* groups fully or mostly agree with the phylogenies of INDA *ET AL.* (2008, 2014). The newly sequenced *F. africana* was reconstructed as sister to the *Leucopoa* + *Subbulbosae* + *Schedonorus-Micropyropsis-Lolium* lineage but with weak support (Fig. 1, S1, page 230). Within the fine-leaved clade, the 'intermediate' *F. pulchella* (Amphigenes) was resolved as sister to the remaining taxa, and the reconstruction of the *Eskia*, American *Vulpia*, *Exaratae*, *Psilurus*/*Vulpia* 4x-6x, *Narduroides*, *Micropyrum*, *Loretia*, *Festuca* + *Wangenheimia* and *Aulaxyper* + *Vulpia* 2x groups were concordant with the phylogenies of CATALAN *ET AL.* (2004) and INDA *ET AL.* (2008). The newly sequenced *F. abyssinica* was reconstructed as sister to the southern American *Hellerochloa,* but with poor support (Fig. 1, S1, page 230).

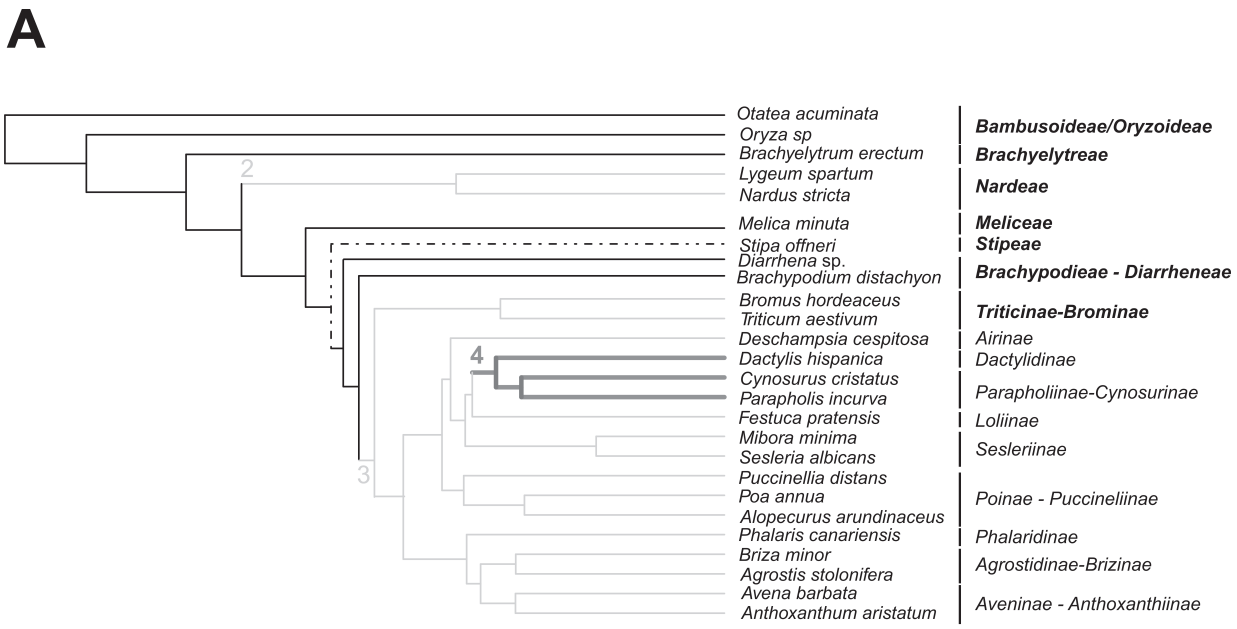
4.2. Rates of diversification and extinction in the Pooideae and Poodae. Evolutionary shifts

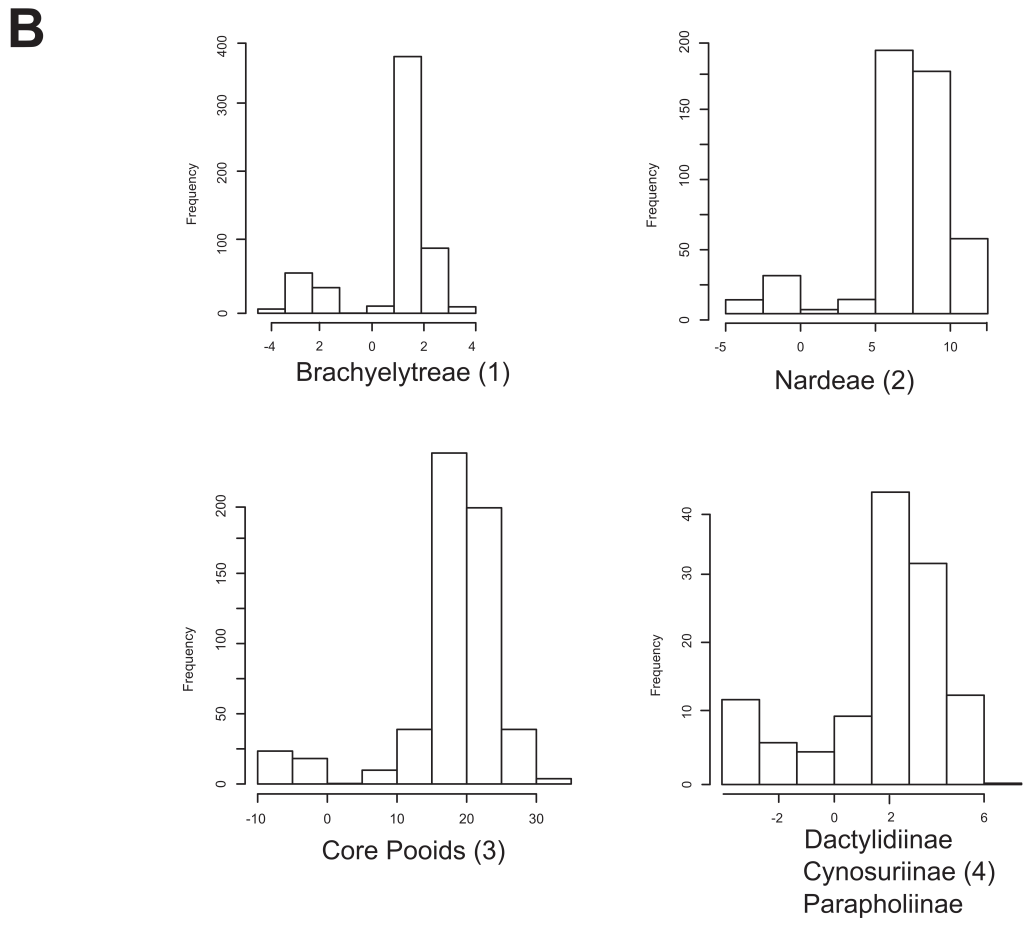
The MEDUSA analysis performed using the time-calibrated consensus tree and the one based on 500 random post burn-in trees yielded largely congruent results. The first analysis, based on a 26-terminal pruned consensus tree (24 Pooideae plus Bambuseae and Ehrarthoideae), including single representatives of each of the ancestral Pooideae tribes plus 16 representatives of the recently evolved Poodae and Triticodae subtribes, recovered three significant (ΔAIC greater than 4.0, gray branches, Fig. 2A) shifts from the background diversification rate (net diversification rate, r = 0.125 spp. My^{-1}): 1) a decrease in Nardeae (r =

0.0151 spp. My^{-1}); 2) an increase in the core pooids, including Triticodae + Poodae (r = 0.2066 spp. My^{-1}); and 3) an increase in Cynosuriinae + Parapholiinae + Dactylidiinae (r = 0.1390 spp. My^{-1}), though it also showed a decrease with respect to the rate of the core pooid clade where they are nested.

Our results based on 500 random post burn-in trees (Fig. 2B) largely coincided with those observed using the pruned consensus tree. Four clades with a mean ΔAIC for a shift in diversification rates greater than 1.0 were observed (Fig. 2B). The detected shifts were: i) a strongly supported increase in diversification rates affecting the lineage comprising the core pooids (Triticodae + Poodae): 485 out of 500 trees showed increased diversification rates [ΔAIC = 23.8; net diversification rate = r = 0.1939 (SD = 0.037) spp. My^{-1}, relative extinction rate = ε = 0.596 (SD = 0.322)]; ii) a poorly supported (detected in 120 out of 500 trees) decrease in diversification rate in the clade including the Cynosuriinae + Parapholiinae + Dactylidiinae: 98 out of 120 trees showed increased diversification rates with respect to the background rates [ΔAIC = 2.56; r = 0.1323 (SD = 0.015) spp. My^{-1}, ε = 0.00105 (SD = 0.00982)]; however these values represent a decrease with respect to the rates of the core pooid clade where the group is nested; iii) a moderately supported decrease in diversification rates affecting the Nardeae (*Nardus* + *Lygeum*): 377 out of 425 trees showed decreased diversification rates [ΔAIC = 3.44; r = 0.014 (SD = 0.003) spp. My^{-1}, ε = 0.054 (SD = 0.204)]; and iv) a very weakly supported decrease in diversification rates in Brachyelytreae: 427 out of 500 trees showed decreased diversification rates [ΔAIC = 1.12; r = 0.0174 (SD = 0.00103) spp. My^{-1}, ε = 0.0381 (SD = 0.005)]. Background rates in this analysis were estimated as r = 0.123 (SD = 0.007) spp. My^{-1}, ε = 0.0037 (SD = 0.035).

Figure 2.- A, Pruned (26 terminals) ultrametric Bayesian consensus tree obtained with BEAST (as in Figure 1; for details see Materials and Methods). Light gray and dark gray branches represent respectively primary (shifts from the background rate) and secondary (shifts from an already modified rate) shifts in the diversification rate detected using the software MEDUSA. Shift numbers are indicated on branches. Numbers as in Figure 4. B, Histograms showing the diversification rate analysis conducted with MEDUSA using 500 random Bayesian trees. Each histogram represents a significant shift in diversification rates. The X-axis represents the amount of change in the diversification rate. The Y-axis represents frequency of the rate (number of trees where the change is registered). Names of lineages correspond to the most updated tribal classification of the Pooideae (Soreng *et al.,* 2007). Shift numbers are indicated in the histograms. Shift numbers as in Figure 4.


A
Otatea acuminata
Oryza sp
Brachyelytrum erectum
Lygeum spartum
Nardus stricta
Melica minuta
Stipa offneri
Diarrhena sp.
Brachypodium distachyon
Bromus hordeaceus
Triticum aestivum
Deschampsia cespitosa
Dactylis hispanica
Cynosurus cristatus
Parapholis incurva
Festuca pratensis
Mibora minima
Sesleria albicans
Puccinellia distans
Poa annua
Alopecurus arundinaceus
Phalaris canariensis
Briza minor
Agrostis stolonifera
Avena barbata
Anthoxanthum aristatum
2
3
4
Bambusoideae/Oryzoideae
Brachyelytreae
Nardeae
Meliceae
Stipeae
Brachypodieae - Diarrheneae
Triticinae-Brominae
Airinae
Dactylidinae
Parapholiinae-Cynosurinae
Loliinae
Sesleriinae
Poinae - Puccineliinae
Phalaridinae
Agrostidinae-Brizinae
Aveninae - Anthoxanthiinae


B
Frequency
Brachyelytreae (1)
Nardeae (2)
Core Pooids (3)
Dactylidiinae
Cynosuriinae (4)
Parapholiinae

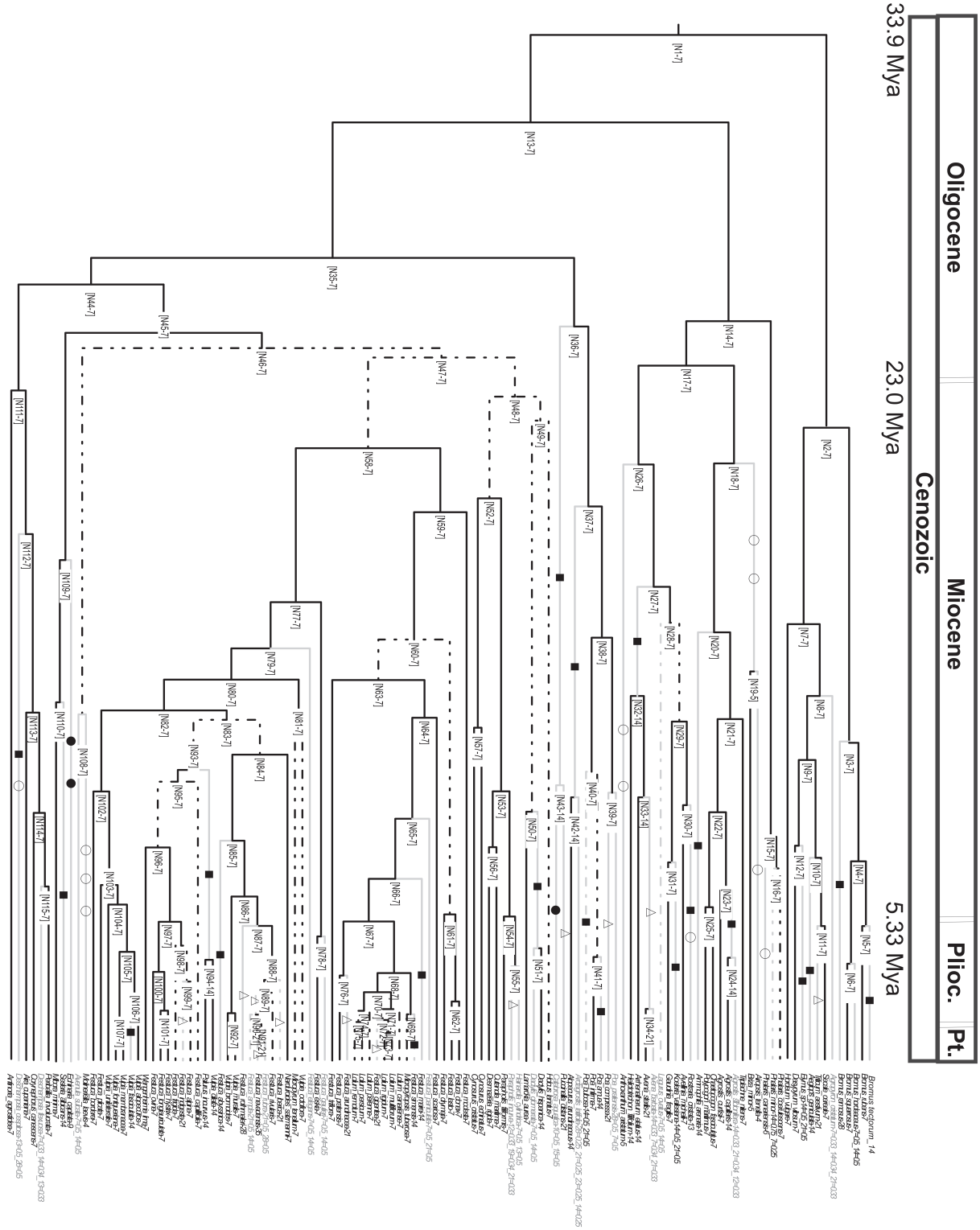

Oligocene
Miocene
Plioc.
Pt.
Cenozoic
33.9 Mya
23.0 Mya
5.33 Mya
Triticinae-Brominae-Hordeinae
Phalaridinae
Brizinae
Agrostidinae
Koelerinae
Aveninae
Anthoxanthiinae
Poinae-Puccinelliinae
Dactylidinae-Holcinae
Parapholiinae-Cynosurinae
Broad-leaved Lolliinae
Fine-leaved Lolliinae
Avenula, Molineriella
Sesleriinae
Airinae

Figure 3.- Pruned ultrametric MCC tree produced with BEAST (as in Figure 1; for details see Materials and Methods). Gray branches represent significant changes in haploid chromosome base number from the background value n = 7 detected using Chromevol. Filled square, haploid chromosome base number transition based on duplication (polyploidy); empty triangle, haploid chromosome base number transition based on addition of a multiple of the base haploid number other than 2 (demipolyploidy); filled circle, single chromosome gain; empty circle, single chromosome loss. Dashed lines represent branches with PP < 0.8. Plioc., Pliocene. Pt., Pleistocene.

4.3 Evolution of chromosome numbers in the Pooideae

The analysis conducted using chromEvol (MAYROSE *ET AL*., 2010) on the core pooids showed that the best-fitting model of chromosome number evolution indicated an underlying haploid chromosome number of 7 (not necessarily the chromosome number at the root of the core pooids; MAYROSE, 2014). The selected model included the following parameters: whole genome duplications, transitions by base number given the base number of the phylogeny (MAYROSE, 2014) and gain and loss of single chromosomes. Our results showed that in the core pooids the polyploidy (including whole genome duplication and base number transition) rate was higher than the rates of either gains or losses of single chromosomes (Polyploidy = 0.02102 mutations/My; single chromosome loss = 0.01057 mutations/My; single chromosome gain = 0.0028 mutations/My). The analyses recovered evidence for 45 chromosome transitions that occurred over the last 20 million years: 33 transitions in the haploid chromosome number (21 duplications, 12 transitions by base number, i.e., by addition of a multiple > 2 of the base haploid number); 9 chromosome losses; and 3 chromosome gains (Fig. 3). Gains and losses of single chromosomes were inferred to be independent of the current chromosome numbers in the lineages. We observed that rates of chromosome mutation were clearly higher in the most recent parts of the tree (Fig. 4).

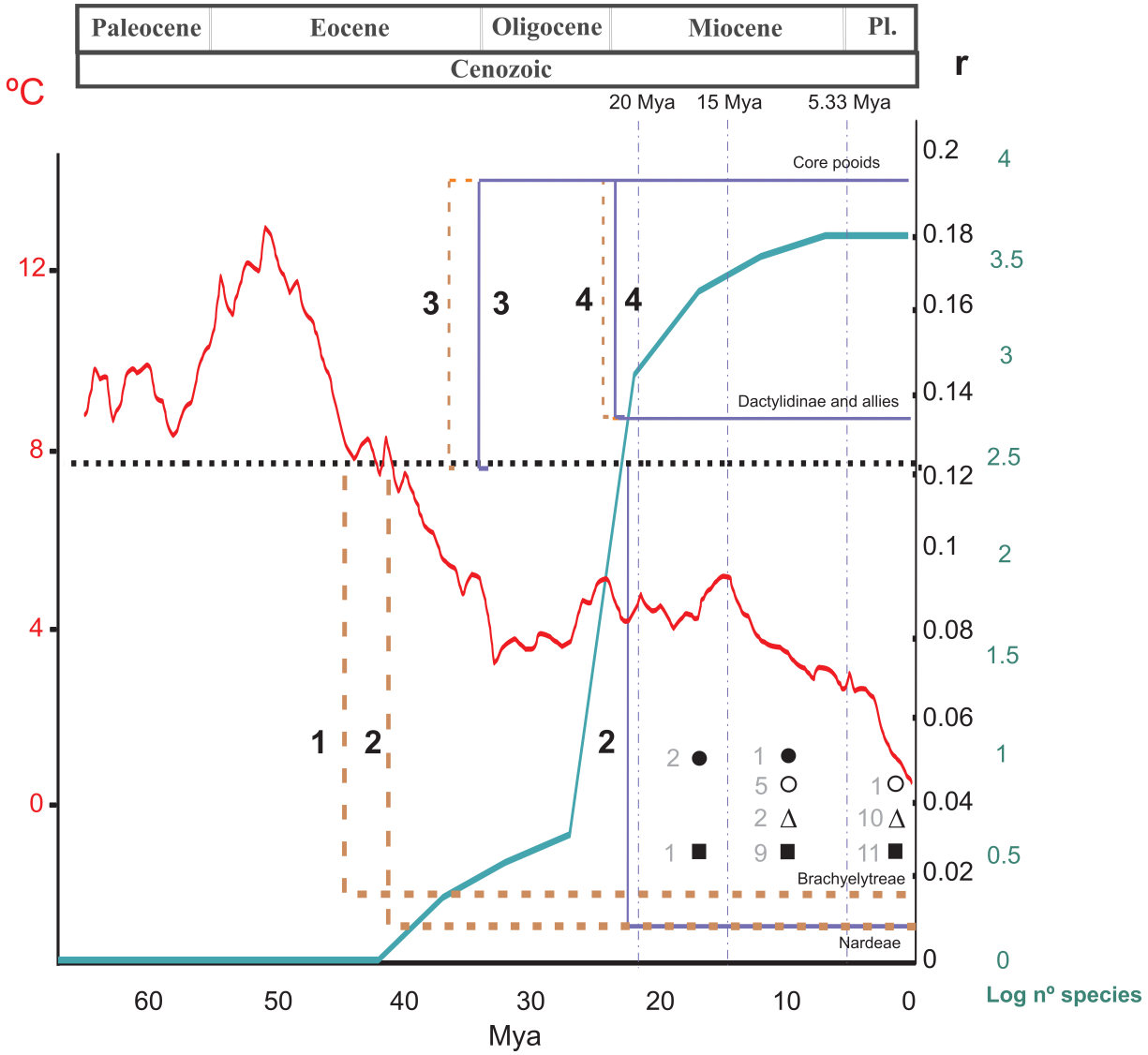


Figure 4.- Summary of the results of the diversification rate analysis and the chromosome evolution analysis across different temporal slices (TS) related to the divergences of the main Pooideae lineages. TSI, before 20 My; TSII, 20-15 My, TSIII, 15-5.33 My, TSIV, 5.33 My-present. 1, 2, 3 and 4, shifts in diversification rates affecting, respectively, the Brachelytreae, the Nardeae, the core pooids (Supertribe Triticodae + Supertribe Poodae) and the Dactylidiinae (plus allies, Cynosuriinae, Parapholiinae and Holcinae). The black dashed line represents the background diversification rate estimated from 500 ultrametric trees randomly selected from the BEAST results (r = 0.123; see Materials and Methods). For each case (1-4), the brown and blue lines represent shifts in diversification rates of, respectively, their stems and crown nodes. The green line represent the number of species (log scale) estimated in the crown nodes of the studied Pooideae tribal and subtribal lineages. The red line represents temperatures during the Cenozoic. Chromosome evolution (in dark gray): Filled square, haploid chromosome base number transition based on duplication (polyploidy); empty triangle, haploid chromosome base number transition based on addition of a multiple of the base haploid number other than 2 (demipolyploidy); filled circle, single chromosome gains; empty circle, single chromosome losses. Numbers refer to the number of events at each temporal window. Pl., Pliocene.

5. Discussion

5.1 Phylogeny and tempo of diversification for the Pooideae and the core pooids

The dates recovered in this work for the early-diverging BEP lineages were largely consistent with previous studies (e.g. VICENTINI *ET AL.*, 2008; BOUCHENAK-KHELLADI *ET AL.*, 2009). Our data showed a sister relationship between the Bambusoideae and the Pooideae that is consistent with ZHANG *ET AL.* (2011). Early diversification within the BEP clade (Erhartoideae / (Bambusoideae / Pooideae) apparently took place between the Mid to Late Paleocene and the Early Eocene. Interestingly, many Erhartoideae are restricted today to wetlands and river banks (ZHENLAN AND PHILLIPS, 2006), whereas early-diverging taxa in the Pooideae [(e.g. Brachyelytreae, Meliceae (*Melica*) and the Bambusoideae] grow mostly in woodlands (e.g. ZHENLAN AND PHILLIPS, 2006; JUDD *ET AL.*, 2008), though others do grow in wetlands [e.g. Nardeae (*Nardus*), Meliceae (*Glyceria*)] or in mesic to xeric places [(e.g. Nardeae (*Lygeum*), Meliceae (*Melica*), Stipeae)]. Some of these potentially inherited ecological preferences (e.g. woodlands) would have concurred with a period of global warming between the Late Palaeocene Thermal Maximum (55 My) and the Early Eocene Climatic Optimum (52 My) (ZACHOS *ET AL.*, 2001), when boreotropical forests were dominant in the Northern Hemisphere (WING *ET AL.*, 2009).

The order of divergence recovered in our phylogeny for the Pooideae (Fig. 1) is consistent with what has been previously published (e.g. SCHNEIDER *ET AL.*, 2009). Brachyelytrae, Nardeae *s.l.* (SCHNEIDER *ET AL.* 2009), Meliceae and Stipeae split in the Middle to Late Eocene (Fig. 1); whereas Diarrheneae and Brachypodieae diverged from their respective common ancestors in the Late Eocene (Fig. 1). Our phylogeny confirms the heretofore controversial position of Diarrheneae (BOUCHENAK-KHELLADI *ET AL.*, 2008; SCHNEIDER *ET AL.*, 2009) as sister to the Brachypodieae + core pooids lineage (SCHNEIDER *ET AL.*, 2009; MINAYA *ET AL.*, 2013). The divergence of several early-diverging lineages in the Pooideae concurs with the opening of grasslands as a result of climate change processes in the late Eocene-Oligocene (EDWARDS *ET AL.*, 2010; STRÖMBERG AND MCINERNEY, 2011).

Diversification rates and chromosome transitions for the Pooideae during the Cenozoic, as recovered in this work, are summarized in Figure 4. Average deep-sea temperatures (BEERLING AND ROYER, 2011) and the increase in species diversity in the Pooideae in the last 65 million years have also been included in this figure. The background diversification rate for the Pooideae (r) was quantified as 0.123 (SD = 0.007) spp. My^{-1}, which

is roughly consistent with the rate estimated for the Poales by MAGALLÓN AND SANDERSON (2001; r = 0.1013). We found two different significant deviations from the background rate of diversification in the Pooideae. One of these deviations affected an early-diverging, highly isolated lineage of the Pooideae: the Nardeae (including the former Lygeae; r = 0.014; SD = 0.003). This shift was dated back to the Middle Eocene-Late Oligocene (stem to crown node, Fig. 4). These dates are mostly prior to the very active diversification of grasses that took place in the Oligocene-Miocene as a result of the adaptation of the Pooideae to open habitats (BOUCHENAK-KHELLADI *ET AL.*, 2010).

Within the core pooids, three lineages were detected within the Triticodae supertribe, corresponding to the subtribes Triticinae, Hordeinae and Brominae (Fig. S1, page 230), as in SCHNEIDER *ET AL.* (2009); our dating suggests an earlier radiation of the Triticinae [14.2 My (HPD 8.5-20.4 My)] compared to that of the Brominae [11 My (HPD 6.1-17.3 My)] in the Miocene (Fig. 1). Our tree also reconstructs a large Poodae lineage that includes former Aveneae, Poeae, Hainardieae, Phalaridae, Phleeae and Seslerieae (*sensu* TZVELEV, 1976) representatives (QUINTANAR *ET AL.*, 2007; DAVIS AND SORENG, 2007). Within Poodae the split between the Aveneae-type and the Poeae-type cpDNA lineages (Fig. 1) reflects the findings of QUINTANAR *ET AL.* (2007) and SAARELA *ET AL.* (2010), and was estimated to have occurred in the early Oligocene [30.6 My (HPD 25.9-34.9 My); Fig. 1]. Some of the traditionally recognized subtribes are strongly supported in our topology (Fig. 1). Phalaridinae is reconstructed as sister to all remaining Aveneae-type cpDNA lineages. The reconstruction of Koeleriinae, traditional Brizinae (e.g. Euro-Asian Brizinae, cf. ESSI *ET AL.*, 2008) and Agrostidinae is consistent with QUINTANAR *ET AL.* (2007). Most of these lineages apparently radiated from the mid Miocene to the Pleistocene (Fig. 1).

Within the Poeae-type cpDNA clade (Fig. 1), the Poa group includes Alopecurinae (*Alopecurus*), which falls within Poinae, as reported by SORENG *ET AL.* (2003). This lineage is reconstructed as sister to Puccinelliinae, as highlighted by GILLESPIE *ET AL.* (2008). The two groups likely split in the late Oligocene [23.9 My (HPD 18.9-29.3 My)], and their respective lineages were estimated to have diverged from the mid to late Miocene to the early Pleistocene. The festucoid group (Fig. 1) encompasses former Aveneae and Seslerieae lineages such as Airinae, Sesleriinae (including *Mibora*), *Avenula* + *Molineriella*, in agreement with QUINTANAR *ET AL.* (2007), but with unclear and poorly supported relationships with respect to the more recently evolved Loliinae, Dactylidinae + Holcinae and Parapholinae

+ Cynosurinae lineages. However, the phylogenetic structure within Loliinae supported the split of the two main broad-leaved and fine-leaved Loliinae clades, as indicated by CATALAN *ET AL.*, (2006) (Figs. 1; S1, page 230). Our divergence time estimations suggest an early Miocene origin for *Festuca* and its close allies (20 My), predating the estimation of INDA *ET AL.* (2008; 13 My).

Diversification within the core pooids was especially active from the Late Oligocene to the Pleistocene, which is consistent with the expansion process of the C3 temperate Eurasian grasses that began in the Early Oligocene (e.g. BOUCHENAK-KHELLADI *ET AL.*, 2010; EDWARDS *ET AL.*, 2010). A clearly significant shift in diversification rates was detected for this group (r = 0.1939; SD = 0.0374), by far the most diverse group in the Pooideae. It was dated back to the Late Eocene-Early Oligocene. Our results indicate a temporal coincidence between the increase in the rate of diversification detected in the core pooids and the drop in global temperatures that took place in the Middle to Late Eocene and the Oligocene (BEERLING AND ROYER, 2011). Interestingly, this increase in diversification of the mostly temperate core pooids occurred before the divergence and diversification of the ungulate families Bovideae and Cervideae in moist Eurasian regions, which took place in the Late Oligocene (MATTHEE AND DAVIS, 2001; BOUCHENAK-KHELLADI *ET AL.*, 2009). By contrast, diversification of tropical, mostly C4, PACCMAD grasses concurred mostly with the diversification of Antilopienae s. lat., Hippotragineae and Alcelaphineae within the Bovidae (BOUCHENAK-KHELLADI *ET AL.*, 2009).

The connection between the development of a cooler, dryer climate in the Oligocene and the diversification of the pooid grasses has been highlighted by several authors in recent years (KELLOGG, 2001; BREDENKAMP *ET AL.*, 2002; STRÖMBERG, 2005; EDWARDS *ET AL.*, 2010; STRÖMBERG AND MCINERNEY, 2011). The same pattern has been discovered for other highly diverse herbaceous groups such as the Cyperaceae (ESCUDERO *ET AL.*, 2012; ESCUDERO AND HIPP, 2013). The diversification of the entirely C3 core pooids during the Oligocene continued during the Miocene and the Pliocene (Figs. 1 and 2) and originated primary grasslands in both hemispheres (BOUCHENAK-KHELLADI *ET AL.*, 2009; EDWARDS *ET AL.*, 2010). Interestingly, the number of diversification shifts detected for the C4 grass lineages in a genus level phylogeny (n = 800) of the PACCMAD group was much higher and occurred recent times (24 shifts during the Pliocene and the Miocene according to BOUCHENAK-KHELLADI *ET AL.*, 2009). The higher number of shifts and earlier times could be explained, at least partially, by differences

in the methodology and evolutionary scale of the analyses (see BOUCHENAK-KHELLADI *ET AL.*, 2009). However, no shift older than 23 (18.2 - 27.8) Mya was detected in PACCMAD despite the fact that the origin of the group was much older (late Eocene); and no shifts later than the boundary between the Eocene and the Oligocene were detected in our analyses despite the fact that 38 (16 tips and 21 nodes) of the 51 (26 tips and 25 nodes) analyzed clades are younger than this boundary. This difference could be explained by the heterogeneous expansion and diversification of the C4 grasses, triggered mostly by local ecological factors and disturbances rather than by changes in atmospheric conditions (OSBORNE AND BEERLING, 2006). According to TIPPLE AND PAGANI (2007) and EDWARDS AND STILL (2008), this ecological heterogeneity in the Miocene mostly affected warm parts of the world, where pooid grasses were less represented. Our results show that the temperate core pooids have presented a high and constant diversification rate influenced mostly by the atmospheric conditions in temperate areas (Fig. 4). We have also observed a gap between the taxonomic diversification in the Pooideae that started in the Late Paleocene-Early Oligocene (Figs. 1, 2) and their rise to ecological dominance today, mostly in the Northern Hemisphere. The same pattern was recovered by STRÖMBERG (2005) with respect to the taxonomic radiation and expansion of open habitat grasses in the Cenozoic in North America. This could be explained by changes in atmospheric CO_2 levels coupled with climatic disturbances and the rise of the herbivores.

Interaction between low CO_2 levels and frequent fires may promote the expansion of grasses (and other herbaceous plants) at the expense of forest trees under higher rainfall conditions (STRÖMBERG, 2005). According to this model, an augmented occurrence of fire or other disturbances (e.g. herbivory) would limit the abundance of trees, and a lowered CO_2 level would act to suppress the post-burn recovery growth rates of trees. KÜRSCHNER AND KVACEK (2009) revealed pronounced CO_2 fluctuations in the Early to Middle Miocene, of such a magnitude that they may have driven much of the temperature variation and major climatic events. Elevated CO_2 levels may also have increased the diversity in C3 plant communities. The CO_2 fluctuations may have influenced the ancient diversity of plant communities indirectly via changes in climate parameters, such as temperature or precipitation, seasonality, and/or directly by affecting plant photosynthetic performance.

5.2 Chromosome evolution in the Pooideae and the core pooids

Chromosome transitions are considered to be key mechanisms in angiosperm evolution (e.g. SOLTIS *ET AL.*, 2009). Different events are included in these mechanisms, mainly polyploidy (including polyploidization and demi-polyploidization *sensu* MAYROSE *ET AL.*, 2010) and gains and losses of single chromosomes (COHLAN *ET AL.*, 2005). Although it is clear that chromosome transitions, especially polyploidy (in the broad sense) are widespread in angiosperm evolution (e.g. SOLTIS *ET AL.*, 2009; MAYROSE *ET AL.*, 2011), their evolutionary significance has been widely disputed. Some studies have detected a relationship between chromosome transitions and diversification (e.g. SOLTIS *ET AL.*, 2009), whereas others support the traditional view that chromosome transitions, mainly polyploidy, are evolutionary dead-ends (ARRIGO AND BARKER, 2012, and references therein, but see also SOLTIS *ET AL.,* 2014).

Our analyses show that the underlying haploid chromosome number (n = 7) is remarkably constant throughout the core pooid phylogeny (Fig. 3). Changes were restricted to middle to shallow levels of the phylogeny (21 Mya-present; Fig. 3), although they occurred in all main clades of the tree. This number is consistent with the base chromosome number (x = 7) reported for most lineages in the core pooids (SHCHAPOVA, 2012). This number has also been found to be the second most common in all of Poaceae (TZVELEV, 1989; SHCHAPOVA, 2012) and represents a derived state in the family (SALSE *ET AL.*, 2008).

Within the core pooids, supported changes in haploid chromosome base numbers were detected from the Early Miocene onwards (Figs. 3, 4), with the most recent changes taking place in the Pleistocene. Polyploidy, including chromosome duplication and base number transitions, is more frequent than disploidy in the core pooids (0.02102 events / My *vs.* 0.01337 events / My, including gains and losses of single chromosomes). Our analysis detected 45 chromosome changes throughout the phylogeny (Figs. 3, 4). More specifically, we detected three single chromosome gains (6.66%), 9 single chromosome losses (20%), 21 haploid chromosome number transitions involving duplication (46.66%) and 12 haploid number transitions not involving duplication (addition of multiples of the base haploid chromosome number; 26.66%). The prevalence of polyploidy in the core pooids has been highlighted by several authors (e.g. Hsiao *et al.*, 1995), with allopolyploidy being especially important in the grasses (e.g. STEBBINS, 1971; LEVY AND FELDMANN, 2002; ROODT AND SPIES, 2003). The characteristics of the analysis conducted prevent us from distinguishing between auto- and allopolyploidy. Chromevol tracks changes along a phylogeny where

relationships are expressed as dichotomies in a phylogenetic tree, and was not designed to analyze reticulate evolution scenarios. This is one of the main criticisms of the method and might affect the precision of our estimates, especially since the effect of polyploidy on topologies is not well understood (SOLTIS *ET AL.*, 2014). However, the method has been successfully used in other plant groups, offering valuable results (e.g. MAYROSE *ET AL.*, 2011; ESCUDERO *ET AL.*, 2014).

Overall, 48.8% of the chromosome changes were dated with a maximum age of 5 million years, whereas 37.7% of them were dated back to a maximum age of 15 million years ago and 13.5% to a maximum date of 21 My (Fig. 3). Stem nodes were used for mutation dating. Interestingly, the results were different depending on the type of chromosome change considered. Thus, 81.8% of the polyploidy events took place between the Late Miocene and the present (11 My), whereas only 25% of the single chromosome gains and losses occurred during this period, with the remaining 75% of single chromosome changes occurring during the Early and Middle Miocene (Fig. 3). Our data indicate that hypothesis of constant chromosome evolution rates can be rejected. The more recent dating for polyploidizations as compared to gains and losses of single chromosomes is consistent with what has been observed in other lineages of flowering plants (ESCUDERO *ET AL.*, 2014).

Among the transitions in base haploid chromosome number (including duplications; Fig. 3), only 10 (30.3%) can be considered neo-polyploidizations *sensu* GUERRA (2008) (i.e. transitions occurring before the diversification of the generic lineages, not the genomic paleoduplication of the ancestral grasses). This fact, together with the lack of a direct relationship between the shifts in diversification rates and polyploidization (Fig. 4) and the higher frequency of these events in shallow parts of the phylogeny would indirectly support the idea that newly formed polyploid lineages in the Pooideae might experience higher extinction risk and fail to persist, as seen in other angiosperm lineages (e.g. FAWCETT *ET AL.*, 2009; MAYROSE *ET AL.*, 2011; ESCUDERO *ET AL.*, 2014; but see SOLTIS *ET AL.*, 2014). However, criticisms of this phylogenetically based method that precludes the analysis of highly reticulate allopolyploidization events (SOLTIS *ET AL.*, 2014) and the finding of remarkable exceptions to this trend, such as the duplication at the root of the subtribe Aveninae, the duplications prior to the split between *Arctagrostis* + *Alopecurus* and *Catabrosa* + *Puccinellia* lineages, and the duplication prior to the diversification of *Dactylis* (Fig. 3), together with the fact that some of the larger (e.g. *Calamagrostis*) or more restricted (e.g. *Elymus*, *Ammophila*)

pooid genera are entirely allopolyploids, support polyploidy as an evolutionary driving force in some specific lineages of the core pooids, as suggested by STEBBINS (1985).

6. Literature cited

Alfaro, M.E., Santinia, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G., Harmong, L.J., 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences USA* 106, 13410-13414.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel metropolis coupled markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC Bioinformatics* 20, 407-415.

Arrigo, N., Barker, M.S., 2012. Rarely successful polyploids and their legacy in plant genomes. *Current Opinion in Plant Biology* 15, 140-146.

Ascherson, P.F., Graebner, P., 1898-1902. *Synopsis der Mitteleuropaischen Flora*, vol, Glumiflorae I. Graminae. Leipzig: Wilhelm Engelmann.

Beerling, D.J., Royer, D.L., 2011. Convergent Cenozoic CO2 history. *Nature Geoscience* 4, 418-420.

Blaner, A., Schneider, J., Röser, M., 2014. Phylogenetic relationships in the grass family (Poaceae) based on the nuclear single copy locus topoisomerase 6 compared with chloroplast DNA. *Systematics and Biodiversity* 12, 111-124.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, F., van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Bouchenak-Khelladi, Y., Verboom, G.A., Hodkinson, T.R., Salamin, N., Francois, O., Chonghaile, G., Savolainen, V., 2009. The origins and diversification of C4 grasses and savanna-adapted ungulates. *Global Change Biology* 15, 2397-2417.

Bouchenak-Khelladi, Y., Verboom, G.A., Savolainen, V., Hodkinson, T.R., 2010. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society* 162, 543-557.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T.G., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10, e1003537.

Bournham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Bredenkamp, G.J., Spada, F., Kazmierczak, E., 2002. On the origin of northern and southern hemisphere grasslands. *Plant Ecology* 163, 209-229.

Brown, R.P., Yang, Z., 2011. Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology* 11, 271.

Cahoon, A.B., Sharpe, R.M., Mysayphonh, C., Thompson, E.J., Ward, A.D., Lin, A., 2010. The complete chloroplast genome of tall fescue (*Lolium arundinaceum* ; poaceae) and comparison of whole plastomes from the family Poaceae. *American Journal of Botany* 97, 49-58.

Catalan, P. Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily based on chloroplast *ndhF* gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Olmstead, R.G., 2000. Phylogenetic reconstruction of the genus *Brachypodium* Beauv. (Poaceae) from combined sequences of chloroplast gene and nuclear ITS. *Plant Systematics and Evolution* 22, 1-19.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and *trnL–F* sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., 2006. Phylogeny and evolution of *Festuca* L. and related genera of subtribe *Loliinae* (Poeae, Poaceae). In: A.K. Sharma, A. Sharma (Eds.), *Plant genome. Biodiversity and evolution*. Science Publishers, Enfield (New Hampshire), pp. 255-303.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new lights on the relationships of Festuca, Lolium, Vulpia and related grasses (Loliinae, Pooideae, Poaceae). In: J.P. Bailey, R.G. Ellis (Eds.), Current Taxonomic Research on the British and European Flora. London: Botanical Society of the British Isles.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., Stace, C.A., 2007. A systematic approach to subtribe Loliinae (Poaceae: Pooideae) based on phylogenetic evidence. *Aliso* 23, 380-405.

Catalan, P., Müller, J., Hasterok, R., Jenkin, G., Mur, L.A.J., Langdon, T., Betekhtin, A., Siwinska, D., Pimentel, M., López-Alvarez, D., 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany* 109, 385-405.

Chiapella, J., 2007. A molecular phylogenetic study of *Deschampsia* (Poaceae. Aveneae) inferred from nuclear ITS and plastid *trnL* sequence data: support for the recognition of *Avenella* and *Vahlodea*. *Taxon* 56, 55-64.

Christin, P.A., Spriggs, E., Osborne, C.P., Strömberg, C.A.E., Salamin, N., Edwards, E.J., 2014. Molecular Dating, Evolutionary Rates, and the Age of the Grasses. *Systematic Biology* 63,153-165.

Clayton, W.D., Renvoize, S.A., 1986. *Genera Graminum, Grasses of the World*. London: Her Majesty's Stationery Office.

Clayton, W.D., Harman, K.T., Williamson, H., 2006. *GrassBase - The Online World Grass Flora*. The Board of Trustees, Royal Botanic Gardens, Kew

Cohlan, A.E., Eichler, E., Olivera, S.G., Paterson, A.H., Stein, L., 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends in Genetics* 21, 673-682.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristic and parallel computing. *Nature Methods* 9, 772.

Davis, J.I., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in GBSSI. *Aliso* 23, 335-348.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Avenae/Poeae complex Pooideae/Poaceae. *Kew Bulettin* 62, 407-424.

Drummond, A.J., Ho, S.Y.W., 2007. *A Rough Guide to BEAST 1. 4*. Small: 1-41.

Drummond, A.J., Rambaut, A., 2007. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29, 1969-73.

Dwivedi, B., Gadagkar, S.R., 2009. Phylogenetic inference under varying proportions of indel induced alignment gaps. *BMC Evolutionary Biology* 9, 211.

Edwards, E.J., Still, C.J., 2008. Climate, phylogeny and the ecological distribution of grasses. *Ecology Letters* 11, 266-276.

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A., C4 grasses consortium. 2010. The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science* 328, 587-591.

Escudero, M., Hipp, A.L., Waterway, M.J., Valente, L.M., 2012. Diversification rates and chromosome evolution in the most diverse angiosperm genus of the temperate zone (Carex, Cyperaceae). *Molecular Phylogenetics and Evolution* 63, 650-655.

Escudero, M., Hipp, A.J., 2013. Shifts in diversification rates and clade ages explain species richness in higher level sedge taxa (Cyperaceae). *American Journal of Botany* 100, 2403-2411.

Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A.M., Pimentel, M., Jiménez-Mejías, P., Valcárcel, V., Vargas, P., Luceño, M., 2014. Karyotypic Changes through Dysploidy Persist Longer over Evolutionary Time than Polyploid Changes. *PlosOne* doi: 10.1371/journal.pone.0085266.

Essi, L., Longhi-Wagner, H.M., de Souza-Chies, T.T., 2008. Phylogenetic analysis of the *Briza* complex (Poaceae). *Molecular Phylogenetics and Evolution* 47, 1018−1029.

Fawcett, J.A., Maere, S., Van de Peer, Y., 2009. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proceeding of the National Academy of Sciences USA 106, 5737-5742.

Gaut, B.S., 2002. Evolutionary dynamics of grass genomes. *New Phytologist* 154, 15-28.

Gillespie, L.J., Soreng, R.J., 2005. A phylogenetic analysis of the bluegrass genus *Poa* based on cpDNA restriction site data. *Systematic Botany* 30, 84-105.

Gillespie, L.J., Archambault, A., Soreng, R.J., 2007. Phylogeny of *Poa* based on *trnT– trnF* sequence data: major clades and basal relationships. In: J.T. Columbus, E.A. Friar, J.M. Porter, L.M. Prince, M.G. Simpson (Eds.), *Monocots: comparative biology and evolution – Poales. Aliso* 23. pp 420-434.

Gillespie, L.J., Soreng, R.J., Bull, R.D., Jacobs, S.W.L., Refulio-Rodriguez, N.F., 2008. Phylogenetic relationships in subtribe Poinae (Poaceae, Poeae) based on nuclear ITS and plastid *trnT-trnL-trnf* sequences. *Botany* 86, 938-967.

Goldblatt, P., Johnson, D.E., 1979. *Index to Plant Chromosome Numbers (IPCN)*. Missouri Botanical Garden, St. Louis.

G.P.W.G., The Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* 88, 373-457.

G.P.W.G., The Grass Phylogeny Working Group, 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist* 193, 304-312.

Guerra, M., 2008. Chromosome number in plant cytotaxonomy: concepts and implications. *Cytogenetic and Genome Research* 120, 339-350.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52, 696-704.

Harmon, L.J., Weir, J.T., Brock, Ch.D, Glor, R.E., Challenger, W., 2008. GEIGER: Investigating evolutionary radiations. *Bioinformatics Applications notes* 24, 129-131.

Ho, S.Y.W., Phillips, M.J., 2009. Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology* 58, 367-380.

Hsiao, C., Chatterton, N.J., Asay, K.H., Hensen, K.B., 1995. Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *Theoretical and Applied Genetics* 90, 389-398.

Hunter, A.M., Orlovich, D.A., Lloyd, K.M., Lee, W.G., Murphy, D.J., 2004. The generic position of *Austrofestuca littoralis* and the reinstatement of *Hookerochloa* and *Festucella* (Poaceae) based on evidence from nuclear (ITS) and chloroplast (trnL-trnF) DNA sequences. *New Zealand Journal of Botany* 42, 253-262.

Inda, L.A., Segarra-Moragues, J.G., Muller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Inda, L.A., Sanmartín, I., Buerki, S., Catalan, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (Festuca subgenus Schedonorus and Lolium). *Journal of Biogeography* 41, 600-614.

Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F., Donoghue, M.J., 2008. *Plant Systematics, A Phylogenetic Approach*. Sunderland: Sinauer Associates Inc.

Kellogg, E.A., 2001. Evolutionary History of the Grasses. *Plant Physiology* 125, 1198-1205.

Kim, E.S., Bolsheva, N.L., Samatadze, T.E., Nosov, N.N., Nosova, I.V., Zelenin, A.V., Punina, E.O., Muravenko, O.V., Rodionov, A.V., 2009. The unique genome of two-

chromosome grasses *Zingeria* and *Colpodium*, its origin, and evolution. *Russian Journal of Genetics* 11, 1329-1337.

Kürschner, W.M., Kvaček, Z., 2009. Oligocene-Miocene CO_2 fluctuations, climatic and palaeofloristic trends inferred from fossil plant assemblages in central Europe. *Bulletin of Geosciences* 84, 189-202.

Levy, A.A., Feldmann, M., 2002. The Impact of Polyploidy on Grass Genome Evolution. *Plant Physiology* 130, 1587-1593.

Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762-1780.

Magallón, S., Castillo, A., 2009. Angiosperm diversification through time. *American Journal of Botany* 96, 349-365.

Manchester, S.R., 2001. Update on the megafossil flora of Florissant Colorado. *Proceedings of the Denver Museum of Nature and Science/series* 4, 137-161.

Matthee, C.A., Davis, S.K., 2001. Molecular Insights into the Evolution of the Family *Bovidae*: A Nuclear DNA Perspective. *Molecular Biology and Evolution* 18, 1220-1230.

Michelangeli, F.A., Davis, J.I., Stevenson, D.W., 2003. Phylogenetic relationships among poaceae and related families as inferred from morphology, inversions in the plastid genome, and sequence data from the mitochondrial and plastid genomes. *American Journal of Botany* 90, 93-106.

Mayrose, I., Barker, M.S., Otto, S.P., 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* 59, 132-144.

Mayrose, I., Zhan, S.H., Rothfelds, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., Otto, S.P., 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333, 1257.

Mayrose, I., 2014. *Chromevol version. 2.0 Manual*. Tel Aviv University.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of Stowaway Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

Müller, K., 2005. SeqState-primer design and sequence statistics for phylogenetic data sets. *Applied Bioinformatics* 4, 65-69.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 24, 581-583.

Osborne, C.P., Beerling, D.J., 2006. Nature's green revolution: the remarkable evolutionary rise of C4 plants. *Philosophical Transactions of the Royal Society B* 361, 173-194.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289-290.

Pimentel, M., Sahuquillo, E., Torrecilla, Z., Popp, M., Catalan, P., Brochmann, C., 2013. Hybridization and long-distance colonization at different time scales: towards resolution of long-term controversies in the sweet vernal grasses (*Anthoxanthum*). *Annals of Botany* 112, 1015-1030.

Pirie, M., Humphreys, A., Galley, C., Barker, N., Verboom, G., Orlovich, D., Draffin, S., Lloyd, K., Baeza, C., Negritto, M., Ruiz, E., Sanchez, J.H., Reimer, E., Linder, H.P., 2008. A novel supermatrix approach improves resolution of phylogenetic relationships in a comprehensive sample of danthonioid grasses. *Molecular Phylogenetics and Evolution* 48, 1106-1119.

Pirie, M., Humphreys, A., Barker, N., Linder, H.P., 2009. Reticulation, data combination, and inferring evolutionary history: an example from Danthonioideae (Poaceae). *Systematic Biology* 58, 612-628.

Popp, M., Mirré, V., Brochmann, C., 2011. A single mid-Pleistocene long distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). *Proceedings of the National Academy of Sciences* USA 108, 6520-6525.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (pooidea, Poaceae) inferred from plastid *TrnT-F* and nuclear ITS sequences. *American Journal of Botany* 94, 1554-1569.

Rabosky, D.L., 2006. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60, 1152-1164.

Rabosky, D.L., Donnellan, S.C., Talaba, A.L., Lovette, I.J., 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's most diverse vertebrate clade. *Proceedings of the Royal Society of London B* 274, 2915-2923.

Rambaut, A., Drummond, A., 2007. *Tracer v. 1.4*. University of Edinburgh. United Kingdom.

Ronquist, F., Huelsenbeck, J.P., van der Mark, P., 2005. *MrBayes 3.1 Manual*. San Diego: University of California at San Diego.

Ronquist, F., Huelsenbeck, J.P., Teslenko M., 2011. MrBayes version 3.2 manual: tutorials and model summaries (http://mrbayes.sourceforge.net/mb3.2_manual.pdf).

R Development Core Team, 2011. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roodt, R., Spies, J.J., 2003. Chromosome studies in the grass subfamily Chloridoideae. I. Basic chromosome numbers. *Taxon* 52, 557-583.

Saarela, J.M., Peterson, P.M., Soreng, R.J., Chapman, R.E., 2003. A Taxonomic Revision of the Eastern North American and Eastern Asian Disjunct Genus *Brachyelytrutn* (Poaceae): Evidence from Morphology, Phytogeography and AFLPs. *Systematic Botany* 28, 674-692.

Saarela, J.M., Liu, Q., Peterson, P.M., Soreng, R.J., Paszko, B., 2010. Phylogenetics of the grass 'Aveneae-type plastid DNA clade' (Poaceae: Pooideae, Poeae) based on plastid and nuclear ribosomal DNA sequence data. In: O. Seberg, G. Petersen, A.S. Barfod, J.

Davis (Eds.), *Diversity, Phylogeny, and Evolution in the Monocotyledons*. Copenhagen: Aarhus University Press. pp. 557-587.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Masood-Quraishi, U., Calcagno, T., Cooke, R., Delseny, M., Feuillet, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell* 20, 11-24.

Sanchez-Ken, J.G., Clarke, L.G., 2010. Phylogeny and a new tribal classification of the Panicoideae *s.l.* (Poaceae) based on plastid and nuclear sequence data and structural data. *American Journal of Botany* 97, 1732-1748.

Schneider, J., Doring, E., Hilu, K.W., Roser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid matK gene-3' trnK exon and nuclear ITS sequences. *Taxon* 58, 405-424.

Schneider, J., 2013. *Molecular phylogenetic analysis and classification of the Pooideae (Poaceae)*. PhD Thesis, Martin Luther Universität, Germany.

Shchapova, A.I., 2012. Evolution of the Basic Chromosome Number in Poaceae Barnh. *Russian Journal of Genetics* 2, 252-259.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49, 369-381.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C., Sankoff, D., dePamphilis, C.W., Wall, P.K., Soltis, P.S., 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96, 336-348.

Soltis, D.E., Segovia-Salcedo, M.C., Jordon-Thaden, I., Majure, L., Miles, N.M., Mavrodiev, E.V., Mei, W., Cortez, M.B., Soltis, P.S., Gitzendanner, M.A., 2014. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytologist* 202, 1105-1117.

Soreng, R.J, Davis, J.I., 2000. Phylogenetic structure in Poaceae subfamily pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. In: S.W.L. Jacobs, J. Everet (Eds.), *Grasses: Systematics and Evolution*. Melbourne: CSIRO Publishing. pp. 61-74.

Soreng, R.J., Peterson, P.M., David, S.E., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the United States National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Monica, A., Voionmaa, M.A., 2007. A Phylogenetic Analysis of Poaceae Tribe Poeae *sensu lato* based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification for the Tribe. *Kew Bulletin* 62, 425-454.

Stebbins, G.L., 1971. *Chromosomal evolution in higher plants*. London: Addison-Wesley.

Stebbins, G.L. 1985. Polyploidy, Hybridization, and the Invasion of New Habitats. *Annals of the Missouri Botanical Garden* 72, 824-832.

Strömberg, C.A.E., 2005. Decoupled taxonomic radiation and ecological expansion of open-habitat grasses in the Cenozoic of North America. *Proceedings of the National Academy of Sciences USA* 102, 11980-11984.

Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. *Annual Review of Earth and Planetary Sciences* 39, 517-544.

Strömberg, C.A.E., McInerney, F.A., 2011. The Neogene transition from C3 to C4 grasslands in North America: assemblage analysis of fossil phytoliths. *Paleobiology* 37, 50-71.

Swofford, D.L., 2002. *PAUP*. Phylogenetic analysis using parsimony* (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Tipple, B.J., Pagani, M., 2007. The Early Origins of Terrestrial C_4 Photosynthesis. *Annual Review of Earth and Planetary Sciences*. 35, 435-461.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved Festuca lineages (Poaceae) based on nuclear ITS sequences. Systematic Botany 27, 241-251.

Torrecilla, P., López Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of ITS and *trnL-F* sequences. *Annals of Missouri Botanical Garden* 91, 124-158.

Tzvelev, N.N., 1976. *Grasses of the Soviet Union*, Vols 1 and 2. Nauka, Leningrad. (Cited from the English translation 1983). Oxonion Press, New Delhi.

Tzvelev, N.N., 1989. The system of grasses (Poaceae) and their evolution. *The Botanical Review* 55, 141-203.

Vicentini, A., Barber, J., Aliscioni, S.S., Giussani, L.M., Kellogg, E.A., 2008. The age of the grasses and clusters of origins of C4 photosynthesis. *Global Change Biology* 14, 2963-2977.

Watson, L., Dallwitz, M.J., 1992 onwards. The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. Version: 12th August 2014.

Wertheim, J.O., Sanderson, M.J., Worobey, M., Bjork, A., 2010. Relaxed Molecular Clocks, the bias- variance trade –off and the quality of phylogenetic inference. *Systematic Biology* 59, 1-8.

Wing, S.L., Herrera, F., Jaramillo, C.A., Gómez-Navarro, C., Wilfe, P., Labandeira, C.C., 2009. Late Paleocene fossils from the Cerrejón Formation, Colombia, are the earliest record of Neotropical rainforest. *Proceedings of the National Academy of Sciences USA* 106,18627-18632.

Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proceedings of the National Academy of Science USA* 84, 9054-9058.

Wu, Z.Q., Ge, S., 2011. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Molecular Phylogenetics and Evolution* 63, 573-578.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K.. 2001. Trends, Rhythms, and aberrations in Global Climate 65 Ma to Present. *Science* 292, 686-693.

Zhang, Y.J., Ma, P.F., Li, D.Z., 2011. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLOS One* 6. e20596.

Zhenlan, W., Phillips, S.M., 2006. *Anthoxanthum* L. In: W. Zhengyi, P.H. Raven, H. Deyuan (Eds.). *Flora of China* Vol 22. Beijing: Science Press. pp. 336-339.

Zucol, A.F., Passeggi, M., Brea, N.I., Patterer, M.A.G., Fernández, P., Colobig, M.M., 2010. Phytolith analysis for the patrok aike lake drilling proyect: sample treatment protocols for the pasado microfossil manual". In: H. Corbella, N.I. Maidana (Eds.), 1ª *Reunión Internodos del Proyecto Interdisciplinario Patagonia Austral y 1er Workshop Argentino del Proyecto Potrok Aike Maar Lake Sediment Archive Drilling Project*. Proyecto Editorial PIPA. Buenos Aires, Argentina. pp. 81-84.

7. Appendix

7.1 Procedures for DNA isolation, amplification, and sequencing

DNA from fresh or silica gel-dried field-collected leaves was extracted following the 2% CTAB procedure of DOYLE AND DOYLE (1990). DNA from herbarium samples was extracted using the NucleoSpin Plant II kit (Macherey-Nagel, Düren, Germany) following the manufacturer's protocol. The *trn*H-*psb*A intergenic spacer was amplified using the primers of TATE AND SIMPSON (2003) and SANG *ET AL.* (1997) following SHAW *ET AL.*'S (2005) procedure. *trn*TL-*trn*LF regions were amplified and sequenced using the primers and procedures from CATALAN *ET AL.* (2004) and QUINTANAR *ET AL.* (2007). Amplification of the plastid coding regions the *ndh*F and the *mat*K gene followed CATALAN *ET AL.* (1997) and DÖRING *ET AL.* (2007), respectively. Amplified products were purified using ExoSap-IT PCR cleanup reagent (Affymetrix, Santa Clara, California). Products were sequenced using the BigDye Terminator Cycle Sequencing Ready Reaction v3.1 kit (Applied Biosystems, Paisley, United Kingdom) on an Applied Biosystems 3710 automated sequencer (Applied Biosystems, Paisley, United Kingdom). Amplification primers were also used for sequencing in all cases except for the *mat*K region (SCHNEIDER *ET AL.*, 2009).

The forward and reverse electropherograms were assembled and edited using CodonCode Aligner v. 4.0 (CodonCode Corporation, Dedham, Massachusetts). All DNA

regions were aligned separately using the MUSCLE algorithm (EDGAR, 2004) as implemented in the software SeaView v. 4.0 (GOUY *ET AL.*, 2010) and manually adjusted when errors were detected. For the alignment of coding regions, the amino acid sequence was used.

7.2. Supplementary Literature

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Molecular Phylogenetics and Evolution* 47, 488-505.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndh*F gene sequences. *Molecular Phylogenetics and Evolution* 8, 150-166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from *ITS* and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca, Lolium, Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British & European flora*. London: BSBI, pp. 45-70.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex (Pooideae, Poaceae). *Kew Bulletin* 62, 407-424.

Doyle, J.J., Doyle, J.L., 1990. Isolation of plant DNA from fresh tissue. *Focus* 12, 13-15.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792-1797.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27, 221-224.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trn*T-F and nuclear *ITS* sequences. *American Journal of Botany* 94, 1554-1596.

Sang, T., Crawford, D.J., Stuessy, T.F., 1997. Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84, 1120-1136.

Schneider, J., Doring, E., Hilu, K.W., Roser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid matK gene-3' trnK exon and nuclear ITS sequences. *Taxon* 58, 405-424.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II. Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92, 142-166.

Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., (Eds.), 2003. Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae. *Contributions from the US National Herbarium* 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

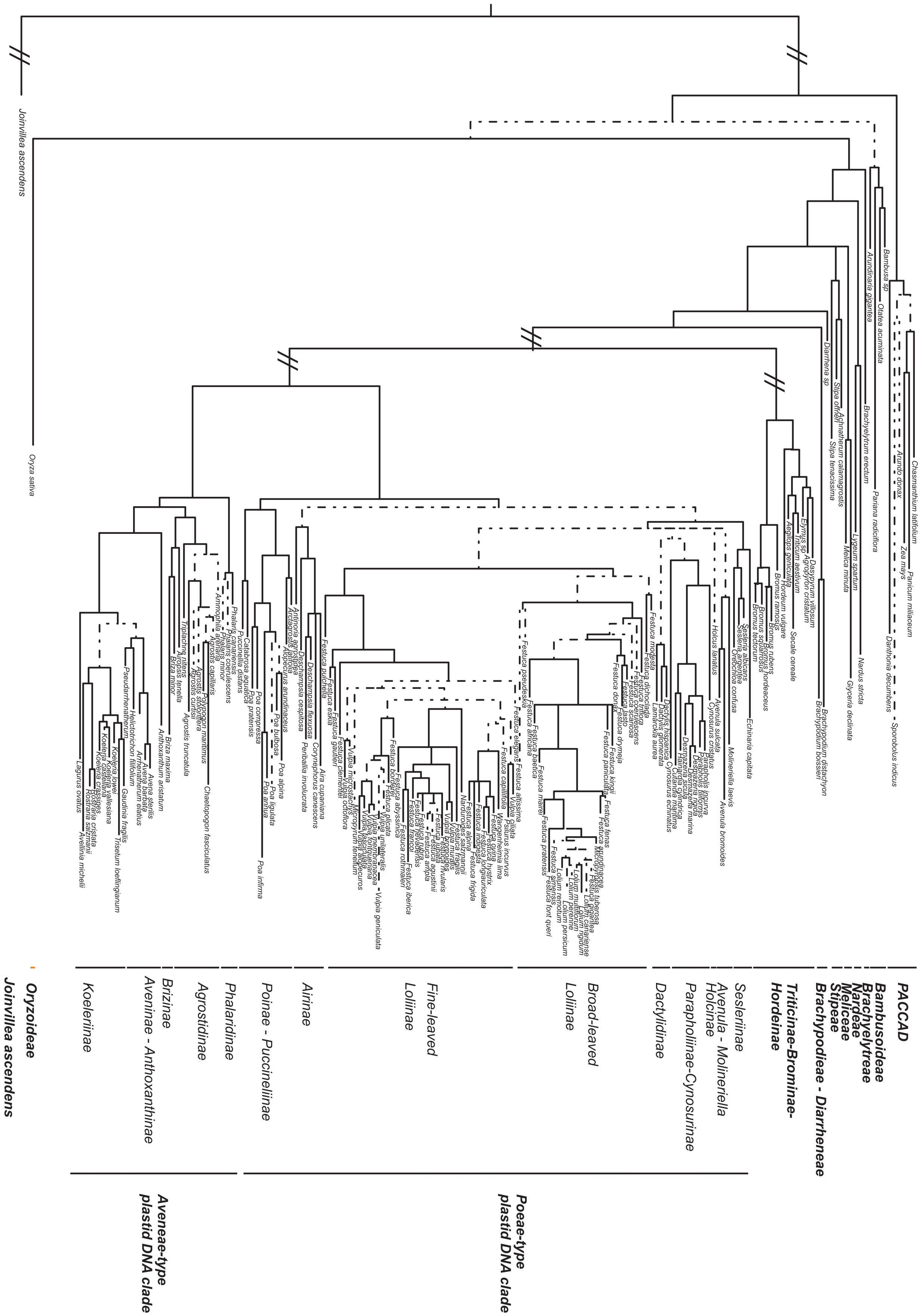
Tate, J.A., Simpson, B.B., 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28, 723-737.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv., and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Figure S1.- Bayesian 50% majority rule consensus tree obtained in the analysis of plastid DNA sequences (*trn*H-*psb*A, *trn*TL, *trn*LF, *ndh*F and *mat*K) from 163 samples representing 152 Pooideae, 6 PACCAD (PACCMAD in text), 3 Bambusoideae and 1 Erhartoideae grass species and *Joinvillea ascendens* as outgroup, constructed with MrBayes. Dashed lines represent branches with posterior probability (PP) < 0.8. // indicates shortened branches.

7.3. Supplementary figures



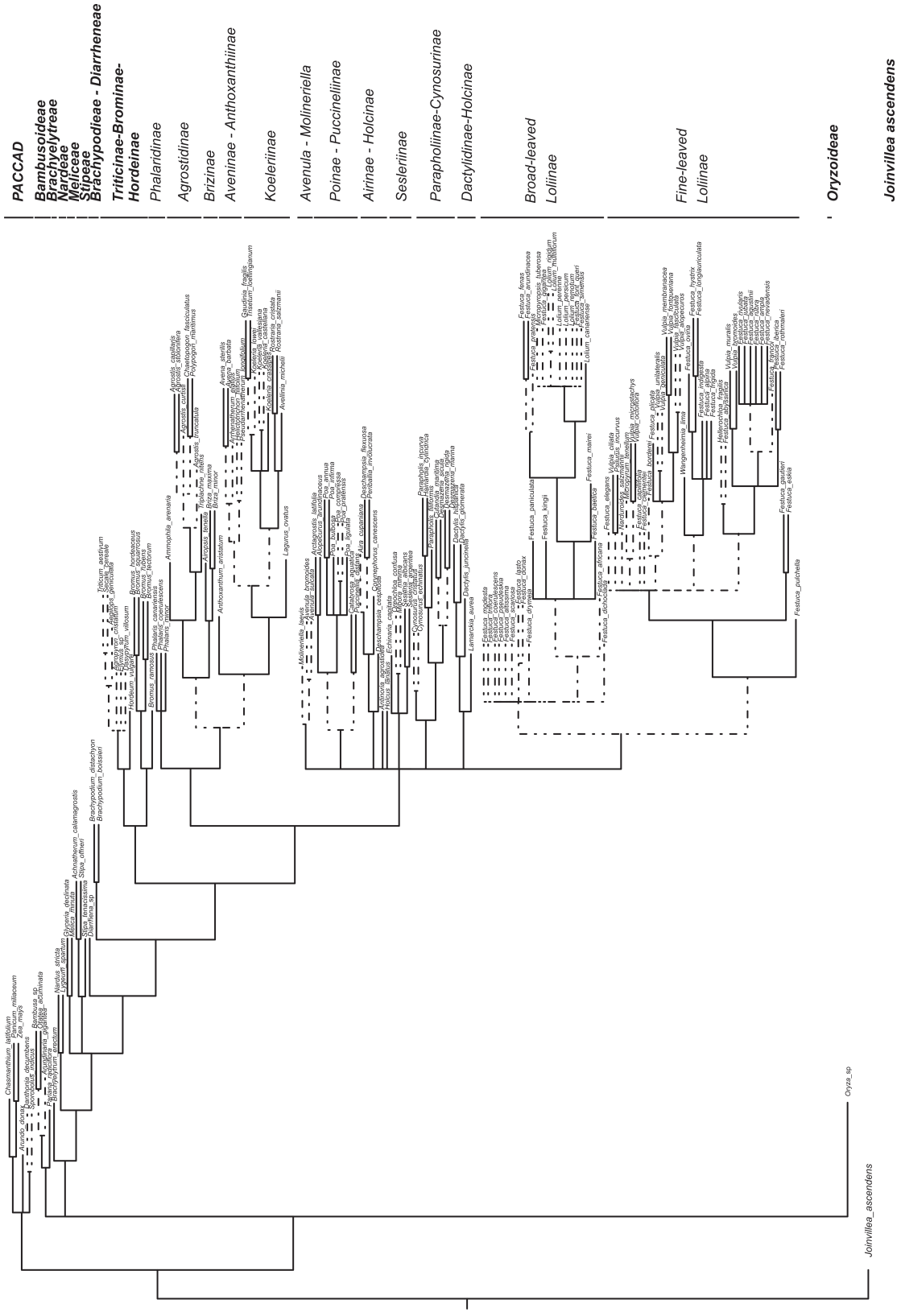


Figure S2. Maximum parsimony consensus tree obtained in the analysis of plastid DNA sequences (*trn*H-*psb*A, *trn*TL, *trn*LF, *ndh*F and *mat*K) from 163 samples representing 152 Pooideae, 6 PACCAD (PACCMAD in text), 3 Bambusoideae and 1 Erhartoideae grass species and *Joinvillea ascendens* as outgroup, constructed with PAUP. Dashed lines represent branches with bootstrap support (BS) < 0.8.

7.4 Supplementary table

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		***Trn*H-*Psb*A**	***mat*K**	***Trn*TL**	***Trn*LF**	***ndh*F**
Joinvilleaceae Tomi. and A.C.Sm.						
Joinvillea ascendens Gaud. ex Brongn and Gris	E. A. Kellogg collection, Genbank	-	AF164380.1	JQ972979	JQ972955	U21973
PACCMAD clade						
Centothecoideae Soderstrom						
Chasmanthium latifolium (Michx.) H.O. Yates	USA, cultivated seeds, G. Sanchez-Ken s.n, Genbank	KJ529116	KJ529308	KJ529257	EF137558	GU359720
Zea mays L.	Genbank	GU575286	X86563	EF541347	EF541269.1	U21985.1
Micraioideae Pilger						
Arundo donax L.	Genbank	HE966493	AF164408	KF169820	GQ869907	U21998
Panicum miliaceum L.	Genbank	-	FR667662.1	JQ972982	JQ972957	FR667670
Danthonioideae Barker and Linder						
Danthonia decumbens (L.) DC	Spain. Los Alcornocales, UZ 31.08, Genbank	KJ529117	KJ529309	JQ972986	EU401189	KJ529423
Chloridoideae Kuth ex Beilschm.						
Sporobolus indicus (L.) R.Br.	Genbank	-	AF144601	JQ972992	EF137569	U21983.1
BEP clade						
Ehrhartoideae Link						
Oryza sativa L.	Genbank	GU575284.1	AF148650.1	DQ415935	EF137577	X15901
Bambusoideae Luerss.						
Arundinaria gigantea (Walter) Muhl.	Genbank	FJ644249	-	GQ468376	EF137522	JX235347
Bambusa vulgaris Schrad. ex J.C. Wendl.	Genbank	FJ644251	FJ970915	FJ644133	JX428443	HE573559
Otatea acuminata (Munro) C.E.Calderon and Soderstr.	Genbank	EF589645.1	EF137436.1	JQ408612	EF137542	AF182350.
Pariana radicifolia Sagot ex Döll	Genbank	FJ644260.1	AF164387.1	FJ644168	EF137543	AF182354.
Pooideae Benth.						
Brachyelytreae Ohwi						
Brachyelytrum erectum (Schreb.)P. Beauv.	USA: Indiana UZ 114.08, Genbank	KJ529118	KJ529310	FJ644136	EU489326.1	KJ529424
Nardeae W.D.J. Koch + Lygeeae J. Presl complex						
Nardus stricta L.	Spain. Granada. UZ 229.07. Genbank	KJ529119	EU489217	JQ972994	EU434097.1	JX438109
Lygeum spartum L.	Spain. Zaragoza: vedado de Peñaflor, UZ 136.07	KJ529120	KJ529311	-	EU434098.1	JX438110
Stipeae Dumort.						
Glyceria declinata Bréb.	Spain. Caceres. UZ 25.07. Genbank	KJ529121	KJ529312	KJ529258	EU223364.1	KJ529425
Melica minuta L.	Spain. Córdoba.UZ 102.07. Genbank	KJ529122	KJ529313	KJ529259	KJ529403	KJ529426
Achnatherum calamagrostis (L.) P. Beauv.	Spain. Huesca: Sra. Guara, UZ 282.07. Genbank	KJ529123	KJ529314	JQ972996	EU200806.1	JX438111
Stipa offneri Breistr.	Spain. Zaragoza. UZ 135.07. Genbank	KJ529124	KJ529315	JQ972997	JQ972961	JX438112

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		***Trn*H-*Psb*A**	***mat*K**	***TrnTL***	***Trn*LF**	***ndh*F**
Stipeae Dumort.						
Stipa tenacissima L.	Spain: Almeria. UZ 180.0 Genbank	-	KJ529316	KJ529260	KJ529404	KJ529427
Brachypodieae Harz + Diarrheneae C.S. Campb.						
Diarrhena americana P. Beauv.	USA. Indiana: Belmont. UZ 116.08, Genbank	KJ529125	FM253123.1	JQ972998	JQ972962	KJ529428
Brachypodium distachyon (L.) P. Beauv.	Spain. Zaragoza. UZ 120.07. Genbank	KJ529126	KJ529317	JQ972999	AF478500.1	KJ529429
Brachypodium boissieri Nyman	Spain. Sra. Nevada. UZ 211.07, Genbank	-	JX438065	JQ973000	JQ972964	JX438113
Triticodae Watson and Dallwitz (Bromeae Dumort. / Triticeae Dumort. Complex)						
Bromus ramosus Huds.	Spain. León, Macizo Central. WGS84. Genbank	KJ529127	KJ529318	KJ529261	KJ529405	KJ529430
Bromus rubens L.	Spain. Almería, Tabernas, Genbank	KJ529128	KJ529319	KJ529262	EU036169.1	KJ529431
Bromus squarrosus L.	Spain. León. WP13. Genbank	KJ529129	KJ529320	KJ529263	KJ529406	KJ529432
Bromus tectorum L.	Spain. León. WP8. Genbank	KJ529130	KJ529321	KJ529264	EU036166.1	KJ529433
Bromus hordeaceus L.	Spain. León. Genbank	KJ529131	KJ529322	-	KJ529407	KJ529434
Triticum aestivum L.	Genbank	GU575287.1	DQ420028.1	JQ973001	AF148757.1	KC912694
Agropyron cristatum (L.) Gaertn.	Spain. Zaragoza. UZ 134.07. Genbank	KJ529132	JX438066	AF519115	AY740792	JX438114
Elymus farctus (Viv.) Runemark ex Melderis	Spain A Coruña. SANT69537 Genbank	KJ529133	KF277162	KJ529265	-	AM849122
Dasypyrum villosum (L.) P. Candargy	Italy. Sardinia. UZ 158.07 Genbank	KJ529134	KJ529323	DQ419995.1	JQ972965	JX438115
Secale cereale L.	Spain. León. Genbank	KJ529135	KJ529324	DQ336856.1	AF478501.1	KJ529435
Hordeum vulgare L.	Genbank	-	AB078138	EU036136	EU036163	HVU22003
Aegilops geniculata Roth	Spain. Zaragoza. UZ 138.07 Genbank	KJ529136	JX438067	EU013843.1	EU013604.1	JX438116
Poodae Watson and Dallwitz (Aveneae Dumort. / Poeae R.Br)						
Aveneae core lineages (Aveneae-type plastid DNA)						
Phalaris canariensis L.	Spain. Valencia. UZ 338.07 Genbank	KJ529137	JX438068	DQ631509.1	DQ631443.1	JX438117
Phalaris coerulescens Desf.	Spain: Badajoz. UZ 50.07. Genbank	KJ529138	KJ529325	KJ529266	KJ529408	KJ529436
Phalaris minor Retz.	Spain. Zaragoza. UZ 137.07 Genbank	KJ529139	-	KJ529267	KJ529409	KJ529437
Ammophila arenaria (L.) Link	Spain. Huelva. UZ 75.07 Genbank	KJ529140	KJ529326	DQ631522.1	DQ631456.1	JX438118
Agrostis capillaris All.	Spain. Lugo. UZ 299.07 Genbank	KJ529141	KJ529327	AY450936.1	JQ972966	JX438119
Agrostis stolonifera L.	Spain. Lugo. UZ 300.07 Genbank	HQ596587	KJ529328	KJ529268	KJ529410	KJ529438
Agrostis curtisii Kerguélen	Spain. Lugo. UZ 314.07 Genbank	KJ529142	KJ529329	KJ529269	-	KJ529439
Neoschischkinia truncatula (Parl.) Valdés and H. Scholz	Spain. Pontevedra. UZ 311.07 Genbank	-	KJ529330	KJ529270	-	KJ529440
Triplachne nitens Link.	Spain. Almeria UZ 365.07 Genbank	KJ529143	KJ529331	DQ336861.1	DQ336836.1	JX438120
Chaetopogon fasciculatus (Link.) Hayek.	Spain. Huelva UZ 77.07. Genbank	KJ529144	-	DQ631523.1	DQ631457.1	JX438121
Polypogon maritimus Willd.	Spain. Badajoz. UZ 52.07 Genbank	KJ529145	KJ529332	DQ336863.1	DQ336838.1	JX438122
Anthoxanthum aristatum Boiss.	Spain. Caceres UZ 24.07 Genbank	KJ529146	KJ529333	DQ631515.1	DQ631449.1	JX438123

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		*Trn*H-*Psb*A	*mat*K	*TrnTL*	*Trn*LF	*ndh*F
Aveneae core lineages (Aveneae-type plastid DNA)						
Avena sterilis L.	Spain. Caceres. UZ 19.07 Genbank	KJ529147	JX438073.1	JQ973002	EU833883.1	JX438124
Avena barbata Pott. ex Link.	Spain. Badajoz. UZ 51.07 Genbank	KJ529148	KJ529334	KJ529271	KJ529411	KJ529441
Helictotrichon filifolium (Lag.) Henrard	Spain. Valencia. UZ 346.07 Genbank	KJ529149	-	DQ336864.1	DQ336839.1	JX438125
Arrhenatherum elatius (L.) P. Beauv. ex. Z. Presl. and C. Presl.	Spain. Badajoz. UZ 55.07 Genbank	FJ395505	KJ529335	DQ336866.1	EF137591.1	KJ529442
Pseudarrhenatheum longifolium Rouy.	Spain. Coruña UZ 298.07 Genbank	KJ529150	JX438075	JQ973003	JQ972967	JX438127
Lagurus ovatus L.	Spain. Huelva. UZ 71.07 Genbank	KJ529151	KJ529336	DQ631530.1	DQ631464.1	KJ529443
Gaudinia fragilis (L.) P. Beauv.	Spain. Caceres. UZ 30.07 Genbank	KJ529152	EF137499.1	DQ631545.1	DQ631478.1	KJ529444
Trisetum loeflingianum (L.) C. Presl.	Spain. Toledo. UZ 10.07 Genbank	KJ529153	JX438076	DQ631539.1	DQ631473.1	JX438128
Koeleria vallesiana (Honck.) Gaudin	Spain. Madrid. UZ 4.07 Genbank	KJ529154	JX438077	DQ631536.1	DQ631468.1	JX438129
Koeleria castellana Boiss. and Reut.	Spain. Madrid. UZ 16.07 Genbank	KJ529155	KJ529337	KJ529272	-	KJ529445
Koeleria crassipes Lange.	Spain. Madrid. UZ 7.07 Genbank	KJ529156	KJ529338	KJ529273	DQ631535	KJ529446
Parafestuca albida (Lowe) E.B. Alexeev	Portugal. Madeira. Genbank	-	DQ786930	DQ336869.1	AF533022	DQ786858
Rostraria cristata (L.) Tzvelev	Spain. Toledo. UZ 9.07 Genbank	KJ529157	JX438078	DQ336879.1	DQ336853.1	JX438130
Rostraria salzmannii (Boiss. and Reut.) Holub	Spain. Castellón. UZ329.07 Genbank	KJ529158	KJ529339	KJ529274	KJ529412	KJ529447
Avellinia michelii (Savi) Parl.	Spain. Cádiz. UZ 98.07 Genbank	KJ529159	KJ529340	DQ631531	DQ631465	KJ529448
Poeae pro parte plus former Aveneae pro parte (Poeae-type plastid DNA)						
Molineriella laevis (Brot.) Rouy	Spain. Toledo UZ 17.07 Genbank	KJ529160	DQ786929	KJ529275	KJ529413	KJ529449
Holcus lanatus L.	Spain. Leon. UZ 291.07 Genbank	KJ529161	KJ529341	DQ631503	DQ631437	KJ529450
Poa bulbosa L.	Spain. Alicante. UZ 334.07 Genbank	KJ529162	KJ529342	EU792472.1	DQ354035.1	KJ529451
Poa infirma Kunth.	Spain. Caceres. UZ 37.07 Genbank	KJ529163	KJ529343	DQ367407.1	AF488773.1	KJ529452
Poa annua L.	Spain. León. WP16. Genbank	KJ529164	KJ529344	DQ353983.1	FJ490809.1	KJ529453
Poa ligulata Boiss.	Spain. Toledo. UZ 12.07 Genbank	KJ529165	KJ529345	KJ529276	-	KJ529454
Poa pratensis L.	Spain. Leon. WP14. LEB Genbank	KJ529166	KJ529346	KJ529277	-	KJ529455
Poa compressa L.	Spain. Lugo. Genbank	KJ529167	KJ529347	DQ354003.1	AY504649.1	KJ529456
Arctagrostis latifolia (R. Br.) Griseb.	Genbank	-	DQ786885.1	DQ353969	DQ353969	DQ786813
Desmazeria marina (L.) Druce	Spain. Alicante. UZ 320.07. Genbank	KJ529168	KJ529348	KJ529278	KJ529414	KJ529457
Catapodium rigidum (L.) C.E. Hubb.	Spain. Zaragoza. UZ 124.07. Genbank	KJ529169	JX438086	JQ973007	-	JX438141
Desmazeria sicula (Jacq.) Dumort.	Genbank	-	HE646576.1	EF584989	EF592948.1	DQ786832
Alopecurus arundinaceus Poir.	Spain. Badajoz. UZ 53.07 Genbank	KJ529170	JX438080	JQ973004	JQ972968	JX438133
Catabrosa aquatica (L.) P. Beauv.	Spain. Huesca. Genbank	KJ529171	JX438061	JQ973005	JQ972969	JX438135
Puccinellia distans (Jacq.) Parl.	Spain. Barcelona. UZ156.07 Genbank	KJ529172	KJ529349	DQ336859.1	AF533024.1	JX438134

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		***Trn*H-*Psb*A**	***mat*K**	***TrnTL***	***Trn*LF**	***ndh*F**
Poeae pro parte plus former Aveneae pro parte (Poeae-type plastid DNA)						
Aira cupaniana Guss.	Spain. Caceres.UZ 21.07 Genbank	-	KJ529350	DQ631508.1	DQ631442.1	KJ529458
Corynephorus canescens (L.) P. Beauv.	Spain. Huelva. UZ 93.07 Genbank	-	KJ529351	DQ631506.1	-	JX438136
Deschampsia cespitosa (L.) P. Beauv.	Spain. Lugo. UZ 305.07. Genbank	FJ395500.1	KJ529352	EF584984.1	AF533026	KJ529459
Avenella flexuosa (L.) Drejer	Spain. Lugo. UZ 313.07 Genbank	KJ529173	JX438083	DQ631505.1	AY237913.1	JX438137
Periballia involucrata Jonka	Spain. Salamanca. UZ 364.07 Genbank	KJ529174	KJ529353	DQ631504.1	DQ631438.1	KJ529460
Airopsis tenella Coss. and Durand	Spain. Huelva. UZ 76.07 Genbank	KJ529175	KJ529354	KJ529279	DQ631445	KJ529461
Helictotrichon sulcatum (J. Gay ex Boiss.) Potztal	Spain. Caceres. UZ 33.07 Genbank	KJ529176	KJ529355	DQ631527.1	DQ631461.1	KJ529462
Avenula bromoides (Gouan) H. Scholz	Spain. Córdoba. UZ 101.07 Genbank	-	KJ529356	KJ529280	DQ631459	-
Mibora minima (L.) Desv.	Spain. Toledo. UZ 14.07 Genbank	-	KJ529357	DQ631520.1	DQ631454.1	DQ786855.1
Briza minor L.	Spain. Caceres, UZ 39.07 Genbank	KJ529177	KJ529358	KJ529281	EU395903.1	KJ529463
Briza maxima L.	Spain. Lugo Genbank	KJ529178	KJ529359	-	EU395901	HE575736
Antinoria agrostidea (DC.) Parl..	Spain. Lugo. UZ 297.07	KJ529179	KJ529360	-	-	KJ529464
Echinaria capitata (L.) Desf.	Spain. Toledo. UZ 13.07. Genbank	KJ529180	KJ529361	DQ631519.1	DQ631453.1	JX438138
Cynosurus echinatus L.	Spain. Lugo. UZ 295.07 Genbank	KJ529181	KJ529362	EF584993	AF533031	KJ529465
Cynosurus cristatus L.	Spain. Leon. UZ 285.07 Genbank	-	KJ529363	JQ973006	EF137599	JX438139
Parapholis incurva (L.) C.E. Hubb.	Spain. Zaragoza. UZ 127.07 Genbank	KJ529182	KJ529364	EF584991	AF533036.1	KJ529466
Parapholis filiformis (Roth) C.E. Hubb.	Spain. Castellón. UZ 326.07 Genbank	KJ529183	KJ529365	KJ529282	KJ529415	KJ529467
Hainardia cylindrica (Willd.) Greuter	Spain. Alicante. UZ 356.07 Genbank	KJ529184	KJ529366	EF584990	AF533035	KJ529468
Dactylis hispanica Roth	Spain. Zaragoza. UZ 116.07 Genbank	KJ529185	KJ529367	EF584994	AF533027.1	KJ529469
Dactylis glomerata L.	Spain. Caceres. UZ 27.07 Genbank	KJ529186	KJ529368	KJ529283	AY327794	KJ529470
Dactylis juncinella Bory	Spain. Granada. UZ 222.07 Genbank	-	KJ529369	-	KJ529416	KJ529471
Lamarckia aurea (L.) Moench	Spain. Caceres. UZ 36.07 Genbank	KJ529187	JX438085.1	EF584995	AF533029	JX438140
Cutandia maritima (L.) Barbey	Spain. Valencia. UZ 324.07 Genbank	KJ529188	KJ529370	EF584987.1	AF487618.1	KJ529472
Loliinae Dumort.						
Broad-Leaved Loliinae						
Festuca modesta Nees ex Steud.	China. Yunnan. US-3420887 Genbank	KJ529189	KJ529371	EF585068	EF592985	JX438143.1
Festuca triflora J.F. Gmel.	Spain. Cádiz. UZ 95.2000 Genbank	KJ529190	JX438088	EF585109	AF533052	JX438144
Festuca coerulescens Desf.	Spain. Cádiz. UZ 19.08 Genbank	KJ529191	JX438089	EF585027.1	JQ972971	JX438145
Festuca pseudoeskia Boiss.	Spain. Granada. UZ-73.2000 Genbank	KJ529192	JX438090	EF585084	AY099000	JX438146
Festuca altissima All.	Norway. Akershus. Genbank	KJ529193	KJ529372	KJ529284	AF478505	-
Festuca scariosa (Lag.) Asch. and Graebn.	Spain. Granada. UZ 219.07. Genbank	KJ529194	KJ529373	EF585100	AY098999	JX438148
Festuca lasto Boiss.	Spain. Cádiz. UZ 146.07 Genbank	KJ529195	KJ529374	EF585060	AY098998	KJ529473

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		***Trn*H-*Psb*A**	***mat*K**	***TrnTL***	***Trn*LF**	***ndh*F**
Broad-Leaved Loliinae						
Festuca drymeja Mert. and W.D.J. Koch	Hungary. Balaton LEI Genbank	KJ529196	KJ529375	EF585034.1	AY098997.1	KJ529474
Festuca donax Lowe	Portugal. Madeira. MS4515 Genbank	-	KJ529376	EF585033	EF592968	KJ529475
Leucopoa kingii (S. Watson) W.A. Weber	USA. Colorado. PC 1.93. UZ Genbank	KJ529197	KJ529377	EF585058.1	AY099004.1	DQ786851.1
Festuca paniculata (L.) Schinz and Thell.	Spain. León. UZ 290.07 Genbank	KJ529198	JX438091.1	JQ973008	AF533046	JX438149
Festuca fenas Lag.	France. BN354. UZ 142.07 Genbank	KJ529199	JX438092	JQ973009	JQ972973	JX438150
Festuca mairei St.-Yves	Morocco. Marrakech. 4064. UAM Genbank	KJ529200	KJ529378	KJ529285	AY098996	KJ529476
Festuca arundinacea Schreb.	Spain: Lugo. JALR 1081 Genbank	KJ529201	JX438093	DQ367405	AF533042	JX438151
Micropyropsis tuberosa Romero-Zarco and Cabezudo	Spain. Huelva. UZ89.07 Genbank	KJ529202	KJ529379	JQ973010	AF533037	JX438152
Festuca pratensis Huds.	Spain. Huesca. Jaca 544795 Genbank	KJ529203	JX438095	JQ973011	AF478503.1	JX438153
Festuca gigantea (L.) Vill.	Norway. Genbank	KJ529204	JX438096	JQ973012	AF533043	JX438154
Lolium canariense Steud.	Cape Verde. LPA 21.399 Genbank	-	KJ529380	EF379014	AY228162	KJ529477
Lolium rigidum Gaudin	Spain. Zaragoza. UZ 18.2000 Genbank	KJ529205	KJ529381	KJ529286	EF378984	KJ529478
Lolium perenne L.	United Kingdom Genbank	KJ529206	JX438097	EF379024	EF378973	JX438155
Lolium multiflorum Lam.	Spain. Alicante UZ 332.07 Genbank	KJ529207	KJ529382	KJ529287	EF378972	KJ529479
Lolium persicum Boiss. and Hohen.	Turkey. Cultivated seeds PI 545661 Genbank	KJ529208	KJ529383	EF379027	EF378975	KJ529480
Lolium remotum Schrank	Sweden. Cultivated seeds PI 233613 Genbank	KJ529209	-	-	EF378979	KJ529481
Festuca font-queri St.-Yves	Morocco. Rif Mountains Genbank	KJ529210	-	KJ529288	AF533044	KJ529482
Festuca africana (Hack.) Clayton	Uganda. Elgon Mnt. MHU1585 Genbank	KJ529212	-	KJ529290	KJ529418	KJ529483
Festuca baetica (Hack.) Richt.	Spain. Granada. UZ 205.07 Genbank	KJ529211	KJ529384	KJ529289	KJ529417	KJ529484
Fine-Leaved Loliinae						
Festuca simensis Hochst. ex A. Rich.	Uganda. Echuya., MHU 1608. Genbank	KJ529213	-	KJ529291	GU573750	KJ529485
Vulpia geniculata (L.) Link	Spain. Sevilla. J29397. Genbank	KJ529214	-	KJ529292	AF478531	-
Festuca elegans Boiss.	Spain. Granada. UZ 231.07 Genbank	KJ529215	KJ529385	EF585038	EF585038	KJ529486
Festuca gautieri (Hack.) K. Richt.	Spain. Granada. UZ 232.07 Genbank	KJ529216	JX438099	EF585044	AF478507	JX438157
Festuca eskia Ramond ex DC.	Spain. Huesca. Genbank	KJ529217	JX438100	EF585040.1	AF478508.1	JX438158
Vulpia ciliata Dumort.	Spain. Zaragoza. UZ 112.07 Genbank	KJ529218	KJ529386	EF585120	AF478534	KJ529487
Psilurus incurvus (Gouan) Schinz and Thell.	Spain. Caceres UZ 31.07 Genbank	KJ529219	JX438059.1	JQ973013	AF478533	JX438159
Narduroides salzmannii (Boiss.) Rouy	Spain. Madrid. UZ 111.07 Genbank	KJ529220	JX438101.1	JQ973014	AF478535	JX438160
Micropyrum tenellum (L.) Link	Spain. Badajoz. UZ 46.07 Genbank	KJ529221	KJ529387	EF585116	AF478534	KJ529488
Festuca capillifolia Dufour ex Roem. and Schult.	Spain. Almeria. UZ 179.07 Genbank	KJ529222	JX438102.1	EF585022.1	AF478511.1	JX438161
Festuca clementei Boiss.	Spain. Granada. UZ 223.07 Genbank	KJ529223	JX438060	EF585025	AF478524	JX438162
Festuca borderei Kerguélen	Spain. Huesca. Genbank	KJ529224	-	KJ529293	AF478510	KJ529489

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		*Trn*H-*Psb*A	*mat*K	*TrnTL*	*Trn*LF	*ndh*F
Fine-Leaved Loliinae						
Festuca plicata Hack.	Spain. Granada. UZ 204.07 Genbank	KJ529225	-	KJ529294	AF478525	KJ529490
Vulpia unilateralis (L.) Stace	Spain. Zaragoza.UZ 114.07. Genbank	KJ529226	KJ529388	EF585130	AY118106	KJ529491
Vulpia membranacea (L.) Dumort.	Spain. Cádiz. UZ 97.07 Genbank	KJ529227	-	EF585124	AY118101	KJ529492
Vulpia fasciculata (Forssk.) Samp	Spain. Barcelon. UZ 15.2000 Genbank	KJ529228	-	EF585121	AF478528	JX438163
Wangenheimia lima (L.) Trin.	Spain. Zaragoza.UZ 113.07 Genbank	KJ529229	JX438103	EF585131.1	AF478536.1	JX438165
Festuca ovina L.	Germany. Thüringen. Genbank	KJ529230	JX438104.1	EF585076.1	JQ972975	JX438166
Festuca indigesta Boiss.	Spain. Granada. UZ 212.07. Genbank	KJ529231	JX438105	EF585054	AF478519	JX438167
Festuca hystrix Boiss.	Spain. Almeria. UZ 185.07 Genbank	KJ529232	KJ529389	EF585051	AF478520	JX438168
Festuca longiauriculata Fuente, Ortúñez and Ferrero	Spain. Almería. UZ 59.2000 Genbank	KJ529233	KJ529390	EF585062	AF478518	KJ529493
Festuca frigida Grossh.	Spain. Granada. UZ 226.07 Genbank	KJ529234	-	DQ631485	AF478521	KJ529494
Festuca fragilis (Luces) B. Briceño	Spain. Mérida. Genbank.	KJ529235	-	KJ529295	EF593014	KJ529495
Vulpia muralis (Kunth) Nees	Spain. Sevilla. PC 1.2002.	KJ529236	KJ529391	EF585126.1	AY118102.1	KJ529496
Vulpia bromoides (L.) Gray	Spain. Lugo.. JALR 01080. Genbank	KJ529237	KJ529392	KJ529296	AY528936	KJ529497
Vulpia alopecuros (Schousb.) Dumort.	Portugal. Algarve. LEI Genbank	KJ529238	KJ529393	EF585117	AF487617	KJ529498
Vulpia fontquerana Melderis and Stace	Spain. Huelva. UZ 91.07 Genbank	KJ529239	-	KJ529297	KJ529419	KJ529499
Vulpia microstachys (Nutt.) Munro	USA. California. Soreng R. 7406 Genbank	KJ529240	KJ529394	KJ529298	EF593015	KJ529500
Vulpia octoflora (Walter) Rydb.	USA. Washington.Peterson P.M. 3263 Genbank	KJ529241	-	KJ529299	EF593016	KJ529501
Festuca rivularis Boiss.	Spain. Granada. UZ 78.2000 Genbank	KJ529242	KJ529395	EF585093	AF478512	KJ529502
Festuca jubata Lowe	Portugal. Madeira.Genbank	KJ529243	-	KJ529300	EF592980	KJ529503
Festuca agustinii Linding.	Spain. Canarias. Genbank	KJ529244	KJ529396	EF585000	EF592949	KJ529504
Festuca francoi Fern. Prieto, C. Aguiar, E. Dias and M. I. Gut.	Portugal. Azores. Sequeira M. 4403 Genbank	KJ529245	-	KJ529301	KJ529420	KJ529505
Festuca rubra L.	UK. Scotland.UZ 150.07 Genbank	KJ529246	JX438106	EF585097.1	EF593001	JX438169
Festuca iberica (Hack.) K. Richt.	Spain. Granada.UZ 218.07 Genbank	-	KJ529397	EF585052.1	AF478516	KJ529506
Festuca ampla Hack.	Spain. Huelva. Genbank	KJ529247	-	EF585007	EF592953	KJ529507
Festuca rothmaleri (Litard.) Markgr.-Dann.	Spain. Madrid. Genbank	KJ529248	-	KJ529302	AF478513	KJ529508
Festuca nevadensis (Hack.) K. Richt.	Spain. Granada., UZ 202.07. Genbank	KJ529249	KJ529398	EF585071	AF478514	KJ529509
Festuca abyssinica Hochst. ex A. Rich.	Uganda. Gabinga. MHU1604. Genbank	KJ529250	JX438107	JQ973016	JQ972976	JX438171.1
Festuca pulchella Schrad.	Switzerland. Bern. JM 7807 Genbank	KJ529251	KJ529399	EF585086	AF519985	KJ529510
Festuca dichoclada Pilg.	Peru. Junin. US-3421417 Genbank	KJ529252	-	KJ529303	EF592967	KJ529511
Sesleria albicans Kit. Ex Schult.	Spain. León. Genbank	KJ529253	KJ529400	KJ529304	KJ529421	KJ529512
Festuca argentea Savi	Spain. Burgos. Genbank	KJ529254	KJ529401	KJ529305	AF533030	KJ529513

Table S1. List of taxa included in this study. Systematic and phylogenetic adscriptions follow those proposed by SORENG *ET AL.* (2003, 2007), TORRECILLA *ET AL.* (2003, 2004), CATALAN *ET AL.* (2006), QUINTANAR *ET AL.* (2007), BOUCHENAK-KHELLADI *ET AL.* (2008), INDA *ET AL.* (2008) and SCHNEIDER *ET AL.* (2009).

Taxa	Source	GenBank accession				
		***Trn*H-*Psb*A**	***mat*K**	***Trn*TL**	***Trn*LF**	***ndh*F**
Fine-Leaved Lolliinae						
Oreochloa confusa (Coincy) Rouy	Spain. Cantabria. Genbank	KJ529255	KJ529402	KJ529306	KJ529422	KJ529514
Festuca alpina Suter	Spain. Huesca. Genbank	KJ529256	-	KJ529307	AF478522	KJ529515

CHAPTER V

PHYLOGENY AND BIOGEOGRAPHY OF THE TEMPERATE LOLIINAE GRASSES IN THE SOUTHERN HEMISPHERE

1. Abstract

- The present-day distribution of biological diversity is the result of historical and ongoing disjunctions and dispersals of species' populations and individuals. Since Loliinae is one of the largest groups of temperate grasses with worldwide distribution, our main aim was to infer the divergence times and biogeographic processes that led to its wide distribution, specifically focusing on the macroevolutionary patterns in the previously almost-unexplored Southern Hemisphere.
- Divergence time reconstructions were estimated using the most comprehensive phylogeny of the subtribe published to date, based on nuclear and plastid markers and fossil calibrations, and applying a relaxed-clock model approach. Biogeographic processes were reconstruted using dispersal-extinction-cladogenesis analysis.
- Our results suggest that the most recent common ancestor of Loliinae originated during the Late Oligocene (24.65 ± 4.95 Ma), predating previous estimates. The respective ancestors of the broad- and fine-leaved Loliinae likely radiated during the Early Miocene (18.91 ± 4.15 and 17.50 ± 3.50 Ma, respectively). Most of the divergences of the main Loliinae lineages occurred from the Middle to the Late Miocene, though the most recent fine-leaved lineages split in the Pliocene.
- The biogeographic scenarios suggest that the Loliinae spread by recurrent long-distance dispersals, neocolonizations and recolonizations, which generally occurred from the Northern to the Southern Hemisphere, but also in the opposite direction and within the Southern Hemisphere. The broad-leaved fescues presented an unequivocal Western Mediterranean ancestor. However, the fine-leaved group showed two independent origins in the Western Mediterrean and in the Patagonian regions. Our results support more thriving dispersions to contiguous regions for the broad-leaved Loliinae, but more successful migrations through less-hospitable regions, such as the Atlantic Ocean, the Bering Strait, the Andes and Southeast Asia, for the fine-leaved lineages.

2. Introduction

Ecological theory predicts that species that are not dispersal-limited should be found where a suitable habitat, not already occupied by a superior competitor, is available (e.g. LAVIN *ET AL.*, 2004; WIENS AND DONOGHUE, 2004; DONOGHUE, 2008; CRISP *ET AL.*, 2009; LINDER *ET AL.*, 2013). Extensive knowledge of the current distributions of organisms is necessary in order to infer the ecogeographical context of the disjunctions and dispersions of their ancestors and to frame the past distributions of their populations, individuals and genes (PIRIE *ET AL.*, 2013). Thus, phylogenetic and biogeographic reconstructions are important tools for understanding the processes governing organismal evolution and geographical distribution. A variety of methods have been proposed for the reconstruction of the spatio-temporal evolution of individual lineages in greater detail (e.g. SANDERSON, 2002; THORNE AND KISHINO, 2002; REE *ET AL.*, 2005; DRUMMOND *ET AL.*, 2006). Among these methods, one of the most widely used is the parametric Dispersal – Extinction Cladogenesis (DEC) model, which allows for the integration of temporal, paleogeographic, paleoclimatic, fossil, molecular dating, and phylogenetic data (cf. REE *ET AL.*, 2005; REE AND SMITH, 2008; SANMARTÍN *ET AL.*, 2008; REE AND SANMARTÍN, 2009; BUERKI *ET AL.*, 2010).

Fossils and other geological data tend to favor vicariance when a land connection route existed at the time of migration (e.g. MILNE AND ABBOTT, 2002; COLEMAN *ET AL.*, 2003; WANG *ET AL.*, 2007). However, the biogeographic history of the Southern Hemisphere biotas cannot be entirely reduced to a simple sequence of allopatric speciations, since several recently evolved groups colonized the Southern Hemisphere after the continental breakup (BAUM *ET AL.*, 1998; SANMARTÍN AND RONQUIS, 2004). Post-Pangean breakup colonizations observed in the Southern Hemisphere may be alternatively explained by long-distance dispersals (LDDs) across land bridges or oceanic barriers (BAUM *ET AL.*, 1998; WATERS *ET AL.*, 2000; BUCKLEY *ET AL.*, 2002; SANMARTÍN AND RONQUIS, 2004; RENNER, 2004; MILNE, 2006). Recent evidence shows that frequent LDD of plants has accumulated, indicating both transoceanic (e.g. BLATTNER, 2006; INDA *ET AL.*, 2008; SMITH AND DONOGHUE, 2010; EMADZADE AND HÖRANDL, 2011) and terrestrial (POPP *ET AL.*, 2011) routes. Recurrent seed dispersion and colonization over sea barriers is apparently more dependent on the existence of favorable ecological conditions on both sides (biome compatibility) than on plants having special mechanisms for LDD (cf. WINKWORTH *ET AL.*, 2002; PIÑEIRO *ET AL.*, 2007; SANMARTÍN *ET AL.*, 2008; CRISP *ET AL.*, 2009). There is an increasing appreciation that much LDD is concomitant, and that there are patterns in both the direction and the frequency of dispersal (LAVIN *ET AL.*, 2004; RENNER,

2004; SANMARTÍN AND RONQUIST, 2004; COOK AND CRISP, 2005). This has led to the development of methods for explicitly testing and explaining these concordant dispersal patterns (CRISP *ET AL.*, 2011), exploring environmental parameters that might facilitate or inhibit LDD. These parameters include wind, which was tested in moss, liverwort, lichen, and pteridophyte floras by MUÑOZ *ET AL.* (2004), in Canarian plants by SANMARTÍN *ET AL.* (2008) and in paper daisies (Asteraceae-Gnaphalieae) by BERGH AND LINDER (2009); and bird migration routes, which were tested in the New Zealand flora by WINKWORTH *ET AL.* (2002) and in the *Empetrum* species by POPP *ET AL.* (2011).

Recent reconstructions indicate that the origin of the Poaceae (the grass-spikelet clade) likely occurred in Gondwana in dry, shady environments during the Late Paleocene (BOUCHENAK-KHELLADI *ET AL.*, 2010). According to these authors, the grasses then dispersed to all continents by approximately 60 Ma. The temperate Pooideae were inferred to have radiated in closed habitats in Eurasia in the Late Eocene (c. 44 Ma) and the core pooids (Triticodae + Poodae) in closed and open Holarctic habitats in the Early Oligocene (38 Ma). All the pooid lineages that have been studied biogeographically show distribution of their most recent common ancestor (MRCA) in Laurasia and posterior colonization of the austral territories, where their divergent lineages adapted to mountain tropical ecosystems and mesic austral landscapes [e.g. Triticeae (BLATTNER, 2006); Loliinae (CATALAN 2006; INDA *ET AL.*, 2008); Stipeae (ROMASCHENKO *ET AL.* 2014)]. The fossil record of the grasses is relatively poor for most of the Cenozoic (JACOBS *ET AL.*, 1999), though several fossil records (e.g., pollen, spikelets, leaf epidermis phytolits) have been confidently attributed to the Oryzeae-Bambusoideae, PACCMAD and Pooideae groups since the Late Cretaceous (STRÖMBERG, 2011; and references therein). Abundant and diverse grass fossils increased from the Oligocene to the Miocene (cf. STRÖMBERG, 2011), pointing to a scenario of ongoing taxonomic diversification within Poaceae, concomitant with the spread of grass-dominated vegetation, during this time (THOMASSON, 1985; BOUCHENAK-KHELLADI *ET AL.*, 2010). A roughly simultaneous taxonomic proliferation and rise to ecological dominance of the grass family, long after its origin, is thought to have been stimulated by changes in global and regional climates towards increased seasonal aridity during the Neogene (WEBB AND OPKYKE, 1995; ZACHOS *ET AL.*, 2001) and by increased grazing pressure from ungulates (BOUCHENAK-KHELLADI *ET AL.*, 2009).

Loliinae is one of the largest subtribes of temperate grasses (> 600 species) and is distributed throughout all the continents except Antarctica (CLAYTON AND RENVOIZE, 1986;

CATALAN, 2006). Previous biogeographic analyses inferred a Eurasian origin for this lineage, which has usually been associated with the geological and climatic changes of the Late Tertiary and Quaternary transition (INDA *ET AL.*, 2008). Loliinae presents an intricate evolutionary history, with remarkable events such as recent hybridization (cf. AINSCOUGH *ET AL.*, 1986; GAUT *ET AL.*, 2000), lineage sorting, and polyploidization (CATALAN, 2006). *Festuca*, a paraphyletic genus with worldwide distribution, is the largest member of Loliinae, encompassing nine subgenera (ca. 370 species in the Northern Hemisphere and ca. 200 sp. in the Southern Hemisphere) and ten allied genera (40 – 50 species), which have been reconstructed into two main clades: broad-leaved and fine-leaved Loliinae (TORRECILLA *ET AL.*, 2003; CATALAN *ET AL.*, 2004; CATALAN, 2006; INDA *ET AL.*, 2008). The poor taxonomic knowledge of many austral *Festuca* taxa has hampered the reconstruction of their spatio-temporal evolution. INDA *ET AL.* (2008) provided the first historical biogeographic study for Loliinae; however, it excluded all the tropical African and South African representatives and many of the South American taxa presented in this research.

The spatiotemporal evolution of Loliinae, a model group for other temperate pooid grasses, can now be studied more deeply thanks to the discovery of fossil evidence that can be assigned to both basal Pooideae and fine-leaved Loliinae lineages, broader sampling of tropical African, South African, South American and Australian Loliinae taxa, which has provided us with a large data set, and powerful DEC-based tools for testing alternative scenarios of neocolonization and recolonization. Thus, the particular aims of this chapter are: (i) to enlarge the phylogeny of the Loliinae, including the widest taxonomic and geographical coverage of this group in the Southern Hemisphere; (ii) to estimate the times of divergence of the broad- and fine-leaved Loliinae lineages using fossil-based calibrations; (iii) to infer the main migration events to and within the Southern Hemisphere according to the tested LDD scenarios; and (iv) to analyze the migration patterns observed between the broad- and the fine-leaved fescues and allies.

3. Material and Methods

3.1 Taxon sampling

Our sampling scheme was designed to cover the historical distribution patterns of the main Loliinae lineages, particularly in the Southern Hemisphere. We sampled the taxonomic and geographic Loliinae groups that were underrepresented in previous analyses (CATALAN *ET AL.*, 2004; CATALAN, 2006; INDA *ET AL.*, 2008), including three new subgenera (*Asperella*,

Mallopetalon, *Xanthochloa*) of *Festuca*, with special emphasis on South American, African, Indian Ocean and South Pacific species (see the appendix of chapter V [page 279] and Table S1 [page 306]. The sampling included *Pseudobromus* (*P. breviligulatus*), a genus of unknown phylogenetic adscription and controversial systematic classification (METCALFE, 1960; CLAYTON AND RENVOIZE, 1986; WATSON AND DALLWITZ, 1992; SORENG *ET AL.*, 2007). In order to encompass the geographic heterogeneity of some austral Loliinae groups, 11 taxonomically complex taxa were represented by more than one population sample [*Festuca abyssinica* (13 samples), *F. argentina* (2), *F. caprina* (2), *F. costata* (8), *F. gilbertiana* (3), *F. macrophylla* (2), *F. parvigluma* (2), *F. pilgeri* (2), *F. scabra* (9), *F. simensis* (2), and *Pseudobromus breviligulatus* (2); (Table S1 (page 306), though they were later reduced to mostly one (up to three) monophyletic representatives for the reconstruction of the dated species-tree.

A total of 178 species (214 accessions) of Loliinae were included in the phylogenetic analysis (Table S1, page 306). We included 18 additional samples (18 species) as outgroups in order to provide reliable calibration points based on fossils for the molecular phylogenetic dating. These outgroups represented the BEP clade: Bambusoideae (sister to Pooideae); Pooideae: Nardeae, Stipeae, Brachypodieae, Triticodae (Triticeae, Bromeae); and Poodae (Poeae cpDNA-type and Aveneae cpDNA-type clades) (Table S1, page 306). We followed the nomenclatural, supertribal, tribal, generic and sub-generic classifications and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages proposed by TORRECILLA *ET AL.* (2003; 2004), CATALAN *ET AL.* (2006), SORENG ET AL. (2007), INDA ET AL. (2008) and CATALAN AND MÜLLER (2012).

3.2 Phylogenetic reconstruction and divergence time analysis

Procedures for DNA extraction, amplification, sequencing and alignment of DNA sequences followed the methods indicated in CATALAN *ET AL.* (2004) (see the appendix of chapter V, page 279). Phylogenetic inference analyses were conducted on the ITS dataset (196 species; 232 accessions) and the plastid (cpDNA) *trn*TL (187 species; 222 accessions) and *trn*LF (194 species; 230 accessions) data sets. The aligned cpDNA matrices were combined into a concatenated *trn*TL-LF data set (189 species; 208 accessions) (Table S2, page 320). Searches were performed for both the independent nuclear (ITS) and plastid (*trn*TL-LF) regions. We performed Bayesian Inference (BI) analyses implemented in MrBayes 3.1.2 (RONQUIST AND HUELSENBECK, 2003; ALTEKAR *ET AL.*, 2004), Maximum Likelihood (ML)

analyses implemented in the online software RAxML 7.2.8, and Maximum Parsimony (MP) analyses implemented in PAUP* 4.0 beta10 (SWOFFORD, 2002).

jModelTest 2 (GUINDON AND GASCUEL, 2003; DARRIBA *ET AL.*, 2012) was used to select the best substitution model for the two datasets based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). TPM1uf+G and SYM+I+G were chosen as optimal models for the cpDNA and ITS data sets, respectively, and were imposed as the closest GTR+G and GTR+I+G models in the respective partitions for the MrBayes and BEAST Bayesian-based searches (see below). Before combining the nuclear and plastid (ITS+*trn*TL-LF) datasets, we checked for conflict among the individual topologies. Incongruence was assessed by looking for nodes that were strongly or well supported (posterior probability support PP>0.95 or bootstrap support BS>75%) in the Bayesian 50% majority rule consensus tree of the nuclear dataset, that were not present in the consensus tree of the cpDNA markers, and vice versa. After removing those taxa that were causing topological conflict between the nuclear and plastid trees (see results for more details), we ran a partitioned Bayesian analysis on the combined ITS – cpDNA (182 species; 218 accessions; (Table S2, page 320) data set, allowing the substitution model and the overall rate of mutation to differ among the partitions.

Absolute divergence times in the Loliinae, Pooideae and BEP outgroup lineages were estimated using a Bayesian relaxed clock, implemented in BEAST v.2.1.3 (BOUCKAERT *ET AL.*, 2014). BEAST analyses were based on a reduced ITS – cpDNA matrix (182 species; 192 accessions; (Table S2, page 320) that excluded those accessions that were causing topological conflict between the nuclear and plastid phylogenies, and included a maximum of three samples per species. The combined ITS – cpDNA dataset was run under a partitioned analysis (plastid versus nuclear), with the selected substitution model unlinked across partitions, a Yule tree prior, and uncorrelated lognormal relaxed molecular clock. The remaining model parameters were left as default prior values except the Euclidean mean, which was assigned a uniform prior distribution and set at 1.0E-4 for the initial and lower values, and 1.0E-1 for the upper value. We ran 50 million generations of Monte Carlo Markov Chains (MCMC), logging parameters every 1000 samples. The impact of the different prior sets on posterior values was assessed by running an extra chain for 50 million generations without data, sampling only the prior, following the method of POPP *ET AL.* (2011). Convergence was analyzed in Tracer (v. 1.5) (RAMBAUT AND DRUMMOND, 2007A), confirming adequate sample size, with ESS values above 650 and plots showing equilibrium. Trees were summarized in a Maximum Clade Credibility (MCC) tree obtained with TreeAnnotator v.2.1.3 (RAMBAUT AND DRUMMOND, 2007B), and

visualized in FigTree v1.3.1, after the exclusion of appropriate burn-in fractions (10% of sampled generations).

We used three of the Pooideae fossils compiled by STRÖMBERG (2011) for calibration. For the calibrations, we selected the upper bound of the geological period in which the fossil was found and assigned it as the minimum age constraint on the crown node of the clade to which the fossil is assumed to belong, based on the existence of synapomorphies between the fossil and the extant members of the clade or to taxonomic similarity to an extant member of the clade (cf. MAGALLÓN AND SANDERSON, 2001). The oldest fossil was a Middle Eocene pooid-type epidermal phytolith found in Patagonia (Argentina) (ZUCOL *ET AL.*, 2010). Therefore, we provided a minimum age for the crown group of Pooideae at 48.6 Ma (log-normal prior distribution with a log-normal mean of 44 Ma and a log-normal standard deviation of 4 Ma). Additionally, we used a stipoid seed macrofossil from Colorado (USA), which is dated from the Late Eocene (MANCHESTER, 2001). Since this fossil was identified as *Stipa* (or a similar genus of grass fruit), we considered that the crown node of Stipeae could be calibrated at a minimum age of 33.9 Ma (mean 36 Ma, S.D. 4 Ma). We also used a *Festuca* leaf macrofossil from Poland, which is dated to the Early Miocene (JUCHNIEWICZ, 1975). This fossil presents a similar cuticle to the contemporary species included in the fine-leaved Loliinae. Consequently, we calibrated the crown node of the fine-leaved Loliinae at a minimum age of 16 Ma (mean 19.5 Ma, S. D. 3.5 Ma).

3.3 Ancestral range reconstruction analysis

We used the parametric DEC approach (REE *ET AL.*, 2005) to investigate the potential biogeographic scenarios of the austral Loliinae. Alternative maximum likelihood DEC models were implemented through Maximum Likelihood Analysis in the Python version of Lagrange v. 20130526 (REE AND SMITH, 2008) using the MCC tree, which contained 174 accessions (164 species), after pruning the 18 outgroup tips with the drop.tip option implemented in R (R DEVELOPMENT CORE TEAM, 2011) (Table S2, page 320). The DEC model makes three simplifying assumptions: i) speciation is dichotomous; ii) lineage divergence can occur within an area (X), or between two areas (XY), for widespread ancestral ranges; iii) cladogenesis results in one of the two daughter species arising in a single region and inheriting its ancestor's range. For ancestors endemic to one area, range inheritance following divergence is therefore identical (REE AND SANMARTÍN, 2009). Lagrange estimated a range inheritance scenario (marginal probability of different splits), and we used the highest marginal relative probability

to infer the direction and the frequency of the biogeographic events along branches, considering that the Lagrange models allow for three types of speciation: sympatric (X/X), vicariance (X/Y) and peripheral isolate speciation (XY\X) (BUERKI *ET AL.*, 2010).

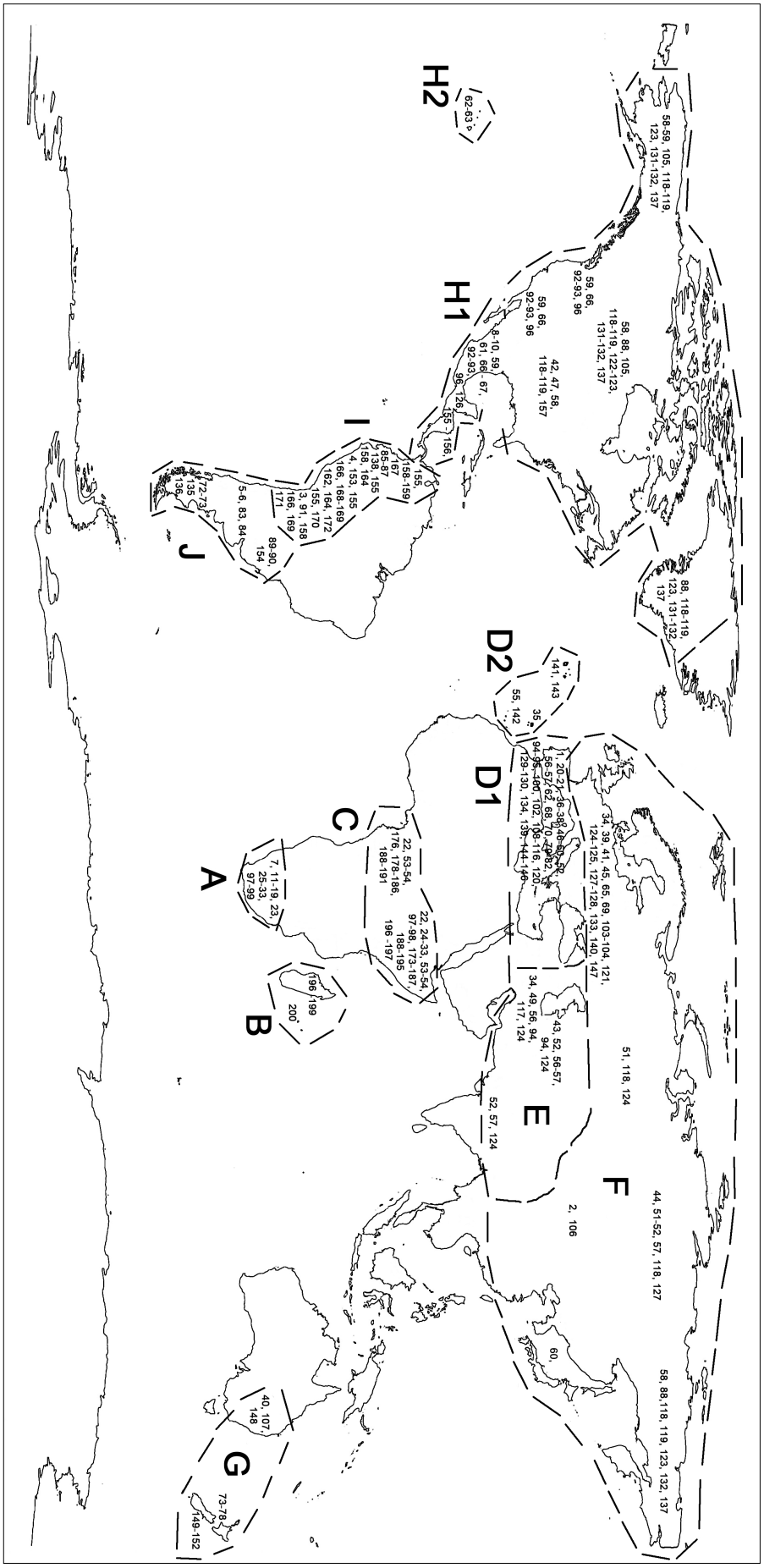
Two alternative models were tested in DEC: the null model (M0), where the dispersal rate among different biogeographic areas was constant throughout time (1.0) and the alternative model (M1), where the phylogeny was divided into four time-slices, and the respective transition matrices were constrained to five different dispersal rates (Table S3, page 321). These dispersal rates were: 1.0 for adjacent or very close regions; 0.75 for close regions with a high probability of connection; 0.50 for regions that could be connected but only through abiotic factors such as wind or ocean currents; 0.25 for regions with little connection; and 0.01 for very distant regions and those very unlikely to be connected (BAKER AND COUVREUR, 2013)

The four temporal windows were selected according to the paleogeographical and paleo wind-stream and oceanic-stream events in the Southern and Northern hemispheres that might have affected the evolution of the main Loliinae lineages (see the appendix of chapter V, page 279). The first time-slice (TSI: Late Oligocene to Middle Miocene, 28.4 - 16.0 Ma) was characterized by general cooling on the entire planet, progressive aridification in Northern Africa (ZACHOS *ET AL.*, 2001, 2008), the appearance of the Benguela current in SW Africa (MARLOW *ET AL.*, 2000; DUPONT *ET AL.*, 2011), the accretion of Mediterranean microplates, the formation of a discontinuous landmass from the proto-Iberian peninsula to Asia Minor (RÖGL, 1999; MEULENKAMP AND SISSINGH, 2003), the rising of the Northern Andes (SIMPSON, 1990), and the transition from closed forest to steppe in the Southern Patagonian Andes (BLISNIUK *ET AL.*, 2005). The second time-slice (TSII: Middle to Late Miocene, 16.1 – 7.2 Ma) was characterized by a major drop in the global climate (MEIJER AND KRIJGSMAN, 2005), the rise of the East African Rift mountains (COHEN *ET AL.*, 1993), the transition from woodlands to grass-dominated savannas in Western and Eastern Africa (JACOBS, 2004), the collision of the Arabian microplate with Eurasia (MEULENKAMP AND SISSINGH, 2003), an increase in aridification in the Mediterranean region (SALVO *ET AL.*, 2010), the closure of the Western Andean Portal (WAP), and the rise of the Central and Northern Andes (TAYLOR, 1991; HOORN *ET AL.*, 1995; HUNGERBUHLER *ET AL.*, 2002). The third time-slice (TSIII: Late Miocene to Pliocene, 7.3 - 2.6 Ma) was characterized by the volcanic formation of the high tropical African and South African mountains (GEHRKE AND LINDER, 2009), the opening of the Red Sea (MEULENKAMP AND SISSINGH, 2003), the Mediterranean Messinian Salinity Crisis, which led to the closure of the Betic and Rifian corridors (KRIJGSMAN, 2002), the closure of the Isthmus of Panama

(BURNHAM AND GRAHAM, 1999), the collision of the Australian and Asian plates, and the rise of mountain chains in Southeast Asia, Northern Australia, New Guinea, and New Zealand (SANMARTÍN AND RONQUIST, 2004). The fourth time-slice (TSIV: Quaternary, 2.61 - 0 Ma) was characterized by the Pleistocene glaciations, which affected the Northern and Southern hemispheres, with ice sheets covering Northern and Central Eurasia, and Southern South America (JOLLY *ET AL.*, 1998; YALDEN AND LARGEN, 1992; RABASSA *ET AL.*, 2000). These glaciations allowed for the Holocene postglacial colonizations in both hemispheres.

Initially, ten Operational Areas (OAs) were defined for the reconstruction of the ancestral ranges of the Loliinae (A to J, Fig. 1). The OAs were selected according to the current distribution ranges of the Loliinae species and the potential historical distributions of their ancestors, delimited by geographical features that could have acted as barriers to dispersal (BUERKI *ET AL.*, 2012). However, in order to increase the accuracy of the historical Loliinae distribution, the number of OAs was later increased to twelve, separating Macaronesia (D2) from the Western Mediterranean region (D1), and Hawaii (H2) from the North American region (H1). Both biogeographic reconstructions (10 vs. 12 OAs) are shown in this study.

Figure 1. Distribution map of the Loliinae samples included in this study, showing the boundaries (dotted lines) of the 12 Operational Areas (OAs) used in the biogeographic analysis. The numbers on the map represent the approximate location and distribution of the species consecutively listed in table S1. OAs: A – Southern Africa; B – Madagascar and Mascarenes; C – High mountains of tropical Africa; D1 – Western Mediterranean; D2 – Macaronesian archipelagos; E – Irano-Turanian-Himalayan region (ITH); F – Eurasia (Eurosiberian region); G – New Zealand and Southwestern Australia; H1 – North America and Mesoamerica; H2 – Hawaiian archipelago; I – Northern Andes; J – Southern Andes and Southern South America.


H2
62-63
H1
59, 66, 92-93, 96
58, 88, 105, 118-119, 122-123, 131-132, 137
42, 47, 58, 118-119, 157
8-10, 59, 61, 66, 67, 92-93, 96, 126, 155, 156
I
158-159
155
167
85-87
138, 155
158, 164
4, 153, 155
166, 168-169
162, 164, 172
155, 170
3, 91, 158
166, 169
171
5-6, 83, 84
89-90, 154
J
72-73
135
136
88, 118-119, 123, 131-132, 137
D2
141, 143
55, 142
35
D1
1, 20-21, 36-38, 48-50, 52, 56-57, 62, 68, 70, 79-82, 94-95, 100, 102, 106-116, 120, 129-130, 134, 139, 144-146
C
22, 53-54, 176, 178-186, 188-191
A
7, 11-19, 23, 25, 33, 97-99
22, 24-33, 53-54, 97-98, 173-187, 188-195, 196-197
B
196
199
200
34, 39, 41, 45, 65, 69, 103-104, 121, 124-125, 127-128, 133, 140, 147
34, 49, 56, 94, 117, 124
43, 52, 56-57, 94, 124
51, 118, 124
E
52, 57, 124
F
2, 106
44, 51-52, 57, 118, 127
60
58, 88, 118, 119, 123, 132, 137
40, 107, 148
G
73-78
149-152

4. Results

4.1 Phylogenetic relationships

The separate nuclear (ITS) and plastid (*trn*TL-LF) markers yielded congruent topologies, with all major clades recovered by the Bayesian (see the appendix of chapter V; Figs. S1, S2; page 289) and Maximum Parsimony (Figs. S1, S2; bootstrap values; page 289) analyses. However, the comparison between the Bayesian inferences of the grasses based on the nuclear versus the plastid data sets indicated that 13 of the species were nested in different lineages: *Castellia tuberculosa, Festuca amethystina, F. borderei, F. caucasica, F. clementei, F. hubsugulica, F. komarovii, F, longipes, F. norica, F. sibirica, F. woronowii, Micropyrum tenellum,* and *Vulpia octoflora* (Figs. S1, S2; page 289; Table S2; page 320). Since the Bayesian phylogenies based on the separate nuclear and plastid genomes were more similar after removing the 13 potentially conflicting species, and since these topological misplacements did not affect the main results and conclusions of this research, we chose to conduct Bayesian (Fig. 2: posterior probability values of Fig. S3; page 296) and MP (Fig. 2: bootstrap values of Fig. S3; page 296) analyses on the joint dataset (ITS+*trn*TL-LF).

Figure 2. Chronogram of Loliinae based on the Maximum Clade Credibility (MCC) tree inferred by BEAST. The three calibration points based on Pooideae fossils (see text) used for the estimates of divergence analyses are indicated by stars, bold numbers and arrows. Mean divergence time estimations are indicated in grey numbers with an asterisk. The grey bars indicate 95% highest posterior density (HPD) intervals for the divergence time estimates of nodes with posterior probabilities above 0.50. Black numbers above and below the branches indicate posterior probability (PP) and bootstrap (BS) support values (Fig. S3; page 296), respectively. The upper time scale bar represents million years ago (Ma). The vertical dotted grey lines indicate the boundaries every 5 million years. Geological time scale is shown at the bottom. Operational areas assigned to each selected species (A to J; Fig. 1), the name of the species, its code number (Fig. 1; Table S1; page 306), and major phylogenetic groups are indicated to the right of the tree.

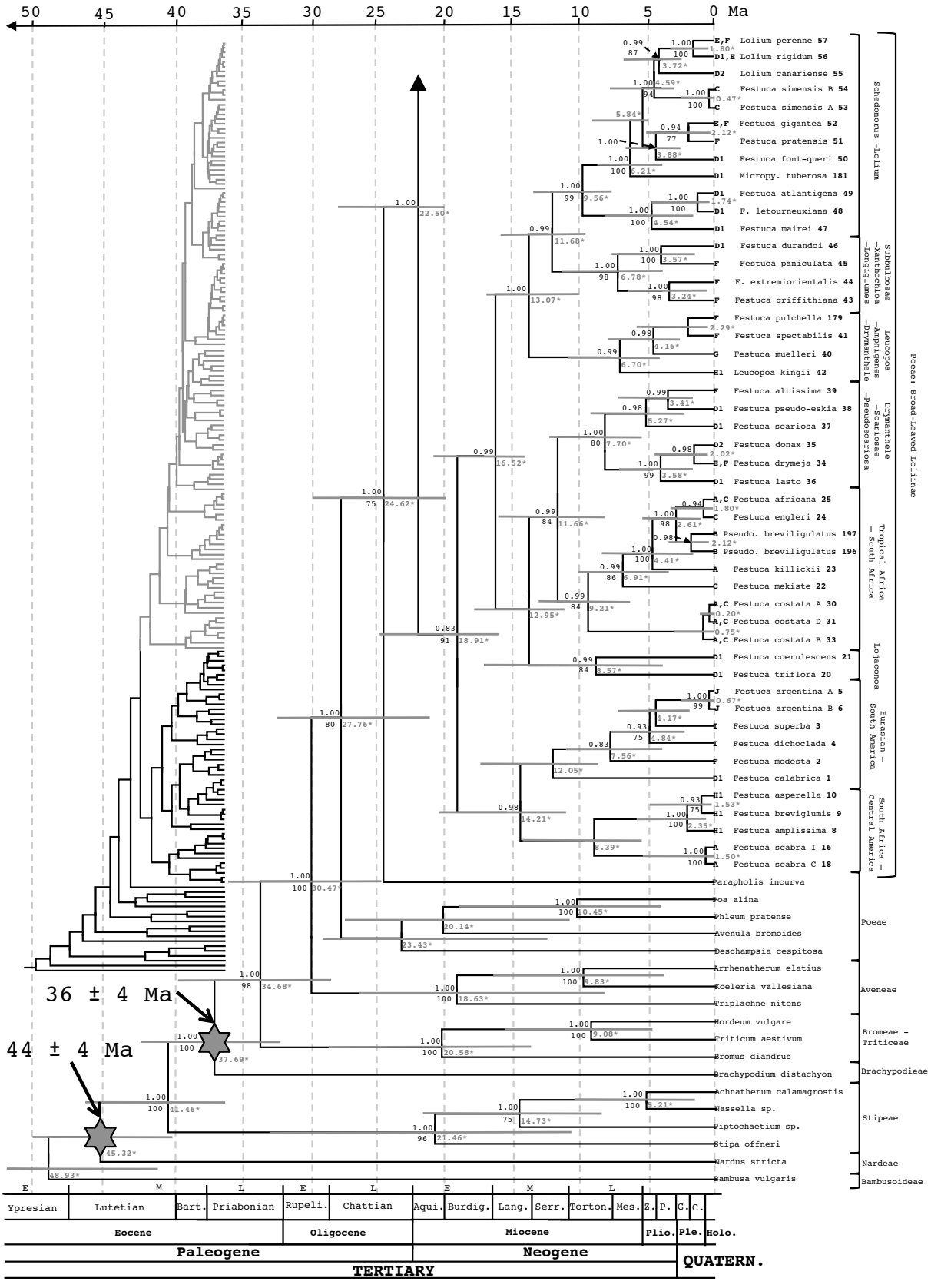


Figure 2. *(Continued)*

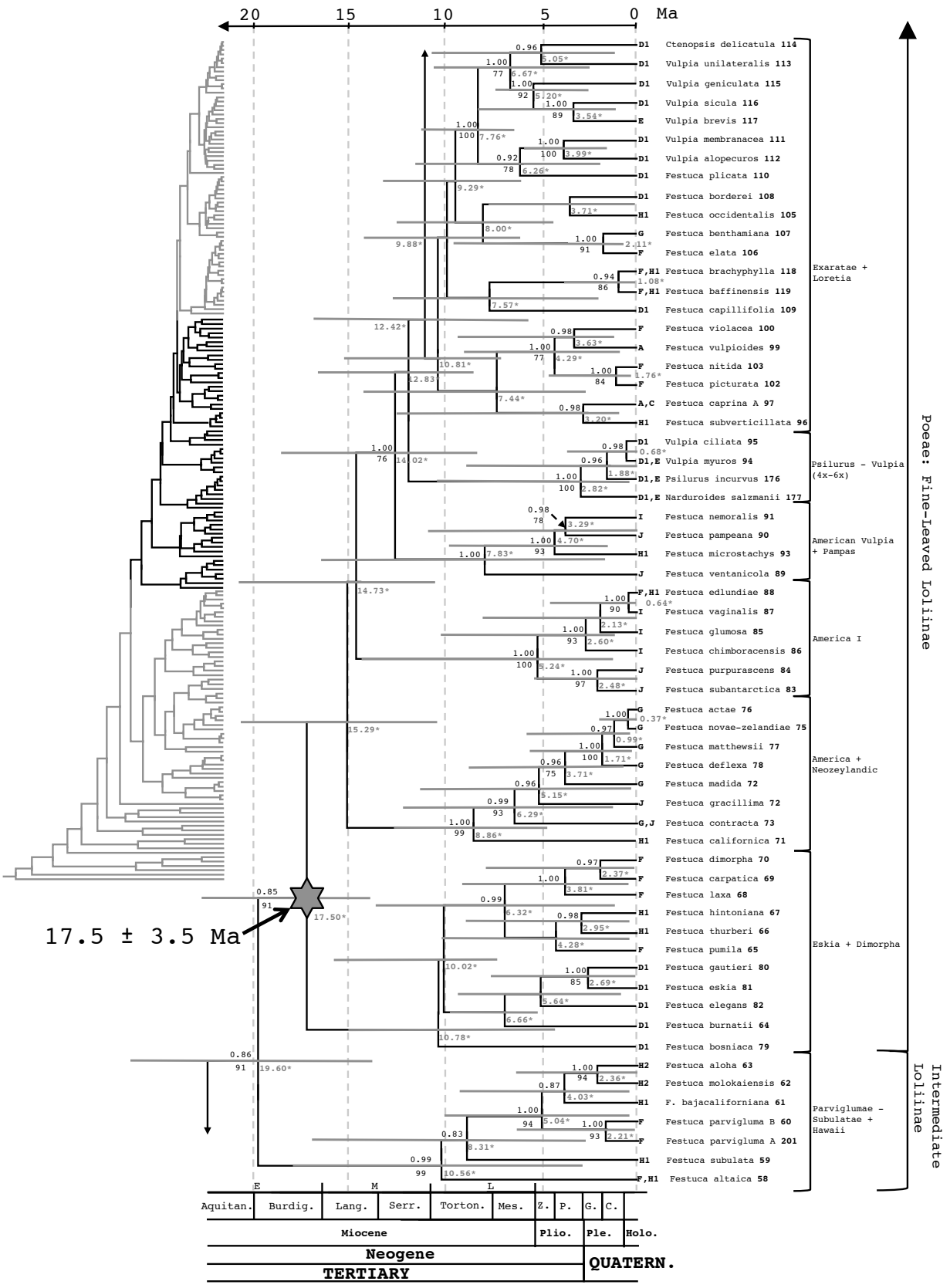


Figure 2. *(Continued)*

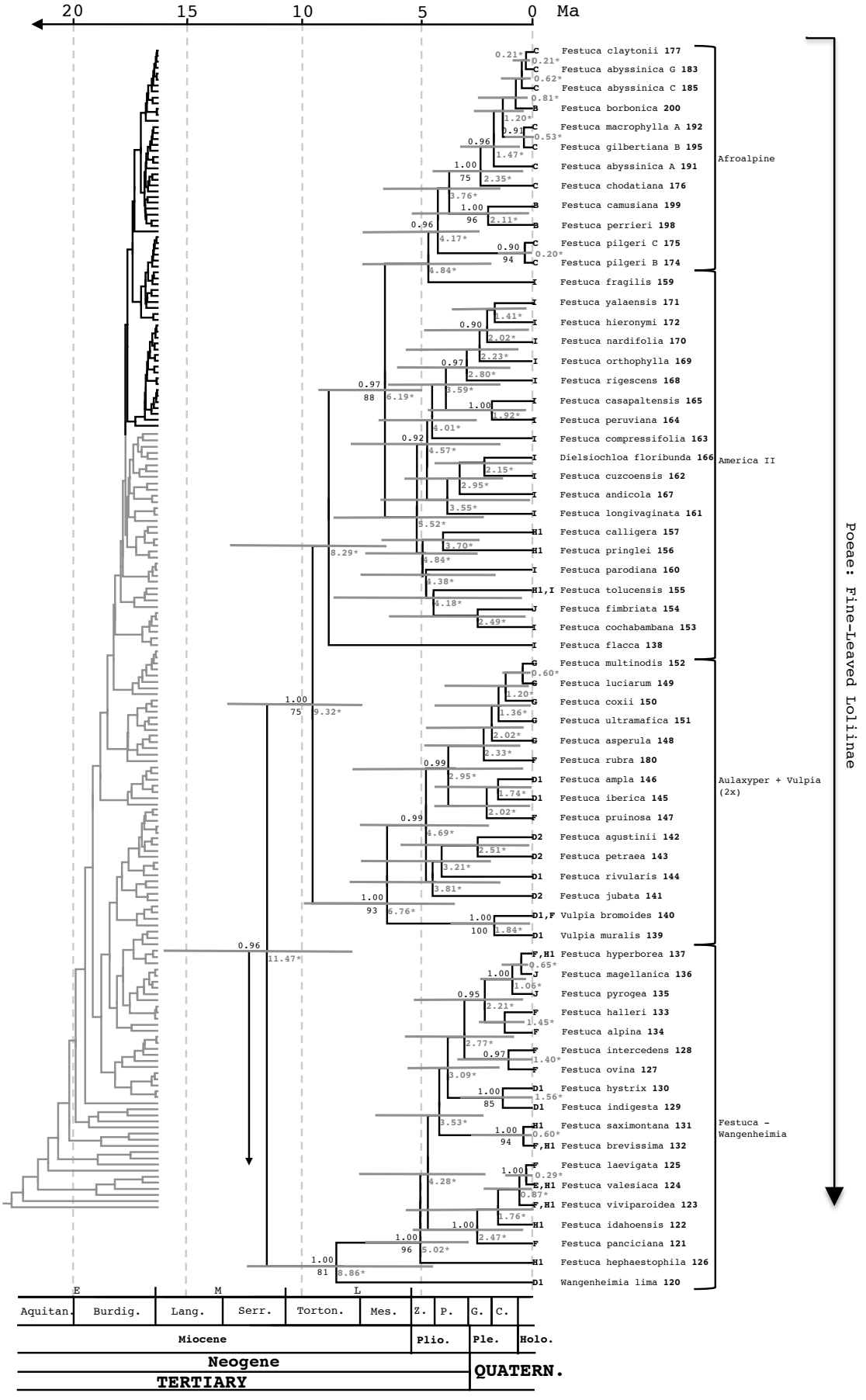


Figure 2. *(Continued)*

The nuclear-and-plastid-based topology (Figs. 2, S3; page 296) was overall better resolved and presented stronger support values for the main divergences than those from the individual analyses (Figs. S1, S2; page 289). Furthermore, the inference based on the combined data set was consistent with previous studies (INDA *ET AL.*, 2008). The MCC (182 taxa; 192 accessions; Fig. 2; Table S2; page 320) and Bayesian (182 taxa; 218 accessions; Table S2; page 320; Fig. S3; page 296) trees were highly congruent to each other, and we will focus on results from the first analysis. Our analysis recovered six new lineages, mostly corresponding to austral Loliinae groups: (i) an early-diverging broad-leaved lineage, which consisted of an enlarged Eurasian – South American clade, showing the early split of the Mediterranean *F. calabrica* (*F.* subgen. *Leucopoa*); (ii) its sister South African – Central American clade, which included the South African *F. scabra* (*incertae sedis*: inc. sed.) and Mexican *F. asperella* (*F.* subgen. *Asperella*); (iii) the well supported Tropical African – South African clade, sister to the Eurasian Drymanthele – Scariosa – Pseudoscariosa clade, represented by *Pseudobromus breviligulatus* (Madagascar), the tropical African *F. engleri* and *F. mekiste*, the South African *F. killickii*, and the tropical African and South African *F. africana* and *F. costata* (inc. sed.); (iv) the strongly supported Hawaiian *F. aloha* / *F. molokaiensis* (inc. sed.) clade, nested within the strongly supported Parviglumae – Subulatae clade (Intermediate lineage); (v) the well supported American Vulpia + Pampas clade, which included the Argentinian *F. nemoralis* (inc. sed.), *F. pampeana* and *F. ventanicola* (*F.* subgen. *Festuca*) plus *Vulpia microstachys* (North American); and (vi) the Afroalpine clade, which collapsed into a polytomy with the American II lineages in the Bayesian tree (Fig. S3; page 296) but was sister to the American II clade (except *F. flacca*) in the MCC tree (Fig. 2). The well supported Afroalpine + American II clade was resolved as sister to the Aulaxyper clade within the more recently evolved fine-leaved clade (Figs. 2; S3; page 296). The Afroalpine clade included the tropical African *F. abyssinica*, *F. chodatiana*, *F. claytonii* (inc. sed.), *F. gilbertiana* and *F. pilgeri* (*F.* subgen. *Festuca*), the Malagasy *F. camusiana* (inc. sed.) and *F. perrieri* (*F.* subgen. *Festuca*), and the Mascarene *F. borbonica* (*F.* subgen. *Festuca*).

Our inferences (Figs. 2, S3; page 296) also resolved the placement of 15 additional austral Loliinae taxa, which were nested within previously described lineages (CATALAN *ET AL.*, 2004; CATALAN, 2006; INDA *ET AL.*, 2008): the Andean *F. superba* (inc. sed.) was placed within the basal Eurasian – South American clade (broad-leaved Loliinae), enlarged in this study; the Australian *F. muelleri* (*F.* subgen. *Drymanthele* sect. *Banksia*) was grouped as sister to the Eurosiberian *F. spectabilis* and *F. pulchella*, within the Leucopoa – Amphigenes –

Drymanthele lineage; the Southwestern Asian *F. griffithiana* (*F.* subgen. *Xantochloa*) was resoved as sister to the Eastern Asian *F. extremiorientalis* (*F.* subgen. *Subulatae* sect. *Longiglumis*), within the Subbulbosae – Xanthochloa – Longiglumes clade; the cicumantarctic *F. contracta* and the Fuegian *F. gracillima* (*F.* subgen. *Festuca*) were nested within the fine-leaved subbasal American + Neozeylandic clade; the Australian *F. benthamiana* (*F.* subgen. *Festuca*) and the South African *F. caprina* (*F.* subgen. *Festuca*) and *F. vulpioides* (inc. sed.) were included within the Exaratae clade; the Patagonian *F. magellanica* and *F. pyrogea* (*F.* subgen. *Festuca*) were grouped with the Holarctic – Alpine *F. hyperborea* within the Festuca clade; the Australian *F. asperula* (inc. sed.) was nested with the New Zealand *F. coxii, F. luciarum*, *F. multinodis* and *F. ultramafica* within the Aulaxyper clade; the Peruvian *F. casapaltensis* (*F.* subgen. *Festuca*), *F. compressifolia* (inc. sed.) and the Argentinian *F. fimbriata* (*F.* subgen. *Mallopetalon*) and *F. hieronymi* (*F.* subgen. *Festuca*) were nested within the American II lineage.

4.2 Divergence times and biogeographic reconstruction

Four biogeographical reconstructions based on the MCC tree (164 species; 174 accessions; see Table S2 (page 320), were performed: DEC stratified – M1 – with 10 (Fig. 3) and 12 OAs (Fig. S4; page 299); and DEC unconstrained – M0 – with 10 (Fig. S5, page 303) and 12 OAs (Fig. S6; page 304). The stratified model (M1) based on 10 OAs showed a better fit for the data (-ln likelihood 402.6) than the alternative reconstructions (M0 10 OAs: 421.8; M1 12 OAs: 452.4; M0 12 OAs: 459.7; Table S2; page 320), and we will refer to this model hereafter. The M1 10 OAs model presented an estimated dispersal rate (dis: 0.04325) that was three times higher than the estimated extinction rate (ext: 0.01506).

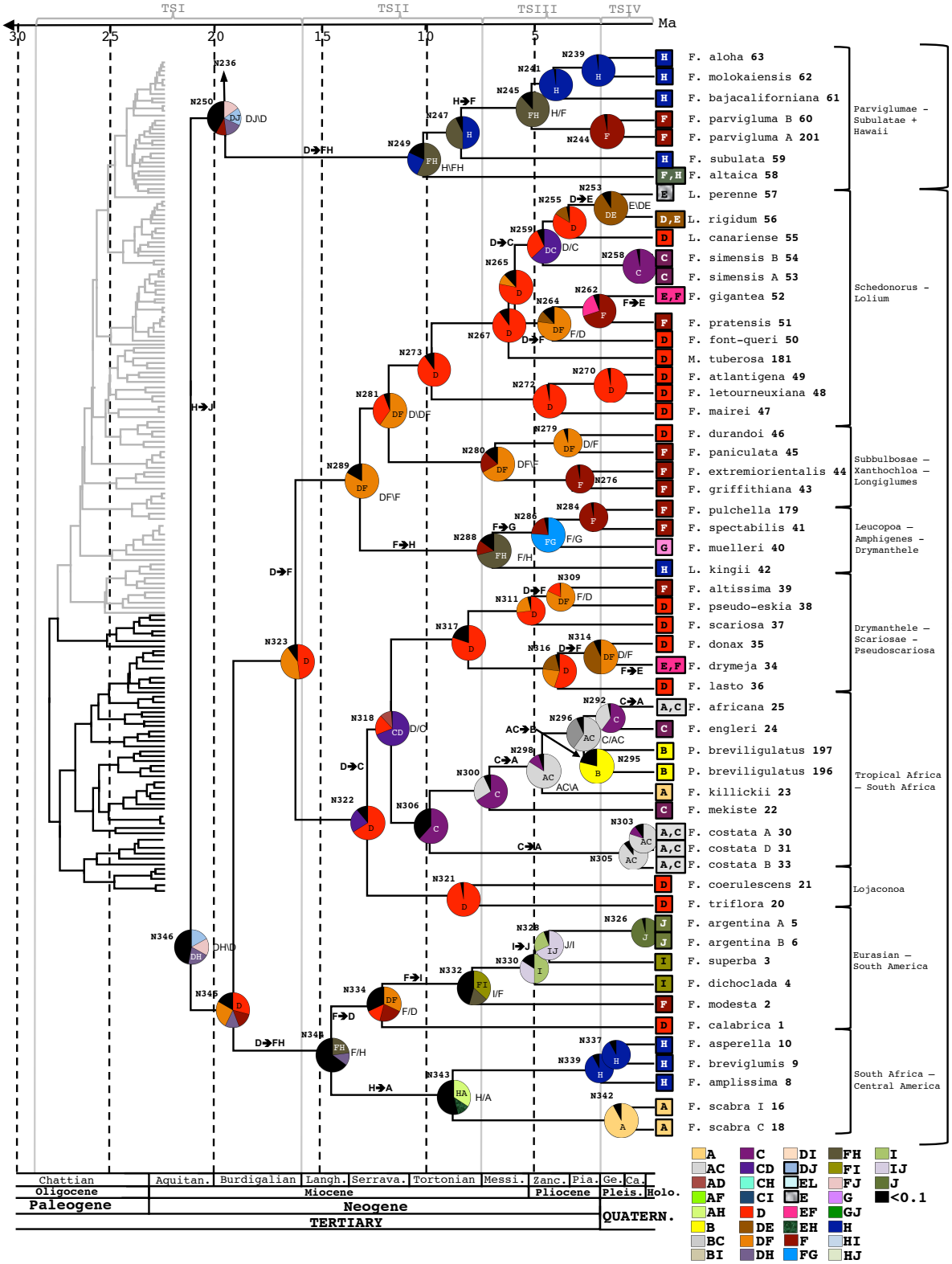


Figure 3. *(Continued)*

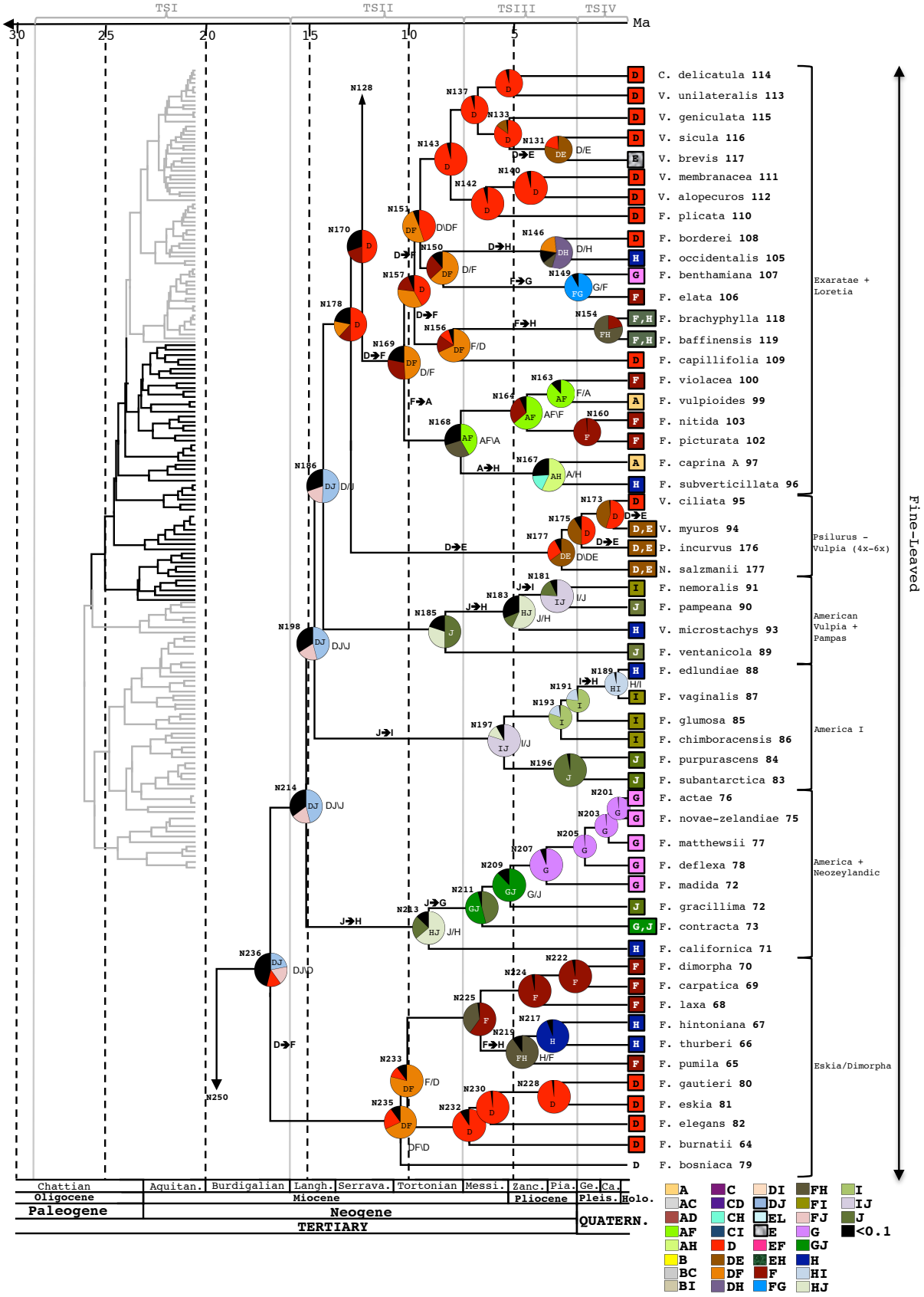


Figure 3. *(Continued)*

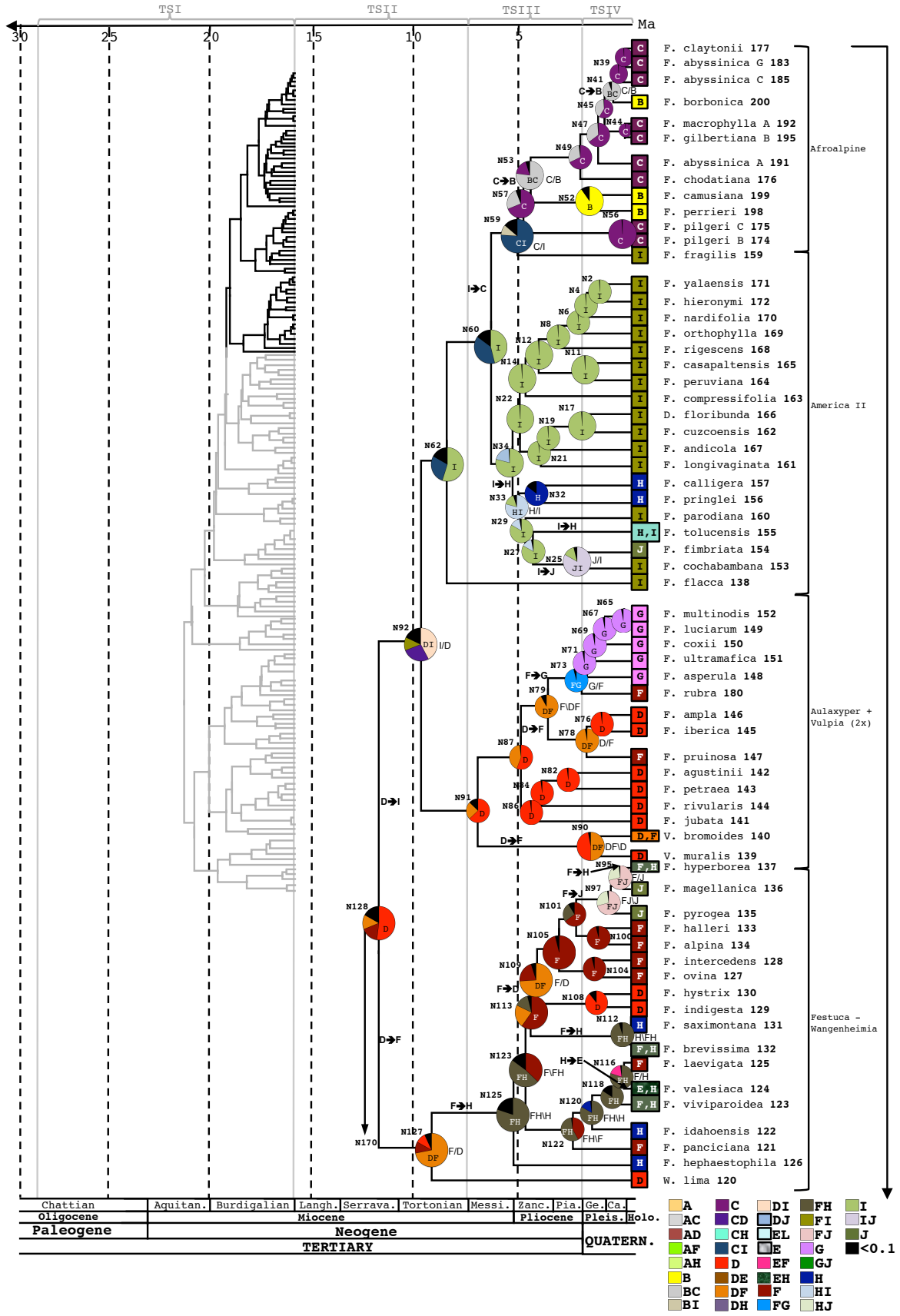


Figure 3. *(Continued)*

Figure 3. Biogeographic reconstruction of the Loliinae inferred from the DEC stratified M1 model based on 10 OAs. The tree presented corresponds to the BEAST MCC chronogram after pruning the 18 out-group taxa (Table S2; page 320). Pie charts at nodes represent ML relative marginal probabilities of ancestral ranges (A to J; Fig. 1), with their color legend at the bottom right-hand corner (black sections of pie charts mean that the sum of ancestral probabilities is lower than 0.1). The letters indicate splits with the ancestral ranges with the highest marginal probabilities. The upper time scale bar represents million years ago (Ma). Geological time scale is shown at the bottom. OAs assigned to each selected species (A to J; Fig. 1), their code numbers (Fig. 1; Table S1; page 306), and the major phyletic lineages are indicated to the right of the tree. Dispersal rate matrices are shown in Table S3 (page 321). The vertical grey lines indicate the boundaries between the four time slices (TSI to TSIV). The number of each node (N) and the inferred biogeographical dispersal (X → Y), vicariance (X/Y), and peripheral isolate speciation (X\XY) events are also indicated.

A complex scenario of successive peripheral speciations, vicariances and dispersals has to be invoked to explain the expansion of the broad- and fine-leaved Loliinae lineages during the Miocene, the Pliocene and the Quaternary (Figs. 2, 3). The split of the stem node of Loliinae was estimated to have occurred during the Late Oligocene (mean age 24.62 Ma) and that of the crown node in the Early Miocene (Node N346; 22.50 Ma) (Figs. 2A, 3A). Although the estimation of the geographic origin of the MRCA of Loliinae showed considerable uncertainty, the Holarctic Western Mediterranean and North America – Mesoamerica (DH) territories were inferred as the areas with the highest marginal probabilities (N346, Fig. 3A). A peripheral isolate speciation apparently led to the crown ancestor of the broad-leaved Loliinae in the Western Mediterranean (D) during the Early-Middle Miocene (N345; 18.91 Ma) and to the crown ancestor of the intermediate plus fine-leaved lineage in the disjunct Western Mediterranean and Southern South America (DJ) areas, after a dispersal from North America to South America and extinction in the former region in the Early-Middle Miocene (N250; 19.60 Ma) (Fig. 3A). Another peripheral isolate speciation was invoked to explain the Middle Miocene (17.50 Ma) origins of the respective crown ancestors of the fine-leaved Loliinae in the Western Mediterranean and Southern South America (N236; DJ) (Fig. 3B), and of the intermediate lineage in Eurosiberia and North America (N249; FH), the latter as a consequence of two consecutive dispersals, from the Western Mediterranean to Asia and from there to North America, via the Bering land bridge, together with extinction in the Mediterranean (Fig. 3A).

Within the broad-leaved lineage, a similar dispersal trend (Western Mediterranean → Eurosiberia → North America, and extinction in the Mediterranean) would explain the origin of the MRCA of the early-diverging Eurasian – South American + Central American – South African clade in regions F and H (Fig.1) during the Mid Miocene (N344; Langhian, 14.21 Ma). This was apparently followed by a vicariance and further dispersals that led to the origins of the respective crown ancestors of the Eurasian – South American group in most of the Palearctic

(regions D and F, Fig. 1) during the Serravallian (N334; 12.05 Ma) and of the Central America – South Africa group in regions H and A (Fig. 1) after a long distance dispersal (LDD) from America to South Africa in the Tortonian (N343; 8.39 Ma). Further vicariances (Mediterranean *F. calabrica*) and dispersals from Asia → Northern South America → Southern South America probably caused the origins of the successive ancestors of the Asian (*F. modesta*), Andean (*F. dichloclada*, *F. superba*) and Patagonian (*F. argentina*) species of the former clade from the Tortonian to the Pliocene (N332, N330, N328; 7.56 – 4.84 – 4.17 Ma); a main transatlantic vicariance likely led to the ancestors of the American (*F. amplissima*, *F. breviglumis*, *F. asperella*) and South African (*F. scabra*) lineages in the transition between the Pliocene and the Pleistocene (N339, N342; 2.35 – 1.5 Ma).

A LDD from the Western Mediterranean, the inferred ancestral area for the MRCA of the remaining broad-leaved lineages, to tropical Africa, followed by vicariance, likely originated the crown ancestor of the Tropical Africa – South Africa clade in the Tortonian (N306; 9.32 Ma). Several dispersals from Central to South Africa, from the Messinian to the Pleistocene, would explain the pan-African distributions of the separate *F. costata* and *F. africana* lineages, and a peripheral isolate speciation would explain the restricted distribution of the South African *F. killickii* (Fig. 3A). Another peripheral isolate speciation event, followed by extinction in South Africa and dispersal from tropical Africa to Madagascar, would explain the origin of the MRCA of *Pseudobromus* on the island at the beginning of the Pleistocene (N295; 2.12 Ma). Two remarkable transcontinental land bridge dispersions, from Eurosiberia to North America (F to H; N289 – N288; 13.07 – 6.70 Ma) and to Australia (F to G; N288 – N286; 6.70 – 4.16 Ma), followed by subsequent vicariances (N288 and N286), explain the respective origins of the North American *Leucopoa kingii* and the Australian *F. muelleri* as the respective sister lineages of the Eurosiberian *F. spectabilis* and *F. pulchella*, within the Leucopoa – Amphigenes – Drymanthele clade. A Pliocene LDD from the Western Mediterranean to tropical Africa (D to C; N265 – N259; 5.84 – 4.59 Ma), followed by vicariance (N259), presumably led to the disjunct distributions of the sister tropical African *F. simensis* and pan-Mediterranean and Eurasian *Lolium* lineages, within the Schedonorus – Lolium clade. A late Neogene series involving Asia → North America colonization, peripheral isolate speciation, North America →Asia recolonization, vicariance, and LDD from Western North America to Hawaii (N249 to N239), was reconstructed for the lineages of the intermediate clade (Figs. 3A, S4A; page 299).

Within the fine-leaved Loliinae an ample Mediterranean and Southern South American area was inferred as the most likely ancestral area for its crown ancestor (Fig. 3B). A Tortonian Mediterranean – Eurosiberian vicariance would explain the respective origins of the Eskia and Dimorpha ancestors. A Mid-Miocene LDD from Southern South America to North America was invoked to explain the distribution of the crown ancestor of the American + Neozeylandic clade in regions H and J (Fig. 1) (N213; 8.86 Ma), and a Tortonian vicariance to explain the sister relationship of the Western North American *F. californica* to the Southern Hemisphere subclade. Within this group, a transoceanic LDD from Patagonia to New Zealand (J to G; N213 – N211; 8.86 – 6.29 Ma) would explain the circumantarctic distribution of *F. contracta*, followed by an early Pliocene vicariance of the Fuegian *F. gracillima* and its sister Neozeylandic group of *F. madida*, *F. deflexa*, *F. matthewsii*, *F. novae-zelandiae* and *F. actae* (Fig. 3B).

A peripheral isolate speciation (N198) followed by dispersal from Southern South America to Northern South America would explain the widespread distribution of the American I ancestor on the continent in the early Pliocene (N197, 5.24 Ma). This was apparently followed by a vicariance, resulting in the distribution of the Patagonian *F. purpurascens* / *F. subantarctica* ancestor in J (N196) and that of the Andean *F. chimboracensis* – *F. glumosa* – *F. vaginalis* in I (N193) at the Pliocene – Pleistocene transition, followed by a recent Pleistocene dispersal from I to H to explain the North American distribution of *F. edlundiae* (Fig. 3B). A mid-Langhian Western Mediterranean / Southern South American vicariance was hypothesized for the ancestor of the core fine-leaved Loliineae (N186). The MRCA of the newly described American Vulpia + Pampas clade was inferred to have originated during the late Tortonian (N185; 7.83 Ma) in J, and a series of two Messinian and Pliocene colonizations from Southern South America to North America (J to H; N185 – N183; 7.83 – 4.70 Ma) and to Northern South America (J to I; N183 – N181; 4.70 – 3.29 Ma), plus two vicariances (N183 and N181), would explain the current distributions of *F. ventanicola* in the Pampas, *V. microstachys* in North and Central America, *F. pampeana* in the Pampas and *F. nemoralis* in the Andes. The next fine-leaved ancestors to diverge were inferred to have originated in the Western Mediterranean in the Serravalian (N178, N170), and separate LDDs to the Eastern Mediterranean and Eurosiberia were invoked to explain the respective widespread distributions of the ancestors of the Psilurus-Vulpia clade in regions D and E (Fig.1) in the late Pliocene (N177) and of the Exaratae-Loretia clade in regions D and F in the Tortonian (N169). The presence of austral Loliinae lineages within the predominantly boreal

Exaratae + Loretia clade was hypothesized to have occurred through four LDDs. Two subsequent LDDs, from Eurosiberia to South Africa (N169 – N168; 10.81 – 7.44 Ma) and from South Africa to North America (N168 – N167; 7.44 – 3.20 Ma), followed by two vicariance events (N167 and N163), would then have led to the current distribution of the sister species *F. subverticillata* (North American) – *F. caprina* (South African), and *F. vulpioides* (South African) – *F. violacea* (Eurosiberian). Another remarkable LDD from the Eurosiberian region to Australia (F to G; N150 – N149; 8.00 – 2.11 Ma), followed by a subsequent vicariance (N149), must be invoked to explain the existence of the current sister taxa *F. elata* (Eurosiberian) – *F. benthamiana* (South Australian) (Fig. 3B).

The more recently evolved fine-leaved Loliinae clade showed a high number of lineages with austral distributions (Fig. 3C); however, its MRCA was hypothetised to have originated in the Western Mediterranean in the Tortonian (N128; 11.5 Ma). A Tortonian LDD to Eurosiberia and a Messinian LDD to North America would explain the predominant Northern Hemisphere distributions of most of the Festuca – Wangenheimia clade lineages. Nonetheless, a recent Pleistocene transcontinental LDD from Eurosiberia to Southern South America (F to J; N101 – N97; 2.21 – 1.06 Ma) was invoked to explain the current Patagonian distributions of *F. magellanica* and *F. pyrogea*, nested within the Holarctic – Alpine *F. hyperborean* clade (Fig. 3C). A Tortonian transatlantic LDD from D to I likely resulted in the widespread Western Mediterranean – Southern South American distribution of the ancestor of the most recent red fescues plus allies group (N92, 9.3 Ma). A later vicariance was invoked to explain the Tortonian Andean and Messinian Western Mediterranean distributions of the respective ancestors of the American II – Afroalpine (N62) and Aulaxyper + Vulpia (2x) (N91) ancestors (Fig. 3C). A complex scenario of Pliocene continental range expansions (DF) followed by a recent Pleistocene LDD from Eurosiberia to Australia and New Zealand (F to G; N79 – N73; 2.95 – 2.33 Ma) and a subsequent vicariance (N73) would explain the current distributions of the Eurosiberian *F. rubra*, Australian *F. asperula* and New Zelander *F. ultramafica*, *F. coxii, F. luciarum*, and *F. multinodis* within the Aulaxyper + Vulpia (2x) clade. Two northward Pliocene colonizations from Northern South America to North America (I to H; N34 – N33; 5.52 – 4.84 Ma and N27; 4.18 – 0 Ma), followed by a vicariance event (N33), would explain the North American distribution of the *F. calligera* / *F. pringlei* ancestor within the American II clade, and the widespread North American – Andean distribution of *F. tolucensis*. Similarly, one southward Pliocene dispersal from the Andean region to Southern South America (I to J; N27 – N25; 4.18 – 2.49 Ma), followed by a vicariance (N25) would explain the Chacoan distribution

of *F. fimbriata* within the American II clade. A transoceanic LDD from Northern South America to tropical Africa (I to C; N60 – N59; 6.19 – 4.17 Ma) was inferred to have occurred in the Messinian-Pliocene transition; this would explain the widespread CI distribution of the Afroalpine + *F. fragilis* clade ancestor (N59, 4.8 Ma). A consequent vicariance likely originated the tropical African distribution of the Afroalpine ancestor (N59; 4.2 Ma). Two recent Pleistocene colonizations from tropical Africa to Madagascar - Mascarenes (C to B; N57 – N53; 4.17 – 3.76 Ma and N45 – N41; 1.20 – 0.81 Ma), followed by their respective vicariances (N53, N41), would explain the distributions of *F. perrieri* and *F. camusiana* in Madagascar and of *F. borbonica* in the Mascarenes (Reunion Island) within the Afroalpine clade, which is dominated by taxa with current tropical African continental distributions.

5. Discussion

5.1 Origin and radiation of the subtribe Loliinae

Our DEC estimates indicate that the ancestral ranges of Loliinae with the highest marginal probabilities were the Western Mediterranean and North American – Mesoamerican regions (Fig. 3). Our results are partially in agreement with the assessments of BOUCHENAK-KHELLADI *ET AL.* (2010), who pointed out that the most likely ancestral areas of the core Pooids were Eurasia and North America, and INDA *ET AL.* (2008), who indicated that the subtribe Loliinae diverged in the pan-Mediterranean – SW Asian region. Similarly, our estimations on the likely origin of the Loliinae ancestor around the Late Oligocene (Fig. 2) are partially in agreement with BOUCHENAK-KHELLADI *ET AL.* (2010), who dated in during the Oligocene. However, our estimations are considerably earlier than those found by previous assessments: for example, CHARMET *ET AL.* (1997) hypothesized that the Loliinae originated during the Late Miocene, and INDA *ET AL.* (2008) indicated that this subtribe diverged in the Middle Miocene. This difference might be attributed to the fossil calibration point used in this study, which was recently published by ZUCOL ET AL. (2010).

Most of the crown nodes of the main clades described in the broad- and fine-leaved Loliinae phylogeny originated during the Middle and Late Miocene (Figs. 2, 3). These results agree with other studies; for example, CHARMET *ET AL.* (1997) dated these divergences to the Tortonian age (Late Miocene), and INDA *ET AL.* (2008; 2014) estimated the broad- and fine-leaved radiations to have occurred between the Serravallian (Middle Miocene) and Tortonian ages. The radiation described here for the temperate Loliinae coincides with the global climatic cooling and drying hypothesized during the Middle and Late Miocene for most temperate

pooids (FLOWER AND KENNETT, 1994; ZACHOS *ET AL.*, 2001). Profound vegetation changes, such as the expansion of grasslands, and major radiations in large mammalian herbivores are associated with these key climatic events (CERLING *ET AL.*, 1997; JACOBS *ET AL.*, 1999; JANIS *ET AL.*, 2000; KELLOGG, 2001; STRÖMBERG, 2002; MACFADDEN, 2005). Consequetly, the Middle and Late Miocene might be considered a laboratory of sorts for strategies and adaptations for the broad- and fine-leaved fescues.

The Western Mediterranean region acted as the focal area of dispesion for the Loliinae from the Late Oligocene to the Middle Miocene (28.4 – 16.0 Ma; TSI; Figs. 3, 4). The complex geologic activity described in the Mediterranean region during this period (e.g. LO PRESTI AND OBERPRIELER, 2009; SALVO *ET AL.*, 2010) probably favored the isolation of the early ancestor of Loliinae, promoting its evolution and speciation and making the proto-Mediterranean Basin a primary center of diversification for the Loliinae lineages during this epoch. Three regions of the planet acted as foci of dispersion for the Loliinae from the Middle Miocene to the Late Miocene (16.1 – 7.2 Ma; TSII; Figs. 3, 4): (i) the Western Mediterranean Basin, (ii) Eurosiberia; and (iii) Southern South American. The gradual aridification of Middle Asia during the Oligocene-Miocene transition (TANG *ET AL.*, 2011) triggered the formation of dry steppe and semidesert landscapes with abundant herbaceous xerophytes, which could have facilitated the colonization and eventually the evolution and posterior spread of several ancestral Loliinae lineages from Eurosiberia to Africa, New Zealand, and North and Central America. The cold, dry climatic conditions of Southern South America combined with the rise of the Southern Patagonian Andes (BLISNIUK *ET AL.*, 2005; BARREDA AND PALAZZESI, 2007), likely allowed for the diversification and expansion of the temperate grasses in the Patagonian steppe during the Miocene.

Four regions of the planet acted as the main foci of dispersion for the Loliinae from the Late Miocene (Messinian) to the Pliocene (7.3 – 2.1 Ma; TSIII; Figs. 3, 4): (i) the Western Mediterranean Basin; (ii) Eurosiberia; (iii) Northern South America; and (iv) Southern South America. The Western Mediterranean continued to be one of the main centers of diversification for the Loliinae clades. The aridity crisis in the Mediterranean Basin allowed for the expansion of taxa from Africa, Asia and Europe. The posterior refilling of the Mediterranean Sea drove the subsequent isolation of the European and African continental plates (KRIJGSMAN *ET AL.*, 1999), which likely led to the speciation and posterior dispersion of the ancestor of the Loliinae. INDA *ET AL.* (2014) describe a similar scenario of diversification for the Schedonorus + Lolium lineage in the Mediterranean Basin during this time. The extreme climatic conditions,

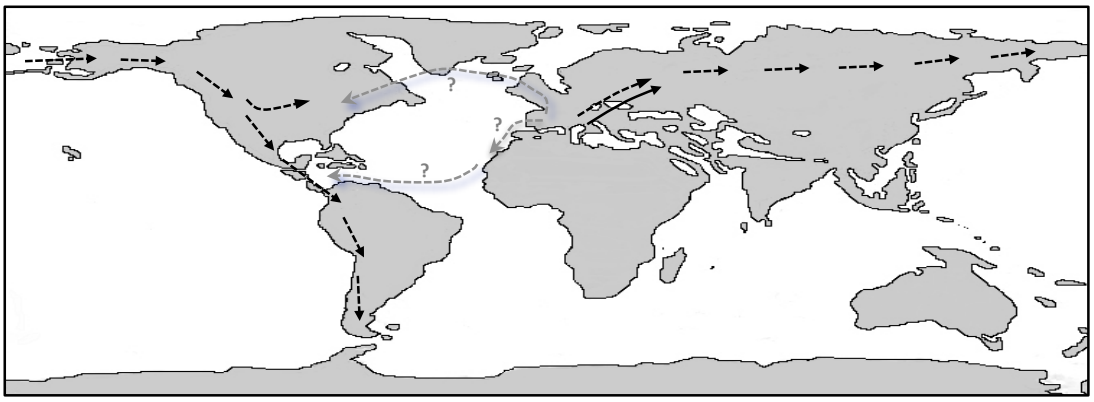
with global climatic cooling and the decrease in overall rainfall that characterized this period, together with the Andean orogeny (MORLEY, 2000), made it possible for new Loliinae lineages to spread in the arid and semiarid habitats in the Eurosiberian, Andean, and Patagonian regions.

The dispersion patterns estimated for the temperate Loliinae during the glacial and interglacial climatic oscillations are likely related to the existence of the treeless tundra and prairie-steppe areas that appeared repeatedly during this time in the Western Mediterranean Basin, Eurosiberia, and the Afroalpine belt (LANG, 1994; HEWITT, 2004; NIETO FELINER, 2011), regions which acted as the main foci of dispersion for the Loliinae during the Quaternary period (2.61 – 0 Ma; TSIV; Figs. 3, 4). The glaciations caused the sea level to drop, which increased the surface area that could be colonized and facilitated dispersions across places like the Bering Strait, the Sunda Shelf and the China Sea, which acted as continental bridges during the Ice Ages (LAFONTAINE AND WOOD, 1988; SANMARTÍN *ET AL.*, 2001; HEWITT, 2004). As a consequence, tundra-adapted species, such as the core fine-leaved Loliinae, were likely able to migrate once again between Siberia and Alaska and across the Andes using the new environments created during the Quaternary glaciations. Glaciations also allowed for the establishment of climatic refugia in parts of the Irano – Turanian – Himalayan region, Patagonia, South Africa, and Madagascar. The Loliinae likely took advantage of these refugia, from which species spread and colonized areas further North in post-glacial times (TABERLET *ET AL.*, 1998).

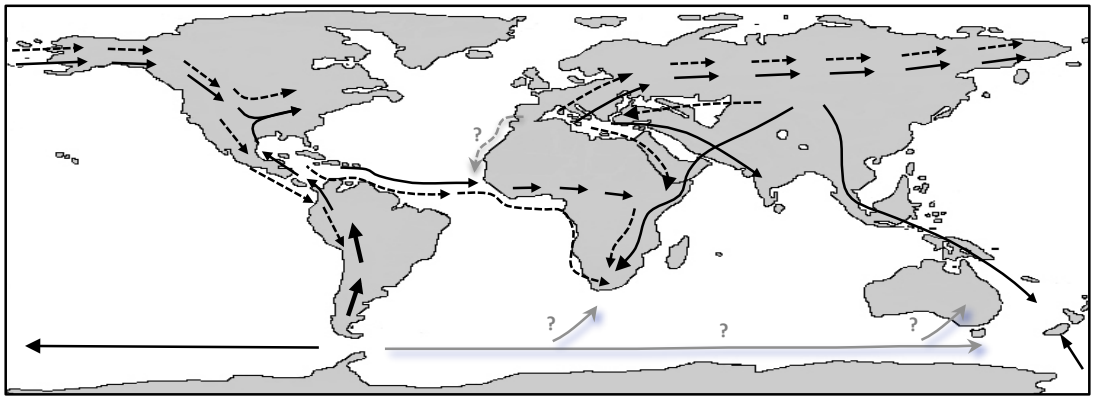
5.2 Colonization of the Southern Hemisphere

The biogeographical analyses conducted in this study allow us to depict a complex scenario of recurrent recolonizations and neocolocizations through continental range expansion, transcontinental landbridges, and trans-Atlantic pathways that generally occurred from the Northern to the Southern Hemisphere, but also took place within the Southern Hemisphere, and from South America and South Africa to North and Central America (Figs. 3, 4). The Southern Hemisphere regions may have acted, in a biogeographic sense, as islands, receiving migrants from the "mainland", in this case the Northern Hemisphere. Since the shortage of Loliinae fossils prevents us from drawing the exact migration route followed to reach the Southern Hemisphere, we have proposed several plausible paths from the main foci of dispersion to the colonized areas. Here, we focus on the migration events that allowed the ancestral lineages of the Loliinae to reach three main austral regions: Southern Africa; South America; and the South Pacific region.

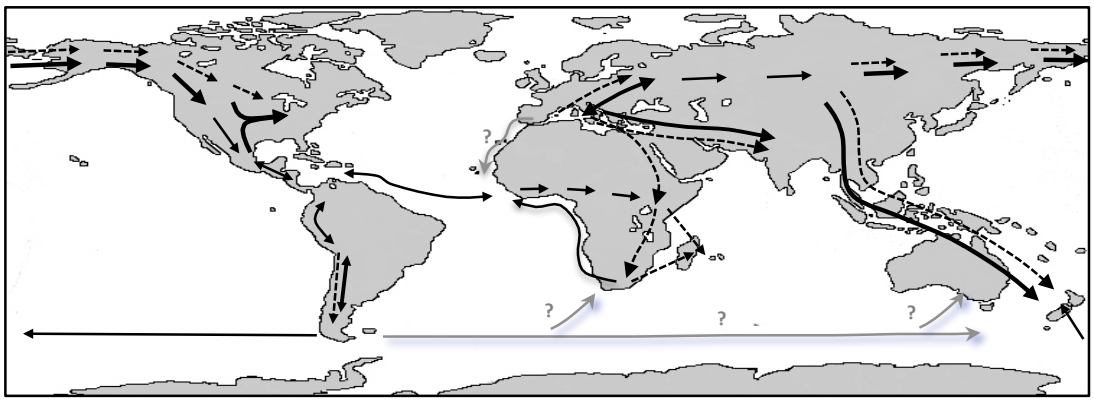
TSI Late Oligocene – Middle Miocene; 28.4 – 16.0 Ma



TSII Middle Miocene – Late Miocene; 16.1 – 7.2 Ma



TSIII Late Miocene (Messinian) – Pliocene; 7.3 – 2.6 Ma



TSIV Quaternary period; 2.61 – 0 Ma

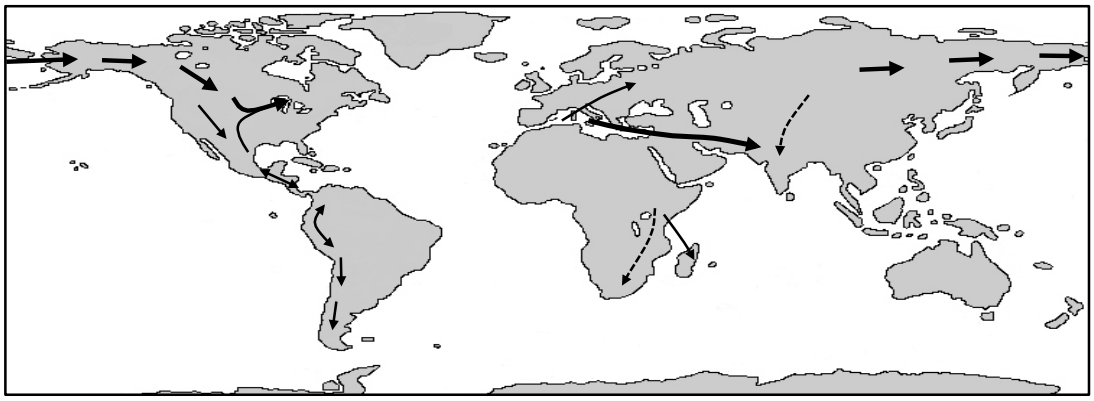


Figure 4. Palaeogeographical configuration of the five continents throughout the four Time Slices (TSI – TSIV) used in the temporally stratified Lagrange model. Adapted from http://cpgeosystems.com/index1.html. Black arrows represent the dispersal events according to Lagrange estimations. The arrows also indicate the direction and the most likely route hypothesized. The width of each arrow is proportional to the frequency of the dispersal. Grey arrows with question marks represent likely alternative migration paths. Continuous and discontinuous arrows represent the dispersion of fine- and broad-leaved ancestors, respectively.

Dispersions to Southern Africa

Eleven migrations of Loliinae lineages were hypothesized to have reached Africa between the Middle Miocene and the Quaternary period (Figs. 3, 4). The spread of the Loliinae observed in Africa may be explained by the slow and progressive aridification that began in Africa during the Oligocene and Early Miocene (c.f. LEVYNS, 1938; 1952; 1964; BLEEKER *ET AL.*, 2002; GALLEY AND LINDER, 2006; BARRES *ET AL.,* 2013), which eventually produced a transition from woodlands to grass-dominated savannas (JACOBS, 2004; ZACHOS *ET AL.*, 2001; 2008).

The existence of trans-Atlantic LDDs is the most likely hypothesis that explains the sister phylogenetic relationships observed between African and American taxa (Figs. 3, 4). Dispersal from America to Africa is not without precedent, recent biogeograpic reconstructions describe similar transatlantic routes from South America to Africa (SIMPSON *ET AL.*, 2006; BARRES *ET AL.*, 2013). Although it appears to have been possible for taxa of moderate to cool temperate climates to cross the Atlanctic using the Bering Strait during the Middle Miocene (TIFFNEY AND MANCHESTER, 2001; TIFFNEY, 2008), a shorter and more plausible route is from Mesoamerica to the Western African coast. Studies demonstrate that pelagic seabirds (e.g. *Calonectris diomedea*, *Puffinus puffinus* and *Puffinus gravis*) travel across the Atlantic Ocean using wind currents to glide (FELICÍSIMO *ET AL.*, 2008), so they are a possible vector. Once the ancestral taxa reached Africa, they could have dispersed from the Western African coast to South Africa using the Benguela Current, which was established during the Miocene (MARLOW *ET AL.,* 2000; DUPONT *ET AL.*, 2011). It is also possible that the Loliinae ancestral clades used the African Easterly Jet (cf. TOMPKINS *ET AL.*, 2005; WU *ET AL.*, 2009) to migrate from Western Africa to the tropical mountains of Eastern Africa, and from there to South Africa.

The existence of LDDs through Arabia is the most likely route postulated to explain the sister phylogenetic relationships observed between African and Eurosiberian taxa (Figs. 3, 4), as previously suggested by BARRES *ET AL.* (2013). The Arabian Peninsula formed an almost-continuous land corridor during the Miocene, extending from the proto-Iberian peninsula to Asia Minor (RÖGL, 1999; KRIJGSMAN, 2002; MEULENKAMP and SISSINGH, 2003). Alternatively,

dispersal events along the Western African coast using the Benguela Current cannot be ruled out.

Dispersions to South America

Twelve migrations of Loliinae lineages were hypothesized to have reached South America between the Late Oligocene and the Quaternary period (Figs. 3, 4). The dispersions hypothesized here through the Andes and Central America are dated around the same time as the closure of the Isthmus of Panama (BRIGGS, 1994; ANTONELLI AND SANMANTÍN, 2011), the rise of the Central and Northern Andes (VAN DER HAMMEN, 1974; SIMPSON, 1975; VAN DER HAMMEN AND CLEEF, 1986; ANTONELLI *ET AL.*, 2009; BELL *ET AL.*, 2012), and the clousure of the Western Andean Portal (TAYLOR, 1991; HOORN *ET AL.*, 1995). These geological events likely enabled the existence of suitable habitats for temperate Loliinae, and thus favored the colonization of the new Andean and Central American areas and the expansion of the Loliinae there. Similar migrations following the Central America and Northern South America route have been previously described by ANTONELLI AND SANMARTÍN (2011).

Dispersions to the South Pacific

Four migrations of Loliinae to the South Pacific region were estimated, one from Southern South America and three from Eurosiberia. All of them took place from the Middle Miocene to the Pliocene (Figs. 3, 4). The dispersion hypothesized from Southern South America to New Zealand and Australia is comparable to other westward migrations mediated by the East Wind Drift (c.f. WINKWORTH *ET AL.*, 2002, SANMARTÍN AND RONQUIS, 2004, SANMARTÍN *ET AL.*, 2008), for example in the genera *Poa* (SORENG, 1990) and *Notochloe* (LINDER *ET AL.*, 2013). The three LDDs estimated from the Eurosiberian region coincided with the collision between Australia and the Asian Plate in the Miocene, which led to the rise of high mountain chains along Malaysia, Northeastern Australia, and the Highlands of New Guinea, and the onset of a drying trend that enabled temperate taxa's colonization of the newly created montane habitats (SANMARTÍN, 2002; SANMARTÍN AND RONQUIS, 2004). Migration from Southeast Asia to Australia has been previously postulated in other angiosperms, there are several examples within temperate lineages that show junction between the Northern Hemisphere and Australia; for example, the Cardueae tribe (BARRES *ET AL.* 2013), and the grasses *Poa* and *Agrostis* (WINKWORTH *ET AL.*, 2002).

5.3 Patterns of dispersion: broad- vs. fine-leaved Loliinae

Our estimates indicate that the broad-leaved fescues had an unequivocal Western Mediterranean ancestor (N345), and their austral lineages are due to migrations from the Northern Hemisphere. Alternative scenarios that explain the presence of the broad-leaved members in the Southern Hemisphere as a result of secondary evolution of the broad-leaved syndrome within the fine-leaved clade, or the historical phenomenon of colonization and hybridization proposed by INDA *ET AL*. (2008), seem to be more the exception than the rule. The fine-leaved group, however, showed two independent origins (N236; Fig. 3), the Western Mediterrean and Southern South America. The Southern South American ancestor of the fine-leaved Loliinae eventually evolved into the American + Neozeylandic (N214; Fig. 3), the American I (N198), and the American Vulpia + Pampas groups (N186). On the other hand, a Mediterranean ancestor was inferred for the Intermediate clade (N250), the Eskia/Dimorpha group (N236), and the core fine-leaved Loliinae (N178), which encompasses the Psilurus – Vulpia (4x – 6x), Exaratae + Loretia and Festuca – Wangenheimia clades, plus the common ancestor of the Aulaxyper + Vulpia (2x) / Afroalpine + American clades.

Our results also suggest that the broad-leaved Loliinae have experienced more thriving dispersions to contiguous OAs than the fine-leaved fescues: e.g. from the Western Mediterranean to Eurosiberia, to the Irano-Turanian-Himalayan region, and to Africa. However, the fine-leaved fescues were able to migrate more successfully when they dispersed across the Atlantic Ocean or through regions such as the Bering Strait, the Andes or Indonesia (Fig. 4). This is likely due to the colonist characteristic of the fine-leaved Loliinae, which are, in general, well adapted to cool and dry ecological conditions. The colonist character of the fine-leaved Loliineae, together with their likely secondary origin in Southern South America, might be the reasons that explain why the fine-leaved lineages have spread more often in South America and the South Pacific, but not in Africa, where broad-leaved ancestors have migrated more frequently. Furthermore, this asymmetric dispersion pattern clarifies why the fine-leaved Loliinae have at least five lineages with most of their current species distributed in the Southern Hemisphere. Meanwhile, the broad-leaved Loliinae have only two lineages with most of their sampled species distributed in the Southern Hemisphere.

6. Literature cited

Ainscough, M.M., Barker, C.M., Stace, C.A., 1986. Natural hybrids between *Festuca* and species of *Vulpia* Section *Vulpia*. *Watsonia* 16, 143-151.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *BMC Bioinformatics* 20, 407-415.

Antonelli, A., Nylander, J.A.A., Persson, C., Sanmartín, I., 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences, U.S.A.* 106, 9749-9754.

Antonelli, A., Sanmartín, I., 2011. Mass extinction, gradual cooling, or rapid radiation? Resonctructing the spatiotemporal evolution of the ancient angiosperm genus *Hedyosmum* (Chloranthaceae) using empirical and simulated approaches. *Systematic Biology* 60, 596-615.

Baker, W.J., Couvreur, T.L.P., 2013. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography* 40**,** 274-285.

Barreda, V., Palazzesi, L., 2007. Patagonian vegetation turnovers during the Paleogene-Early Neogene: origin of arid-adapted floras. *Botanical Review* 73, 31-50.

Barres, L., Sanmartin, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R., 2013. Reconstructing the Evolution and biogeographic history of tribe Cardueae (Compositae). *American Journal of Botany* 100, 867-882.

Baum, D.A., Small, R.L., Wendel, J.F., 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Systematic Biology* 47, 181-207.

Bell, C.D., Kutschker, A., Arroyo, M.T.K., 2012. Phylogeny and diversification of Valerianaceae (Dipsacales) in the southern Andes. *Molecular Phylogenetics and Evolution* 63, 724-737.

Bergh, N.G., Linder, H.P., 2009. Cape diversification and repeated out-of-southern-Africa dispersal in paper daisies (Asteraceae–Gnaphalieae). *Molecular Phylogenetics and Evolution* 51, 5-18.

Blattner, F.R., 2006. Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist* 169, 603-614.

Bleeker, W., Franzke, A., Pollmann, K., Brown, A.H.D., Hurka, H., 2002. Phylogeny and biogeography of Southern Hemisphere high-mountain Cardamine species (Brassicaceae). *Australian Systematic Botany* 15, 575-581.

Blisniuk, P.M., Stern, L.A., Chamberlain, C.P., Idelman, B., Zeitler, K., 2005. Climatic and ecological changes during Miocene surface uplift in the southern Patagonian Andes. *Earth and Planetary Science Letters* 230, 125–142.

Briggs, J.C., 1994. The genesis of Central America: biology versus geophysics. *Global Ecology and Biogeography Letters* 4, 169-172.

Bouchenak-Khelladi, Y., Verboom, G.A., Hodkinson, T.R., Salamin, N., Francois, O., Chonghaile, G.N., Savolainen, V., 2009. The origins and diversification of C4 grasses and savanna-adapted ungulates. *Global Change Biology* 15, 2397-2417.

Bouchenak-Khelladi, Y., Verboom, G.A., Savolainen, V. and Hodkinson, T.R., 2010. Biogeography of the grasses (Poaceae): a phylogenetic approach to reveal evolutionary history in geographical space and geological time. *Botanical Journal of the Linnean Society* 162, 543-557.

Bouchenak-Khelladi, Y., Slingsby, J.A., Verboom, G.A., Bond, W.J., 2014. Diversification of C4 grasses (Poaceae) does not coincide with their ecological dominance. *American Journal of Botany* 101, 300-307.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T.G., Wu, C-H, Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST2: A software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10, e1003537.

Buckley, T.R., Arensburger, P., Simon, C., Chambers, G.K., 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. *Systematic Biology* 51, 4-18.

Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N., Sanmartín, I., 2010. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography* 38, 531-550.

Buerki, S., Jose, S., Yadav, S.R., Goldblatt, P., Manning, J.C., Forest, F., 2012. Contrasting biogeographic and diversification patterns in two Mediterranean-type ecosystems. *PLoS ONE* 7, e39377.

Burnham, R.J., Graham, A., 1999. The history of neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden* 86, 546-589.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe *Loliinae* and allies (Poeae, Pooideae) inferred from *ITS* and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., 2006. Phylogeny and evolution of *Festuca* L. and related genera of subtribe *Loliinae* (Poeae, Poaceae). In: A.K. Sharma, A. Sharma (Eds.), *Plant Genome. Biodiversity and Evolution*. Science Publishers, Enfield, NH, US, pp. 255-303.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British and European flora. Botanical Society of the British Isles*, London, pp. 45-70.

Catalan, P., Müller, J., 2012. *Festuca* L. In: A.M. Anton, F.O. Zuloaga (Eds.), *Flora de Argentina.* Vol. 3, tomo II. CONICET. Buenos Aires**,** pp. 219-250.

Charmet, G., Ravel, C., Balfourier, F., 1997. Phylogenetic analysis in the *Festuca – Lolium* complex using molecular markers and ITS rDNA. *Theoretical and Applied Genetics* 94, 1038-1046.

Cerling, T.E., Harris, J.M., MacFadden, B.J., Leacey, M.G., Quade, J., Eisenmann, V., Ehleringer, J.R., 1997. Global vegetation change through the Miocene/Pliocene boundary. *Nature 389*, 153-158.

Clayton, W.D., Renvoize, S.A., 1986. *Genera Gramineum: Grasses of the World. Kew Bulletin*. Series XIII, Royal Botanic Gardens, Kew.

Cohen, A.S., Soreghan, M.J., Scholz., C.A., 1993. Estimating the age of formation of lakes: An example from Lake Tanganyika, East African Rift system. *Geology* 21, 511-514.

Coleman, M., Liston, A., Kadereit, J.W., Abbott, R.J., 2003. Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert *Senecio* (Asteraceae). *American Journal of Botany* 90, 1446-1454.

Cook, L.G., Crisp, M.D., 2005. Directional asymmetry of long-distance dispersal and colonization could mislead reconstructions of biogeography. *Journal of Biogeography* 32, 741-754.

Crisp, M.D., Arroyo, M.T.K., Cook, L.G., Gandolfo, M.A., Jordan, G.J., McGlone, M.S., Weston, P.H., Westoby, M., Wilf, P., Linder, H.P., 2009. Phylogenetic biome conservatism on a global scale. *Nature* 458, 754-756.

Crisp, M.D., Trewick, S.A., Cook, L.G., 2011. Hypothesis testing in biogeography. *Trends in Ecology and Evolution* 26, 66-72.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest2: more models, new heuristic and parallel computing. *Nature Methods* 9, 772.

Donoghue, M.J., 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences USA* 105, 11549-11555.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4, 699-710.

Dupont, L.M., Linder, H.P., Rommerskirchen, F., Schefub, W., 2011. Climate-driven rampant speciation of the Cape flora. *Journal of Biogeography* 38, 1059-1068.

Emadzade, K., Hörandl, E., 2011. Northern Hemisphere origin, transoceanic dispersal, and diversification of *Ranunculeae* DC. (Ranunculaceae) in the Cenozoic. *Journal of Biogeography* 38, 517-530.

Felicísimo, A.M., Muñoz, J., González-Solis, J., 2008. Ocean surface winds drive dynamics of transoceanic aerial movements. *PLoS ONE* 3, e2928.

Flower, B.P., Kennett, J.P., 1994. The middle Miocene climatic transition: East Antarctic ice sheet development, deep ocean circulation and global carbon cycling. *Palaeogeography, Palaeoclimatology, Palaeoecology* 108, 537-555.

Galley, C., Linder, H.P., 2006. Geographical affinities of the Cape flora, South Africa. *Journal of Biogeography* 33, 236-250.

Gaut, B.S., Tredway, L.P., Kubik, C., Gaut, R.L., Meyer, W., 2000. Phylogenetic relationships and genetic diversity among members of the *Festuca-Lolium* complex (Poaceae) based on *ITS* sequence data. *Plant Systematics and Evolution* 224, 33-53.

Gehrke, B., Linder, H.P., 2009. The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of the Royal Society B* 276, 2657-2665.

Guindon, S., Gascuel, O., 2003. A simple, fasta and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* 52, 696-704.

Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B* 359, 183-195.

Hoorn, C., Guerrero, J., Sarmiento, G.A., Lorente, M.A., 1995. Andean tectonics as a cause of changing drainage patterns in Miocene northern South America. *Geology* 23, 237-240.

Hungerbuhler, D., Steinmann, M., Winkler, W., Seqard, D., Eguez, A., Peterson, D.E., Helg, U., Hammer, C., 2002. Neogene stratigraphy and Andean geodynamics of southern Ecuador. *Earth-Science Reviews* 57:75-124.

Inda, L.A., Segara-Moragues, J.G., Muller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae (Poaceae, Pooideae) grasses in the northern and southern hemispheres. *Molecular Phylogenetics and Evolution* 46, 932-957.

Inda, L.A., Sanmartín, I., Buerki, S., Catalan, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (Festuca subgenus Schedonorus and Lolium). *Journal of Biogeography* 41, 600-614.

Jacobs, B.F., Kingston, J.D., Jacobs, L.L., 1999. The origin of grass dominated ecosystems. *Annals of the Missouri Botanical Garden* 86, 590-644.

Jacobs, B.F., 2004. Paleobotanical studies from tropical Africa: relevance to the evolution of forest, woodland, and savanna biomes. *Philosophical Transactions of the Royal Society B* 359, 1573-1583.

Janis, C.M., Damuth, J., Theodor, J.M., 2000. Miocene ungulates and terrestrial primary productivity: where have all the browsers gone. *Proceedings of the National Academy of Sciences USA 97*, 7899-7904.

Jolly, D., Harrison, S.P., Damnati, B., Bonnefille, R., 1998. Simulated climate and biomes of Africa during the Late Quaternary: comparison with pollen and lake status data. *Quaternary Science Reviews* 17, 629-657.

Juchniewicz, K., 1975. Flora Kopalna Turowa Kolo Swietle Analizy Nablonkowey. *Prace Muzeum Ziemi* 24, 65-134.

Kellogg, E.A., 2001. Evolutionary history of the grasses. *Plant Physiology* 125, 1198-1205.

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S., 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* 400, 652-655.

Krijgsman, W., 2002. The Mediterranean: *mare nostrum* of earth sciences. *Earth and Planetary Science Letters* 205, 1-12.

Lafontaine, J.D., Wood, D.M., 1988. A zoogeographic analysis of the Noctuidae (Lepidoptera) of Beringia, and some inferences about past Beringian habitat. *Memoirs of the Entomological Society of Canada* 144, 109-123.

Lang, G., 1994. Quartäre Vegetationsgeschichte Europas. Fischer, Jena.

Lavin, M., Schrire, B.P., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C.E., Matos, A.B., Wojciechowski, M.F., 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, 1509-1522.

Levyns, M.R., 1938. Some evidence bearing on the past history of the Cape flora. *Transactions of the Royal Society of South Africa*. 26, 404-424.

Levyns, M.R., 1952. Clues to the past in the Cape flora of today. *South African Journal of Science* 49, 155-164.

Levyns, M.R., 1964. Presidential address, migrations and origin of the Cape flora. *Transactions of the Royal Society of South Africa* 37, 85-107.

Linder, H.P., Antonelli, A., Humphreys, A.M., Pirie, M.D., Wüest, R.O., 2013. What determines biogeographical ranges? Historical wanderings and ecological constraints in the danthonioid grasses. *Journal of Biogeography* 40, 821-834.

Lo Presti, R.M., Oberprieler, C., 2009. Evolutionary histoy, biogeography and eco-climatological differentiation of the genus *Anthemis* L. (Compositae, Anthemideae) in the circum-Mediterranean area. *Journal of Biogeography* 36, 1313-1332.

MacFadden, B.J., 2005. Terrestrial mammalian herbivore response to declining levels of atmospheric CO2 during the Cenozoic: Evidence from North American horses (family Equidae). In: J.R. Ehleringer, T.E. Cerling, M.D. Dearing (Eds.), *A history of atmospheric CO2 and its effects on plants, animals, and ecosystems*. Ecological Studies 177. Springer Verlag, Berlin. pp. 273–292.

Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762-1780.

Manchester, S.R., 2001. Update on the megafossil flora of Florissant, CO. In: E. Evanoff, K.M. Gregory-Wodzicki, K.R. Johnson (Eds.), *Fossil flora and stratigraphy of the Florissant Formation, Colorado*. Denver, pp. 137-162.

Marlow, J.R., Lange, C.B., Wefer, G. Rosell-Mele, A., 2000. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* 290, 2288-2291.

Meijer, P.T., Krijgsman, W., 2005. A quantitative analysis of the desiccation and re-filling of the Mediterranean during the Messinian Salinity Crisis, *Earth and Planetary Science Letters* 240, 510-520.

Metcalfe, C.R., 1960. *Anatomy of the Monocotyledons*. Vol. 1. Gramineae. Clarendon Press, Oxford.

Meulenkamp, J.E., Sissingh, W., 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the Afrian-Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology* 196, 209-228.

Milne, R.I., Abbott, R.J., 2002. The origin and evolution of Tertiary relict floras. *Advances in Botanical Research* 38, 281-314.

Milne, R.I., 2006. Northern Hemisphere plant disjunctions: a window on tertiary land bridges and climate change? *Annals of Botany* 98, 465-472.

Morley, R.J., 2000. Origin and evolution of tropical rain forests. New York: Wiley.

Muñoz, J., Felicisimo, A.M., Cabezas, F., Burgaz, A.R. and Martinez, I., 2004. Wind as a long-distance dispersal vehicle in the Southern Hemisphere. *Science* 304, 1144-1147.

Nieto Feliner, G., 2011. Southern European glacial refugia: a tale of tales. *Taxon* 60, 362-372.

Piñeiro, R., Fuertes-Aguilar, J., Draper-Munt, D., Nieto-Feliner, G., 2007. Ecology matters: Atlantic-Mediterranean disjunction in the sand-dune shrub *Armeria pungens* (Plumbaginaceae). *Molecular Ecology* 16, 2155-2171.

Pirie, M.D., Humphreys, A.M., Antonelli, A., Galley, C., Linder, H.P., 2013. Model uncertainty in ancestral area reconstruction: A parsimonious solution? *Taxon* 61, 652-664.

Popp, M., Mirre, V., Brochmann, C., 2011. A single Mid-Pleistocene long-distance dispersal by a bird can explain the extreme bipolar disjunction in crowberries (*Empetrum*). *Proceedings of the National Academy of Sciences, U.S.A.* 108, 6520-6525.

R Development Core Team, 2011. R: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Rabassa, J., Coronato, A., Bujalesky, G., Salemme, M., Roig, C., Meglioli, A., Heusser, C., Gordillo, S., Roig, F., Borromei, A., Quattrocchio, M., 2000. Quaternary of Tierra del Fuego, Southernmost South America: an updated review. *Quaternary International* 68, 217-240.

Rambaut, A., Drummond, A.J., 2007a. Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Rambaut A., Drummond A.J., 2007b. TreeAnnotator, version 1.5.4. Available from http://beast.bio.ed.ac.uk/Main_Page

Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299-2311.

Ree, R.H., Smith, S.A., 2008. A maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57, 4-14.

Ree, R.H., Sanmartín, I., 2009. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* 35, 428-449.

Renner, S.S., 2004. Plant dispersal across the tropical Atlantic by wind and sea currents. *International Journal of Plant Sciences* 165, S23-S33.

Rögl, F., 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geologica Carpathica* 50, 339-349.

Romaschenko, K., Garcia-Jacas, N., Peterson, P.M., Soreng, R.J., Vilatersana, R., Susanna, A. 2014. Miocene-Pliocene speciation, introgression, and migration of *Patis* and *Ptilagrostis* (Poaceae: Stipeae). *Molecular Phylogenetics and Evolution* 70, 244-259.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *BMC Bioinformatics* 19, 1572-1574.

Salvo, G., Ho, S.Y.W., Rosenbaum, G., Ree, R., Conti, R., 2010. Tracing the temporal and spatial origins of island endemics in the Mediterranean region: A case study from the citrus family (*Ruta*, L., Rutaceae). *Systematic Biology* 59, 705-722.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalizaed likelihood approach. *Molecular Biology and Evolution* 19, 101-109.

Sanmartín, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society* 73, 345-390.

Sanmartín, I., 2002. A paleogeographic history of the southern hemisphere. Available at Digital.CSIC: http://hdl.handle.net/10261/34831.

Sanmartín, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: post versus animal patterns. *Systematic Biology* 53, 216-243.

Sanmartín, I., van der Mark, P., Ronquist, F., 2008. Inferring dispersal: a Bayesian, phylogeny-based approach to island biogeography, with special reference to the Canary Islands. *Journal of Biogeography* 35, 428-449

Simpson, B., 1975. Pleistocene changes in the flora of the high tropical Andes. *Paleobiology* 1, 273-294.

Simpson, B.B., Todzia, C.A., 1990. Patterns and processes in the development of the high Andean flora. *American Journal of Botany* 77, 1419-1432.

Simpson, B., Larkin, L., Weeks, A., McDill, J., 2006. Phylogeny and biogeography of *Pomaria* (Caesalpinioideae: Leguminosae). *Systematic Botany* 31, 792-804.

Smith, S.A., Donoghue, M.J., 2010. Combining historical biogeography with niche modeling in the Caprifolium clade of Lonicera (Caprifoliaceae, Dipsacales). *Systematic Biology* 59, 322-341.

Soreng, R.J. 1990. Chloroplast-DNA phylogenetics and biogeography in a reticulating group: study in *Poa* (Poaceae). *American Journal of Botany* 77, 1383-1400.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Strömberg, C.A.E., 2002. The origin and spread of grass-dominated ecosystems in the late Tertiary of North America: preliminary results concerning the evolution of hypsodonty. *Palaeogeography, Palaeoclimatology, Palaeoecology* 177, 59-75.

Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. *Annual review of Earth and Planetary Sciences* 39, 517-544.

Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony and other methods. Version 4b10. Sunderland (MA): Sinauer Associates.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7, 453-464.

Tang, Z.H., Ding, Z.L., White, P.D., Dong, X.X., Ji, J.L., Jiang, H.C., Luo, P., Wang, X., 2011. Late Cenozoic central Asian drying inferred from a palynological record from the northern Tian Shan. *Earth and Planetary Science Letters* 302, 439-447.

Taylor, D.W., 1991. Paleobiogeographic relationships of Andean angiosperms of Cretaceous to Pliocene age. *Palaeogeography, Palaeoclimatology, Palaeoecology* 88, 69-84.

Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* 51, 689-702.

Tiffney, B.H., Manchester, S.R., 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences* 162, S3-S17.

Tiffney, B.H., 2008. Phylogeography, fossils, and Northern Hemisphere biogeography: the role of physiological uniformitarianism. *Annals of the Missouri Botanical Garden* 95, 135-143.

Thomasson, J.R., 1985. Miocene fossil grasses: possible adaptation in reproductive bracts (lemma and palea). *Annals of the Missouri Botanical Garden* 72, 843-851.

Tompkins, A.M., Diongue-Niang, A., Parker, D.J., Thorncroft, C.D., 2005. The African easterly jet in the ECMWF integrated forecast system: 4D-Var analysis. *Quarterly Journal of the Royal Meteorological Society* 131, 2861-2885.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear ITS sequences. *Systematic Botany* 27, 241-251.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and *trn*L-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Van der Hammen, T., 1974. The Pleistocene changes of vegetation and climate in tropical South America. *Journal of Biogeography* 1, 3-26.

Van der Hammen, T., Cleef, A.M., 1986. *High Altitude Tropical Biogeography*. Oxford University Press, New York, NY, USA.

Wang, W., Chen Z-D, Liu Y., Li, R-Q, Li J-H., 2007. Phylogenetic and biogeographic diversification of Berberidaceae in the Northern Hemisphere. *Systematic Botany* 32, 731-742.

Watson, L., Dallwitz, M.J., 1992 onwards. The grass genera of the world: descriptions, illustrations, identification, and information retrieval; including synonyms, morphology, anatomy, physiology, phytochemistry, cytology, classification, pathogens, world and local distribution, and references. Version: 12th August 2014. http://delta-intkey.com

Waters, J.M., Dijkstra, L.H., Wallis, G.P., 2000. Biogeography of a southern hemisphere freshwater fish: How important is marine dispersal? *Molecular Ecology* 9, 1815-1821.

Webb, S.D., Opdyke, N.D., 1995. Global climatic influence on Cenozoic land mammal faunas. In: S. Stanley, J. Kennett (Eds.), *Effects of Past Global Change on Life, Geophysical Monographs*. National Academy of Sciences Press, Washington D.C., pp. 184-208.

Wiens, J.J., Donoghue, M.J., 2004. Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution* 19, 639-644.

Winkworth, R.C., Wagstaff, S.J., Glenny, D., Lockhart, P.J., 2002. Plant dispersal N.E.W.S. from New Zealand. *Trends in Ecology and Evolution* 17, 514-520.

Wu, M.C., Schubert, S.D., Suarez, M.J., Koster, R.D., Pegion, P.J., 2009. African Easrely Jet: structure and maintenance. *American Meteorological Society*, 4459-4480.

Yalden, D.W., Largen, M.J., 1992. The endemic mammals of Ethiopia. *Mammal Review* 22, 115-150.

Zachos, J.C., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686-693.

Zachos, J.C., Dickens, G.R., Zeebe, R.E., 2008. An Early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279-283.

Zucol, A.F., Brea, M., Bellosi, E., 2010. Phytolith studies in Gran Barranca (central Patagonia, Argentina): Middle and Late Eocene. In: R.H. Madden, A.A. Carlini, M.G. Vucetich, R.F. Kay (Eds.), *The paleontology of Gran Barranca: Evolution and environmental change through the Middle Cenozoic of Patagonia*. Cambridge University Press, pp. 317-340.

7. Appendix

7.1 Taxon sampling

We followed the distribution areas of the Loliinae and their close allies as traditionally defined by major taxonomic and geographical studies (e.g. TUTIN, 1980; SORENG *ET AL.*, 2003; CHEN *ET AL.*, 2006). However, due to the relative lack of sampling in some poorly studied areas, several biogeographical regions could not be represented in the present study (e.g. India, Indonesia, and New Guinea). This study includes a total of 179 species of Loliinae and 215 accessions (Table S1; page 306): 154 species belonging to *Festuca*, 12 to *Vulpia*, 3 to *Lolium*, and 10 to various *Festuca*-related genera (*Castellia, Ctenopsis*, *Dielsiochloa* [the so-called avenoid Dielsiochloa; cf. QUINTANAR *ET AL.*, 2007], *Leucopoa*, *Micropyropsis, Micropyrum*, *Narduroides, Pseudobromus, Psilurus, Wangenheimia*). Of the 215 Loliinae accessions (excluding the out-groups), 103 were collected from the Southern Hemisphere. Deliberate effort was made to cover different regions of Africa, including 53 accessions. Twenty-seven samples were from tropical Africa: 7 from Ethiopia, 8 from Kenya, 4 from Tanzania, and 8 from Uganda; twenty-seven samples were from South Africa and Madagascar: 4 from Madagascar, 1 from Reunion Island, and 22 from South Africa. Thirty-six South American samples were used in this research. Twenty were from the Andes: 4 from Bolivia, 2 from Colombia, 4 from Ecuador, 7 from Peru, and 3 from Venezuela; and sixteen were from the Southern Cone: 15 from Argentina and 1 from South Georgia Island. Finally, the South Pacific region was represented by thirteen species: 9 from New Zealand and 4 from Southwestern Australia. Several other traditionally poorly represented areas were sampled in this study. Nine

accessions were collected from the Middle East and Asia (Afghanistan, China, Japan, Mongolia, Russia, and Cyprus); 7 accessions from Mexico; 2 accessions from Hawaii; and 1 accession from Alaska. The remaining 93 accessions were from Macaronesia (Azores, Canary Islands, and Madeira Islands), the Western and Central Mediterranean, Eurosiberia (from the North of Spain to Siberia), and North America (Canada, the United States, and Greenland).

7.2 Procedures for DNA isolation, amplification, sequentiation and phylogenetic inference

The 253 new sequences included for the first time in this study were obtained from herbarium vouchers or freshly collected from the field and dried in silica gel. The sequences of the remaining analyzed species were retrieved for previously studied taxa. The Macherey-Nagel Nucleospin Plant II kit was used for DNA extraction from finely ground leaf samples, following the manufacturer's instructions. The nuclear (ITS) and two plastid (*trn*TL and *trn*LF) regions were amplified and sequenced using the primers and procedures from CATALAN *ET AL.* (2004) and QUINTANAR *ET AL.* (2007). After amplification, the PCR products were purified with Exosap-IT following the manufacturer's instructions. Purified PCR products were sequenced by Macrogen, Inc. (Seoul, South Korea) with the same primers that were used for PCR amplification. Complementary strands were assembled and verified using Sequencher version 4.2.2 (Gene Codes Corporation 1991 – 2004). The obtained sequences were aligned and adjusted manually using MacClade 4 version 4.08 OS X (MADDISON AND MADDISON, 2008).

Three datasets were obtained for the three different regions that were sequenced: nuclear ITS and plastid *trn*TL and *trn*LF. These three datasets were analyzed separately and combined using parsimony and Bayesian phylogenetic approaches. Bayesian analyses were run with two million generations for the nuclear marker and one million generations for the plastid DNA dataset and the combined (ITS+*trn*TLF) dataset using the Monte Carlo Markov Chain (MCMC) algorithm. Trees were sampled every 1,000 generations, and 25% of the generations were discarded as burn-in once stability in the likelihood values was attained. MP analysis was based on heuristic searches of 10,000 random-order-entry trees, with TBR branch swapping and saving no more than 10 trees of length equal to or shorter than 10 per replicate. The most parsimonious trees were used to compute the respective strict consensus trees. Branch support was estimated through 1,000 bootstrap replicates (FELSENSTEIN, 1985) using the TBR-M (Tree Bisection Reconstruction swapping, MULPARS OFF) strategy of DEBRY AND OLMSTEAD (2000) as a method to reduce computational time. Clades with bootstrap support values (BS) of

75-100% (*PAUP 4b.10-86-macosx*) or Posterior Probability support values (PPS) of 90-100% (MrBayes 3.1.2) were considered moderately to strongly supported.

7.3 Biogeographic reconstruction

Although several previous analyses found similarities between the biogeographical scenarios drawn by DEC and Bayes-DIVA (e.g. BUERKI *ET AL.*, 2010; INDA *ET AL.*, 2014), we choose to use DEC because: (i) Bayes-DIVA reconstructed "impossible" dispersal events that did not conform to paleogeographical history; (ii) DEC is more flexible in estimating the range inheritance by peripheral isolation speciation, which gives more alternative scenarios and reduces the frequency of terminal dispersals, but increases the uncertainty in ancestral geographical ranges; and (iii) DIVA would never infer extinction events unless explicit geographical constraints were used to modify the original cost assignment rules. The DEC model explicitly incorporates extinction into the model as a parameter in the Q matrix that produces range contraction, and in this respect it is more realistic than DIVA (REE *ET AL.*, 2005; SANMARTÍN, 2007; NYLANDER *ET AL.*, 2008). DEC is a continuous-time stochastic model for geographical range evolution in discrete areas, with maximum likelihood (ML) parameters estimated from dispersal rates between areas (range expansion) and local extinction within areas (range contraction), according to exponential rate parameters. The probability of dispersal and extinction events along any branch of the phylogeny is dependent on evolutionary time. The model allows dispersal rates between areas to be constrained based on paleogeographical information and assumptions about the long-distance dispersal ability of species (REE AND SMITH, 2008).

Two alternative dispersal models were tested in DEC (M0-M1). The unconstrained model (M0) assumes that the spatial arrangement of plants has no effect on biogeographical patterns of evolution, thus assuming that dispersal rates were equal between all geographical areas (i.e. setting all possible transitions to equal values). The M1 model assumes that the geological evolution of the Earth during the last 23 Ma influenced the connectivity of biogeographical areas, and therefore there were patterns of dispersal/vicariance among ancestral areas.

For the stratified M1 model, the phylogeny was divided into four time-slices, which reflect the main paleogeographical changes during the history of Loliinae. The oldest time-slice comprises the Late Oligocene to Middle Miocene, between 28.4 and 16.0 Ma (TSI). A general cooling began on the entire planet at the start of the Miocene. A slow and progressive

aridification commenced on the African continent, specifically in Northern Africa where dry areas appeared (ZACHOS *ET AL.*, 2001; ZACHOS *ET AL.*, 2008). This event may have facilitated migrations of arid-adapted plants between Southwestern Africa and the Horn of Africa. Furthermore, the Benguela current through the Southwestern African coast also appeared during the Miocene (MARLOW *ET AL.*, 2000; DUPONT *ET AL.*, 2011). In the Mediterranean Basin, the progressive accretion of the microplates located between the Parathethys and Tethys Seas caused the formation of a more-or-less continuous landmass during this period, extending from the proto-Iberian peninsula to Asia Minor (RÖGL, 1999; MEULENKAMP AND SISSINGH, 2003). This new geological configuration allowed biotic exchange between the Eastern and Western proto-Mediterranean Basin, resulting in the circum-Mediterranean distribution of many terrestrial organisms (cf. STEININGER *ET AL.*, 1985; OOSTERBROEK AND ARNTZEN, 1992; PALMER AND CAMBEFORT, 2000; SANMARTÍN, 2003; OBERPRIELER, 2005; MANSION *ET AL.*, 2008; YESSON *ET AL.*, 2009). In South America, the Northern Andes of Colombia and Ecuador underwent a progressive rise during this period (SIMPSON, 1990). In the Southern Patagonian Andes there was a transition from a closed forest to Patagonian steppe with the concomitant expansion of C4 grasses (BLISNIUK *ET AL.*, 2005).

The second time-slice captured the Late and Middle Miocene periods, from 16.1 to 7.2 Ma (TSII). During this period there was a transition from woodlands to grass-dominated savannas in Western and Eastern Africa (JACOBS, 2004). The East African Rift System started to develop, with initiation of the Central Tanganyika Basin, at about 12 to 10 Ma (COHEN *ET AL.*, 1993). In the Eastern Mediterranean, the rotation of the African plate led to the closure of the connection between the Tethys and the Indian Ocean by causing the collision of the Arabian microplate with Eurasia in the Middle Miocene. Disruption of this seaway had a profound effect on oceanic circulation patterns and caused a major drop in global climate (MEIJER AND KRIJGSMAN, 2005) and an increase in aridification in the Mediterranean region (SALVO *ET AL.*, 2010). In South America, the Western Andean Portal (WAP) started to rise during this period, and marine incursions ended, in connection with the rise of the Eastern Cordilleras of the Central and Northern Andes (TAYLOR, 1991; HOORN *ET AL.*, 1995; HUNGERBUHLER *ET AL.*, 2002). The closure of the WAP enabled the dispersion of plant lineages from the Northern to the Central Andes (ANTONELLI *ET AL.*, 2009).

The third time-slice covers the Messinian and the Pliocene periods, from 7.3 to 2.6 Ma (TSIII). During this period the formation of the high mountains in tropical and Southern Africa began (GEHRKE AND LINDER, 2009). In addition, the Red Sea (Sudan) and the Gulf of Aden

were formed, marking the tectonic boundary between the African and Arabian plates (SANMARTÍN, 2003). In the Mediterranean area, the Messinian Salinity Crisis at the end of the Miocene (KRIJGSMAN, 2002) established a land bridge between Northwest Africa and the Iberian Peninsula, allowing circum-Mediterranean lineages to reach the Northern Tethys platform (SANMARTÍN, 2003; OBERPRIELER, 2005). In Central America, the Panama Island Arc connected North and South America. The connection of the two continents clearly served as a major conduit for plant and animal migration in the recent past (BURNHAM AND GRAHAM, 1999). There was a series of collisions between Australia and the Asian Plate in the Late Miocene – Pliocene, which resulted in the rise of mountain chains in Southeast Asia, Northern Australia, New Guinea, and New Zealand (SANMARTÍN AND RONQUIS, 2004).

The fourth time-slice covers the Quaternary period, from 2.61 to 0 Ma (TSIV). The Pleistocene Glaciations produced arid areas in the Northern Hemisphere, which were associated with a substantial reduction in forest cover (JOLLY *ET AL.*, 1998). However, the Mediterranean region acted as a climatic refuge from which species spread and colonized Central and Northern Europe in post-glacial times (TABERLET *ET AL.*, 1998). Secondary contacts among species and the potential for hybridization and polyploid speciation arose during the warmest interglacial phases (BARTON AND HEWIIT, 1989). The afro-alpine vegetation belt was 1000-1500 m lower during cooler periods in the Pleistocene, making the connection of the alpine habitats more likely (YALDEN AND LARGEN, 1992). In South America, the Patagonian Ice Sheet covered the whole Southern third of Chile and adjacent areas of Argentina during this period. Although the Western coast of Patagonia was largely glaciated, several authors have pointed out the possible existence of an ice-free refuge for some plant species (RABASSA *ET AL.*, 2000).

The OAs chosen in the biogeographical scenarios drawn by DEC were (Fig. 1): A – Southern Africa: coast and Eastern region and Cape of Good Hope. B – Madagascar and Mascarenes (Reunion Island). C – High mountains of tropical Africa: Ethiopia, Kenya, Tanzania, Cameroon, Equatorial Guinea, Uganda and Zimbabwe. D1 – Western Mediterranean. D2 – Macaronesian archipelagos: Azores, Canary Islands, and Madeira Islands. E – Irano-Turanian-Himalayan region (ITH): includes all the uplifted landmasses and mountain ranges that resulted from the collision of the Indian and Arabian plates against Eurasia in the Eocene and Middle Miocene, respectively (SANMARTÍN, 2003; WANG *ET AL.*, 2009). The region extends from the Zagros Mountains and the Caucasus in the West through the Iranian plateau, the Hindu Kush, Tien Shan and Kunlun Shan Mountains to the Altay Mountains in the Northeast, and the Tibetan Plateau and the Himalayan mountain range in the Southeast. F – Eurasia:

Central and Northern Eurasian territories, part of the Southern European territories, Central and East Asia: China, Japan and South Korea. G – New Zealand and Southwestern Australia: New South Wales. H1 – North America: the United States, Canada, Greenland and Mesoamerica: from Mexico to Panama. H2 – Hawaiian archipelago. I – Northern Andes: Bolivia, Colombia, Ecuador, Peru, Venezuela, Northern Chile and Northwestern Argentina. J – Southern Andes and Southern South America: rest of Argentina and Chile (Central Andes, Pampas, Patagonia).

7.4 Supplementary literature

Antonelli, A., Nylander, J.A.A., Persson, C., Sanmartín, I., 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences, U.S.A.* 106, 9749-9754.

Barton, N.H., Hewitt, G.M., 1989. Adaptation, speciation and hybrid zones. *Nature* 341, 497-503.

Blisniuk, P.M., Stern, L.A., Chamberlain, C.P., Idelman, B., Zeitler, K., 2005. Climatic and ecological changes during Miocene surface uplift in the southern Patagonian Andes. *Earth and Planetary Science Letters* 230, 125-142.

Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N., Sanmartín, I., 2010. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. *Journal of Biogeography* 38, 531-550.

Burnham, R.J., Graham, A., 1999. The history of neotropical vegetation: new developments and status. *Annals of the Missouri Botanical Garden* 86, 546-589.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe *Loliinae* and allies (Poeae, Pooideae) inferred from *ITS* and *trn*L-F sequences. *Molecular Phylogenetics and Evolution* 31, 517-541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: J. Bailey, R.G. Ellis (Eds.), *Current taxonomic research on the British and European flora.* Botanical Society of the British Isles, London, pp. 45-70.

Chen, S., Li, D., Zhu, G., *et al*. (24 co-authors), 2006. Poaceae. In: Z.Y. Wu, P.H. Raven, D.Y. Hong (Eds.), *Flora of China 22*. Beijing and Missouri Botanical Garden Press, St. Louis. Science Press, pp. 1-733.

Cohen, A.S., Soreghan, M.J., Scholz., C.A., 1993. Estimating the age of formation of lakes: An example from Lake Tanganyika, East African Rift system. *Geology* 21, 511-514.

DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. *Systematic Biology* 49, 171-179.

Dupont, L.M., Linder, H.P., Rommerskirchen, F. Schefub, E., 2011. Climate-driven rampant speciation of the Cape flora. *Journal of Biogeography* 38: 1059-1068.

Farris, J.S., 1989. The retention index and the rescaled consistency index. *Cladistics* 5, 417-419.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783-791.

Gehrke, B., Linder, H.P., 2009. The scramble for Africa: pan-temperate elements on the African high mountains. *Proceedings of the Royal Society B* 276, 2657-2665.

Hoorn, C., Guerrero, J., Sarmiento, G.A., Lorente, M.A., 1995. Andean tectonics as a cause of changing drainage patterns in Miocene northern South America. *Geology* 23, 237-240.

Hungerbuhler, D., Steinmann, M., Winkler, W., Seqard, D., Eguez, A., Peterson, D.E., Helg, U., Hammer, C., 2002. Neogene stratigraphy and Andean geodynamics of southern Ecuador. *Earth-Science Reviews* 57:75-124.

Inda, L.A., Sanmartín, I., Buerki, S., Catalan, P., 2014. Mediterranean origin and Miocene–Holocene Old World diversification of meadow fescues and ryegrasses (*Festuca* subgenus Schedonorus and *Lolium*). *Journal of Biogeography* 41, 600-614.

Jacobs, B.F., 2004. Paleobotanical studies from tropical Africa: relevance to the evolution of forest, woodland, and savanna biomes. *Phylosophical Transactions of the Royal Society B* 359, 1573-1583.

Jolly, D., Harrison, S.P., Damnati, B., Bonnefille, R., 1998. Simulated climate and biomes of Africa during the Late Quaternary: comparison with pollen and lake status data. *Quaternary Science Reviews* 17, 629-657.

Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of Anurans. *Systematic Zoology* 18, 1-32.

Krijgsman, W., 2002. The Mediterranean: *mare nostrum* of earth sciences. *Earth and Planetary Science Letters* 205, 1-12.

Maddison, D.R., Maddison, W.P., 2008. *MacClade 4.08 version for OS X*. Sinauer Associates, Inc.

Mansion, G., Rosenbaum, G., Schoenenberger, N., Bacchetta, G., Rosselló, J.A., Conti, E., 2008. Phylogenetic analysis informed by Geological history supports multiple, sequential invasions of the Mediterranean basin by the angiosperm family Araceae. *Systematic Biology* 57, 269-285.

Marlow, J.R., Lange, C.B., Wefer, G. Rosell-Mele, A., 2000. Upwelling intensification as part of the Pliocene-Pleistocene climate transition. *Science* 290, 2288-2291.

Meijer, P.T., Krijgsman, W., 2005. A quantitative analysis of the desiccation and re-filling of the Mediterranean during the Messinian Salinity Crisis, *Earth and Planetary Science Letters* 240, 510-520.

Meulenkamp, J.E., Sissingh, W., 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the Afrian-Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology* 196, 209-228.

Nylander, J.A.A., Olsson, O., Alström, P., Sanmartín, I., 2008. Accounting for phylogenetic uncertainity in biogeography: a Baysian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology* 57, 257-268.

Oberprieler, C., 2005. Temporal and spatial diversification of circum-Mediterranean Compositae-Anthemideae. *Taxon* 54, 951-966.

Oosterbroek, P., Arntzen, J.W., 1992. Area-cladograms of circum-Mediterranean taxa in relation to Mediterranean palaeogeography. *Journal of Biogeography* 19, 3-20.

Palmer, M., Cambefort, Y., 2000. Evidence for reticulate palaeogeography: beetle diversity linked to connection-disjunction cycles of the Gibraltar Strait. *Journal of Biogeography* 27, 403-416.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid *trn*T-F and nuclear *ITS* sequences. *American Journal of Botan* 94, 1554-1596.

Rabassa, J., Coronato, A., Bujalesky, G., Salemme, M., Roig, C., Meglioli, A., Heusser, C., Gordillo, S., Roig, F., Borromei, A., Quattrocchio, M., 2000. Quaternary of Tierra del Fuego, Southernmost South America: an updated review. *Quaternary International* 68-71, 217-240.

Ree, R.H., Moore, B.R., Webb, C.O., Donoghue, M.J., 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution* 59, 2299-2311.

Ree, R.H., Smith, S.A., 2008. A maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57, 4-14.

Rögl, F., 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geologica Carpathica* 50, 339-349.

Salvo, G., Ho, S.Y.W., Rosenbaum, G., Ree, R., Conti, R., 2010. Tracing the temporal and spatial origins of island endemics in the Mediterranean region: A case study from the citrus family (*Ruta*, L., Rutaceae). *Systematic Biology* 59, 705-722.

Sanmartín, I., 2003. Dispersal vs. vicariance in the Mediterranean: historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography* 30, 1883-1897.

Sanmartín, I., Ronquist, F., 2004. Southern hemisphere biogeography inferred by event-based models: post versus animal patterns. *Systematic Biology* 53, 216-243.

Sanmartín, I., 2007. Event-based biogeography: integrating patterns, processes, and time. In: M.C. Ebach, R. Tangney (Eds.*), Biogeography in a changing world*. CRC Press, Boca Raton, Fl., pp. 135-159.

Simpson, B.B., Todzia, C.A., 1990. Patterns and processes in the development of the high Andean flora. *American Journal of Botany* 77, 1419-1432.

Soreng, R.J., Peterson, P.M., David, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O. (Eds.). 2003. *Catalogue of New World grasses (Poaceae): IV. Subfamily Pooideae*. Contributions from the U.S. National Herbarium. pp. 48, 1-730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. *Kew Bulletin* 62, 425-454.

Steininger, F.F., Rabeder, G., Rögl, F., 1985. Land mammal distribution in the Mediterranean Neogene: a consequence of geokinematic and climatic events. In: D.J. Stanley, F.C. Wezel (Eds.), *Geological evolution of the Mediterranean basin*. New York: Springer-Verlag. pp. 559-571.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7, 453-464.

Taylor, D.W., 1991. Paleobiogeographic relationships of Andean angiosperms of Cretaceous to Pliocene age. *Palaeogeography, Palaeoclimatology, Palaeoecology* 88, 69-84.

Torrecilla, P., López-Rodríguez, J.A., Stancik, D., Catalan, P., 2003. Systematics of *Festuca* L. sects. *Eskia* Willk., *Pseudatropis* Kriv., *Amphigenes* (Janka) Tzvel., *Pseudoscariosa* Kriv. and *Scariosae* Hack. based on analysis of morphological characters and DNA sequences. *Plant Systematics and Evolution* 239, 113-139.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and trnL-F sequences. *Annals of the Missouri Botanical Garden* 91, 124-158.

Tutin, T.G., Heywood, V.H., Burges, N.A., Moore, D.M., Valentine, D.H., Walters, S.M., Webb, D.A. (Eds.), 1980. *Flora Europaea. Vol. 5, Poaceae*. Cambridge University Press, Cambridge, UK.

Wang, Y.J., Susanna, A., von Raab-Straube, E., Milne, R., Liu, J.Q., 2009. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) trigered by uplifts of the Qinghai-Tibetan Plateau. *Biological Journal of the Linnean Society* 97, 983-903.

Yalden, D.W., Largen, M.J., 1992. The endemic mammals of Ethiopia. *Mammal Review* 22, 115-150.

Yesson C., Toomey, N.H., Culham, A., 2009. Cyclamen: time, sea and speciation biogeography using a temporally calibrated phylogeny. *Journal of Biogeography* 36, 1234-1252.

Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686-693.

Zachos, J.C., Dickens, G.R., Zeebe, R.E., 2008. An Early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279-283.

Figure S1. Bayesian inference of the Loliinae subtribe based on the nuclear (ITS) marker. Numbers above and below branches correspond to posterior probability and bootstrap support, respectively. Species highlighted in bold and with an asterisk were removed in the concatenated nuclear and plastid data matrix. Scale bar indicates number of expected substitutions per site. Sections and assemblages are represented on the right side of the figure.

7.5. Supplementary figures

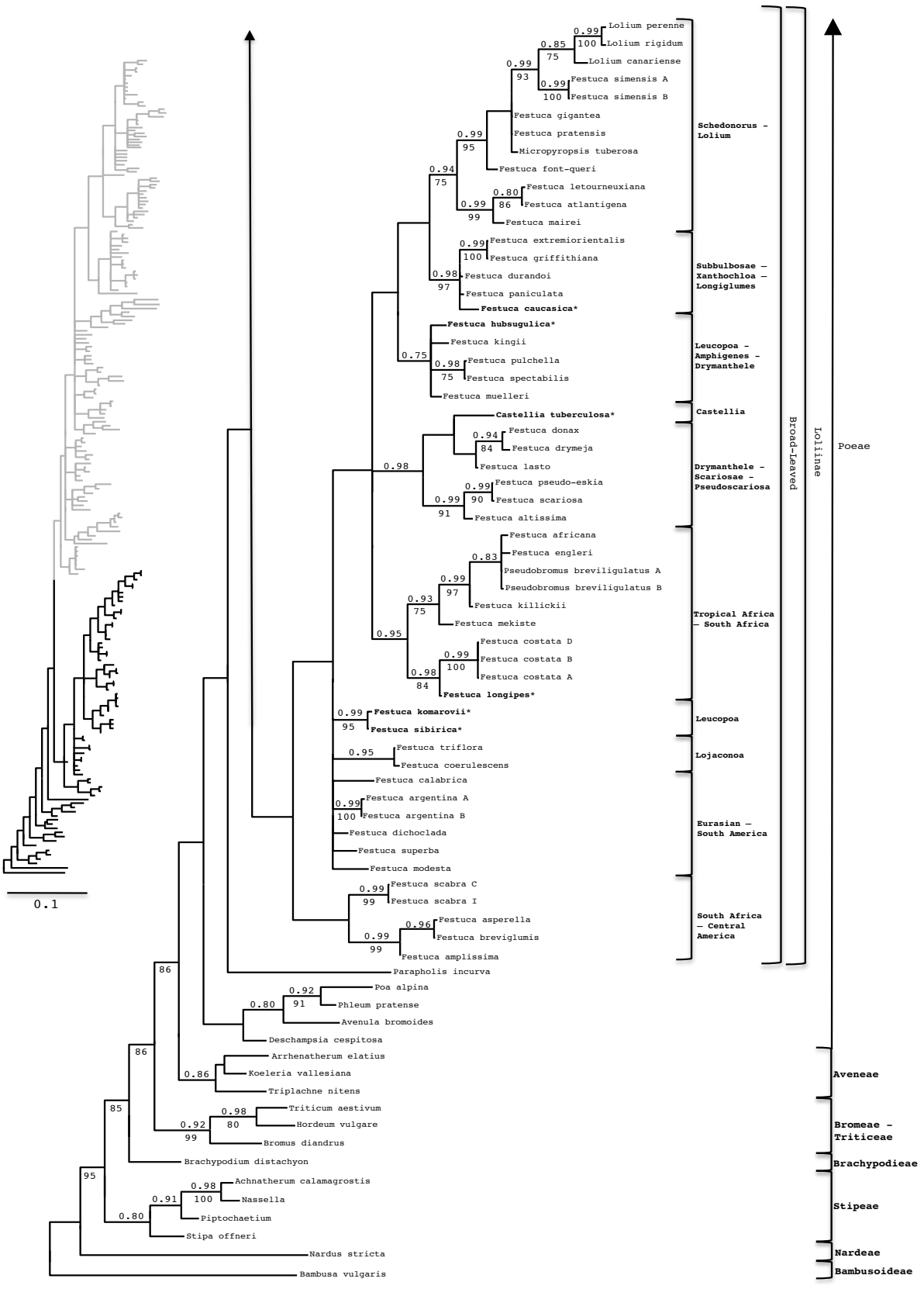


Figure S1. *(Continued)*

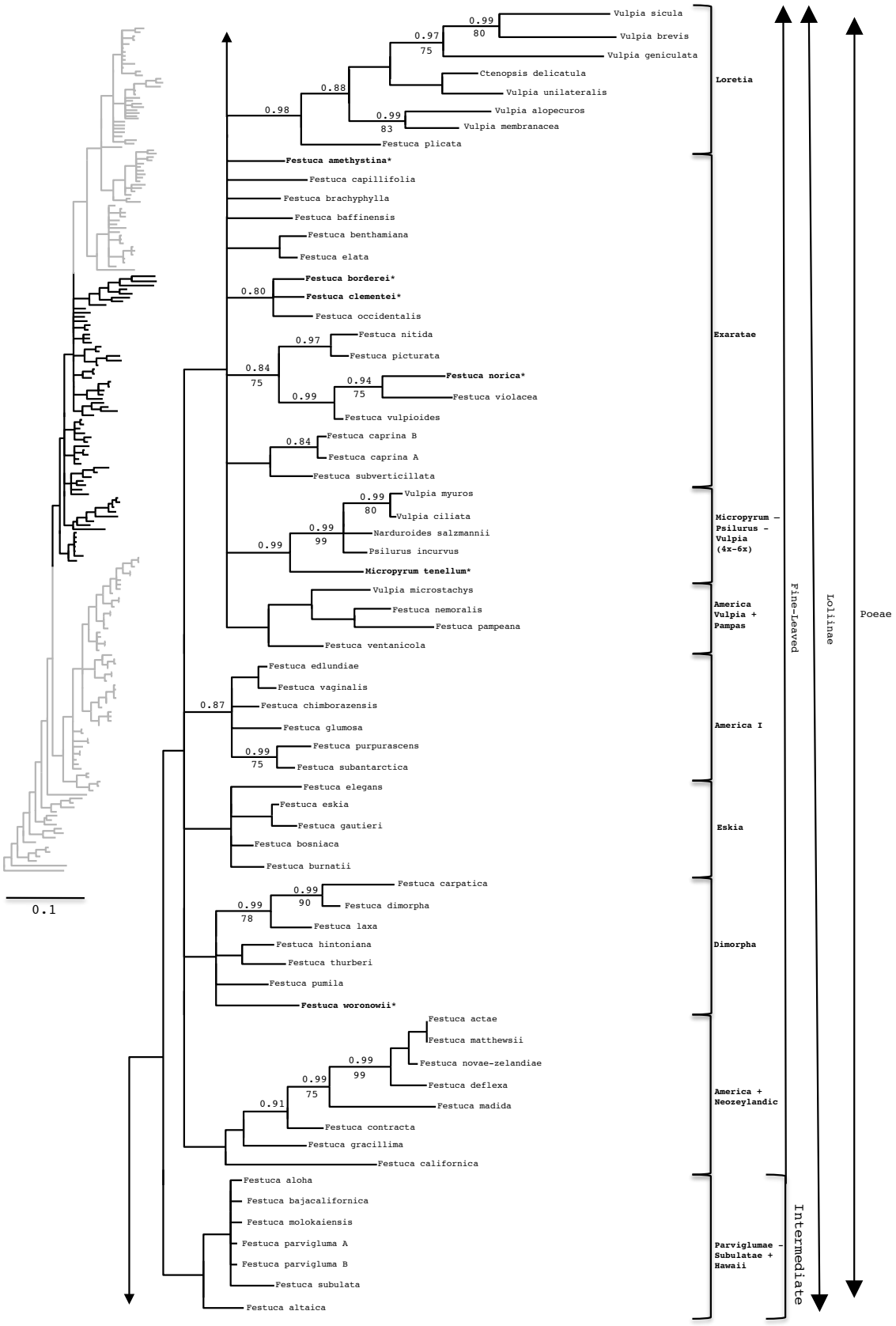


Figure S1. *(Continued)*

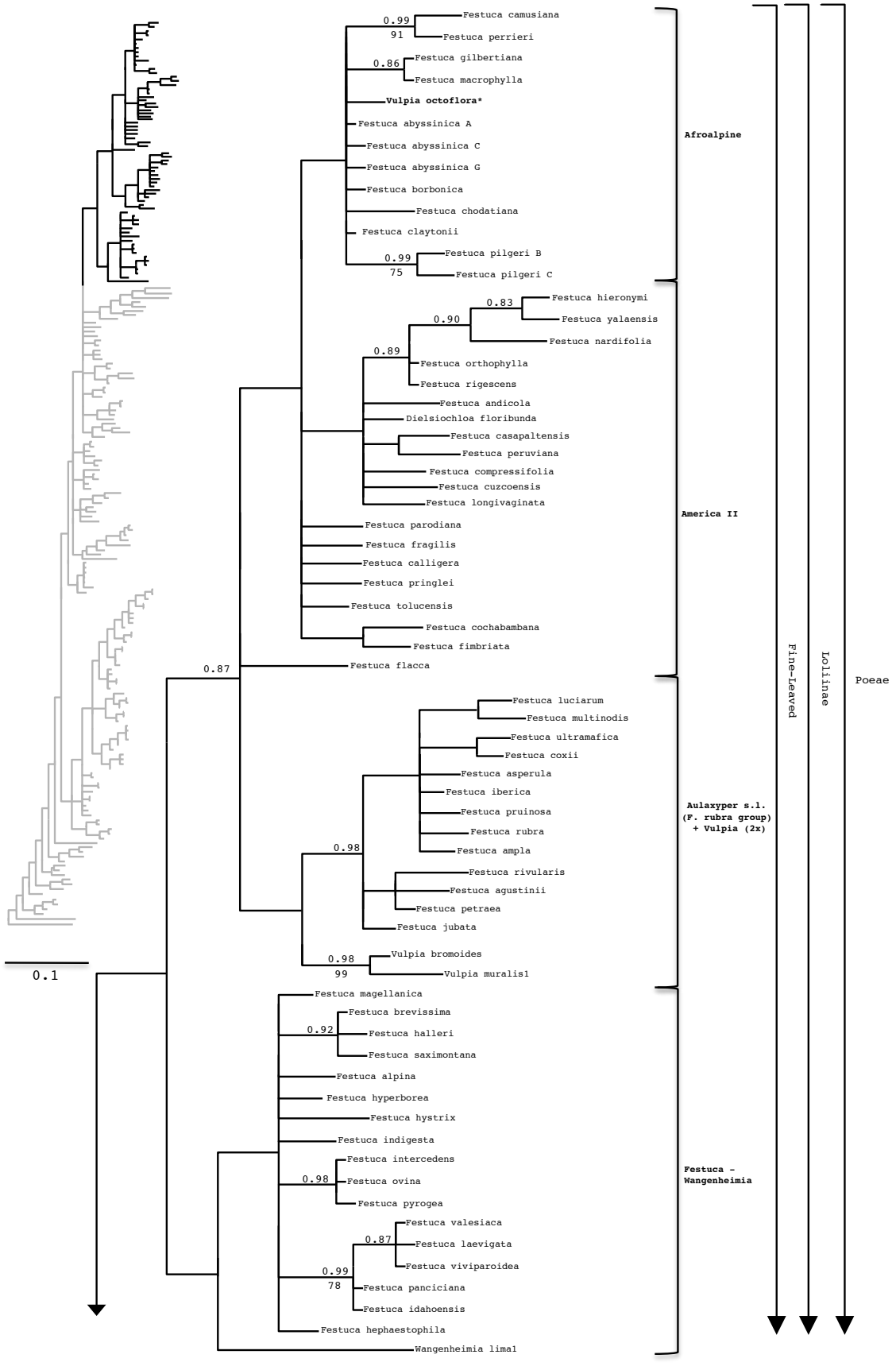


Figure S1. *(Continued)*

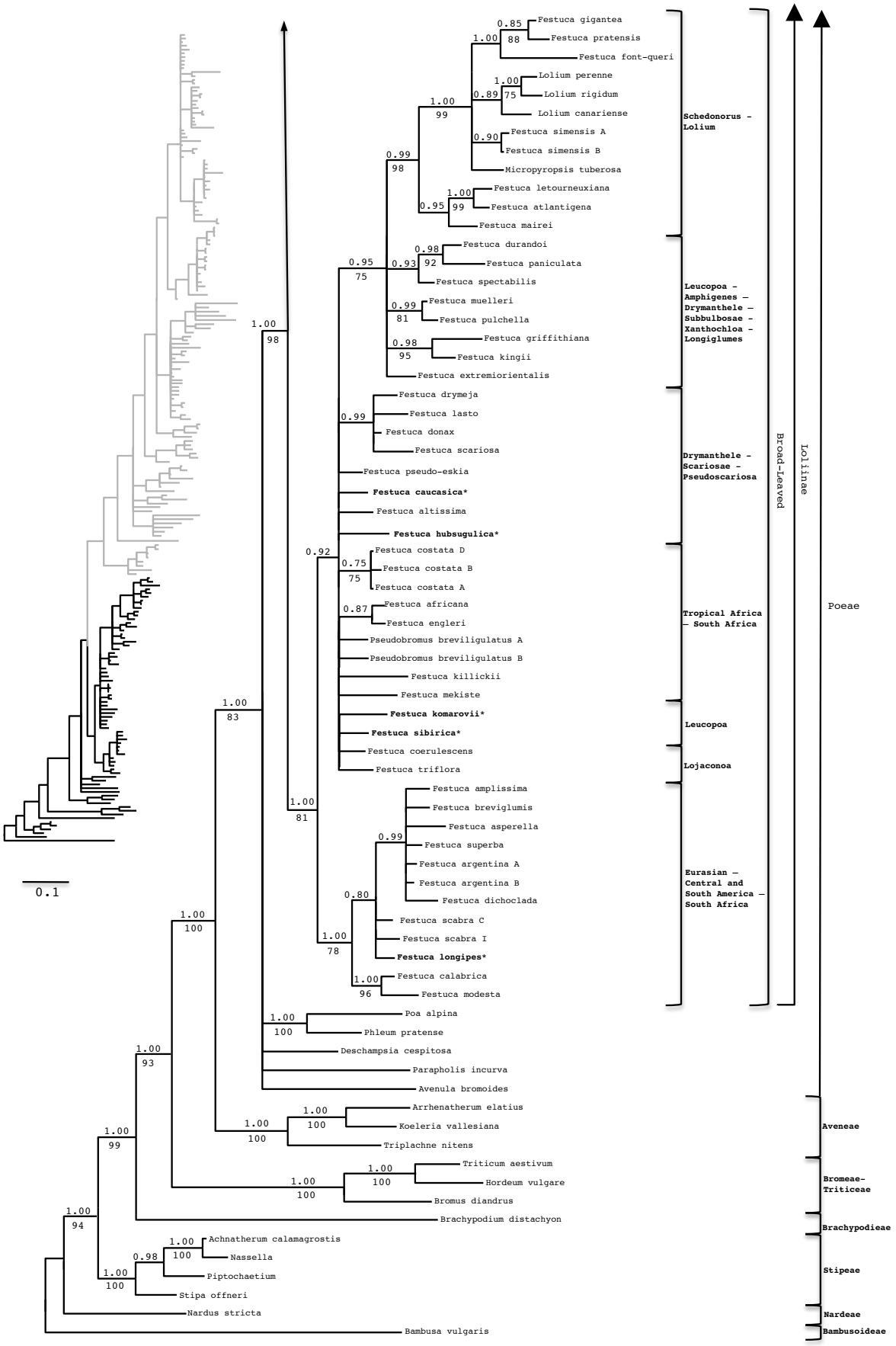


Figure S2. *(Continued)*

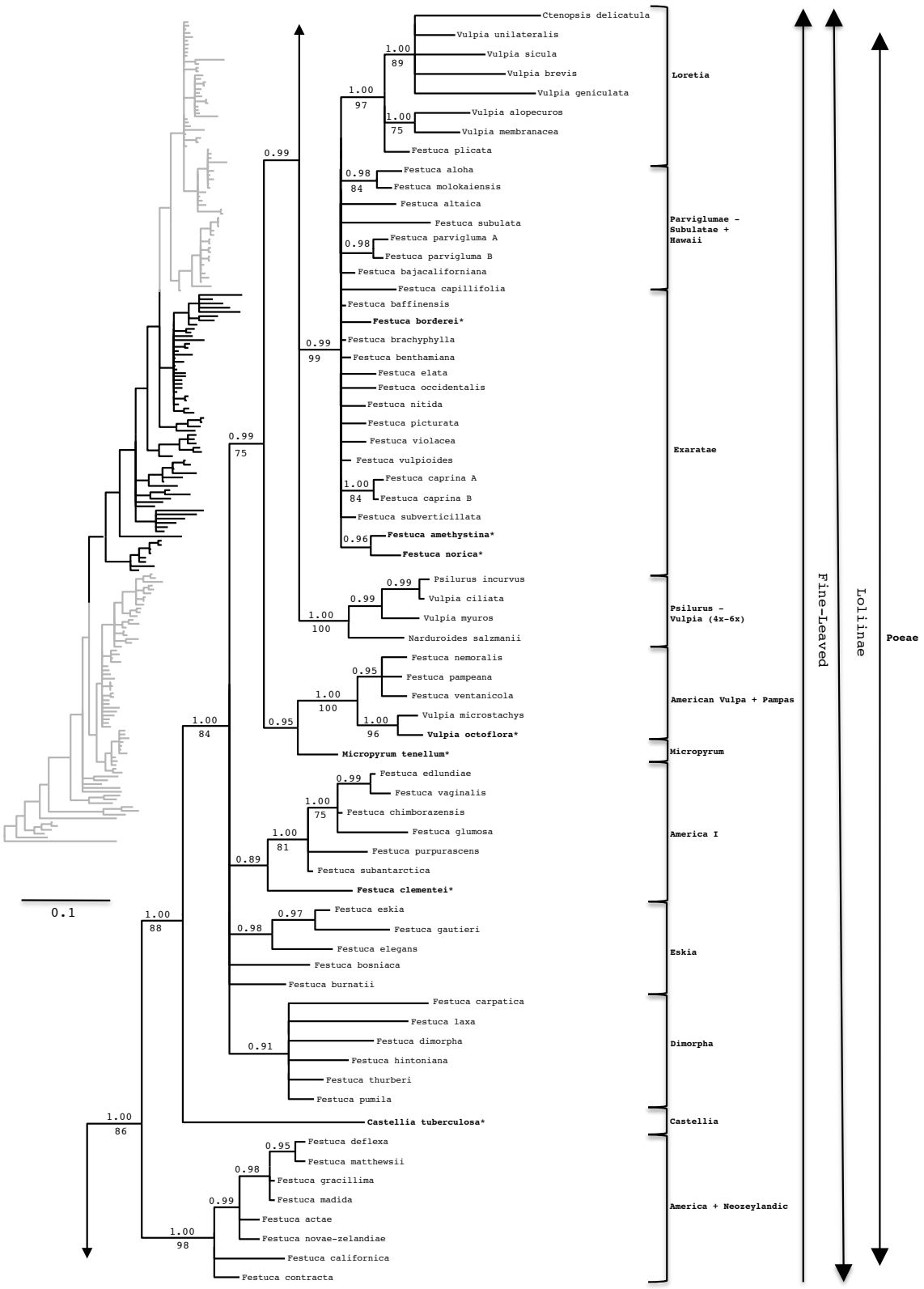


Figure S2. *(Continued)*

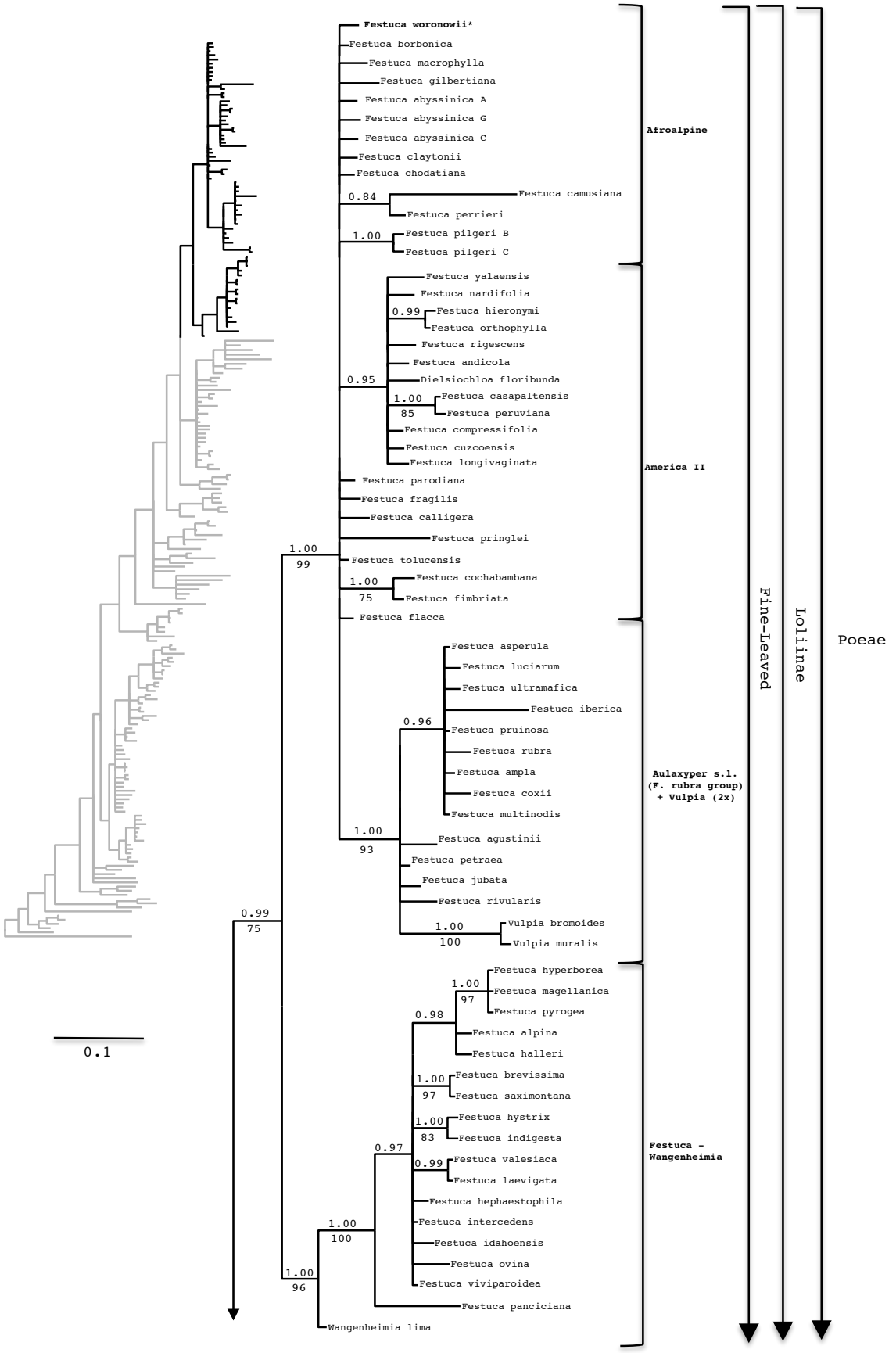


Figure S2. *(Continued)*

Figure S2. Bayesian inference of the Loliinae subtribe based on the plastid (*trn*TL-LF) region. Numbers above and below branches correspond to posterior probability and bootstrap support, respectively. Species highlighted in bold and with an asterisk were removed in the concatenated nuclear and plastid data matrix. Scale bar indicates number of expected substitutions per site. Sections and assemblages are represented on the right side of the figure.

Figure S3. Bayesian phylogenetic reconstruction of the Loliinae based on the combined nuclear (ITS) and plastid (*trn*TL-LF) markers. Species number follows those in table S1 (page 306). Scale bar indicates number of expected substitutions per site. Numbers above and below the branches indicate the posterior probability (PP) and the bootstrap (BS) support values, respectively. Sections and assemblages are represented on the right side of the figure.

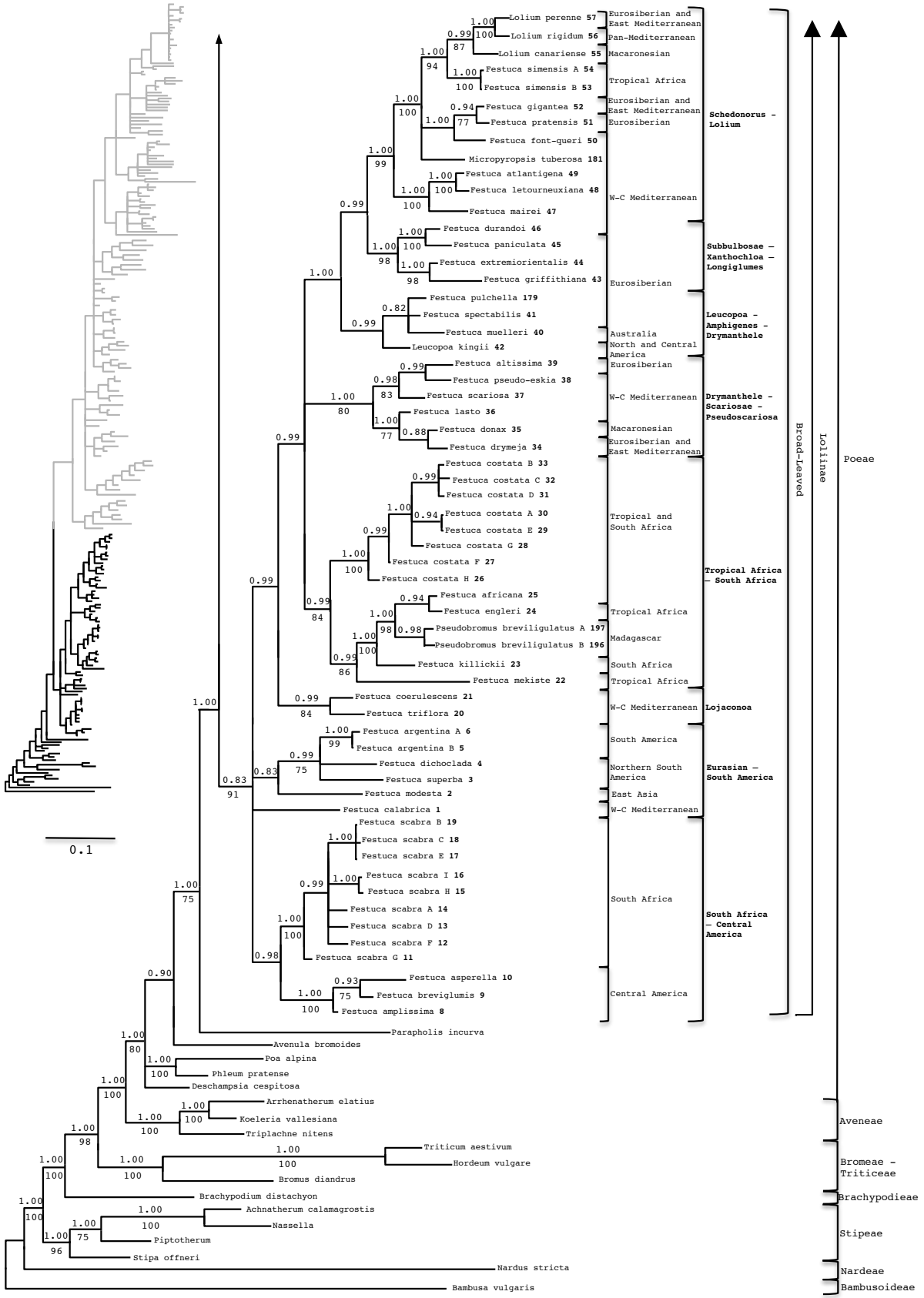


Figure S3 *(Continued)*

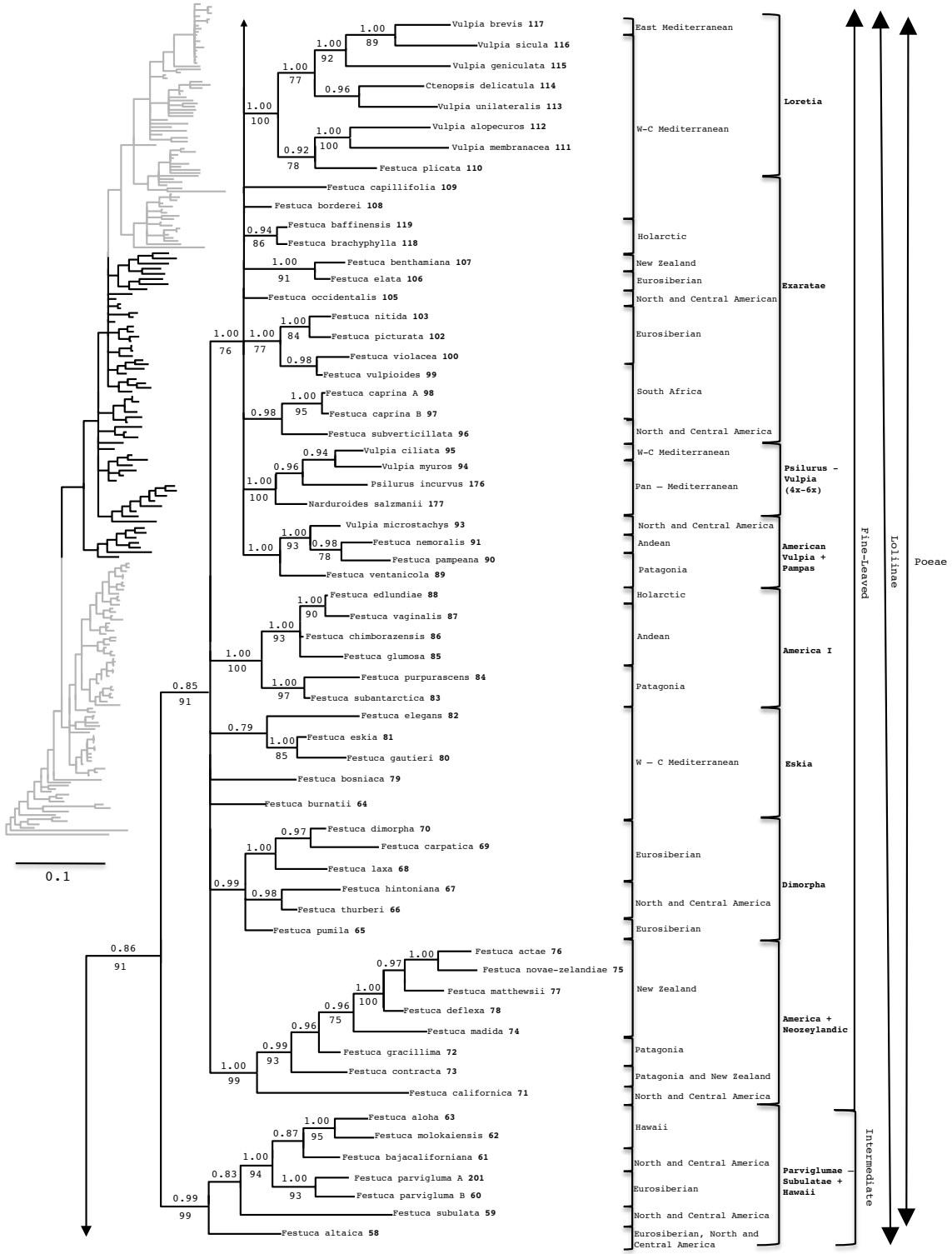


Figure S3. *(Continued)*

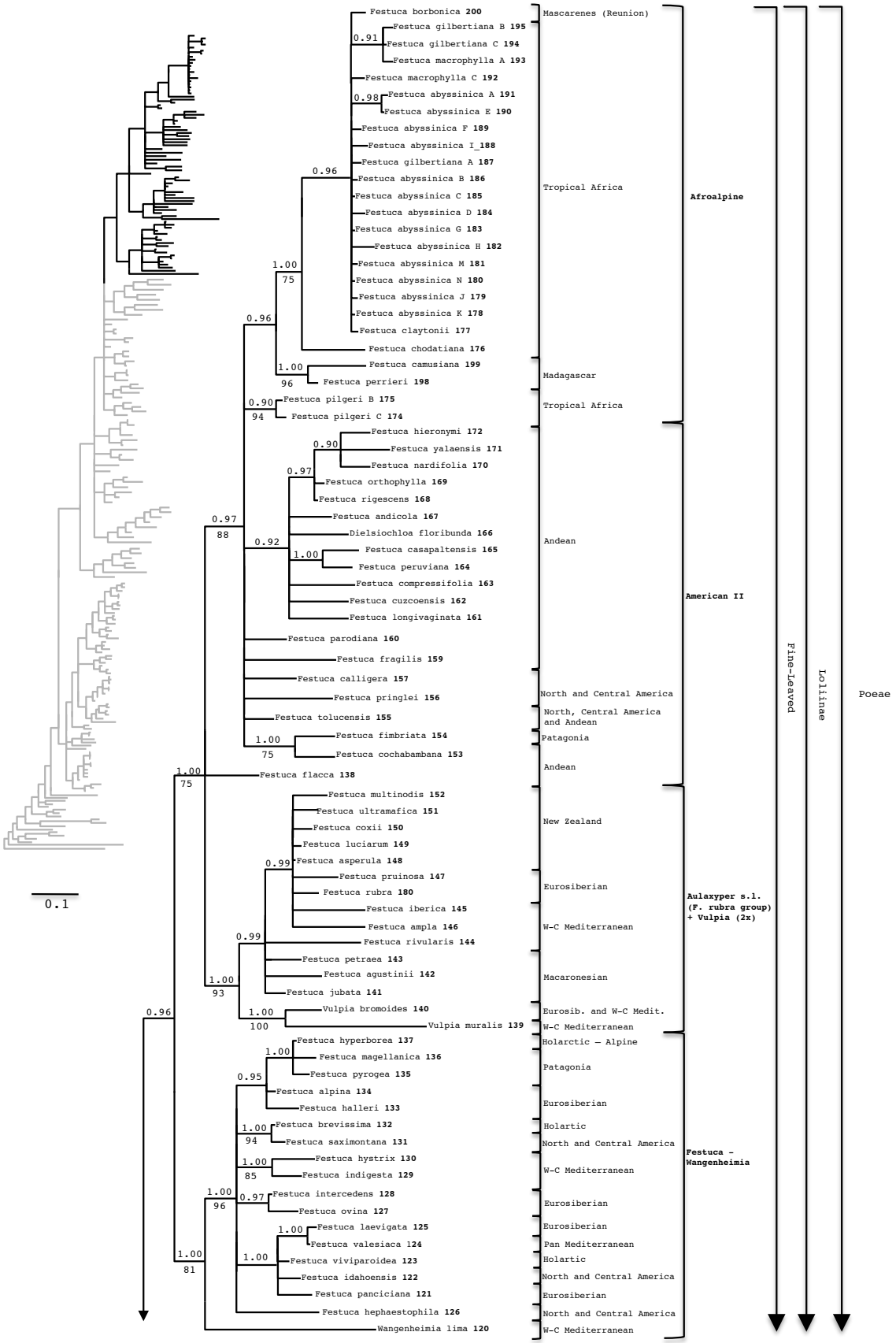


Figure S3. *(Continued)*

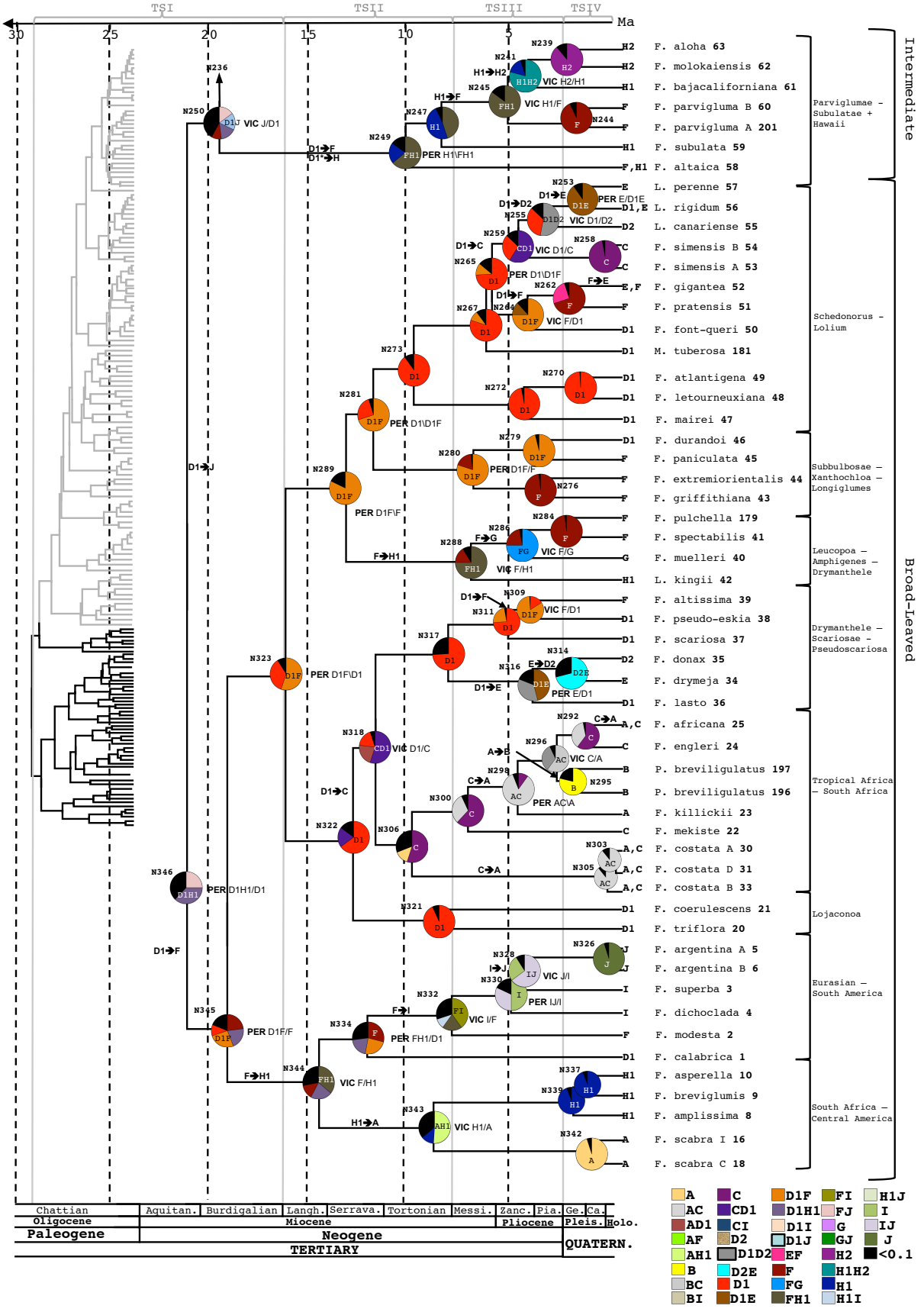


Figure S4. *(Continued)*

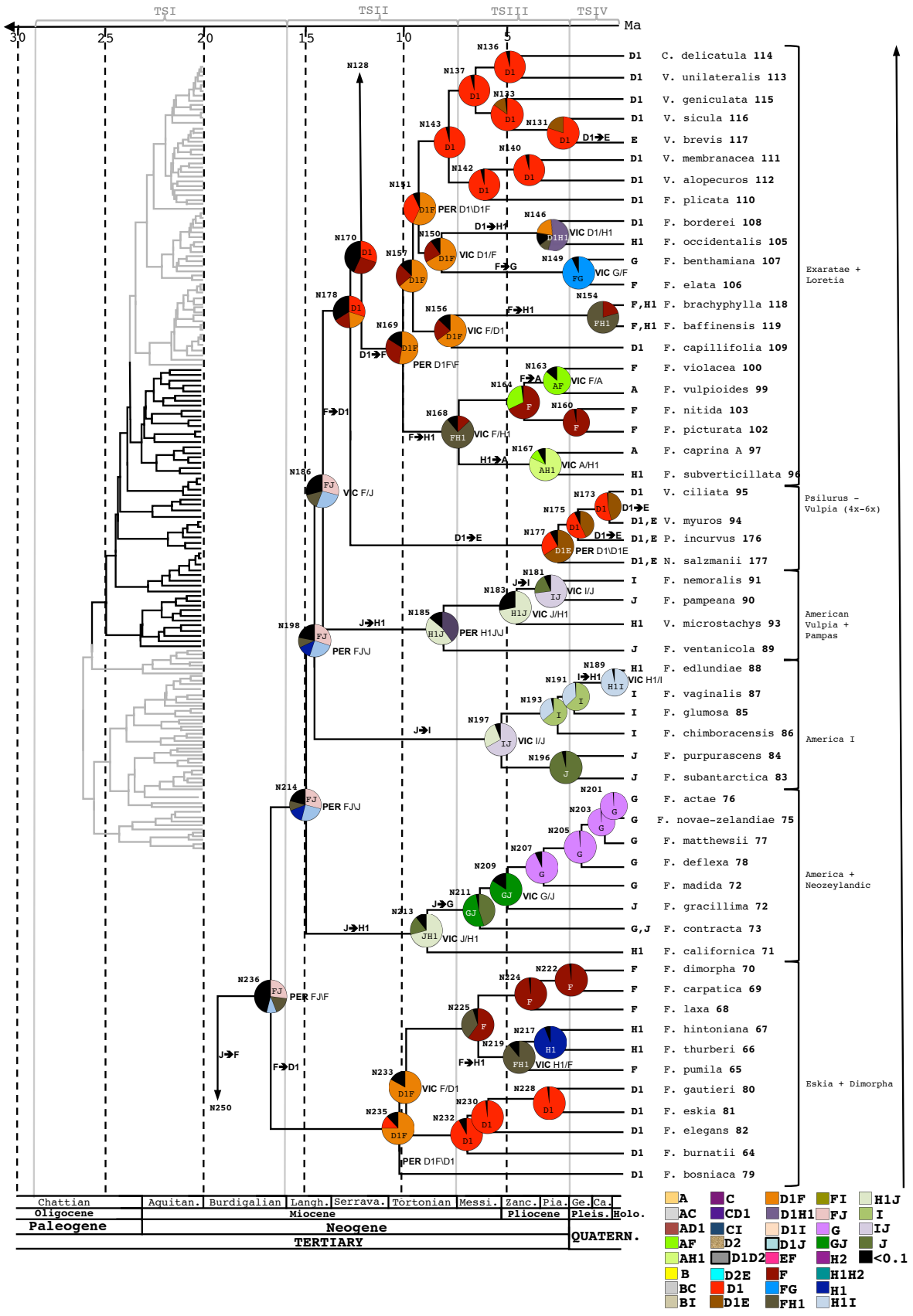


Figure S4. *(Continued)*

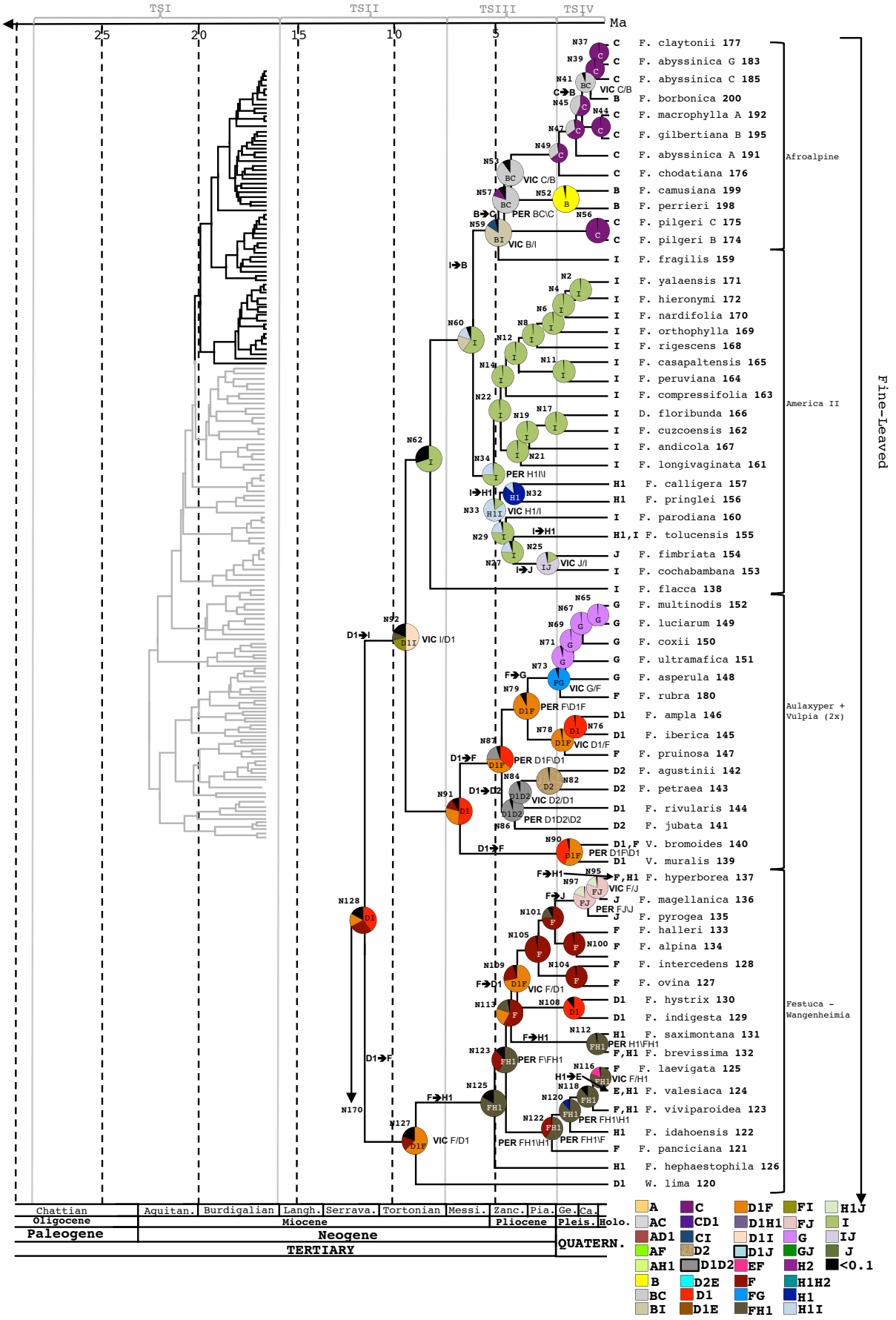


Figure S4. *(Continued)*

Figure S4. Biogeographic reconstruction for the Loliinae subtribe inferred using the DEC stratified M1 models based on 12 OAs. The tree presented corresponds to the BEAST MCC chronogram (Fig. 2; Table S2, page 320) after pruning the out-group taxa. Pie charts at nodes represent ML relative marginal probabilities of ancestral ranges, with their color legend at the bottom right-hand corner (black sections of pie charts mean that the sum of ancestral probabilities is lower than 0.1). The letters indicate split with the ancestral range with a higher marginal probability. The upper time scale bar represents million years ago (Ma). Geological time scale is shown at the bottom. OAs assigned to each selected species (A to J; Fig. 1), the name of the species, its number (Fig. 1; Table S1, page 306), and the major clades established appear to the right of the tree. The number of each node (N) is indicated above the branches. Dispersal rate matrices are shown in table S3 (page 321). The vertical grey lines indicate the boundaries between the four time slices (TSI to TSIV). The number of each node (N), the inferred biogeographical dispersal (X → Y), vicariance (X/Y), and peripheral isolate speciation (X\XY) events are also indicated.

Figure S5. Biogeographic reconstruction for the Loliinae subtribe inferred using the DEC unconstrained M0 based on 10 OAs. The tree presented corresponds to the BEAST MCC chronogram (Fig. 2; Table S2, page 320) after pruning the out-group taxa. Pie charts at nodes represent ML relative marginal probabilities of ancestral ranges, with their color legend in the upper left-hand corner (black sections of pie charts mean that the sum of ancestral probabilities is lower than 0.1). The upper time scale bar represents million years ago (Ma). Geological time scale is shown at the bottom. OAs assigned to each selected species (A to J; Fig. 1), the name of the species, its number (Table S1, page 306), and the major clades established appear to the right of the tree. The number of each node (N) is indicated above the branches.

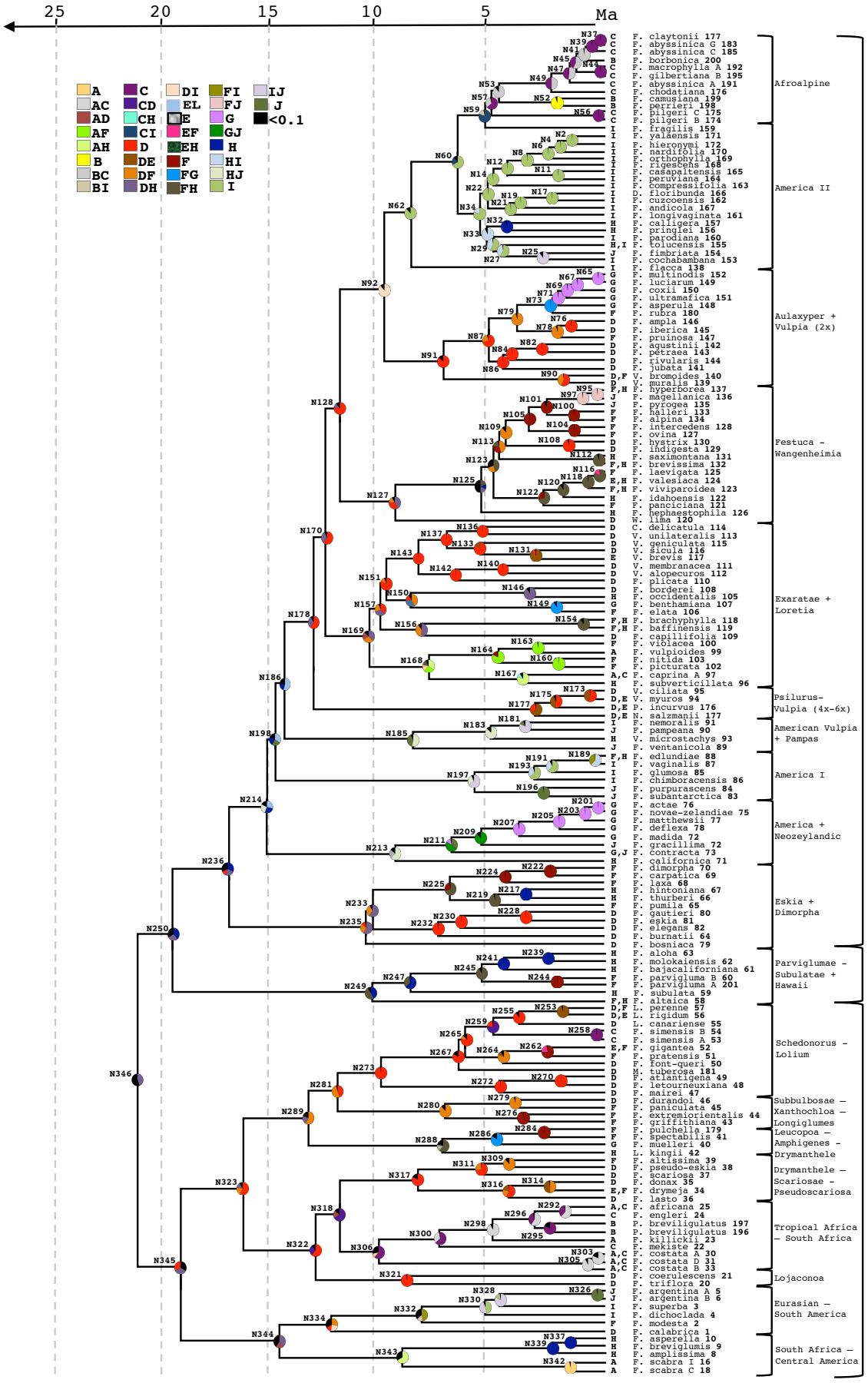

25
20
15
10
5
Ma
A
AC
AD
AF
AH
B
BC
BI
C
CD
CH
CI
D
DE
DF
DH
DI
EL
E
EF
EH
F
FG
FH
FI
FJ
G
GJ
H
HI
HJ
I
IJ
J
<0.1
Afroalpine
America II
Aulaxyper + Vulpia (2x)
Festuca - Wangenheimia
Exaratae + Loretia
Psilurus- Vulpia (4x-6x)
American Vulpia + Pampas
America I
America + Neozeylandic
Eskia + Dimorpha
Parviglumae - Subulatae + Hawaii
Schedonorus - Lolium
Subbulbosae – Xanthochloa – Longiglumes
Leucopoa – Amphigenes - Drymanthele
Drymanthele – Scariosae - Pseudoscariosa
Tropical Africa – South Africa
Lojaconoa
Eurasian – South America
South Africa – Central America
Fine-Leaved
Intermediate
Broad-Leaved

Figure S5

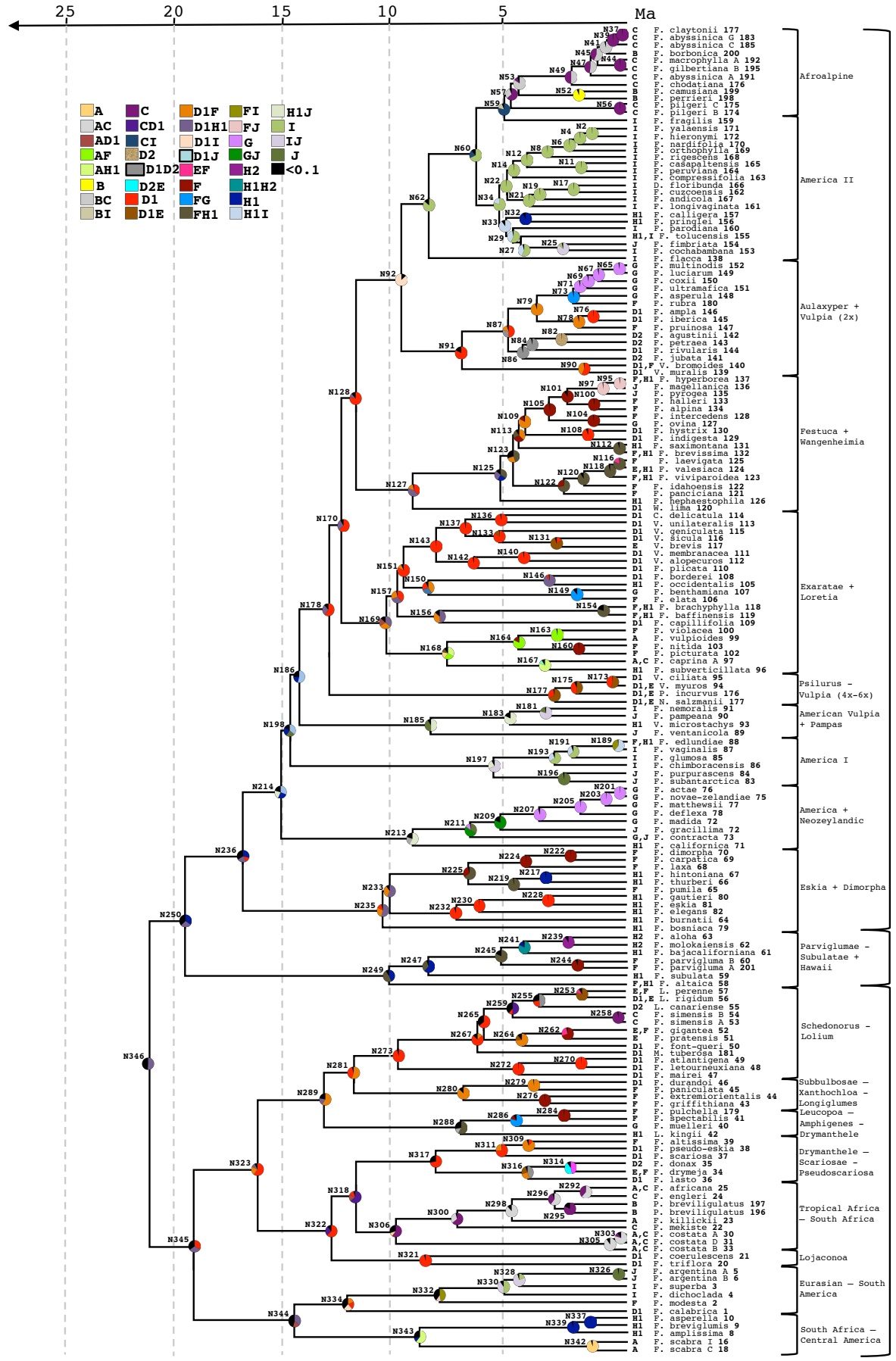

25
20
15
10
5
Ma
A
AC
AD1
AF
AH1
B
BC
BI
C
CD1
CI
D2
D1D2
D2E
D1
D1E
D1F
D1H1
D1I
D1J
EF
F
FG
FH1
FI
FJ
G
GJ
H2
H1H2
H1
H1I
H1J
I
IJ
J
<0.1
C F. claytonii 177
C F. abyssinica G 183
C F. abyssinica C 185
B F. borbonica 200
C F. macrophylla A 192
C F. gilbertiana B 195
C F. abyssinica A 191
C F. chodatiana 176
B F. camusiana 199
B F. perrieri 198
C F. pilgeri C 175
C F. pilgeri B 174
I F. fragilis 159
I F. valaensis 171
I F. hieronymi 172
I F. nardifolia 170
I F. orthophylla 169
I F. rigescens 168
I F. casapaltensis 165
I F. peruviana 164
I F. compressifolia 163
I D. floribunda 166
I F. cuzcoensis 162
I F. andicola 167
I F. longivaginata 161
H1 F. calligera 157
H1 F. pringlei 156
I F. parodiana 160
H1,I F. tolucensis 155
J F. fimbriata 154
I F. cochabambana 153
I F. flacca 138
G F. multinodis 152
G F. luciarum 149
G F. coxii 150
G F. ultramafica 151
G F. asperula 148
F F. rubra 180
D1 F. ampla 146
D1 F. iberica 145
F F. pruinosa 147
D2 F. agustinii 142
D2 F. petraea 143
D1 F. rivularis 144
D2 F. jubata 141
D1,F V. bromoides 140
D1 V. muralis 139
F,H1 F. hyperborea 137
J F. magellanica 136
J F. pyrogea 135
F F. halleri 133
F F. alpina 134
F F. intercedens 128
G F. ovina 127
D1 F. hystrix 130
D1 F. indigesta 129
H1 F. saximontana 131
F,H1 F. brevissima 132
F F. laevigata 125
E,H1 F. valesiaca 124
F,H1 F. viviparoidea 123
F F. idahoensis 122
F F. panciciana 121
H1 F. hephaestophila 126
D1 W. lima 120
D1 C. delicatula 114
D1 V. unilateralis 113
D1 V. geniculata 115
D1 V. sicula 116
E V. brevis 117
D1 V. membranacea 111
D1 V. alopecuros 112
D1 F. plicata 110
D1 F. borderei 108
H1 F. occidentalis 105
G F. benthamiana 107
F F. elata 106
F,H1 F. brachyphylla 118
F,H1 F. baffinensis 119
D1 F. capillifolia 109
F F. violacea 100
A F. vulpioides 99
F F. nitida 103
F F. picturata 102
A,C F. caprina A 97
H1 F. subverticillata 96
D1 V. ciliata 95
D1,E V. myuros 94
D1,E P. incurvus 176
D1,E N. salzmanii 177
I F. nemoralis 91
J F. pampeana 90
H1 V. microstachys 93
J F. ventanicola 89
F,H1 F. edlundiae 88
I F. vaginalis 87
I F. glumosa 85
I F. chimboracensis 86
J F. purpurascens 84
J F. subantarctica 83
G F. actae 76
G F. novae-zelandiae 75
G F. matthewsii 77
G F. deflexa 78
G F. madida 72
J F. gracillima 72
G,J F. contracta 73
H1 F. californica 71
F F. dimorpha 70
F F. carpatica 69
F F. laxa 68
H1 F. hintoniana 67
H1 F. thurberi 66
F F. pumila 65
H1 F. gautieri 80
H1 F. eskia 81
H1 F. elegans 82
H1 F. burnatii 64
H1 F. bosniaca 79
H2 F. aloha 63
H2 F. molokaiensis 62
H1 F. bajacalifornica 61
F F. parvigluma B 60
F F. parvigluma A 201
H1 F. subulata 59
F,H1 F. altaica 58
E,F L. perenne 57
D1,E L. rigidum 56
D2 L. canariense 55
C F. simensis B 54
C F. simensis A 53
E,F F. gigantea 52
E F. pratensis 51
D1 F. font-queri 50
D1 M. tuberosa 181
D1 F. atlantigena 49
D1 F. letourneuxiana 48
D1 F. mairei 47
D1 F. durandoi 46
F F. paniculata 45
F F. extremiorientalis 44
F F. griffithiana 43
F F. pulchella 179
F F. spectabilis 41
G F. muelleri 40
H1 L. kingii 42
F F. altissima 39
D1 F. pseudo-eskia 38
D1 F. scariosa 37
D2 F. donax 35
E,F F. drymeja 34
D1 F. lasto 36
A,C F. africana 25
C F. engleri 24
B P. breviligulatus 197
B P. breviligulatus 196
A F. killickii 23
C F. mekiste 22
A,C F. costata A 30
A,C F. costata D 31
A,C F. costata B 33
D1 F. coerulescens 21
D1 F. triflora 20
J F. argentina A 5
J F. argentina B 6
I F. superba 3
I F. dichoclada 4
F F. modesta 2
D1 F. calabrica 1
H1 F. asperella 10
H1 F. breviglumis 9
H1 F. amplissima 8
A F. scabra I 16
A F. scabra C 18
Afroalpine
America II
Aulaxyper + Vulpia (2x)
Festuca + Wangenheimia
Exaratae + Loretia
Psilurus - Vulpia (4x-6x)
American Vulpia + Pampas
America I
America + Neozeylandic
Eskia + Dimorpha
Parviglumae - Subulatae + Hawaii
Schedonorus - Lolium
Subbulbosae — Xanthochloa - Longiglumes
Leucopoa — Amphigenes - Drymanthele
Drymanthele — Scariosae - Pseudoscariosa
Tropical Africa — South Africa
Lojaconoa
Eurasian — South America
South Africa — Central America
Fine-Leaved
Intermediate
Broad-Leaved

Figure S6

Figure S6. Biogeographic reconstruction for the Loliinae subtribe inferred using the DEC unconstrained M0 based on 12 OAs. The tree presented corresponds to the BEAST MCC chronogram (Fig. 2; Table S2, page 320) after pruning the out-group taxa. Pie charts at nodes represent ML relative marginal probabilities of ancestral ranges, with their color legend in the upper left-hand corner (black sections of pie charts mean that the sum of ancestral marginal probabilities is lower than 0.1). The upper time scale bar represents million years ago (Ma). Geological time scale is shown at the bottom. OAs assigned to each selected species (A to J; Fig. 1), the name of the species, its number (Table S1, page 306), and the major clades established appear to the right of the tree. The number of each node (N) is indicated above the branches.

7.6 Supplementary tables

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank. The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al.* (2003; 2004), Catalan *et al.* (2006), and Soreng *et al.* (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Poeae R. Br.					
***Loliinae* Dumort.**					
Broad-leaved Loliinae					
South African – Central American clade					
Festuca amplissima Rupr. **8**	Mexico: Chihuahua: Barranca del Cobre; Peterson P.M.,Catalan P. 17573	H1	xxxx	EF592954	EF585008
Festuca asperella E.B. Alexeev **10**	Mexico: Mexico DF; Morelos boundary. Trott et al. 142, MO 2744225	H1	xxxx	xxxx	xxxx
Festuca breviglumis Swallen **9**	Mexico: Jalisco: Ciudad Guzman; Peterson P.M., Rosales O.16078.	H1	AY367952	EF592960	EF585017
Festuca longipes Stapf	South Africa: Cape Province: Encobo district. Catalan & Pimentel. UZ- SA031	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _A **14**	South Africa: Western Cape: Cape Town, Catalan & Pimentel, UZ-SA008	A	JQ972954	JQ972978	JQ973017
Festuca scabra Vahl _B **17**	South Africa: Western Cape: Kranskop. Catalan & Pimentel, UZ-SA010	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _C **18**	South Africa: Western Cape: Stellenbosch. Catalan & Pimentel, UZ-SA011	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _D **13**	South Africa: Western Cape: Somerset West. Catalan & Pimentel, UZ-SA013	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _E **19**	South Africa: Western Cape: Ceres, Guido's Pass. Catalan & PimentelUZ-SA014	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _F **12**	South Africa: Western Cape: Oudtshoorn, Swartberg Pass. Catalan & Pimentel, UZ-SA016	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _G **11**	South Africa: Western Cape: Mossel Bay - Vleesbaai. Catalan & Pimentel, UZ-SA017	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _H **15**	South Africa: KwaZulu Natal: Cathedral Peak, Organ Pipes Pass, Catalan & Pimentel, UZ- SA039	A	xxxx	xxxx	xxxx
Festuca scabra Vahl _I **16**	South Africa: KwaZulu Natal: Underberg, Sani Pass, Catalan & Pimentel, UZ-SA047	A	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Eurasian – South American clade					
Festuca argentina (Speg.) Parodi_A **5**	Argentina: Santa Cruz. Lago Argentino; Peterson P.M. 17158	J	EF584923	EF592957	EF585012
Festuca argentina (Speg.) Parodi_B **6**	Argentina: Rio Negro: Bariloche: Segunda Angostura. Catalan UZ-02.10	J	xxxx	xxxx	xxxx
Festuca calabrica Huter, Porta & Rigo **1**	Italy: Calabria: Prov. Cosenza. Muller, J. 10838	D1	xxxx	xxxx	xxxx
Festuca dichoclada Pilg. **4**	Peru: Junin: Tarma: Maraynioc; Peterson P.M., Tovar O.14056, US-3421417	I	EF584933	EF592967	EF585031
Festuca modesta Nees ex Steud. **2**	China: Yunnan: Fugong (Bijiang); Soreng R.J.& al. 5227 US-3420887	F	EF584953	EF592985	EF585068
Festuca superba Parodi ex Türpe **3**	Argentina: Jujuy. Yala: Laguna Rodeo. Catalan & al. UZ-356.08	I	xxxx	xxxx	xxxx
***F*. sect. *Lojaconoa* Gand.**					
Festuca coerulescens Desf. **21**	Spain: Cadiz: Jerez de la Frontera; Catalan P. & al., UZ-91.2000	D1,E	AF538363	AF533051	EF585027
Festuca triflora J.F. Gmel. **20**	Spain: Cadiz Grazalema: Bco. Ballesteros; Catalan P. & al.,UZ 95.2000	D1,E	AF538362	AF533052	EF585109
F*. subgen. *Leucopoa					
Festuca komarovii Krivot.	Russia: Vostochnyi Sayan. Malyschev, LE	E, F	xxxx	xxxx	-
Festuca sibirica Hack. ex Boiss.	Mongolia: Eastern aimak. Kamelin et al. No. 1171, LE	E, F	xxxx	xxxx	-
Tropical African – South African clade					
Festuca africana (Hack.) Clayton **25**	Kenya: Mt. Kenya, Sirimon; Namaganda 1738a	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_A **30**	South Africa: Eastern Cape: Amathole Mnts., Stuttenheim 2, Catalan & Pimentel, UZ-SA024	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_B **33**	South Africa: Eastern Cape: Amathole Mnts, Cathcart. Catalan & Pimentel, UZ-SA027	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_C **32**	South Africa: Eastern Cape: Amathole Mnts., Hogsback 1, Catalan & Pimentel, UZ-SA030	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_D **31**	South Africa: Eastern Cape: Amathole Mnts, Hogsback 2. Catalan & Pimentel, UZ-SA033	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_E **29**	South Africa: Eastern Cape: Amathole Mnts, Stuttenheim 1, Catalan & Pimentel, UZ- SA034	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_F **27**	South Africa: Eastern Cape: Elliot, Barkly Pass. Catalan & MPimentel, UZ-SA036	A,C	xxxx	xxxx	xxxx
Festuca costata Nees_G **28**	South Africa: KwaZulu Natal: Cathedral Peak, Organ Pass, Catalan & Pimentel, UZ-SA040	A,C	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Tropical African – South African clade					
Festuca costata Nees_H **26**	South Africa: KwaZulu Natal: Underberg, Castle View. Catalan & Pimentel, UZ-SA048	A,C	xxxx	xxxx	xxxx
Festuca engleri Pilg. **24**	Kenya: Mt. Kenya, Sirimon; Namaganda 1739	C	xxxx	xxxx	xxxx
Festuca killickii Kenn.-O'Byrne **23**	South Africa: KwaZulu Natal: Cathedral Peak, Rainbow Canyon, Central Drakenberg. Catalan & Pimentel, JACA s.n., SA041	A	xxxx	xxxx	xxxx
Festuca mekiste Clayton **22**	Kenya: Mt. Elgon National Park. Carvalho 4521	C	xxxx	xxxx	xxxx
Pseudobromus breviligulatus Stapf ex A. Camus_A **196**	Madagascar: Andringitra National Park., Vorontsoba F1221	B	xxxx	xxxx	xxxx
Pseudobromus breviligulatus Stapf ex A. Camus_B **197**	Madagascar: Andringitra National Park. Vorontsoba F1222	B	xxxx	xxxx	xxxx
***Festuca* subgen. *Drymanthele* – *F.* sect. *Scariosae* Hack. – *F.* sect. *Pseudoscariosa* Krivot.**					
Festuca altissima All. **39**	France: Pyrenees: Aspe; Catalan P. UZ s.n.	F	AF303411	AF478505	EF585003
Festuca donax Lowe **35**	Portugal: Madeira: Porto Moniz; Catalan & Sequeira, UZ-MS4515	D2	EF584935	EF592968	EF585033
Festuca drymeja Mert. & W.D.J.Koch **34**	Hungary: Balaton; Stace C.A.	E,F	AF303425	AY098997	EF585034
Festuca lasto Boiss. **36**	Spain: Cadiz: Sierra Bermeja; Stace C.A.	D1	AF303418	AY098998	EF585060
Festuca pseudo-eskia Boiss. **38**	Spain: Granada: Sierra Nevada: Collado del Diablo, Catalan P. & al., UZ-73.2000	D1	AF519979	AY099000	EF585084
Festuca scariosa Lag. ex Willk. **37**	Spain: Almeria: Sierra Filabres: Las Menas; Catalan P.& al., UZ- 62.2000	D1	AF519978	AY098999	EF585100
Festuca* subgen. *Leucopoa* – *F.* subgen. *Amphigenes* Janka – Australian *F.* subgen. *Drymanthele					
Festuca hubsugulica Krivot.	Russia: Vostochnyi Sayan: Pogranichnyi mountain range. Pezhemskyi No. 99 III, LE	F	xxxx	xxxx	-
Festuca muelleri Vickery **40**	Australia: Capital Territory: Mt Bimberi, Streimann 67177, CBG 67177	G	xxxx	xxxx	xxxx
Festuca pulchella Schrad. **179**	Switzerland: Bern; Muller J. 7807 UZ	F	AF519980	AF519985	EF585086
Festuca spectabilis Jan **41**	Italy: Lombardia: Bergamo: Passo della Presolana; Muller J.8229	F	AF519977	AF519984	EF585102
Leucopoa kingii (S. Watson) W.A. Weber **42**	USA: Colorado: Boulder Co: Flat Irons; Catalan P. UZ-1.93,	H1	AF303410	AY099004	EF585058

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
***Festuca* sect. *Subbulbosae* Nyman ex Hack. – *F.* sect. *Xanthochloa* (Krivot.) Tzvelev – *F.* sect. *Longiglumes* S.L.Lu**					
Festuca caucasica Hack. & Trautv.	Russia: Caucasus, Balkaria, Sukan; E Bush and N Bush No. 26, LE	F	xxxx	-	xxxx
Festuca durandoi Clauson **46**	Spain: Segovia: Riaza; Lopez Rodrıguez J.A. UZ-6-6-00,	D1	AF543514	AF533047	EF585035
Festuca extremiorientalis Ohwi **44**	Russia: Krasnoyarsk. Pen'kovskaya & N Sozinova, LE	F	xxxx	xxxx	xxxx
Festuca griffithiana (St.-Yves) Krivot. **43**	Afghanistan: Hajigak pass. Furse 8489, LE	F	xxxx	xxxx	xxxx
Festuca paniculata (L.) Schinz **45**	France: Mont Aigoual; Cebolla & Rivas-Ponce, UAM	F	AF303407	AF533046	EF585077
***Festuca* subgen. *Schedonorus* (P. Beauv.) Peterm. – *Lolium* L.**					
Festuca arundinacea subsp. *atlantigena* (St.-Yves) Auquier **49**	Morocco: High Atlas Mnts,(cultivated seed in Belgium, ABYBN 1304)	D1	AJ240155	EF378956	EF378996
Festuca arundinacea var. *letourneuxiana* Torrecilla & Catalán **48**	Morocco: High Atlas Mnts (cultivated seed in UK, ABY-BN 400.1967U)	D1	EF379056	EF378960	EF378998
Festuca mairei St.-Yves **47**	Morocco: Marrakech: Oukaimeden; Cebolla & Rivas-Ponce 4064, UAM	D1	AF303424	AY098996	EF379006
Festuca font-queri St.-Yves **50**	Morocco: Rif Mountains; Mejias J., US.	D1	AF303404	AF533044	DQ631486
Festuca gigantea (L.) Vill. **52**	Spain: Navarra: Arce; Aizpuru & Catalan, UZ s. n.	E,F	AF303416	AF533043	EF379003
Festuca pratensis (Huds.) P. Beauv. **51**	England: Wilshire: Calne; Stace C.A.	F	AF303421	AF478503	EF379007
Festuca simensis Hochst. ex A.Rich._A **53**	Uganda: Echuya, Kabale, Kisoro boarde. Namaganda 204K	C	GU573754	GU573746	GU573749
Festuca simensis Hochst. ex A.Rich._B **54**	Kenya: Mt. Kenya. Namaganda 1750	C	xxxx	xxxx	xxxx
Lolium canariense Steud. **55**	Spain: Canary Isles: Tenerife: Las Cañadas, UZ s. n.	D2	AY228161	AY228162	EF379014
Lolium perenne L. **57**	UK: Wales (cutivated seeds PI 619001 USDA-Pullman)	E,F	AF303401	AF478504	EF379024
Lolium rigidum Gaudin **56**	Spain: Zaragoza: Vedado de Peñaflor; Catalan P. & al., UZ-18.2000,	D1,E	AF532944	AF533039	EF379034
Micropyropsis tuberosa Romero-Zarco & Cabezudo **181**	Spain. Huelva: Almonte. Catalan P. & al, UZ-89.07,	D1	JQ972944	AF533037.1	JQ973010
***Castellia* Tineo**					
Castellia tuberculosa (Moris) Bor	Spain: Cadiz: Sanlucar de Barrameda: La Algaida; Lopez Rodrıguez J.A.,. UZ s.n	D1	AF522954.1	AF533053	EF584996

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Intermediate Loliinae					
***Festuca* sect. *Parviglumae* S.G. Aiken & Xiang Chen – *F.* subgen. *Subulatae* (Tzvel.) Alexeev + Hawaian clade**					
Festuca aloha Catalán, Soreng & P.M.Peterson **63**	USA: Hawaii: Kauai; PTBG 17678, R. Wood 1704	H2	GQ162206	GQ162208	-
Festuca altaica Trin. **58**	Canada: Yukon Territory: Teslin Lake; Soreng R.J. 5996, US	F,H1	AF532952	AF533055	EF585002
Festuca bajacaliforniana Gonz.-Led. & S.D.Koch **61**	Mexico. Peterson, 5287, US	H1	xxxx	xxxx	xxxx
Festuca molokaiensis Soreng, Peterson & Catalán **62**	USA: Hawaii: Molokai, BISH728771.	H2	GQ162207	GQ162210	xxxx
Festuca parvigluma Steud._A **201**	China: Xizhang (Tibet): Miling; Soreng R.J. & al.5576, US-3420901	F	EF584960	EF592992	EF585079
Festuca parvigluma Steud._B **60**	Japan: Kyoto: Kawarajiri: Kameoka-shi. Tsugaru S. & Sawada M., 18032, MO-4945850	F	xxxx	EF592992	xxxx
Festuca subulata Trin. **59**	USA: Oregon: Clatsop County: Saddle Mnts; Wilson B. 10512	H1	AF532953	AF533056	EF585104
Fine-leaved Loliinae					
***Festuca* sect. *Eskia* Willk.**					
Festuca bosniaca Kummer & Sendtn. **79**	Bosnia-Herzegovina: Troglav: Sajkovacko zdrlo. Schönswetter, UZ-6.08	D1	xxxx	xxxx	xxxx
Festuca burnatii St.-Yves **64**	Spain: Cantabria: Picos de Europa; Catalan P. &Torrecilla P., UZ-44.2001	D1	AY099007	AY099002	EF585019
Festuca elegans Boiss. **82**	Spain. Granada: Baza; Cebolla & Rivas-Martinez s.n.,UAM	D1	AF303406	AF478509	EF585038
Festuca eskia Ramond ex DC. **81**	Spain: Huesca: Pyrenees: Benasque; Catalan P.,Mirones V. UZ	D1	AF303412	AF478508	EF585040
Festuca gautieri (Hack.) K.Richt. **80**	Spain: Girona: Pyrenees: Nuria; CatalaÅLn P., Mirones V. UZ	D1	AF303414	AF478507	EF585044
***Festuca* sect. *Dimorpha* Schreb.**					
Festuca carpatica F.Dietr. **69**	Slovak Republic: VysokeL Tatry Mts.: Tisovnice; Marhold K. s.n.	F	AY099006	AY099001	EF585023
Festuca dimorpha Guss. **70**	France: Alpes de Haute-Provence Col des Champs; Korneck D. s.n, Herb. Muller J. 10969	F	AF519982	AF519987	EF585032
Festuca hintoniana E.B.Alexeev **67**	Mexico: Nuevo Leon: Sierra Madre Oriental; Peterson ,Knowless M.B. 13343, US	H1	EF584945	EF592977	-
Festuca laxa Host **68**	Slovak Republic: Veliki Stador. Gutermann et al. 38483	F	xxxx	xxxx	xxxx
Festuca pumila Chaix **65**	Slovak Republic: Veliki Stador. Gutermann 38499, UZ 4.08	F	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
***Festuca* sect. *Dimorpha* Schreb.**					
Festuca thurberi Vasey **66**	USA: New Mexico: Sangre Cristo Mts; Allred K. 8257	H1	EF584975	EF593007	EF585107
Festuca woronowii Hack.	Armenia: Akhyryan District: Shiraksky. Tzvelev N. & Tcherepanov S. No. 173	E	xxxx	xxxx	-
American + Neozeylandic clade					
Festuca actae Connor **76**	New Zealand: Canterbury: Banks peninsula: Lake Forsyth; Lloyd K.M. 57589	G	AY524829	AY528949	EF584998
Festuca californica Vasey **71**	USA: Oregon: Benton County; Wilson B. 7014	H1	AF532956	AF533054	EF585020
Festuca contracta Kirk **73**	South Georgia Islands. HU Peter s/n, JE.	G,J	xxxx	xxxx	-
Festuca deflexa Connor **78**	New Zealand: Poverty Basin: Mt Owen: Nelson; Lloyd K.M.57588	G	AY524838	AY528942	EF585030
Festuca gracillima Hook. f. **72**	Argentina: Tierra de Fuego. Catalan. UZ- 482.08	J	xxxx	xxxx	xxxx
Festuca madida Connor **74**	New Zealand: Central Otago: Rock & Pillar Range; Lloyd K.M. 57627	G	AY524833	AY528943	EF585065
Festuca matthewsii (Hack.) Cheeseman **77**	New Zealand: Fiorland: Tahake valley; Lloyd K.M. 57938	G	AY524836	AY528948	EF585066
Festuca novae-zelandiae (Hack.) Cockayne **75**	New Zealand: Fiorland: Lloyd K.M. 57940	G	AY524832	AY528941	EF585073
American I clade					
Festuca chimborazensis E.B.Alexeev **86**	Ecuador: Cotopaxi: Chaupi; Stancik D. 4034, US-3428932	I	EF584930	EF592963	EF585024
Festuca edlundiae S.G. Aiken, L.L. Consaul & Lefk. **88**	Canada: Nunavut: BaÅLthurst island: Polar Bear Pass; Gillespie., Consaul L.L. 6990	F,H1	EF584936	EF592969	EF585036
Festuca glumosa Hack. ex E.B. Alexeev **85**	Ecuador: Imbabura Cayambe: Volcano Cayambe; Standik., US-3428930	I	EF584940	EF592973	EF585046
Festuca purpurascens Banks & Sol. ex Hook. f. **84**	Argentina: Santa Cruz: Lago Argentino; Peterson P.M. 17147	J	EF584964	EF592996	EF585087
Festuca subantarctica Parodi **83**	Argentina: Santa Cruz. Peterson P.M. 17163	J	EF584973	EF593004	EF585103
Festuca vaginalis (Benth.) Laegaard **87**	Ecuador: Pichincha Pifo; Stancik. US-3428946	I	EF584977	EF593010	EF585111
American Vulpia + Pampas clade					
Festuca nemoralis Türpe **91**	Argentina: Jujuy: Yala: Laguna Rodeo. Catalan 358.08 & J. Müller.	I	xxxx	xxxx	xxxx
Festuca pampeana Speg. **90**	Argentina: Buenos Aires: Sierra de la Ventana. Catalan UZ-428.08	J	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
American Vulpia + Pampas clade					
Festuca ventanicola Speg. **89**	Argentina: Buenos Aires: Sierra de la Ventana. Catalan UZ- 418.08	J	xxxx	xxxx	xxxx
Vulpia microstachys (Nutt.) Munro **93**	USA: California: W of San Luis Obispo; Soreng R.J. 7406	H1	EF584981	EF593015	EF585125
Vulpia octoflora (Walter) Rydb**.**	USA: Washington: Okanogan; Peterson P.M. 3263	H1	EF584982.1	EF593016	EF585128
Psilurus–Vulpia (4x-6x) clade					
Narduroides salzmannii (Boiss.) Rouy **177**	Spain. Madrid: Arganda. UZ 111.07,	D1,E	AF478497.1	AF478535.1	JQ973014
Psilurus incurvus (Gouan) Schinz & Thell. **176**	Spain: Malpartida de Plasencia. UZ 31.07,	D1,E	JQ972945.1	AF478533.1	JQ973013.1
Vulpia ciliata Dumort. **95**	Spain: Zaragoza: Vedado de Peñaflor; Catalan P. UZ-19.2002	D1	AY118094	AY118104	EF585120
Vulpia myuros (L.) C.C. Gmel. **94**	USA: Washington: King Co: Seattle: Lake Forest Park; Catalan P. UZ-54.2001	D1,E	AY118092	AY118103	EF585127
Exaratae group					
Festuca amethystina L.	Germany: Bayern; Muller J. 6966	F	EF584919	EF592950	EF585004
Festuca baffinensis Polunin **119**	Canada: Northwest: Masik River Valley; Gillespie L.J. & al.7116	F,H1	EF584925	EF592958	EF585013
Festuca benthamiana (Benth.) Vickery **107**	Australia: South Australia: region 8: Northern Lofty. R Bates 16474, CANB 503274	G	xxxx	xxxx	-
Festuca borderei Kerguélen **108**	Spain: Huesca: Pyrenees: Vallibierna; CatalaÅLn P., UZ-2000	D1	AF303403	AF478510	EF585015
Festuca brachyphylla Schult. & Schult. f. **118**	USA: Alaska: North Slope Borough: Prudhoe Bay; Soreng R.J. 6243 US	F,H1	EF584927	xxxx	EF585016
Festuca capillifolia Dufour **109**	Spain: Jaen: Cazorla; Cebollada , Rivas-Ponce M. A.	D1	AF303419	AF478511	EF585022
Festuca caprina Nees_A **97**	South Africa: Eastern Cape. Catalan & Pimentel, UZ- SA026	A,C	xxxx	xxxx	xxxx
Festuca caprina Nees_B **98**	South Africa: Eastern Cape. Catalan & Pimentel, UZ- SA036	A,C	xxxx	xxxx	-
Festuca clementei Boiss.	Spain: Granada: Sierra Nevada: Veleta; Catalan P & al., UZ-881.2000	D1	AF478482	AF478524	EF585025
Festuca elata Keng ex E.B. Alexeev 1**06**	China: Yunnan, Lushui, Gaoligong Shan; Soreng R.J.& al.5268, US-3420890	F	EF584937	EF592970	EF585037
Festuca nitida Kit. ex Schult. **103**	Montenegro: Prutas: Hangmulde nordöstlich unter dem Gipfe. Gutermann et al. 38891, UZ	F	xxxx	xxxx	xxxx
Festuca norica (Hack.) K. Richt.	Italy: Trentino-Alto Adige; Muller J. 8422	F	EF584955	EF592987	EF585072
Festuca occidentalis Hook. **105**	USA: Oregon: Klamath: Winema; Soreng R.J. 5971, US	H1	EF584956	EF592988	EF585074

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Exaratae group					
Festuca picturata G. Pils **102**	Austria: Tirol, Zillertaler Alpen: Plauener Hütte. Schonswetter & Katharina, UZ 11.08	F	xxxx	xxxx	xxxx
Festuca subverticillata (Pers.) E.B. Alexeev **96**	USA: West Virginia: Grant Co.; Peterson P.M., Saarela J.M.15784, US	H1	EF584974	EF593006.1	EF585106
Festuca violacea Gaudin **100**	Switzerland: Bern; Muller J. 7907	F	EF584979	EF593012	EF585113
Festuca vulpioides Steud. **99**	South Africa: Eastern Cape: Amathole Mnts. Catalan & Pimentel, UZ-SA032	A	xxxx	xxxx	xxxx
Loretia clade					
Ctenopsis delicatula (Lag.) Paunero **114**	Spain: Madrid: Garganta de Los Montes; Lopez Rodrıguez J.A. UZ s.n.	D1	AF478499	AF478537	EF584997
Festuca plicata Hack. **110**	Spain: Granada: Sierra Nevada: Dornajo; Catalan P. & al.,UZ-86.2000	D1	AF478483	AF478525	EF585082
Vulpia alopecuros (Schousb.) Dumort. **112**	Portugal: Algave: Meia Praia: Lagos; Stace C.A., LEI	D1	AF478491	AF487617	EF585117
Vulpia brevis Boiss. & Kotschy **117**	Cyprus: SE Nicosia; Stace C.A., LEI	E	AF478489	AF478530	EF585118
Vulpia geniculata (L.) Link **115**	Spain: Sevilla: Constantina; JACA J29397	D1	AF478490	AF478531	EF585123
Vulpia membranacea (L.) Dumort. **111**	Spain: Cadiz: Sanlucar de Barrameda: La Algaida; Catalan P., UZ-8.2002,	D1	AY118090	AY118101	EF585124
Vulpia sicula Link **116**	France: Corse: Ponte Leccia; JACA 366589	D1	AY118089	AY118100	EF585129
Vulpia unilateralis (L.) Stace **113**	Spain: Zaragoza: Vedado de Peñaflor; Catalan P., UZ-18.2002	D1	AY118095	AY118106	EF585130
Festuca– Wangenheimia clade					
Festuca alpina Suter **134**	Spain: Huesca: Pyrenees: Vallibierna; Catalan P., UZ-2002	F	AF303415	AF478522	EF585001
Festuca brevissima Jurtzev **132**	USA: Alaska: Denali Borough: Alaska Range; Soreng R.J.6021 US	F,H1	EF584928	EF592961	EF585018
Festuca halleri All. **133**	Switzerland: Valais; Muller J. 8032	F	EF584942	EF592975	EF585047
Festuca hephaestophila Nees ex Steud. **126**	Mexico: Nuevo Leon: Sierra Madre Oriental; Peterson P.M.,Knowless M.B. 13347, US	H1	EF584943	EF592976	EF585048
Festuca hyperborea Holmen ex Fred. **137**	Canada, Northwest Territories, Prince Patrick Island; Gillespie L.J., Consaul L.L. 6893	F,H1	EF584946	EF592978	EF585050
Festuca hystrix Boiss. **130**	Spain: Almerıa: Sierra de Gador; Catalan P. & al. UZ- 31.2000	D1	AF478480	AF478520	EF585051
Festuca idahoensis Elmer **122**	USA: Oregon: Deschutes County; Wilson B. 533	H1	EF584947	AF533064	EF585053

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Festuca – Wangenheimia clade					
Festuca indigesta Boiss. **129**	Spain: Granada: Sierra Nevada; Catalan P. & al. UZ-43.2000	D1	AF303426	AF478519	EF585054
Festuca intercedens (Hack.) Lüdi ex Bech. **128**	Switzerland: Valais: Muller J. 7987	F	EF584948	EF592979	EF585055
Festuca laevigata Gaudin **125**	Italy: Lombardia. Muller J. 8267	F	EF584950	EF592981	EF585059
Festuca magellanica Lam. **136**	Argentina: Tierra de Fuego: Ctra. Haberton. Catalan & A. Sarria UZ-461.08	J	xxxx	xxxx	xxxx
Festuca ovina L. **127**	Germany: Thuringen: Saale-Holzland-Kreis; Muller J. 6879	F	AF532959	AF533063	EF585076
Festuca panciciana K. Richt. **121**	Bosnia-Herzegovina: Troglav: Sajkovacko zdrlo. Gutermann et al., UZ s.n.	F	JQ972951	JQ972974	JQ973015
Festuca pyrogea Speg. **135**	Argentina: Tierra de Fuego: Cabo San Pablo. Catalan & A. Sarria UZ-475.08	J	xxxx	xxxx	xxxx
Festuca saximontana Rydb. **131**	USA: Oregon, Soreng R. J. 6021-1	H1	EF584969	EF593002	EF585098
Festuca valesiaca Schleich. ex Gaudin **124**	Germany: Thuringen; Muller J. 6939	E,F	EF584978	EF593011	EF585112
Festuca viviparoidea Krajina ex Pavlick **123**	Canada Northwest Territories, Banks Iisland: Nelson Head; Gillespie L.J. & al. L 7204	F,H1	EF584980	EF593013	EF585114
Wangenheimia lima (L.) Trin. **120**	Spain: Zaragoza: Vedado de Peñaflor; Catalan P. & al., UZ-17.2000	D1	AF478498	AF478536	EF585131
***Micropyrum* Link**					
Micropyrum tenellum (L.) Link	Spain: Segovia: Navafrıa; Lopez Rodrıguez J.A, UZs.n.,	D1	AF478494	AF478534	EF585116
***Vulpia* sect. *Vulpia* (2x)**					
Vulpia bromoides (L.) Gray **140**	Spain: Lugo: Lancara; Lopez Rodrıguez J.A. 01080, UZ	D1,F	AF478485	AF487616	EF585119
Vulpia muralis (Kunth) Nees **139**	Spain: Zaragoza: Actur; Pyke & Catalan UZ-11.2000	D1	AY118091	AY118102	EF585126
Aulaxyper clade					
Festuca agustinii Linding. **142**	Spain: Canarias: Tenerife: Anaga: Bailadero; A.Santos	D2	AY099005	AY099003	EF584999
Festuca ampla Hack. **146**	Spain: Cadiz: Grazalema; Lopez Rodrıguez J.A, 1326, UZ	D1	EF584921	EF592953	EF585007
Festuca asperula Vickery **148**	Australia: NWS Australia. Hosking JR; CANB 502693	G	xxxx	xxxx	xxxx
Festuca coxii (Petrie) Hack. **150**	New Zealand: Chatham Island: The Pinnacles; Lloyd K.M.57956	G	AY524825	AY528937	EF585028
Festuca iberica (Hack.) K. Richt. **145**	Spain: Granada: Sierra Nevada: Borreguiles de S. Juan;Catalan P. & al., UZ-77.2000	D1	AY118087	AF478516	EF585052
Festuca jubata Lowe **141**	Portugal: Madeira: Pico das Torres; Sequeira & Catalan s.n.	D2	EF584949	EF592980	EF585056

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Aulaxyper clade					
Festuca luciarum Connor **149**	New Zealand: Maungaharuru Range: Hawkes Bay; LloydK.M. 57621	G	AY524828	AY528939	EF585064
Festuca multinodis Petrie & Hackel **152**	New Zealand: Wellington: Baring Head; Lloyd K.M. 57958	G	AY524827	AY528940	EF585069
Festuca petraea Guthnick **143**	Portugal: Azores: St. Maria, Maia; Sequeira M. 4393	D2	EF584962	EF592994	EF585081
Festuca pruinosa (Hack.) Patzke **147**	Spain: Pontevedra: Cangas de Morrazo, Donon; Sauquillo& Pimentel	F	EF584963	EF592995	EF585083
Festuca rivularis Boiss. **144**	Spain: Granada: Sierra Nevada: Borreguiles de S. Juan; Catalan P. & al., UZ-78.2000	D1	AF478475	AF478512	EF585093
Festuca rubra L. **180**	Finland: A.Kosonen (Hb.Univ.Oulu) JACA JA-474496	F	EF584968	EF593001	EF585097
Festuca ultramafica Connor **151**	New Zealand: Nelson, Windy Point; Lloyd K.M. 57629	G	AY524826	AY528938	EF585110
American II clade					
Dielsiochloa floribunda (Pilg.) Pilg. **166**	Bolivia: Dpto. La Paz, Cumbre near La Paz , Muller J., MA 721312	I	DQ539563	DQ631428	DQ631494
Festuca andicola Kunth **167**	Ecuador: Loja Saraguro; Stancik D. 3780 US-3428937	I	EF584922	EF592955	EF585009
Festuca calligera (Piper) Rydb. **157**	USA: New Mexico: Sangre Cristo Mts; Allred K. 8262	H1	EF584929	EF592962	EF585021
Festuca casapaltensis Ball **165**	Peru: Lima: Canta: Cordillera Viuda. Peterson 20289, Soreng, and Romashchenko. US	I	xxxx	xxxx	xxxx
Festuca cochabambana E.B.Alexeev **153**	Bolivia: Cochabamba: Prov. Chapare; Muller J. 9277	I	EF584931	EF592964	EF585026
Festuca compressifolia J.Presl **163**	Peru: Huancavelica: Tayacaja. Peterson P. M. 14144 & O. Tovar, MO 5750428	I	xxxx	xxxx	xxxx
Festuca cuzcoensis Stančík & P.M.Peterson **162**	Peru: Cuzco: Calca; Peterson P.M. & Refulgio-Rodriguez 16582	I	EF584932	EF592966	EF585029
Festuca fimbriata Nees **154**	Argentina: Misiones: Dpto. Apóstoles. Marcelo Kostling 44, UZ 498.08	J	xxxx	xxxx	xxxx
Festuca flacca Hack. ex E.B.Alexeev **138**	Ecuador: Pichincha: Amaguana: Volcano Pasochoa; Stancik D., US-3428939	I	EF584938	EF592971	EF585041
Festuca fragilis (Luces) B. Briceño **159**	Venezuela: Merida: Paramo de Piedras Blancas; Catalan P., MERC	I	AF532960	AF533059	EF585115
Festuca hieronymi Hack. **172**	Argentina: Córdoba: Yacanto de Calamuchita. Catalan P. UZ-383.08	I	xxxx	xxxx	xxxx
Festuca longivaginata Tovar **161**	Peru: Ancash: Pallasca: Huandoval; Peterson., Refulio Rodriguez 13922, US-3423002	I	EF584951	EF592983	EF585063

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
American II clade					
Festuca nardifolia Griseb. **170**	Argentina: Salta Abra: El Acay; Peterson P.M. 10379	I	EF584954	EF592986	EF585070
Festuca orthophylla Pilg. **169**	Bolivia: Tarija: Aviles: Copacabana; Muller J. 9245	I	EF584957	EF592989	EF585075
Festuca parodiana (St.-Yves) Nicora **160**	Bolivia: Tarija: Prov. O'Connor; (cult., Jena Univ.), Muller J.9261	I	EF584959	EF592991	EF585078
Festuca peruviana Infantes **164**	Peru: Lima: Canta: Cordillera Viuda. Peterson P.M. 20293	I	xxxx	xxxx	xxxx
Festuca pringlei St.-Yves **156**	Mexico: Zacatecas. Peterson P.M. 21440, US	H1	xxxx	xxxx	xxxx
Festuca rigescens (J. Presl) Kunth **168**	Peru: Ancash: Recuay: Cordillera Blanca; Peterson.,Refulio Rodriguez, US-3423058	I	EF584966	EF592999.1	xxxx
Festuca tolucensis Kunth **155**	Venezuela: Merida: Paramo de Piedras Blancas; Catalan P., UZ-99.2000, MERC	H1,I	EF584976	EF593038	EF585108
Festuca yalaensis Joch.Müll. & Catalán **171**	Argentina: Jujuy: Yala: Laguna Rodeo. Catalan & Müller, UZ-358.08	I	GQ849279	GQ849280	GQ849281
Afroalpine clade					
Festuca abyssinica Hochst. ex A. Rich._A **191**	Uganda: Gabinga. Namaganda 192Vg, MHU1604	C	JQ972952	JQ972976	JQ973016
Festuca abyssinica Hochst. ex A. Rich._B **186**	Kenya: Mt Kenya. Namaganda 1741a	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._C **185**	Tanzania: Kilimanjaro. Afroalp PopSpecNo TZ-0252-3	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._D **184**	Uganda: Rwenzori. Afroalp PopSpecNo UG-2547-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._E **190**	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._F **189**	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-4	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._G **183**	Tanzania: Meru. Afroalp PopSpecNo TZ-0441-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._H **182**	Tanzania: Kilimanjaro. Afroalp PopSpecNo TZ-0095-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._I **188**	Ethiopia: Bale. Afroalp PopSpecNo ET-0941-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._J **179**	Tanzania: Meru. Afroalp PopSpecNo TZ-0441-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._K	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-3	C	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and sub-generic classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Afroalpine clade					
Festuca abyssinica Hochst. ex A. Rich._A **191**	Uganda: Gabinga. Namaganda 192Vg, MHU1604	C	JQ972952	JQ972976	JQ973016
Festuca abyssinica Hochst. ex A. Rich._B **186**	Kenya: Mt Kenya. Namaganda 1741a	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._C **185**	Tanzania: Kilimanjaro. Afroalp PopSpecNo TZ-0252-3	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._D **184**	Uganda: Rwenzori. Afroalp PopSpecNo UG-2547-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._E **190**	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._F **189**	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-4	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._G **183**	Tanzania: Meru. Afroalp PopSpecNo TZ-0441-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._H **182**	Tanzania: Kilimanjaro. Afroalp PopSpecNo TZ-0095-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._I **188**	Ethiopia: Bale. Afroalp PopSpecNo ET-0941-1	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._J **179**	Tanzania: Meru. Afroalp PopSpecNo TZ-0441-5	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._K 1**78**	Uganda: Muhavura. Afroalp PopSpecNo UG-2070-3	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._M **181**	Kenya: Aberdare. Afroalp PopSpecNo KN-0704	C	xxxx	xxxx	xxxx
Festuca abyssinica Hochst. ex A. Rich._N **180**	Ethiopia: Bale. Afroalp PopSpecNo ET-0006-x	C	xxxx	xxxx	xxxx
Festuca borbonica Spreng. **200**	Mascarenes: Reunion Island, Besnard	B	xxxx	xxxx	xxxx
Festuca camusiana St.-Yves **199**	Madagascar: Andringitra National Park, Vorontsoba, F.84	B	xxxx	xxxx	xxxx
Festuca chodatiana (St.-Yves) E.B.Alexeev **176**	Uganda: Elgon Mnt. Namaganda 322E	C	xxxx	xxxx	xxxx
Festuca claytonii E.B.Alexeev **177**	Uganda: Elgon Mnt. Namaganda 271E	C	xxxx	xxxx	xxxx
Festuca gilbertiana E.B. Alexeev ex S.M. Phillips B **195**	Ethiopia: Simen, Namaganda	C	xxxx	xxxx	-
Festuca gilbertiana E.B. Alexeev ex S.M. Phillips C **194**	Ethiopia: Simen, Namaganda	C	xxxx	xxxx	xxxx
Festuca gilbertiana E.B. Alexeev ex S.M Phillips A**187**	Ethiopia: Simen, Namaganda	C	xxxx	xxxx	xxxx
Festuca macrophylla Hochst. ex A.Rich. A **193**	Ethiopia: Simen. Afroalp PopSpecNo ET-0459-2	C	xxxx	xxxx	xxxx

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	Source	OAs	GenBank accession		
			ITS	*trn*LF	*trn*TL
Afroalpine clade					
Festuca macrophylla Hochst. ex A.Rich. C **192**	Ethiopia: Simen. Afroalp PopSpecNo ET-0144-2	C	xxxx	xxxx	xxxx
Festuca perrieri A. Camus **198**	Madagascar: Andringitra National Park, Vorontsoba, F.85	B	xxxx	xxxx	xxxx
Festuca pilgeri St.-Yves_B **174**	Kenya: Mt Kenya. Afroalp PopSpecNo KN-1053-3	C	xxxx	-	xxxx
Festuca pilgeri St.-Yves_C **175**	Kenya: Mt Kenya. Afroalp PopSpecNo KN-0833	C	xxxx	xxxx	xxxx
OUTGROUPS					
Poeae R.Br – Aveneae Dumort.					
Arrhenatherum elatius (L.) P. Beauv. ex J. Presl & C. Presl	Spain. Badajoz: Embalse de la Tentudis. UZ 55.07, GenBank		FM179388.1	EF137591.1	DQ336866.1
Avenula bromoides (Gouan) H. Scholz	Spain. Zaragoza: Vedado de Peñaflor. UZ 133.07,		JQ972940	DQ631459.1	DQ631525.1
Deschampsia cespitosa (L.) P.Beauv.	USA. Colorado: Boulder Co: Rocky Mnts; Catalan P. UZ		AF532929	AF533026	EF584984
Koeleria vallesiana (Honck.) Gaudin	Spain. Madrid: Torrelaguna-El Verrueco. UZ 4.07		DQ539601.1	DQ631468.1	DQ631536.1
Parapholis incurva (L.) C.E. Hubb.	Spain. Zaragoza: Vedado de Peñaflor. UZ 127.07, GenBank		FM179422.1	AF533036.1	DQ631491.1
Phleum pratense L.	Spain. Ciudad Real: Fuencaliente, GenBank		HQ600524.1	DQ353964.1	DQ353964.1
Poa alpina L.	Andorra: Ordino. Llamas, Acedo & Alonso, 62, LEB		EU792390	AY504635	DQ353986
Triplachne nitens (Guss.) Link	Spain. Almeria: Cabo de Gata. UZ 365.07		DQ336816.1	DQ336836.1	DQ336861.1
Bromeae Dumort. – Triticeae Dumort.					
Bromus diandrus Roth	GenBank		AY367936.1	AB732924.1	EU036144.1
Hordeum vulgare L.	GenBank		FJ593180.1	KF600708.1	KF600708.1
Triticum aestivum L.	GenBank		DQ981410.1	AB732940.1	JQ973001.1
Brachypodieae Harz					
Brachypodium distachyon (L.) P.Beauv.	Spain: Caceres. UZ 28.07		AF303399	AF478500	DQ336855
Stipeae Dumort.					
Stipa offneri Breistr.	GenBank		FN434569.1	JQ972961.1	JQ972997.1
Piptochaetium J. Presl sp.	GenBank		FJ461290.1	GU192025.1	DQ887480.1
Nassella (Trin.) E. Desv. sp.	GenBank		EU489159.1	DQ887450.1	KC904358.1
Achnatherum calamagrostis (L.) P. Beauv.	GenBank		GU254638.1	GU254940.1	JQ972996.1

Table S1. Loliinae and outgroup taxa included in the study: Data on source of samples, adscription to selected Operational Areas (OAs), and Genbank accession codes of the nuclear and plastid DNA sequences used in the phylogenetic analysis. Accessions with bold numbers after the name of the species correspond to those used in the biogeographic analyses. Each number is used in Figure 1 to represent the current location and distribution of each species. xxxx indicates sequences to be submitted to GenBank**.** The nomenclatural, supertribal, tribal, generic and subgeneric classification and phylogenetic adscriptions of BEP, Pooideae and Loliinae taxa and lineages are those proposed by Torrecilla *et al*. (2003; 2004), Catalan *et al*. (2006), and Soreng *et al*. (2007). UZ: University of Zaragoza herbarium. sp. indicates that sequences from more than one species have been collected for that genus.

Taxon	**Source**	**OAs**	**GenBank accession**		
			ITS	***trn*LF**	***trn*TL**
Nardeae W.D.J. Koch					
Nardus stricta L.	GenBank		EU489143.1	EU434097.1	JQ972994.1
Bambusoideae Luerss.					
Bambusa vulgaris Schrad. ex J.C. Wendl.	GenBank		FJ410319.1	JX428443.1	FJ644133.1

Table S2. Tree statistics of Maximum Parsimony (MP), BEAST and Lagrange analyses. Consistency index following Kluge and Farris (1969), Retention Index following Farris (1989). * Bayesian phylogeny based on the ITS+*trn*TLF data set excluding those accessions that were causing topological conflict between the nuclear and plastid phylogenies. ** BEAST analyses were based on an ITS+*trn*TLF reduced matrix that excluded those accessions that were causing topological conflict between the nuclear and plastid phylogenies, and included a maximum of three samples for each species. *** Lagrange analyses were based on the MCC (Maximum Clade Credibility) tree after the exclusion of the 18 outgroups.

	ITS	***trn*TL**	***trn*LF**	***trn*TLF**	**Bayesian ITS+cpDNA***	**BEAST ITS+cpDNA****	**Lagrange analyses*****
no. species	196	187	194	189	182	182	164
no. accessions	232	222	230	208	218	192	174
Aligned nucleotide positions	603	770	840	1945	2548	2548	-
Parsimony Informative characters	277	-	-	484	750	-	-
Length (L)	1990	-	-	1740	3521	-	-
Consistency Index (CI)	0.334	-	-	0.650	0.405	-	-
Retention Index (RI)	0.698	-	-	0.872	0.796	-	-
12 OAs M0 Likelihood value (-Ln)	-		-	-	-	-	459.7
12 OAs M1 Likelihood value (-Ln)	-	-	-	-	-	-	452.4
10 OAs M0 Likelihood value (-Ln)	-	-	-	-	-	-	421.8
10 OAs M1 Likelihood value (-Ln)	-	-	-	-	-	-	**402.6**
12 OAs M0 / M1 Extinction rate (e)	-	-	-	-	-	-	e_{M0}=0.01096 /e_{M1}=0.01596
12 OAs M0 / M1 Dispersion rate (d)		-		-	-	-	d_{M0}=0.009826/d_{M1}=0.02469
10 OAs M0 / M1 Extinction rate (e)		-		-	-	-	e_{M0}=0.006831/e_{M1}=**0.01506**
10 OAs M0 / M1 Dispersion rate (d)	-	-	-	-	-	-	d_{M0}=0.01108 /d_{M1}=**0.04325**

Table S3. Dispersal rate matrices reflecting the palaeogeographic connectivity among the Operational areas (OAs) selected in each historical scenario (time slices TSI, TSII, TSIII, TSIV). OAs: A – southern Africa. B – Madagascar. C – High mountains of tropical Africa. D1 – Western Mediterranean. D2 – Macaronesian archipelagos. E – Eastern Mediterranean areas and Southwest Asia. F – Eurasia. G – New Zealand and southwestern Australia. H1 – North America and Mesoamerica. H2 – Hawaiian archipelago. I – northern Andes. J – southern Andes and southern South America (see Fig. 1 for more details on each OA). Dispersal rates between the areas: distant regions and those very unlikely to be connected = 0.01; regions with sparse connection = 0.25; for regions that could be connected but only through abiotic factors such as wind or ocean currents = 0.5; close regions with elevated probability of connection = 0.75; areas adjacent or very close = 1)

TSI (Late Oligocene - Middle Miocene; 28.4 - 16 Ma)

Ten Operational Areas

	A	B	C	D	E	F	G	H	I	J
A	-	0.50	0.75	0.25	0.25	0.01	0.01	0.01	0.01	0.01
B		-	0.75	0.01	0.25	0.01	0.01	0.01	0.01	0.01
C			-	0.25	0.75	0.25	0.01	0.01	0.01	0.01
D				-	0.75	1.0	0.01	0.01	0.01	0.01
E					-	1.0	0.50	0.01	0.01	0.01
F						-	0.50	0.75	0.01	0.01
G							-	0.01	0.01	0.01
H								-	0.50	0.25
I									-	0.75
J										-

Twelve Operational Areas

	A	B	C	D1	D2	E	F	G	H1	H2	I	J
A	-	0.50	0.75	0.25	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
B		-	0.75	0.01	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
C			-	0.25	0.01	0.75	0.25	0.01	0.01	0.01	0.01	0.01
D1				-	0.50	0.75	1.0	0.01	0.01	0.01	0.01	0.01
D2					-	0.01	0.01	0.01	0.01	0.01	0.01	0.01
E						-	1.0	0.50	0.01	0.01	0.01	0.01
F							-	0.50	0.75	0.25	0.01	0.01
G								-	0.01	0.01	0.01	0.01
H1									-	0.25	0.50	0.25
H2											0.01	0.01
I											-	0.75
J												-

TSII (Middle Miocene to Late Miocene; 16.1 - 7.2 Ma)

Ten Operational Areas

	A	B	C	D	E	F	G	H	I	J
A	-	0.50	0.75	0.25	0.25	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.25	0.01	0.01	0.01	0.01	0.01
C			-	0.50	0.75	0.01	0.01	0.01	0.01	0.01
D				-	0.75	1.0	0.01	0.01	0.01	0.01
E					-	1.0	0.25	0.50	0.01	0.01
F						-	0.25	0.75	0.01	0.01
G							-	0.01	0.01	0.01
H								-	0.75	0.25
I									-	1.00
J										-

Twelve Operational Areas

	A	B	C	D1	D2	E	F	G	H1	H2	I	J
A	-	0.50	0.75	0.25	0.25	0.25	0.01	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
C			-	0.50	0.25	0.75	0.01	0.01	0.01	0.01	0.01	0.01
D1				-	0.75	0.75	1.0	0.01	0.01	0.01	0.01	0.01
D2					-	0.25	0.01	0.01	0.01	0.01	0.01	0.01
E						-	1.0	0.25	0.50	0.01	0.01	0.01
F							-	0.25	0.75	0.25	0.01	0.01
G								-	0.01	0.01	0.01	0.01
H1									-	0.25	0.75	0.25
H2										-	0.01	0.01
I											-	1.00
J												-

Table S3. Dispersal rate matrices reflecting the palaeogeographic connectivity among the Operational areas (OAs) selected in each historical scenario (time slices TSI, TSII, TSIII, TSIV). OAs: A – southern Africa. B – Madagascar. C – High mountains of tropical Africa. D1 – Western Mediterranean. D2 – Macaronesian archipelagos. E – Eastern Mediterranean areas and Southwest Asia. F – Eurasia. G – New Zealand and southwestern Australia. H1 – North America and Mesoamerica. H2 – Hawaiian archipelago. I – northern Andes. J – southern Andes and southern South America (see Fig. 1 for more details on each OA). Dispersal rates between the areas: distant regions and those very unlikely to be connected = 0.01; regions with sparse connection = 0.25; for regions that could be connected but only through abiotic factors such as wind or ocean currents = 0.5; close regions with elevated probability of connection = 0.75; areas adjacent or very close = 1)

TSIII Late Miocene (Messinian) – Pliocene (7.3 – 2.6 Ma)

Ten Operational Areas

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
A	-	0.50	0.75	0.25	0.25	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.01	0.01	0.01	0.01	0.01	0.01
C			-	0.5	0.75	0.01	0.01	0.01	0.01	0.01
D				-	1.0	1.0	0.01	0.01	0.01	0.01
E					-	1.0	0.25	0.01	0.01	0.01
F						-	0.50	0.50	0.01	0.01
G							-	0.01	0.01	0.01
H								-	1.00	0.50
I									-	1.00
J										-

Twelve Operational Areas

	A	**B**	**C**	**D1**	**D2**	**E**	**F**	**G**	**H1**	**H2**	**I**	**J**
A	-	0.50	0.75	0.25	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
C			-	0.5	0.01	0.75	0.01	0.01	0.01	0.01	0.01	0.01
D1				-	0.75	1.0	1.0	0.01	0.01	0.01	0.01	0.01
D2					-	0.25	0.01	0.01	0.01	0.01	0.01	0.01
E						-	1.0	0.25	0.01	0.01	0.01	0.01
F							-	0.50	0.50	0.25	0.01	0.01
G								-	0.01	0.01	0.01	0.01
H1									-	0.25	1.00	0.50
H2										-	0.25	0.01
I											-	1.00
J												-

TSIV Quaternary period (2.61 – 0 Ma)

Ten Operational Areas

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
A	-	0.50	0.75	0.25	0.25	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.25	0.01	0.01	0.01	0.01	0.01
C			-	0.25	0.75	0.01	0.01	0.01	0.01	0.01
D				-	0.75	1.0	0.01	0.01	0.01	0.01
E					-	1.0	0.50	0.50	0.01	0.01
F						-	0.75	1.0	0.01	0.01
G							-	0.01	0.01	0.01
H								-	1.0	0.75
I									-	1.0
J										-

Twelve Operational Areas

	A	**B**	**C**	**D1**	**D2**	**E**	**F**	**G**	**H1**	**H2**	**I**	**J**
A	-	0.50	0.75	0.25	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
B		-	0.50	0.01	0.01	0.25	0.01	0.01	0.01	0.01	0.01	0.01
C			-	0.25	0.01	0.75	0.01	0.01	0.01	0.01	0.01	0.01
D1				-	0.75	0.75	1.0	0.01	0.01	0.01	0.01	0.01
D2					-	0.25	0.01	0.01	0.01	0.01	0.01	0.01
E						-	1.0	0.50	0.50	0.01	0.01	0.01
F							-	0.75	1.0	0.25	0.01	0.01
G								-	0.01	0.01	0.01	0.01
H1									-	0.25	1.0	0.75
H2										-	0.25	0.01
I											-	1.0
J												-

GENERAL CONCLUSIONS

1. The complex evolutionary dynamics of the LCNG *ß-amylase* and the mitochondrial *rps3* gene in the Poaceae family have been demonstrated through the detection of a plethora of sequences with likely pseudogenization, paralogy, homeology, recombination and misplacement events, which could be interpreted as indicative of relaxation of selection pressure.

2. The Miniature Inverted Repeat Transposable Elements (MITEs) found in the introns and intergenic spacers of the *ß-amylase, Xylose isomerase* (*xly*), *Barley leucine zipper* (*blz-1*), *Nucellin* (*nuc*) and *Disrupted meiotic cDNA* (*dmc1*) genes throughout the Pooideae lineages make the interpretation of the evolutionary dynamics of these nuclear genes still more intricate. Different waves of acquisition followed by multiple losses and horizontal transfers appears to be the most likely hypothesis for their existence within the temperate grasses studied.

3. The identification and evaluation of genes and sequences under relaxation of selection pressure and recombination has been demonstrated to be a key step for using molecular markers to disentangle the phylogenetic history of the Poaceae. This dissertation provides a number of striking examples that show that these singular evolutionary patterns appear more frequently in the grasses than previously thought. The commonly applied model of gene evolution in plants is currently limited to speciation, duplication, and loss events. However nature is much more complicated, and molecular events such as horizontal gene transfer also affect the evolution of angiosperms, as demonstrated in the studied grasses.

4. Other grass phylogenies presented in this dissertation, based on nuclear (ITS, *ß-amylase*) and plastid (*mat*K, *ndh*F, *trnH-psbA*, *trn*TL and *trn*LF) markers, were largely congruent with previously published results and rendered an optimal and well-supported tree at all taxonomic levels, with a complex and consistent internal structure. Within the more extensively sampled BEP clade, a sister relationship was recovered for Bambusoideae and Pooideae, while Diarrheneae was recovered as sister to the Brachypodieae + core pooids clade. We also recovered five new lineages, mostly corresponding to austral Loliinae

groups: Eurasian – South American clade, South African – Central American clade, tropical African – South African clade, American Vulpia + Pampas clade, and the Afroalpine clade.

5. Our fossil-calibrated dated analysis of Pooideae indicated that the earliest diversification within the BEP clade took place between the Middle Paleocene and the Early Eocene, and seems to be related to a period of global warming. Nevertheless, the development of a cooler, dryer climate during the Late Eocene-Oligocene period concurs with the opening of grasslands and likely explains the increase in the rate of diversification detected in the core pooids. The Loliinae originated during the Late Oligocene, and most of the divergences of the main Loliinae lineages occurred from the Middle to the Late Miocene, though the most recent fine-leaved lineages split in the Pliocene.

6. The biogeographic scenarios reconstructed for the Loliinae suggest that most ancestors spread by recurrent long-distance dispersals, neocolonizations, and recolonizations, which generally occurred from the Northern to the Southern Hemisphere, but also in the opposite direction and within the Southern Hemisphere. The broad-leaved fescues presented an unequivocal Western Mediterranean ancestor. However, the fine-leaved group showed two independent origins: in the Western Mediterrean, and in the Patagonian regions.

CONCLUSIONES GENERALES

1. La compleja y singular dinámica evolutiva del gen nuclear copia simple (LCNG) *ß-amylase* y del gen mitocondrial *rps3* en Poaceae ha sido corroborada a través de la detención de multitud de secuencias bajo pseudogenización, paralogía, homeología, recombinación y desplazamientos filogenéticos.
2. La existencia de MITEs (Miniature Inverted Repeat Transposable Elements) en regiones intrónicas y espacios intergénicos de los genes *ß-amylase, Xylose isomerase* (*xly*), *Barley leucine zipper* (*blz-1*), *Nucellin* (*nuc*) y *Disrupted meiotic cDNA* (*dmc1*) nos ayuda a interpretar la compleja dinámica evolutiva del genoma nuclear de las gramíneas, el cual ha resultado ser mucho más complejo e intrincado de lo esperado. La hipótesis más plausible que explica la existencia de dichos elementos transponibles dentro de las Loliinae templadas es su adquisición independiente en diferentes momentos, seguida de múltiples pérdidas y transferencias horizontales.
3. La identificación y evaluación de genes y secuencias bajo recombinación y con una presión evolutiva negativa es un paso clave en el uso de marcadores moleculares como fuente de datos en reconstrucciones filogenéticas. La existencia de dichos singulares eventos evolutivos ha resultado ser mucho más común de lo esperado en gramíneas. Quizá estemos en los albores de entender que el modelo de evolución génica en plantas, actualmente limitado a especiaciones, duplicaciones y pérdidas de genes, es mucho más amplio debiéndose incorporar a la rutina investigadora fenómenos como microconversiones, o transferencias horizontales de genes entre especies alejadas filogenéticamente.
4. Las filogenias presentadas en esta tesis, basadas en marcadores nucleares (ITS, *ß-amylase*) y plastídicos (*mat*K, *ndh*F, *trnH-psbA*, *trn*TL and *trn*LF), fueron en general altamente congruentes con los trabajos publicados hasta la fecha. Cabe destacar las nueva relación observada entre Brachypodieae + core pooids clade, y los cinco nuevos linages descritos correspondientes mayormente a Loliinae Australes: el clado Euroasiatico – Sur Americano, el clado sur Africano – centro Americano, el clado de Africa tropical – sur Africano, el clado de Vulpias Americanas + Pampas, y el clado Afroalpino.
5. Las dataciones de Pooideae calibradas con fósiles indican que la diversificación más temprana dentro del clado BEP ocurrió entre el Paleoceno medio y el Eoceno temprano. Dichas diversificaciones parecen estar relacionadas con un periodo de calentamiento

global. Sin embargo, el desarrollo de un clima más frío y seco durante el Eoceno tardío y el Oligoceno favoreció la diversificación de las Pooideae debido probablemente a la formación de grandes espacios abiertos en forma de prados y praderas. Nuestras estimaciones indican que la subtribu Loliinae se originó durante el Oligoceno tardío, y divergió fundamentalmente durante el Mioceno medio y el tardío.

6. Los escenarios biogeográficos reconstruidos para la subtribu Loliinae sugieren que la mayor parte de sus ancestros se dispersaron mediante neocolonizaciones, recolonizaciones y migraciones a larga distancia desde el hemisferio norte al sur, pero también en dirección opuesta e incluso dentro del hemisferio sur. Las "festucas de hoja ancha" presentaron un inequívoco ancestro común en el Mediterráneo occidental. Sin embargo, el grupo de las "festucas de hoja fina" muestran dos orígenes independientes: el Mediterráneo occidental y la región Patagónica.

Minaya, M., Pimentel, M., Mason-Gamer, R., Catalan, P., 2013. Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses. *Molecular Phylogenetics and Evolution* 68, 106-118.

Molecular Phylogenetics and Evolution 68 (2013) 106–118

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Distribution and evolutionary dynamics of *Stowaway* Miniature Inverted repeat Transposable Elements (MITEs) in grasses

Miguel Minaya [a], Manuel Pimentel [b], Roberta Mason-Gamer [c], Pilar Catalan [a,*]

[a] Departmento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior Huesca, Universidad Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain
[b] Department of Plant and Animal Biology and Ecology, Facultad de Biología, Universidad de Coruña, 15071 A Coruña, Spain
[c] UIC Biological Sciences, SEL 1008 M/C 067, 840 West Taylor Street, Chicago, IL 60607, United States

ARTICLE INFO

Article history:
Received 17 September 2012
Revised 2 March 2013
Accepted 6 March 2013
Available online 16 March 2013

Keywords:
Beta amylase
MITEs
Poaceae
BEP and PACCMAD clades
Temperate grasses
Transposable elements

ABSTRACT

The occurrence of *Stowaway* MITEs and their potential footprints in the grasses was assessed within an explicit phylogenetic framework. An organismal tree was used to analyze the distribution and evolutionary dynamics of these elements and their potential excision footprints in the fourth intron of the *β-amylase* gene and in other introns of several nuclear genes across the Poaceae. Megablast and discontiguous megablast searches in the Entrez nucleotide database were performed for the *β-amylase*, *blz-1*, *dmc1*, *nuc*, and *xly* genes MITEs. These elements and their potential footprints were distributed in introns and intergenic spacers of many other nuclear genes throughout the BEP lineages; however, they were absent in the studied PACCMAD lineages.

A plausible underlying dynamic of successive acquisitions and deletions of *β-amylase Stowaway* MITEs in the temperate grasses could be explained by three alternative hypotheses: (i) a single early acquisition of a palindrome element, similar to *Tc1-Mariner*, in the fourth intron of the *β-amylase* gene in the ancestor of the Pooideae, followed by multiple independent losses, (ii) multiple independent acquisitions of MITEs in non-related pooid lineages or (iii) different waves of acquisition of MITEs, followed by multiple losses and horizontal transfers in the temperate grasses. This last hypothesis seems to fit best with the evidence found to date.

1. Introduction

Miniature Inverted repeat Transposable Elements (MITEs) are a heterogeneous group of small, non-autonomous mobile elements. They have between a few dozen and a few hundred base pairs, and are found in both prokaryotic and eukaryotic genomes. Numerically, they are the most abundant of all types of transposable elements in plants, with copy numbers in the thousands (Feschotte et al., 2002a, 2003; Jiang et al., 2004). MITEs are closely associated with euchromatic single-copy genes (Lu et al., 2012, and references therein). The close physical association between MITEs and plant genes has caused some authors to hypothesize that MITEs might play a role in gene regulation (Bureau and Wessler, 1994; Bureau et al., 1996; Mao et al., 2000; Wicker et al., 2007; Zhang et al., 2004). Recently, Kuang et al. (2009) revealed a novel mechanism by which MITEs could affect gene regulation via a small RNA pathway. They also identified several MITEs that have the capacity to alter the structure of the genes.

MITEs likely originated from defective remnants of a subset of much longer autonomous class II DNA transposons (Feschotte and Wessler, 2002; Feschotte et al., 2002a,b, 2003, 2005; Macas et al., 2005; Menzel et al., 2006; Turcotte and Bureau, 2002; Zhang et al., 2004) that lost their transposase-encoding ORFs due to internal deletions (Flavell et al., 1994). MITEs have no coding capacity, and thus their mobility depends on the activity *in trans* of transposases encoded by cognate full-length autonomous class II transposons (Feschotte et al., 2005; Jiang et al., 2003, 2004) that recognize a common terminal inverted repeated (TIR) sequence. These specific transposases catalyze both the insertion and the excision of their associated transposons (Feschotte et al., 2002a,b; Jiang et al., 2003; Kikuchi et al., 2003; Nakazaki et al., 2003; Yang et al., 2007). Thus, MITEs should be expected to excise with a certain frequency, and—as is the case in class II transposons—leave footprints upon excision. Neither substitutions nor deletions lead to compensatory changes; hence, the highly stable secondary structure of the elements may gradually be reduced (Feschotte et al., 2002b; Petersen and Seberg, 2000). Consequently, the dynamics underlying the acquisition, loss, and high copy number of MITEs are still obscure (Mason-Gamer, 2007).

* Corresponding author. Fax: +34 974 239 302.
E-mail addresses: miguel.minay@gmail.com (M. Minaya), mpimentel@udc.es (M. Pimentel), robie@uic.edu (R. Mason-Gamer), pcatalan@unizar.es (P. Catalan).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.03.005

Plant *Stowaway* elements of the *Mariner* family are short non-autonomous sequences derived from class II DNA transposons (Wicker et al., 2007). These elements vary considerably in length, from 59 to 323 bp (Bureau and Wessler, 1994; Feschotte et al., 2002b, 2003, 2005; Macas et al., 2005; Turcotte and Bureau, 2002), and can account for up to 2% of the entire genome (e.g., *Oryza*; Mao et al., 2000). *Stowaway* MITEs are flanked by TIRs with the almost invariant 5′-CTCCTCCC···GGGAGGAG-3′ pattern and, like all *Tc1-Mariner* elements, always produce a TA target site duplication (TSD) upon insertion (Wicker et al., 2007). Class II transposable elements can increase their numbers by transposing during chromosome replication (Greenblatt and Brink, 1962). Alternatively, they can exploit gap repair following excision to create an extra copy at the donor site (Nassif et al., 1994). Transposition is mediated by a transposase that recognizes the TIRs and cuts both strands at each end (Wicker et al., 2007). An interaction between *Mariner*-like transposases and *Stowaway* MITEs' TIRs was detected in rice (Feschotte et al., 2005), giving insight into the mechanism that is likely responsible for the evolutionary dynamics of these mobile elements. Recently, Yang et al. (2009) demonstrated that the high transposition activity achieved by rice MITEs was caused by their competitive use of the transposases encoded by related autonomous class II transposons. Comparative genomic analysis of *Oryza sativa* genomes (subsp. *japonica* and *indica*) have shown evidence of several amplification bursts of MITEs at different evolutionary times, as well as their important regulatory role in gene expression and their potential influence in species diversification (Lu et al., 2012).

The distribution and long-term evolution of *Stowaway* MITEs can be traced by examining their sequences in the context of a well-known phylogeny, such as the Triticeae (Mason-Gamer, 2007; Petersen and Seberg, 2000, 2009). Mason-Gamer (2007) found up to seven non-related *Stowaway* MITEs in different Triticeae genera at the fourth intron of the *β-amylase* gene. Four of them showed greater similarity to elements in other genes or in other Triticeae genomes than to each other. The polyphyletic distribution of the element was also confirmed by the detection of five distinct putative excision footprints and two distinct deletions of excision-flanking regions in both related and unrelated species, revealing a complex history of gains and losses of *Stowaway β-amylase* MITEs in the wheat tribe (Mason-Gamer, 2007).

In order to further investigate the distribution and evolutionary dynamics of the *β-amylase Stowaway* MITEs, an explicit phylogenetic framework was used to analyze the evolutionary patterns of sequence diversity and presence vs. absence of *Stowaway* MITEs—and their potential footprints—at the fourth intron of the *β-amylase* gene in a broader representation of grass lineages, with the temperate pooids sampled more exhaustively. For this analysis, an organismal tree constructed from nuclear (*β-amylase,* ITS) and plastid (*mat*K, *ndh*F, *trn*TL, *trn*LF) sequences, was used as reference. Our study aimed to test whether these mobile elements and their footprints reflect the organismal phylogeny, or could have been acquired from other non-related lineages through horizontal gene transfer (HGT) events or from intragenomic mobility. Additionally, we assessed the presence versus absence of other *Stowaway* MITEs in the *blz*1, *dmc*1, *nuc* and *xyl* genes of the Triticeae (Petersen and Seberg, 2000, 2009) in the current genomic data repositories (e.g. Entrez nucleotide database). Our goal was to compare the patterns of distribution of different MITEs in the major BEP and PACCMAD lineages of Poaceae. Particularly, we wanted to test whether the presence of *β-amylase* MITEs in the BEP clade and their absence in the PACCMAD clade (see Section 3) would be congruent with similar distributions of different MITEs from other nuclear genes, and whether this would add support to the existence of an 'evolutionary barrier' for the transference of nuclear MITEs between representatives of these two main grass lineages.

2. Material and methods

2.1. Taxon sampling

The sample includes a broad representation of Pooideae taxa, especially of the more recently evolved Aveneae/Poeae and Bromeae/Triticeae core-pooid lineages, which show high levels of divergence and evolutionary rates (Catalan, 2006; Catalan et al., 2006; GPWG, 2001; Quintanar et al., 2007; Mason-Gamer, 2005). Several representatives of the Ehrhartoideae and of the PACCMAD clade were also included in the study, to determine whether the *β-amylase Stowaway* MITEs are found in more distantly related grass lineages.

The sampling covered 117 species representing 82 genera and six subfamilies (GPWG, 2001; Soreng et al., 2007, and http://www.tropicos.org): Centothecoideae: 1 species, Panicoideae: 6, Danthonioideae: 4, Chloridoideae: 5, Ehrhartoideae: 2, and Pooideae: 100. Within Pooideae the sampling included the following tribes: Lygeeae/Nardeae: 2 species, Stipeae: 3, Diarrheneae: 1, Brachypodieae: 3, Bromeae/Triticeae: 22, and Aveneae/Poeae: 69. Thirteen species, including both diploid and polyploid taxa, were cloned (not shown): *Brachypodium distachyon* (9), *Festuca abyssinica* (2), *F. altaica* (7), *F. arundinacea* (7), *F. capillifolia* (9), *F. elegans* (7), *F. fenas* (7), *F. hystrix* (5), *F. panciciana* (2), *F. paniculata* (10), *F. rubra* (5), *F. scariosa* (8) and *Micropyropsis tuberosa* (2), and 28 sequences were downloaded from GenBank. Source or voucher information and GenBank accession numbers of all analyzed samples are indicated in Appendix A.

2.2. DNA sequencing and alignment

Doyle and Doyle's (1987) CTAB method was used to isolate DNA from silica-gel-dried leaves or from fresh materials for most of the studied samples. For herbarium samples, DNA extracts were obtained using the DNAeasy® Plant Mini Kit (QIAGEN Ltd., Wets Crawley, UK) procedure.

A 1360-nucleotide region, extending from exons 2 to 5 of the low-copy nuclear *β-amylase* gene, which encodes a ubiquitous tissue enzyme, was amplified and sequenced using the primers 2a-for, 2b-for, 2c-for, 2d-for, 2g-for, 3a-for, 3a-bac, 4b-for, 4a-bac, 5a-bac and 5b-bac, following the procedures indicated in Mason-Gamer (2005).

Amplification reactions were performed using a reaction mix containing 0.5 μl of forward and reverse primer (10 mM), 1 μl of 10× PCR buffer, 3 μl of $MgCl_2$ (50 mM), 0.3 μl of *Taq* DNA polymerase (5 U/μl), 1 μl of deoxynucleoside triphosphate set PCR grade (10 mM), between 0.5 and 1 μl of template DNA, and an aliquot of sterile ultrapure water (MilliQ) added to achieve a final volume of 10 μl. The amplification parameters consisted of an initial denaturing step of 1 min at 94 °C, followed by four cycles of 45 s denaturing at 94 °C, 2 min annealing at 65 °C, and 1 min extension at 72 °C, followed by 29 cycles of 30 s denaturing at 94 °C, 40 s annealing at 65 °C, and 40 s extension at 72 °C, with a final extension step of 7 min at 72 °C. All PCR products were cleaned using ExoSAP-IT® following the protocol indicated by the manufacturer.

PCR products of 14 species were cloned using pGEM-T Easy vectors (Promega), and transformed into *E. coli* JM109 competent cells (Promega) following the manufacturer's protocol, except that all reactions were halved. Between 2 and 10 cloned fragments were amplified directly from white colonies using the same primers and recipe as were used for the original PCR. However, subsequent sequencing was done from direct PCR products because: (i) all the clones from single individuals, either diploids or polyploids, were monophyletic, (ii) they showed low intraindividual variety, and (iii) there was an absence of clonal sequence variation at the fourth intron of the *β-amylase* gene.

β-amylase products were used as templates for the cycle sequencing reaction using the ABI Big-Dye Terminator Cycle Sequencing Kit® (Applied Biosystems). The external and internal primers listed above were used in this step. The sequencing reactions were prepared with 2 μl of Premix Ready Reaction, 2 μl of primer (5 mM), between 1 and 3 μl of purified DNA, and an aliquot of sterile ultrapure water (MilliQ) to achieve a final volume of 10 μl. PCR was performed with one denaturing cycle of 1 min at 96 °C, 24 cycles of 10 s denaturing at 96 °C, 5 s annealing at 50 °C, and 4 min extension at 60 °C, then a termination step of 4 min at 60 °C. Forward and reverse fragments were generated for each sample and contigs were assembled and manually adjusted using the program Sequencher® 4.2.2 (Genes Codes Corporation, Ann Arbor, Michigan, USA).

Alignment of the *β-amylase* sequences was done with Protein Multiple Sequence Alignment (MUSCLE) Software version 3.7 (Edgar, 2004) (http://www.ebi.ac.uk/Tools/msa/muscle/). The alignments were manually adjusted using the program MacClade 4.08 OS X (Maddison and Maddison, 2008). Amino acid translations were used to guide the nucleotide alignments. Alignments of the fourth *β-amylase* intron from the newly generated grass sequences were inspected visually and compared with some previously published *β-amylase* sequences in *Triticeae,* in which a collection of *Stowaway* MITEs had been found (Mason-Gamer, 2007). Following alignment, a concatenated data matrix of sequences from the six genes yielded 5485 nucleotide positions after removal of the highly variable or homoplasious positions (including the *β-amylase Stowaway* elements).

2.3. Organismal tree reconstruction

Phylogenetic distribution of the presence and absence of *Stowaway β-amylase* transposable elements and their target site footprints was assessed using a reference multigenic organismal tree of the Poaceae. The Poaceae organismal tree was constructed using the nuclear single-copy *β-amylase* gene, the multicopy ribosomal *ITS*1-5.8S-*ITS*2 region, and the plastid *mat*K, *ndh*F, *trn*TL and *trn*LF genes from a collection of 116 selected grass taxa plus the close relative *Joinvillea* (Joinvillaceae) included as an out-group (Appendix A). The ITS and *trn*TL–*trn*LF regions were amplified and sequenced using the primers and procedures from Catalan et al. (2004) and Quintanar et al. (2007); the *ndh*F gene protocols followed Catalan et al. (1997); and the *mat*K protocols followed Döring et al. (2007).

The potential combinability of the data sets was assessed using the Partition Homogeneity (PH) test (Incongruence Length Difference of Farris et al., 1994), implemented in PAUP 4.0 beta 10 (Swofford, 2002). The PH test was conducted through heuristic searches of 100 random-order-entry replicates, with TBR and MulTrees on to estimate if the separate and combined nuclear and plastid data sets were significantly different from random partitions of the same sizes.

Bayesian and Maximum parsimony (MP) based searches were performed for both the independent and the combined data sets using, respectively, *MrBayes 3.1.2* (Altekar et al., 2004; Ronquist and Huelsenbeck, 2003) and *PAUP** 4.0 beta10 (Swofford, 2002). All gaps were treated as missing data. Bayesian analysis of both the independent and the combined data sets was conducted according to the GTR+I+gamma model, which was selected as the optimal model for each partition using the tests of goodness of fit for alternative nucleotide substitution models performed through the Hierarchical Likelihood Ratio Test (hLRTs), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in jModelTest 2 (Darriba et al., 2012; Guindon and Gascuel, 2003). Two independent analyses were run with one million generations each using the Markov Chain of Monte Carlo (MCMC) algorithm. Trees were sampled every 1000 generations, and 25% of the generations were discarded as burn-in once stability in the likelihood values was attained. A half-compatible consensus Bayesian tree was computed from the 750 posterior probability saved trees. MP analysis was based on heuristic searches of 10,000 random-order-entry trees, with TBR branch swapping and saving no more than 10 trees of length equal to or shorter than 10 per replicate. The most parsimonious trees were used to compute the respective strict consensus trees. Branch support was estimated through 1000 bootstrap replicates (Felsenstein, 1985) using the TBR-M (Tree Bisection Reconstruction swapping, MULPARS OFF) strategy of DeBry and Olmstead (2000) as a method to reduce computational time. Following the criteria of Mason-Gamer and Kellogg (1997) and Inda et al. (2008), clades with bootstrap support values (BS) of 75–100% (*PAUP 4b.10-86-macosx*) or Posterior Probability support values (PPS) of 90–100% (*MrBayes 3.1.2*) were considered moderately to strongly supported.

2.4. Distribution and dynamics of the Stowaway MITEs

Megablast and discontiguous Megablast searches were performed in the Entrez nucleotide database using as queries the *Stowaway*-like sequences found in the *β-amylase* gene (Mason-Gamer, 2005, 2007, and current study). The searches aimed to target other potentially homologous *Stowaway* sequences present in other grass representatives or in other genes, and to retrieve information on their respective genomic locations. Similar searches were performed using as query sequences the *Stowaway*-like elements described by Petersen and Seberg (2000, 2009) in the *blz1, dmc1, nuc*, and *xyl* genes of the Triticeae in order to determine whether other described TEs show similar inter-tribal distribution patterns in the grass family as those found in the fourth intron of the *β-amylase* gene (see Section 3). In all cases we used a query coverage >80%, threshold *E*-values > 1e−08, and sequence similarity values >69% to compare their potential homologies. Searches were performed both including and excluding the TSD and the conserved TIR. The specific *Stowaway*-bearing genes were examined visually at each specific locus. The secondary structure of these elements was inspected through their minimum-energy folding values using MFOLD (Zuker, 2003) and their free energy values following SantaLucia (1998).

3. Results

3.1. The organismal tree of grasses

Evolutionary dynamics of the mobile *Stowaway* MITEs could only be confidently inferred within a solid phylogenetic framework and from their remaining conserved termini (Mason-Gamer, 2007; Petersen and Seberg, 2009). The Poaceae organismal tree (Fig. 1) was constructed by combining sequence data from the nuclear *β-amylase* (117 taxa/1143 aligned nucleotides) and ITS (112/609) genes, and the plastid *mat*K (99/1259), *ndh*F (102/710), *trn*TL (98/789) and *trn*LF (102/975) genes. GenBank accession numbers for all sequences are indicated in Appendix A.

The independent phylogenetic searches conducted with the six separate *β-amylase*, ITS, *mat*K, *ndh*F, *trn*TL and *trn*LF data sets yielded highly congruent topologies (results not shown except the *β-amylase* tree; Fig. 2). The PH test detected incongruence among the cpDNA data sets, except for *mat*K–*trn*TL ($P = 0.4$) and *ndh*F–*trn*TL ($P = 0.3$). Partition incongruence was also found between the nuclear *β-amylase* and ITS data ($P = 0.01$), and between the plastid and nuclear data ($P = 0.01$). Despite this, the separate topologies (e.g. *β-amylase* tree; Fig. 2) were congruent with the tree based on the six concatenated nuclear and plastid genes (Fig. 1), with incongruence observed mostly among some terminal tips. Furthermore,

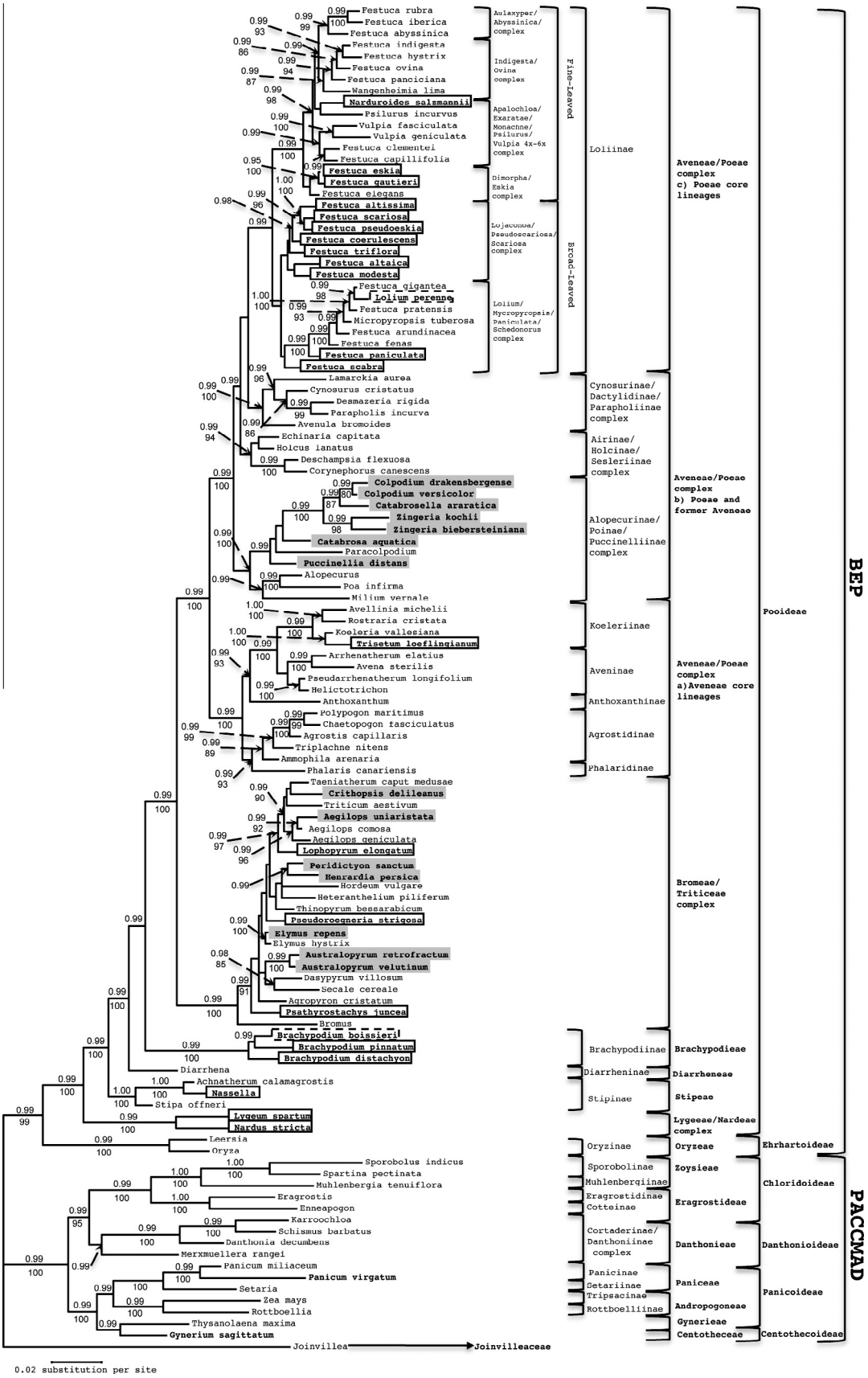


Fig. 1. Phylogenetic distribution of the *β-amylase Stowaway* MITEs and their potential footprints in the Poaceae organismal tree. Bayesian Inference estimation was based on nuclear (ITS, *β-amylase*) and plastid (*trn*TL, *trn*LF, *ndh*F, *mat*K) gene sequences. Bayesian posterior probabilities of ⩾95% are shown above the nodes, and MP bootstrap support of ⩾75% is shown below the nodes. Solid line boxes indicate species with putative degraded excision footprints within the two conserved TIR sequences of 10-bp. Gray boxes indicate species with *Stowaway* MITE footprints among the TA---GTA motifs. Dotted line boxes indicate species lacking the the TA---GTA motifs along with portions of the flanking intron. Black boldfaced species in the PACCMAD clade have different kinds of insertions in the same position.

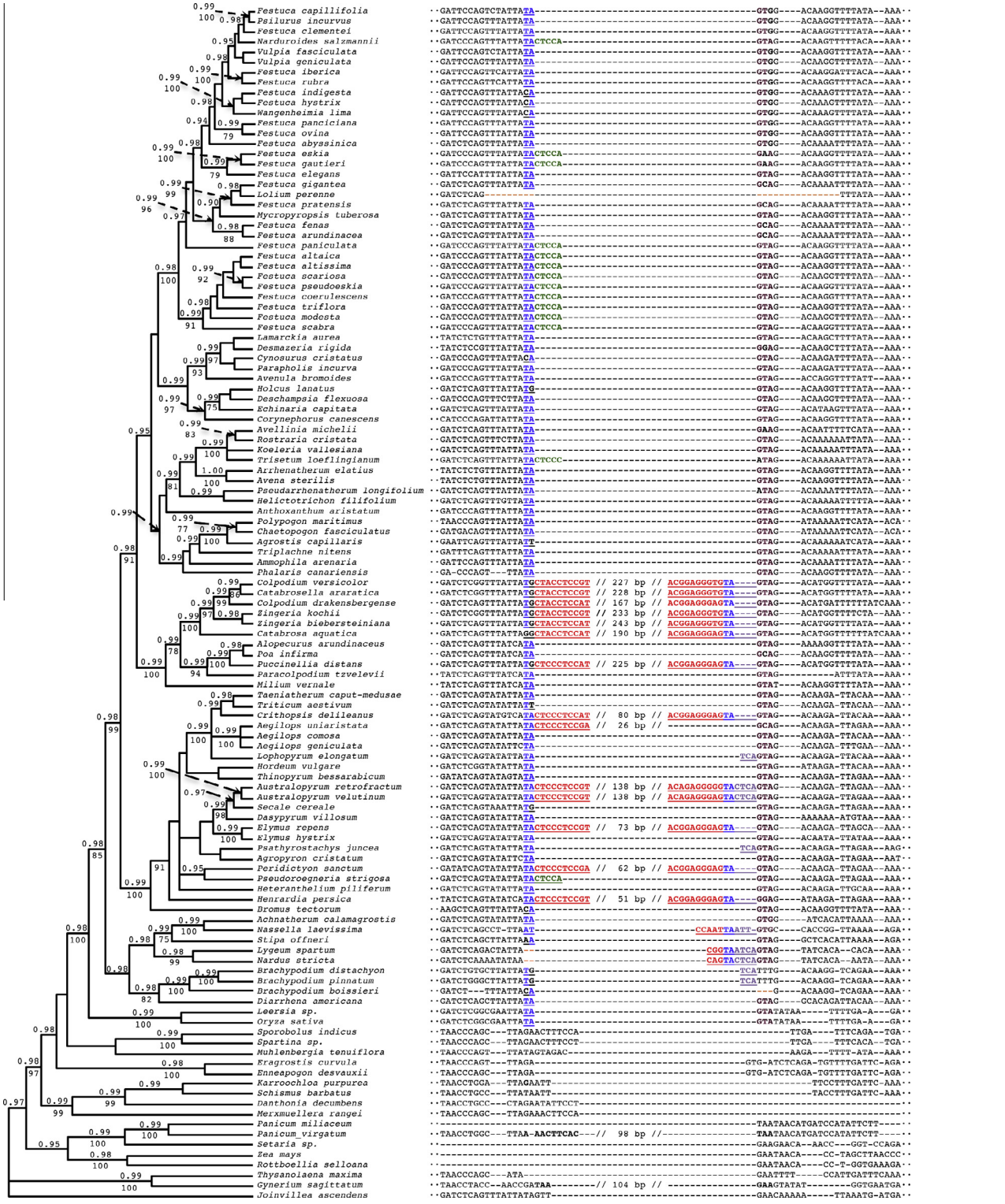


Fig. 2. *β-amylase* tree and alignment of the fourth intron region of the gene containing *Stowaway* elements and their footprints in the studied Poaceae species. Halfcompatible bayesian consensus heuristic tree based on a reduced unambiguous *β-amylase* data set (see text for details of the analyses); values above and below branches represent PPS and BS values, respectively. The data set shows: (i) the 2-bp target site (TA) underlined in blue boldface; (ii) the 10 bp TIR sequence underlined in red boldface; (iii) the likely degraded excision footprints adjacent to the 5′ flanking regions of the TIR sequence underlined in green boldface; (iv) the likely degraded excision footprints adjacent to the 3′ flanking regions of the TIR sequence underlined in purple boldface; (v) deletions including the target site in red boldface; (vi) the 3-bp (GTA) motif downstream from the *Stowaway* elements (common in the BEP clade) in brown boldface, and (vii) the lengths of the internal portions of the inserted elements inside double slashes.

those topological conflicts were not well supported. Since the incongruence did not affect the main results and conclusions of this research, we chose to conduct phylogenetic analyses on the combined concatenated data set of the six sequenced genes, which had a total length of 5485 characters, of which 2274 (42.45%) were parsimony informative. The BI tree (Fig. 1) and MP tree (not shown) from the combined data set are highly congruent with one another and with those obtained for the family by previous authors.

3.2. β-amylase Stowaway MITEs in the Poaceae

According to Wicker's et al. (2007) classification, the insertions found at the fourth intron of the *β-amylase* gene in the BEP clade of the Poaceae were classified in the class II (DNA transposon) – subclass 1, TIR order, *Tc1-Mariner* superfamily as non-autonomous *Stowaway* MITEs (but see also Seberg and Petersen (2009) and Wicker et al. (2009) for a discussion on the potential classification system of transposable elements) based on: (i) their two conserved TIR sequences of 10-bp (5′-CTMCCTCCRW···ACRGAGGGWG-3′), and the characteristic 2-bp TSD "TA" (Fig. 2); (ii) their high similarity (69) 80–100%) to other MITEs deposited in the Entrez nucleotide database (overall >80% Blast matches) (Appendix B); and (iii) their hairpin-like secondary structure (Fig. 3). The target site 5′-TA-3′ and the 5′-GTA-3′ motif were found in all the 100 species sampled within the BEP clade. *Stowaway* MITEs and other potential excision footprints within the 5′-TA---GTA-3′ region were found in 37 of the 100 species sampled within the BEP clade (Figs. 1 and 2; Appendix B). Within the PACCMAD clade, 2 out of the 16 surveyed species showed a different kind of insertion in the same position where the BEP-clade *Stowaway* MITEs resided (e.g. *Gynerium sagittatum* and *Panicum virgatum*, insertions of 104 bp and 98 bp, respectively; Fig. 2). However, the absence of typical *Stowaway* features such as the flanked 2-bp target site preference (TA), and similar TIR sequences, ruled out the possibility that these insertions could be the remnants of *Stowaway* MITE footprints.

In a previous study, Mason-Gamer (2007) described a collection of three main *Stowaway* MITE types that were found within the fourth intron of the *β-amylase* gene in several Triticeae species (Figs. 1 and 2; Appendix B). Two *Australopyrum* species, *A. retrofractum* AY821692, and *A. velutinum* AY821693.1, showed the same 138 bp MITE. A second type was shared by the partially degraded *A. uniaristata* AY821691.1; 26 bp (94.7% identity in the 5′-end region) and *Peridictyon sanctum* AY821714.1, 62 bp (89% identity with the former). The third type was shared by *Crithopsis delileana* AY821694.1, 80 bp, and *Henrardia persica* AY821703.1, 51 bp (94.4% identity with the former).

Most of the *β-amylase Stowaway* MITEs found by Mason-Gamer (2007) formed almost perfect hairpin-like secondary structures that had very low free energies (Fig. 3; type A). Our analyses indicated that the different MITEs could be classified into two main types of *β-amylase* MITEs within the Triticeae. The Blast searches performed in the Entrez nucleotide database showed that the identical insertion of *Australopyrum retrofractum* and *A. velutinum* (138 bp; dG = −67.44) was also present in other genes of other Triticeae species (e.g. 1st intron of the *HvPKABA1* gene, 6th intron of the *p5cdh* gene and an intergenic spacer in *Hordeum vulgare*; 3rd intron of the *vp1D* gene and an intergenic spacer in *Triticum aestivum*; 1st intron the *phosphate synthase II* gene in *Triticum urartu*; intergenic spacers of *Aegilops tauschii, Triticum monococcum* and *T. turgidum*; Appendix B). Apart from the Triticeae, similar elements were also found in other genes of Poeae s.l. taxa (e.g. 2nd intron of the *fl* gene in *Avena agadiriana, A. hirtula, A. insularis,* and *A. lusitanica*). However, the Blast searches did not find similarities among these MITEs and those found at the same *β-amylase* site in other Triticeae taxa. Conversely, the insertions found at the fourth intron of the *β-amylase* gene in *Aegilops uniaristata* (26 bp; dG = −4.53), *Crithopsis delileanus* (80 bp; dG = −36.78), *Elymus repens* (73 bp; dG = −30.31), *Henrardia persica* (51 bp; dG = −40.30) and *Peridictyon sanctum* (62 bp; dG = −44.90) were similar to each other (Appendix B) but not to the *Australopyrum* ones. They were respectively similar to sequences found at other loci in other Triticeae species (e.g., 4th intron of the *sey1* gene in *Secale cereale*; intergenic spacer sequence in *Hordeum vulgare* and *Triticum aestivum*) and Brachypodieae (e.g., 2nd intron of the ribosomal subunit 8E gene in *Brachypodium sylvatium*; intergenic spacer in *B. distachyon*) taxa.

In this study a new *Stowaway* MITE was found at the 4th intron of the *β-amylase* gene in *Puccinellia distans* (Genbank JX536564, 225 bp), a member of the Alopecurinae/Poinae/Puccinelliinae clade that is nested within the large Poeae s.l. clade (Figs. 1 and 2; Appendix B). This insertion is highly similar to other *β-amylase* 4th intron MITE sequences in Alopecurinae/Poinae/Puccinelliinae in Genbank, such as *Catabrosa aquatica* (HE565906, 190 bp), *Catabrosella araratica* (HE565912, 228 bp), *Colpodium drakensbergense* (HE565915, 167 bp), *C. versicolor* (HE565911, 227 bp), *Zingeria biebersteiniana* (HE565930, 243 bp) and *Z. kochii* (HE565927, 233 bp). Most of these Alopecurinae/Poinae/Puccinelliinae *Stowaway* MITEs formed stem-and-loop secondary structures with small additional insertions (Fig. 3; type B). Although they showed low free energies, the additional insertions disturbed, at least in part, their hairpin-like structures. Consequently, almost all of these TEs were longer and less stable than those detected in the Triticeae (Mason-Gamer, 2007). The exception was the *Zingeria biebersteiniana* insertion (243 bp; dG = −37.13), which showed a stem-and-loop secondary structure similar to the *Australopyrum* element (Fig. 3A). The remaining Alopecurinae/Poinae/Puccinelliinae MITEs showed identities >80% (Fig. 3; type B). This group included *P. distans* (225 bp; dG = −37.82), *C. aquatica* (190 bp; dG = −26.34), *C. araratica* (228 bp; dG = −44.16), *C. versicolor* (227 bp; dG = −44.16), and *Z. kochii* (233 bp; dG = −43.93) (Appendix B). The *Stowaway* MITEs of *C. araratica*, *C. versicolor* and *Z. kochii* were also highly similar (>80% identity) to those found within the Ehrhartoideae/Oryzeae clade (e.g. 5th intron of the *Ogl12g0016G16_1* gene in *Oryza glaberrima*; intergenic spacer in *O. sativa* and *O. minuta*). The *C. aquatica* insertion had the highest free energy (dG = −26.34) of all insertions found within the Alopecurinae/Poinae/Puccinelliinae group. The insertion found in *C. drakensbergense* (167 bp; dG = −14.02) did not form a stem-and-loop secondary structure as almost one third of the sequence has been lost (see *β-amylase* alignment, Appendix C). Additionally, this sequence showed a contiguous 232 bp insertion located 36 bp downstream of the 3′-end TA target duplication site (Appendix C). This insertion, flanked by a duplicated 5′-AGAAGTTGCATCT-3′ sequence on its 5′ and 3′ ends could correspond to a piece of another mobile element, though it does not fold into a hairpin-like structure.

Potential *Stowaway* MITE footprints have also been observed at the fourth intron of the *β-amylase* gene in BEP clade grasses. These footprints consist of short insertions and/or deletions that could indicate the potential presence of a TE in the common ancestor. This is consistent with the proposed mechanism of excision of the *Tc1/mariner* element in rice (Yang et al., 2006), and could help to explain the evolutionary dynamics of the *Stowaway* MITEs. The *Stowaway* MITE footprints fell into six categories (Fig. 2): (i) the target site 5′-TA-3′ and the 5′-GTA-3′ motif were present in nearly all representatives of the Pooideae. Mutations at these target sites, which could alter excision–insertion events (Fig. 2; underlined blue boldface), have been also occasionally found in Triticeae (Mason-Gamer, 2007). Several mutations in the preferred target sites were observed in distinct pooid lineages (e.g. *Bromus tectorum* 'CA···GTA'; *Secale cereale* 'T**G**···GTA'; *Triticum aestivum* 'T**T**···GTA'; *Agrostis capillaris* 'T**T**···GTA'; *Pseudarrhenatherum longifolium* 'TA···**A**TA'; *Avellinia michelli* 'TA···G**A**A'; *Poa infirma* 'TA···G**C**A'; *Desmazeria rigida* 'TA···G**G**A'; *Stipa tenacissima* and *S. offneri* '**A**A···GTA'; *Brachypodium distachyon* and *B. pinnatum* T**G**; *Lolium/Micropyropsis/Festuca* subgen. *Schedonorus* (except *Micropyropsis tuberosa*) 'TA···G**C**A'; *Festuca eskia/F. gautieri*'TA···G**A**A'; all fine-leaved *Festuca* (except the *F.* sects. *Dimorpha/Eskia* group and *Narduroides salzmannii*) share the GT**G** extreme; all the *Festuca* sect. *Festuca* taxa, except *F. panciciana* and *F. ovina,* share the **C**A extreme); (ii) Small insertions of 2–5 bp adjacent to the 5′-TA duplication site have been detected in 14 different pooid species. They correspond to short-TIR remnants, with sequences varying from

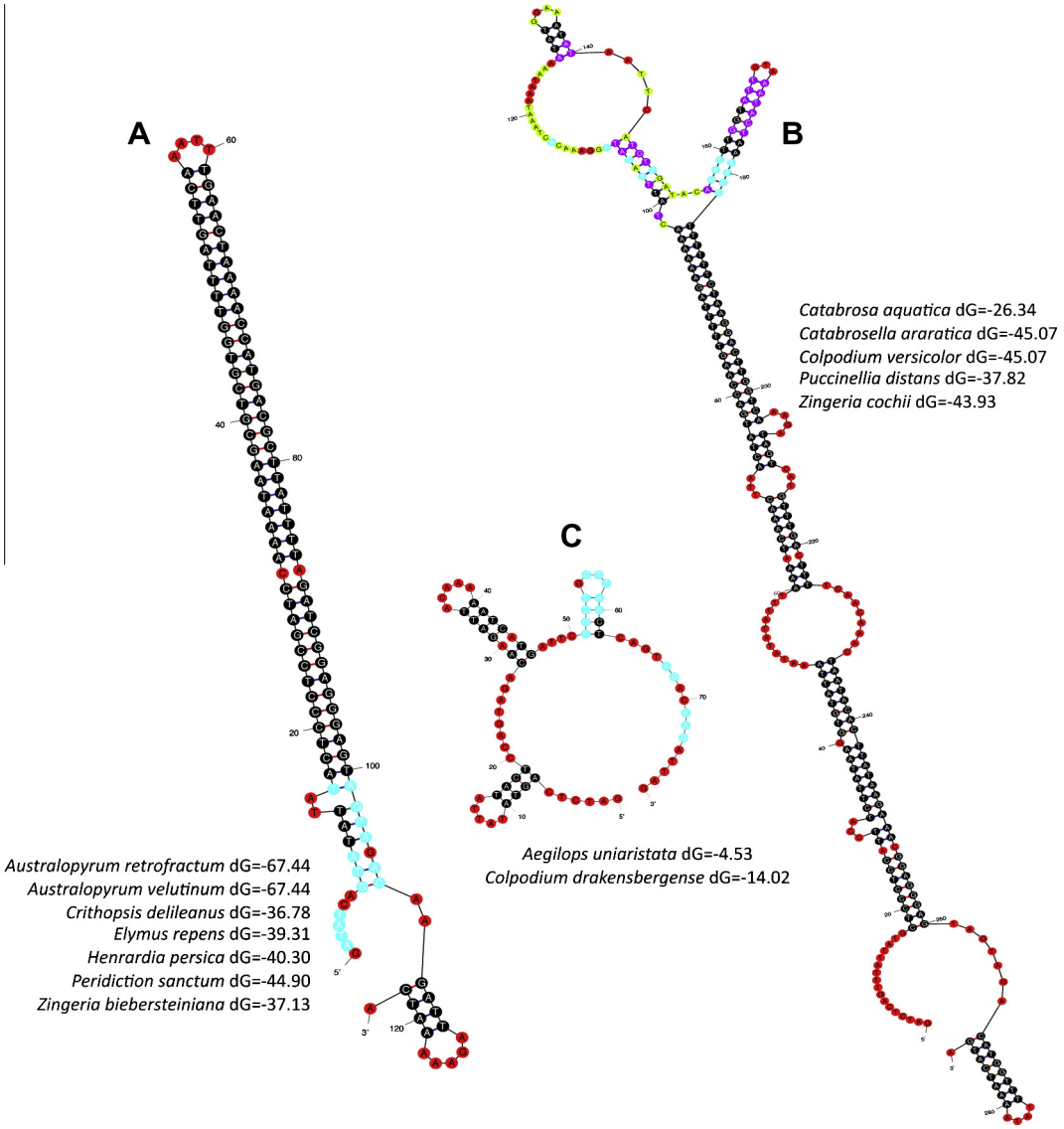


Fig. 3. The secondary structures of the three types of *Stowaway* MITEs found at the fourth intron of the *β-amylase* gene in grasses. Type A: Stem-and-loop secondary structure; Type B: Stem-and-loop secondary structure with small additional insertions; Type C: Additional small insertions that do not form a stem-and-loop secondary structure. The free energy is listed after the name of each species. The color of the nucleotide represents the probability that the base is single-stranded in the computed holdings. Red: 0.999%, green: 0.90, blue: 0.50, purple: 0.022, and black: 0.001. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the 5′-CTCCA-3′ (in *Pseudoroegneria strigosa*, *Festuca scabra*, *F. modesta*, *F. triflora*, *F. coerulescens*, *F. pseudeskia*, *F. scariosa*, *F. altissima*, *F. altaica*, *F. paniculata*, *F. gautieri*, *F. eskia* and *Narduroides salzmani*) to 5′-CTCCC-3′ (in *Trisetum loeflingianum*); (iii) Small insertions of 2–9 bases adjacent to the 5′-GTA-3′ motif were detected in seven species. They likely correspond to flanking regions of TIR and short-TIR remnants with sequences varying from: 5′-TCA-3′ (in *Brachypodium pinnatum, B. distachyon, Psathyrostachys juncea* and *Lophopyrum elongatum*), to 5′-CAGTACTCA-3′ (*Nardus stricta*), 5′-CGGTAATCA-3′ (*Lygeum spartum*), 5′-CCAATTAATT-3′ (*Nassella laevissima*); (iv) A 5′-GTA-3′ deletion in *Brachypodium boissieri*; (v) Two 5′-TA-3′ deletions in *Nardus stricta* and *Lygeum spartum*; and (vi) Two small deletions in, respectively, the 5′ (9 bp) and 3′ (15 bp) ends of the intron sequence flanking the empty element site were found in *Lolium perenne*. It is not clear whether all these deletions (iv, v, vi) are associated with element loss.

3.3. Triticeae Stowaway blz1, dmc1, nuc, and xyl MITEs found in other genes and in other pooid lineages

The Stowaway-like elements described in the *blz1, dmc1, nuc*, and *xyl* genes of Triticeae (Petersen and Seberg, 2000, 2009) were occasionally present in the same loci though they were mostly distributed in intergenic spacers and in the introns of other nuclear genes in the newly searched species (Appendix B). These insertions were almost exclusively found in representatives of the Triticeae; however one of them (*xyl* MITE) was also detected in representatives of other Pooideae lineages.

The 68–73 bp TE detected in the first intron of the *xyl* gene was only found in this locus in the 26 originally studied diploid *Hordeum* species (Petersen and Seberg, 2009; Appendix B). Similar sequences were found, however, in other loci of species outside of *Hordeum*. These sequences were mainly observed within representatives of Triticeae (e.g., introns of the *pdha1* (*Aegilops*), *stpk-v*

(*Dasypyrum*), and *mtk4* (*Triticum*) genes, and in intergenic spacers (*Triticum*)), but they were also found in representatives of Brachypodieae (e.g., introns of *fha domain protein* and *methyl sulfoxide reductase* genes (*Brachypodium distachyon, B. sylvaticum*) and of Loliinae (e.g., intergenic spacer, *Lolium perenne*) (Appendix B)). All the sequences showed a similar hairpin-like secondary structure (cf. Petersen and Seberg, 2009 and results not shown).

The 96–104 bp TE detected in the third intron of the *blz1* gene was found only in this locus in three species of *Hordeum* (Petersen and Seberg, 2009; Appendix B). Similar sequences were found at different nuclear loci and at one putative mitochondrial locus in other representatives of Triticeae (e.g., introns of the *psy-B1* (*Aegilops, Triticum*), *psi-acc-2* (*Triticum*), and *dmc1* (*Heteranthelium*) and mitochondrial *putative neurolysin precursor* (*Triticum*) genes, and intergenic spacers in *Triticum* (Appendix B)). All these sequences showed a hairpin-like secondary structure similar to that described by Petersen and Seberg (2009) for their long *blz1-A* MITE. No matches were recovered for their short *blz1-B* MITE (47–49 bp), which was a defective TE, having one conserved TIR sequence and lacking a TSD (Petersen and Seberg, 2009).

The 164 bp TE detected in the fifth intron of the *nuc* gene in *Hordeum vulgare* (AY485643, Appendix B) was only found in this locus in this species. Similar sequences were found in other loci in other species of Triticeae (e.g., non-coding sequences of *β-amylase* (*Australopyrum*), and introns of *dp62 hypothetical protein* (*Hordeum*), *pdha1* (*Secale*), *pmm-d1* (*Aegilops*), and *sucrose phosphate synthase II* (*Triticum*) genes, and in intergenic spacers in *Triticum* (Appendix B)). These sequences also showed an almost perfect hairpin-like secondary structure (cf. Petersen and Seberg, 2009, and results not shown).

Petersen and Seberg (2000, 2009) found two types of MITEs in the fifth intron of the *dmc1* gene: a short insertion of 59–76 bp, and a long insertion of 283 bp. The short element was found in representatives of *Australopyrum* and *Taeniatherum* (Petersen and Seberg, 2000) and has been also found in other representatives of *Hordeum* (Appendix B). We found similar sequences of this element in other loci of other Triticeae representatives and of *Hordeum* (e.g., introns of the *acc-1* and *acc-2* (*Triticum*), *bkini2* (*Hordeum*), *cpk9* (*Aegilops*), *pgk1* (*Agropyron, Kengyilia*), *pina* (*Aegilops*), *psy-b1* (*Triticum*), *putative expressed protein* (*Secale*), *rga2* (*Triticum*), and *vp-1b* (*Triticum*) genes and in intergenic spacers in *Hordeum* (Appendix B)). The long *dmc1* element was recorded in *Heteranthelium piliferum*, with potential footprints detected in *Dasypyrum* (Petersen and Seberg, 2000). We found similar sequences in other loci of other Triticeae representatives (e.g., introns of the *adp-glucose pyrophosphorylase small subunit* (*Aegilops*), *putative ankyrin repeat protein* (*Hordeum, Triticum*), and *wknox1b* (*Triticum*) genes and in intergenic spacers in *Triticum* (Appendix B). Those sequences also showed a hairpin-like secondary structure (results not shown), similar to that of *H. piliferum* (Petersen and Seberg, 2000).

4. Discussion

4.1. Evolutionary dynamics of the Stowaway MITEs in grasses

The occurrence of *Stowaway* MITEs and their potential footprints in the grasses was assessed within a robust Poaceae phylogeny, under the assumption that the tree obtained from the combined analysis of the nuclear and plastid genes (Fig. 1) represented a good estimate of the true evolutionary history of the family.

The 116 Poaceae species sampled fell into the two traditionally described clades PACCMAD and BEP (e.g. GPWG, 2001; Sánchez-Ken et al., 2007; Bouchenak-Khelladi et al., 2008) (Fig. 1). The monophyletic PACCMAD clade was split into two strongly supported sister lineages, the Centothecoideae-Panicoideae and the Danthonioideae-Chloridoideae. The successive divergences within each of these clades were in agreement with previously published phylogenies (e.g. Barker, 1997; Barker et al., 1995; Bouchenak-Khelladi et al., 2008; Duvall and Morton, 1996; Mathews et al., 2000; Sánchez-Ken et al., 2007; Soreng and Davis, 1998). The BEP clade included members of the Ehrhartoideae and Pooideae. The topological resolution within the Ehrhartoideae and the Pooideae groups was in agreement with previous studies (e.g., Bouchenak-Khelladi et al., 2008; Davis and Soreng, 2007; GPWG, 2001; Schneider et al., 2009; Soreng and Davis, 2000; Soreng et al., 2003; Triplett and Clark, 2010). The Pooideae clade showed the successive divergences of the more ancestral Lygeeae/Nardeae, Stipeae, Diarrheneae and Brachypodieae lineages, and the separation of the more recently evolved Bromeae/Triticeae and Aveneae/Poeae 'core pooid' lineages. The divergences within the Aveneae/Poeae lineage agreed with those proposed by Döring et al. (2007), Quintanar et al. (2007), and Schneider et al. (2009). Quintanar et al. (2007) used an intertribal grouping system to reflect the morphological and evolutionary heterogeneity of the large Poeae s.l. clade that consisted of two strongly supported sister groups, the "core-Aveneae" lineage and the "core-Poeae + Poeae *pro parte* + former Aveneae" lineage. The results of the successively enlarged phylogenies of Loliinae obtained by Catalan et al. (2004, 2006), Inda et al. (2008), Torrecilla and Catalan (2002), and Torrecilla et al. (2004) were confirmed in our analyses; the Loliinae were split into two main groups, the more ancestral broad-leaved Loliinae and the more recently evolved fine-leaved Loliinae (Fig. 1).

Seven types of *Stowaway* MITEs were found within the Pooideae, confirming the polyphyletic origins of these elements in some cases. The apparent homogeneity observed in the TEs located in the *blz1, dmc1, nuc*, and *xyl* genes of the Triticeae (Petersen and Seberg, 2000, 2009), and occasionally in other loci of other pooids ('*xyl*' element in *Brachypodium* and *Lolium*), contrasts with the heterogeneity among the elements in the fourth intron of the *β-amylase* gene in different BEP lineages and in other loci of the temperate grasses. Three different types of *Stowaway* MITEs have been detected in the *β-amylase* gene, two of them in the Triticeae, and one in the Alopecurinae/Poinae/Puccinelliinae. It was not possible to align their disparate sequences, and megablast and discontiguous megablast searches in the Entrez nucleotide database showed their presence in a variety of genes throughout the BEP clade (Appendix B), thus confirming their polyphyletic nature.

Although the heterogeneous and uneven distributions of four of the seven types of *Stowaway* MITEs would preclude any stem-based vertically-inherited transmission, most of these elements were found within the same phylogenetic lineage (e.g., *bzl*1, *dmc*1, and *nuc* MITEs in Triticeae; *β-amylase Puccinellia*-type MITEs in Alopecuriinae/Poiinae/Puccinelliinae; *β-amylase Australopyrum*-type and *Aegilops–Peridictyon*-type MITEs in Triticeae; Appendix B), with a few cases of apparent transmissions from/to less-related grass groups (e.g. *xyl*-type Triticeae MITEs in Brachypodieae and Poeae (*Lolium*); *β-amylase Aegilops–Peridictyon*-type MITEs in Brachypodieae; *β-amylase Australopyrum*-type MITEs in Poeae (Avena), and *β-amylase* Alopecuriinae/Poiinae/Puccinelliinae MITEs in *Oryza* (Appendix B). This suggests a high intragenomic mobility and a more restricted interspecific and intergeneric transference of these elements. Remarkably, none of the seven BEP MITEs has been found in any loci of any representative of the PACCMAD grasses. This is consistent with the strong evolutionary divergence and reproductive and genomic isolation of these two main grass lineages, which separated between 50 and 70 Mya (Vicentini et al., 2008; Salse et al., 2008).

4.2. Phylogenetic distributions of the β-amylase Stowaway MITEs in the BEP clade

The reconstruction of the evolutionary history of the two types of *β-amylase* 4th intron Triticeae MITEs is difficult to decipher. One potential hypothesis would begin with the transference of an 84 bp TE from a more basal Brachypodieae species to some members of the more recently evolved Triticeae (e.g. *P. sanctum*). From these species the MITE could have been transferred to other Triticeae (*C. delileanus*, *E. repens*, and *H. persica*), and to other loci of other representatives of the wheat tribe (*H. vulgare*, *S. cereale* and *T. aestivum*). These transfers would have involved, or been followed by, the putative acquisition of new nucleotides in some cases or the loss of others (e.g. *A. uniaristata*), including the loss of their stem-and-loop secondary structures. By contrast, another hypothesis must be invoked to explain the distribution of the *Australopyrum*-type insertion. This element has not been found in the basal Pooideae and, therefore, could have been transferred from the Triticeae clade to the Poeae (*Avena*) clade or vice versa.

The evolution of the Alopecurinae/Poinae/Puccinelliinae *Stowaway* MITE could be explained by alternative hypotheses. This large insertion could have been transferred from the more basal Ehrhartoideae clade (*Oryza* spp) to the Alopecurinae/Poinae clade (*C. araratica*, *C. versicolor*, *Z. kochii*), after which it could have been transferred within this group (to *Z. biebersteiniana*) and to the close Puccinelliinae (*P. distans*) group. The TE found in *Z. biebersteiniana* is the only one found within the Poeae s.l. clade that shows a perfect stem-and-loop secondary structure. The insertions found in *C. aquatica* (190 bp) and *C. drakensbergense* (167 bp) (sequence similarity to *C. aquatica*: 86%; query coverage: 100%; e-value: 5e−81) might have resulted from additional transferences from *C. araratica*, *C. versicolor* and *Z. kochii*. During these processes, the *C. aquatica* and *C. drakensbergense* TEs could have experienced deletions, thus reducing their similarities to other Alopecurinae/Poinae/Puccinelliinae *Stowaway* MITEs.

All the potential intertribal transferences invoked to explain the disparate distributions of the three types of *β-amylase Stowaway* MITEs involve largely isolated, divergent lineages which show no evidence of current crossing. The oldest split corresponds to that of Ehrhartoideae (Oryza) and Pooideae, dated at 50–47 Mya, followed by those of Brachypodieae and the 'core-pooid' clade, 30–25 Mya, and of Triticeae and Poeae s. l. (including Aveneae), c. 25 Mya (Vicentini et al., 2008). The sexual isolation of these lineages would preclude any recent intertribal transfer, and their respective long evolutionary distances would require complex evolutionary scenarios to explain the observed distribution of their MITEs (see comments below).

4.3. Acquisition of the β-amylase Stowaway MITEs in the Poaceae

The *β-amylase* elements had to be mobilized from ancestral sequences in accordance with the mechanisms proposed to explain the potential independent losses of MITEs. Izsvák et al. (1998) proposed a model in which the stable secondary structures of MITEs (e.g., Fig. 3 types A and B) could dislocate from the replication complex. This model does not, however, exclude the possibility that MITEs could also be mobilized by trans-complementary transposases in a cross-mobilization process, but these elements would not be able to copy or excise if they have any mutation or deletion in the TIR region, or if they lose their secondary structure (Petersen and Seberg, 2000, 2009; Santiago et al., 2002). In *Oryza sativa*, the transpositionally active MITEs contain internal sequences that enhance transposition (Yang et al., 2009), and the structure of these mobile elements shows integrity for their TIR and TSD regions (Lu et al., 2012).

Because the *β-amylase* insertions are in an intron, it is unlikely that they have been subjected to selection pressure. Most of the *β-amylase Stowaway* MITEs found by Mason-Gamer (2007) formed almost perfect hairpin-like secondary structures that had very low free energies (Fig. 3; type A). Accumulation of mutations, insertions, and deletions in the original element would result in the observed diversity detected among the present *Stowaway* MITEs. These mechanisms could be invoked to explain the insertions without stem-and-loop secondary structures (Fig. 3; type C) in, for example, *Aegilops uniaristata* and *Colpodium drakensbergense*. Despite the large number of *Stowaway* MITEs and other types of MITEs studied, MITE mobility is an open question, because their transposase binding sites and transposition enhancer regions are almost always degraded. As a consequence, estimation of their absolute or relative age is difficult, because the dynamics underlying their gain, loss, and high copy number are not yet well understood (Feschotte et al., 2002b; Petersen and Seberg, 2000, 2009; Lu et al., 2012).

All the compiled information provides evidence that both gains and losses of *Stowaway* MITEs at the *β-amylase* locus may have occurred repeatedly in the BEP clade of Poaceae. The plausible underlying dynamic of acquisitions and deletions of these elements and their footprints could be explained by three, non-mutually-exclusive, alternative hypotheses (Fig. 4).

4.3.1. A single early acquisition of a β-amylase TE in the Pooideae followed by multiple independent losses

This first hypothesis favors a single early insertion of a palindrome element, similar to *Tc1-Mariner*, in the fourth intron of the *β-amylase* gene in the ancestor of the BEP clade, followed by multiple independent losses (Fig. 4a). This potentially autonomous element was flanked by its TDS (5′-TA-3′) and the 5′-GTA-3′ motif, explaining the wide distribution of the 5′-TA---GTA-3′ footprints across the pooids. This early insertion could have been vertically transmitted to successive generations. This hypothesis would explain the broad phylogenetic distribution of the 5′-TA---GTA-3′ region (and derived mutated sequences) in almost all the analyzed BEP members (Figs. 1 and 2). By contrast, the lack of target sites, *Stowaway* MITEs, and TE footprints in the PACCMAD members sequenced to date suggest that these insertions were never present at the same locus in that clade. However, early acquisition by the BEP ancestor does not explain the observed dissimilarities between the three types of *β-amylase Stowaway* MITEs detected in the Triticeae and Alopecurinae/Poinae/Puccinelliinae. These MITEs are more similar to elements found in other parts of the same genomes, or in other species' genomes, than they are to each other (Appendix B), raising doubts about their potential monophyletic origin.

Additionally, the single early acquisition scenario would require at least 23 independent partial losses to explain the absence of *β-amylase Stowaway* insertions in many of the surveyed species, and the presence of the different types of footprints found in the BEP clade (Fig. 2). Losses of footprint-like sites and the adjacent intron regions in *Brachypodium boissieri* and *Lolium perenne* might additionally be interpreted as secondary parallel deletions. Shared insertions adjacent to the TA target site (CTCCM) in 14 pooid taxa (Fig. 2) and adjacent to the GTA motif in 7 pooid taxa could be explained by an acquisition followed by later independent losses, by multiple independent acquisitions, or by a mixture of both processes.

4.3.2. Independent insertions of three different Stowaway MITEs at the same locus in the β-amylase gene in the Triticeae and in the Alopecurinae/Poinae/Puccinelliinae lineages

The second hypothesis favors multiple independent insertions of different *Stowaway* elements at the same *β-amylase* locus in

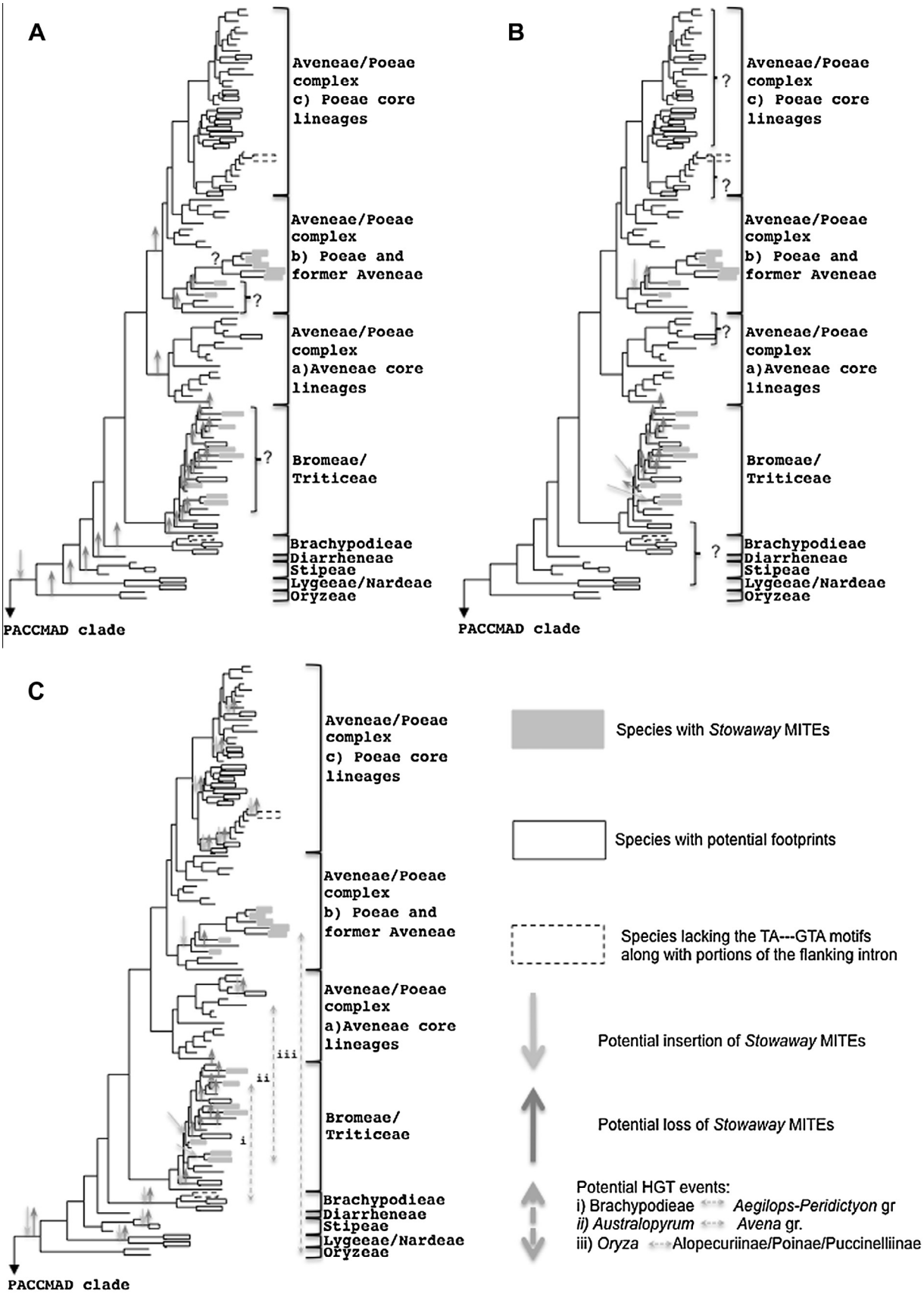


Fig. 4. Alternative hypotheses on the potential evolutionary dynamics of gains, losses and transferences of *β-amylase Stowaway* MITEs across the studied Poaceae lineages. (a) single early acquisition hypothesis; (b) three independent acquisitions hypothesis; (c) different acquisitions followed by multiple losses and different horizontal transfers hypothesis (see text for more detailed explanations on each hypothesis). Light and dark gray arrows represent potential gains and losses of MITEs, respectively. Question marks indicate uncertain events. Gray dashed arrows represent potential HGT events for the *Aegilops–Peridictyon*-type (i), *Australopyrum*-type (ii) and Alopecuriinae/Poinae/Puccinelliinae-type (iii) MITEs. A double arrow indicates uncertainty in the direction of the transference.

non-related Pooideae lineages (Fig. 4b). This scenario would require at least three distinct insertions, which correspond to the three types of *Stowaway* MITEs analyzed in this study (Appendix B). This hypothesis would explain the polyphyletic distribution of the homoplasious insertions, their differences in sequence diversity and size, and their marked similarities to elements found in non-homologous sites in other genomes. This hypothesis is also consistent with the high number of MITE copies found in the grass genomes, where they are clearly abundant (Jiang et al., 2004). This theory of multiple insertions at homologous sites has traditionally been viewed as the result of an active history of amplification and insertion of MITEs (Feschotte et al., 2002a,b; Mason-Gamer, 2007; Wessler et al., 1995) and other types of transposable elements (Bringaud et al., 2002; Penton et al., 2002; Saville et al., 1999; Walker et al., 1997) in grasses. However, this hypothesis does not explain the existence of the TDS (5′-TA-3′) and the 5′-GTA-3′ motif and derived sequences found in almost all members of the Pooideae.

Although the MITEs' very short target site (TA) is extremely frequent, empirical observations suggest that it is the preferred insertion site for this type of transposable element (Bureau et al., 1996; Mason-Gamer, 2007; Petersen and Seberg, 2009; Wessler, 2006). Since all grass MITEs found in the fourth intron of *β-amylase* to date are in the Pooideae, and the 5′-TA---GTA-3′ region is widespread only in the BEP clade, it is plausible to hypothesize that this motif is important in making the fourth intron of *β-amylase* a MITE hotspot.

4.3.3. A single early acquisition and multiple losses, followed by three independent insertions, additional losses, and horizontal transfer

The third hypothesis depicts a more complex scenario of a single early acquisition of a *β-amylase Stowaway* MITE in the ancestor of the Pooideae followed by one or a few early losses. This would explain the wide distribution of 5′-TA---GTA-3′ footprints. A series of more recent independent insertions, losses, and potential horizontal transfers (Fig. 4c) is required to explain their current distribution. *Stowaway* MITEs are a highly dynamic group of rapidly evolving sequences, and their history appears to have involved several temporally separated waves of amplification (Feschotte et al., 2003; Lu et al., 2012). Because of this, and since neither of the previous two hypotheses completely explained the evolutionary dynamic of the *β-amylase Stowaway* MITEs found in the grasses, a third hypothesis which draws from both of them, is proposed. According to this hypothesis, two different waves of element acquisitions, accompanied by multiple losses and horizontal gene transfers, might have taken place in the history of the Pooideae. This scenario could be explained by five consecutive steps:

(1) An element with a palindromic organization and a hairpin-like secondary structure (Feschotte and Wessler, 2002) inserted into the fourth intron of the *β-amylase* gene in the ancestor of the BEP group. This insertion probably took place c. 55 Mya, at the time of the divergence between the BEP and PACCMAD lineages (Vicentini et al., 2008). This element likely inserted with a mechanism similar to the *Tc1-Mariner* superfamily, with analogous TIR and TSD sequences.
(2) The initial transposable element underwent one or more early excisions using a trans-system transposase that recognized the TIRs, leaving 5′-TA---GTA-3′ footprints, which are currently widespread throughout the Ehrhartoideae/Pooideae clade (there is no data on Bambusoideae).
(3) At least three independent insertions of orthologous *Stowaway* MITEs may have occurred in members of the Triticeae (e.g., in the *Aegilops–Peridictyon* group and in the *Australopyrum* group), and in members of the Alopecurinae/Poinae/Puccinelliinae lineage, within the previously created flanking site 5′-TA---GTA-3′ (Fig. 2). It is also possible that some of the present-day elements are direct descendents of the initial insertion, and have been degraded by the accumulation of mutations, insertions, and deletions, with some even losing their stem-and-loop secondary structure, as in the case of *C. drakensbergense* (Fig. 3; type C).
(4) Subsequent deletions are needed to explain such observations as partial elements (e.g., *Ae. uniaristata*) and the existence and distribution of apparent atypical footprints within the widespread 5′-TA---GTA-3′ motif (e.g., the green boldface sequences in Fig. 2). These events probably followed the same mechanism of cross-mobilization described above. Little is known about the methods of transposition *in trans* and their requirements, because the vast majority of MITEs identified so far are probably remnants of ancient insertion events. However, it is plausible that some of those hypothetical deletions or transpositions left the collection of potential *Stowaway* footprints found today (Fig. 2).
(5) Hybridization, polyploidization, and lineage sorting may have contributed to the distribution of *Stowaway* MITEs and their footprints throughout the Pooideae. These processes have often been used to explain the incongruence between the phylogeny of the nuclear and plastid genes in the grasses (Catalan, 2006; Catalan et al., 2004, 2006; GPWG, 2001; Kellogg et al., 1996; Soreng et al., 2007). Horizontal transfer of mobile elements has been documented in grasses (Diao et al., 2006), as has the non-sexual HGT of a functional copy of the *pgiC* gene between two non-related and sexually isolated Poeae lineages (Ghatnekar et al., 2006; Vallenback et al., 2008, 2010). These findings demonstrate that both non-expressed and expressed sequences can be transferred among non-related grasses. Horizontal transfer could explain the unquestionable disconnect found between the evolutionary relationships of the inserted *Stowaway β-amylase* elements and the relationships among the species and genera that harbor them (Fig 1). Nonetheless, this could also be explained by recurrent transfer of ancestrally variable elements across the genome. If a genome has a wide collection of elements inherited from an early ancestral population, with some or all types shared across some or all taxa, then moving them around within genomes could also cause an apparent disconnect between element types and the organismal phylogeny. Ongoing wide genome sequencing research in several model pooid species and in other grass representatives will likely help to decipher the potential intergenomic vs. intragenomic mobilities of the transposable elements and the evolutionary dynamics of the *β-amylase* MITEs.

Acknowledgments

Elizabeth Kellogg, Lynn Clark, Lynn Gillespie and Jeff Saarela provided grass materials and DNA samples for us. We thank Majid Sharifi, Pedro Torrecilla, Diane McCarthy, Michael Jorgensen, Mira Markova, and Brandon Fonnemann for their advice with the *β-amylase* cloning and sequencing, Emily Lemonds for linguistic assistance, and an anonymous reviewer for valuable comments to an earlier version of the manuscript. The work was supported by two consecutive Spanish Ministry of Science and Innovation research grant projects CGL2006-00319 and CGL2009-12955-C02-01. M.M. was funded by a Spanish Ministry of Science and Innovation PhD FPI fellowship. MP acknowledges a Galician Government postdoctoral contract (Ángeles Alvariño program). Phylogenetic analyses were conducted at the cluster computer system of the BIFI Research Institute (University of Zaragoza, Spain).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2013.03.005.

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P., Ronquist, F., 2004. Parallel metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20, 407–415.

Barker, N.P., 1997. The relationships of *Amphipogon*, *Elytrophorus*, and *Cyperochloa* (Poaceae) as suggested by *rbcL* sequence data. Telopea 7, 205–213.

Barker, N.P., Linder, H.P., Harley, E.H., 1995. Polyphyly of Arundinoideae (Poaceae): evidence from *rbcL* sequence data. Syst. Biol. 20, 423–435.

Bouchenak-Khelladi, Y., Salamin, N., Savolainen, V., Forest, V., Van der Bank, M., Chase, M.W., Hodkinson, T.R., 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. Mol. Phylogenet. Evol. 47, 488–505.

Bringaud, F., Biteau, N., Melville, S.E., Hez, S., El-Sayed, N.M., Leech, V., Berriman, M., Hall, N., Donelson, J.E., Baltz, T., 2002. A new, expressed multigene family containing a hot spot for insertion of retroelements is associated with polymorphic subtelomeric regions of *Trypanosoma brucei*. Eukaryot. Cell 1, 137–151.

Bureau, T.E., Wessler, S.R., 1994. Stowaway: a new family of inverted repeat elements associated with genes of both monocotyledonous and dicotyledonous plants. Plant Cell. 6, 907–916.

Bureau, T.E., Ronald, P.C., Wessler, S.R., 1996. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc. Natl. Acad. Sci. 93, 8524–8529.

Catalan, P., 2006. Phylogeny and evolution of Festuca L. and related genera of subtribe Loliinae (Poeae, Poaceae). In: Sharma, A.K., Sharma, A. (Eds.), Plant Genome. Biodiversity and Evolution. Science Publishers, Enfield (New Hampshire), pp. 255–303.

Catalan, P., Kellogg, E.A., Olmstead, R.G., 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequences. Mol. Phylogenet. Evol. 8, 150–166.

Catalan, P., Torrecilla, P., López Rodríguez, J.A., Olmstead, R.G., 2004. Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and trnL-F sequences. Mol. Phylogenet. Evol. 31, 517–541.

Catalan, P., Torrecilla, P., López-Rodríguez, J.A., Müller, J., 2006. Molecular evolutionary rates shed new light on the relationships of *Festuca*, *Lolium*, *Vulpia* and related grasses (Loliinae, Pooideae, Poaceae). In: Bailey, J., Ellis, R.G. (Eds.), Current Taxonomic Research on the British and European Flora. Botanical Society of the British Isles, London, pp. 45–70.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest2: more models, new heuristic and parallel computing. Nat. Methods 9, 772.

Davis, J.I., Soreng, R.J., 2007. A preliminary phylogenetic analysis of the grass subfamily Pooideae (Poaceae), with attention to structural features of the plastid and nuclear genomes, including an intron loss in GBSSI. Aliso 23, 335–348.

DeBry, R.W., Olmstead, R.G., 2000. A simulation study of reduced tree-search effort in bootstrap resampling analysis. Syst. Biol. 49, 171–179.

Diao, X., Freeling, M., Lisch, D.R., 2006. Horizontal transfer of a plant transposon. PLoS Biol. 4, 119–128.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex Pooideae, Poaceae. Kew Bull. 62, 407–424.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Duvall, M.R., Morton, B.R., 1996. Molecular phylogenetics of Poaceae: an expanded analysis of rbcL sequence data. Mol. Phylogenet. Evol. 5, 352–358.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1994. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Feschotte, C., Wessler, S.R., 2002. Mariner-like transposases are widespread and diverse in flowering plants. Proc. Natl Acad. Sci. USA 99, 280–285.

Feschotte, C., Jiang, N., Wessler, S.R., 2002a. Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. 3, 329–341.

Feschotte, C., Zhang, X., Wessler, S.R., 2002b. Miniature inverted-repeat transposable elements MITEs and their relationship with established DNA transposons. In: Craig, N.L., Craigie, R., Gellert, M., Lambovitz, A.M. (Eds.), Mobile DNA II. American Society of Microbiology Press, Washington, DC, pp. 1147–1158.

Feschotte, C., Lakshmi, S., Wessler, S.R., 2003. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with Stowaway miniature inverted repeat transposable elements MITEs. Genetics 165, 747–758.

Feschotte, C., Osterlund, M.T., Peeler, R., Wessler, S.R., 2005. DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITEs. Nucleic Acids Res. 33, 2153–2165.

Flavell, A.J., Pearce, S.R., Kumar, A., 1994. Plant transposable elements and the genome. Curr. Opin. Genet. Dev. 4, 838–844.

Ghatnekar, L., Jaarola, M., Bengtsson, B.O., 2006. The introgression of a functional nuclear gene from *Poa* to *Festuca ovina*. Proc. Roy. Soc. Ser. B: Biol. 273, 395–399.

G.P.W.G., The Grass Phylogeny Working Group, 2001. Phylogeny and subfamilial classification of the grasses. Ann. Missouri Bot. Gard. 88, 373–457.

Greenblatt, I.M., Brink, R.A., 1962. Twin mutations in medium variegated pericarp maize. Genetics 47, 489–501.

Guindon, S., Gascuel, O., 2003. A simple, fasta and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52, 696–704.

Inda, L.A., Segarra-Moragues, J.G., Müller, J., Peterson, P.M., Catalan, P., 2008. Dated historical biogeography of the temperate Loliinae Poaceae, Pooideae grasses in the northern and southern hemispheres. Mol. Phylog. Evol. 46, 932–957.

Izsvák, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H., Hackett, P.B., 1998. Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. J. Mol. Evol. 48, 13–21.

Jiang, N., Bao, Z., Zhang, X., Hirochika, H., Eddy, S.R., McCouch, S.R., Wessler, S.R., 2003. An active DNA transposon family in rice. Nature 421, 163–167.

Jiang, N., Feschotte, C., Zhang, X., Wessler, S.R., 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements MITEs. Curr. Opin. Plant Biol. 7, 115–119.

Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When gene trees tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 21, 321–347.

Kikuchi, K., Terauchi, K., Wada, M., Hirano, H.Y., 2003. The plant MITE mPing is mobilized in anther culture. Nature 421, 167–170.

Kuang, H., Padmanabhan, C., Li, F., Kamei, A., Bhaskar, P.B., Ouyang, S., Jiang, J., Buell, C.R., Baker, B., 2009. Identification of miniature inverted-repeat transposable elements MITEs and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res. 19, 42–56.

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., Kuang, H., 2012. Miniature Inverted-Repeat Transposable Elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. Mol. Biol. Evol. 29, 1005–1017.

Macas, J., Koblízková, A., Neumann, P., 2005. Characterization of Stowaway MITEs in pea *Pisum sativum* L. and identification of their potential master elements. Genome 48, 831–839.

Maddison, D.R., Maddison, W.P., 2008. MacClade 4.08 version for OS X. Sinauer Associates Inc., Sunderland (MA), USA.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, S.A., Wing, R.A., 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. 10, 982–990.

Mason-Gamer, R.J., 2005. The β-amylase of grasses and phylogenetic analysis of the Triticeae (Poaceae). Am. J. Bot. 92, 1045–1058.

Mason-Gamer, R.J., 2007. Multiple homoplasious insertions and deletions of a Triticeae (Poaceae) DNA transposon: a phylogenetic perspective. BMC Evol. Biol. 7, 92–102.

Mason-Gamer, R.J., Kellogg, E.A., 1997. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. 45, 524–545.

Mathews, S., Tsai, R.C., Kellogg, E.A., 2000. Phylogenetic structure in the grass family (Poaceae): evidence from the nuclear gene phytochrome B. Am. J. Bot. 87, 96–107.

Menzel, G., Dechyeva, D., Keller, H., Lange, C., Himmelbauer, H., Schmidt, T., 2006. Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L. Chromosome Res. 14, 831–844.

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., Tanisaka, T., 2003. Mobilization of transposons in the rice genome. Nature 421, 170–172.

Nassif, N., Penney, J., Pal, S., Engels, W., Gloor, G., 1994. Efficient copying of nonhomologous sequences from ectopic sites via P-element-indiced gap repair. Mol. Cell. Biol. 14, 1613–1625.

Penton, E.H., Sullender, B.W., Crease, T.J., 2002. Pokey, a new DNA transposons in *Daphnia Cladocera*: Crustacea. J. Mol. Evol. 55, 664–673.

Petersen, G., Seberg, O., 2000. Phylogenetic evidence for excision of Stowaway miniature inverted-repeat transposable elements in Triticeae (Poaceae). Mol. Biol. Evol. 17, 1589–1596.

Petersen, G., Seberg, O., 2009. Stowaway MITEs in *Hordeum* (Poaceae): evolutionary history, ancestral elements and classification. Cladistics 25, 198–208.

Quintanar, A., Castroviejo, S., Catalan, P., 2007. Phylogeny of the tribe Aveneae (Pooideae, Poaceae) inferred from plastid trnT-F and nuclear ITS sequences. Am. J. Bot. 94, 1554–1596.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Salse, J., Bolot, S., Throude, M.I., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuilleta, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20, 11–24.

Sánchez-Ken, J.G., Clark, L.G., Kellogg, E.A., Kay, E.E., 2007. Reinstatement and emendation of subfamily Micrairoideae (Poaceae). *Syst. Bot.* 32, 71–80.

SantaLucia, J.R., 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. 95, 1460–1465.

Santiago, N., Herráiz, C., Goñi, J.R., Messeguer, X., Casacuberta, J.M., 2002. Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis thaliana*. *Mol. Biol. Evol.* 19, 2285–2293.

Saville, K.J., Warren, W.D., Atkinson, P.W., O'Brochta, D.A., 1999. Integration specificity of the hobo element of *Drosophila melanogaster* in dependent on sequences flanking the integration site. Genetica 105, 133–147.

Schneider, J., Döring, E., Hilu, K.W., Röser, M., 2009. Phylogenetic structure of the grass subfamily Pooideae based on comparison of plastid matK gene-3'trnK exon and nuclear ITS sequences. Taxon 58, 405–424.

Seberg, O., Petersen, G., 2009. Correspondence: a unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 10, 276. http://dx.doi.org/10.1038/nrg2165-c3.

Soreng, R.J., Davis, J.I., 1998. Phylogenetics and character evolution in the grass family (Poaceae): simultaneous analysis of morphological and chloroplast DNA restriction character sets. Bot. Rev. Lancaster 64, 1–85.

Soreng, R.J., Davis, J.I., 2000. Phylogenetic structure in Poaceae subfamily Pooideae as inferred from molecular and morphological characters: misclassification versus reticulation. In: Jacobs, S.W.L., Everett, J. (Eds.), Grasses: Systematics and Evolution. CSIRO, Melbourne, pp. 61–74.

Soreng, R.J., Peterson, P.M., Davidse, G., Judziewicz, E.J., Zuloaga, F.O., Filgueiras, T.S., Morrone, O., 2003. Catalogue of new world grasses Poaceae: IV. subfamily Pooideae. Contr. U.S. Natl. Herb. 48, 1–730.

Soreng, R.J., Davis, J.I., Voionmaa, M.A., 2007. A phylogenetic analysis of Poaceae tribe Poeae sensu lato based on morphological characters and sequence data from three plastid-encoded genes: evidence for reticulation, and a new classification of the tribe. Kew. Bull. 62, 425–454.

Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis using Parsimony and Other Methods. Version 4b10. Sinauer Associates, Sunderland MA.

Torrecilla, P., Catalan, P., 2002. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (Poaceae) based on nuclear ITS sequences. Syst. Bot. 27, 241–251.

Torrecilla, P., López-Rodríguez, J.A., Catalan, P., 2004. Phylogenetic relationships of *Vulpia* and related genera (Poeae, Poaceae) based on analysis of its and trnL-F sequences. Ann. Missouri Bot. Gard. 91, 124–158.

Triplett, J.K., Clark, L.G., 2010. Phylogeny of the temperate bamboos (Poaceae: Bambusoideae: Bambuseae) with an emphasis on *Arundinaria* and allies. Syst. Bot. 35, 102–120.

Turcotte, K., Bureau, T., 2002. Phylogenetic analysis reveals Stowaway-like elements may represent a fourth family of the IS630-Tc1-mariner superfamily. Genome 45, 82–90.

Vallenback, P., Jaarola, M., Ghatnekar, L., Bengtsson, B., 2008. Origin and timing of the horizontal transfer of a *PgiC* gene from *Poa* to *Festuca ovina*. Mol. Phylogenet. Evol. 46, 890–896.

Vallenback, P., Ghatnekar, L., Bengtsson, B., 2010. Structure of the Natural Transgene PgiC2 in the Commo Grass *Festuca ovina*. Plos One, e13529. http://dx.doi.org/10.1371/journal.pone.0013529.

Vicentini, A., Barber, J.C., Aliscioni, A.A., Giussani, L.M., Kellogg, E.M., 2008. The age of the grasses and clusters of origins of C4 photosynthesis. Glob. Change Biol. 14, 2693–2977.

Walker, E.L., Eggleston, W.B., Demopoulos, D., Kermicle, J., Dellaporta, S.L., 1997. Insertions of a novel class of transposable elements with a strong target site preference for the r locus of maize. Genetics 146, 681–693.

Wessler, S.R., 2006. Transposable elements and the evolution of eukaryotic genomes. Proc. Natl. Acad. Sci. 21, 17600–17601.

Wessler, S.R., Bureau, T.E., White, S.E., 1995. LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. 5, 814–821.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, B., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, B., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2009. Correspondence: Reply: a unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 10, 276. http://dx.doi.org/10.1038/nrg2165-c4.

Yang, G., Weil, C.F., Wessler, S.R., 2006. A rice Tc1/Mariner-like element transposes in yeast. Plant Cell 18, 2469–2478.

Yang, G., Zhang, F., Hancock, C.N., Wessler, S.R., 2007. Transposable of the rice miniature inverted repeat transposable element mPing in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. 104, 10962–10967.

Yang, G., Nagel, D.H., Feschotte, C., Hancock, C.N., Wessler, S.R., 2009. Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science 325, 1391–1394.

Zhang, X., Jiang, N., Feschotte, C., Wessler, S.R., 2004. PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. Genetics 166, 971–986.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.