Rubén Sancho Cohen

Genome annotation, comparative genomics and evolution of the model grass genus Brachypodium (Poaceae)

Departamento

Ciencias Agrarias y del Medio Natural

Director/es

Catalán Rodríguez, Pilar
Contreras Moreira, Bruno

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza

1542

Tesis Doctoral

GENOME ANNOTATION, COMPARATIVE GENOMICS AND EVOLUTION OF THE MODEL GRASS GENUS BRACHYPODIUM (POACEAE)

Autor

Rubén Sancho Cohen

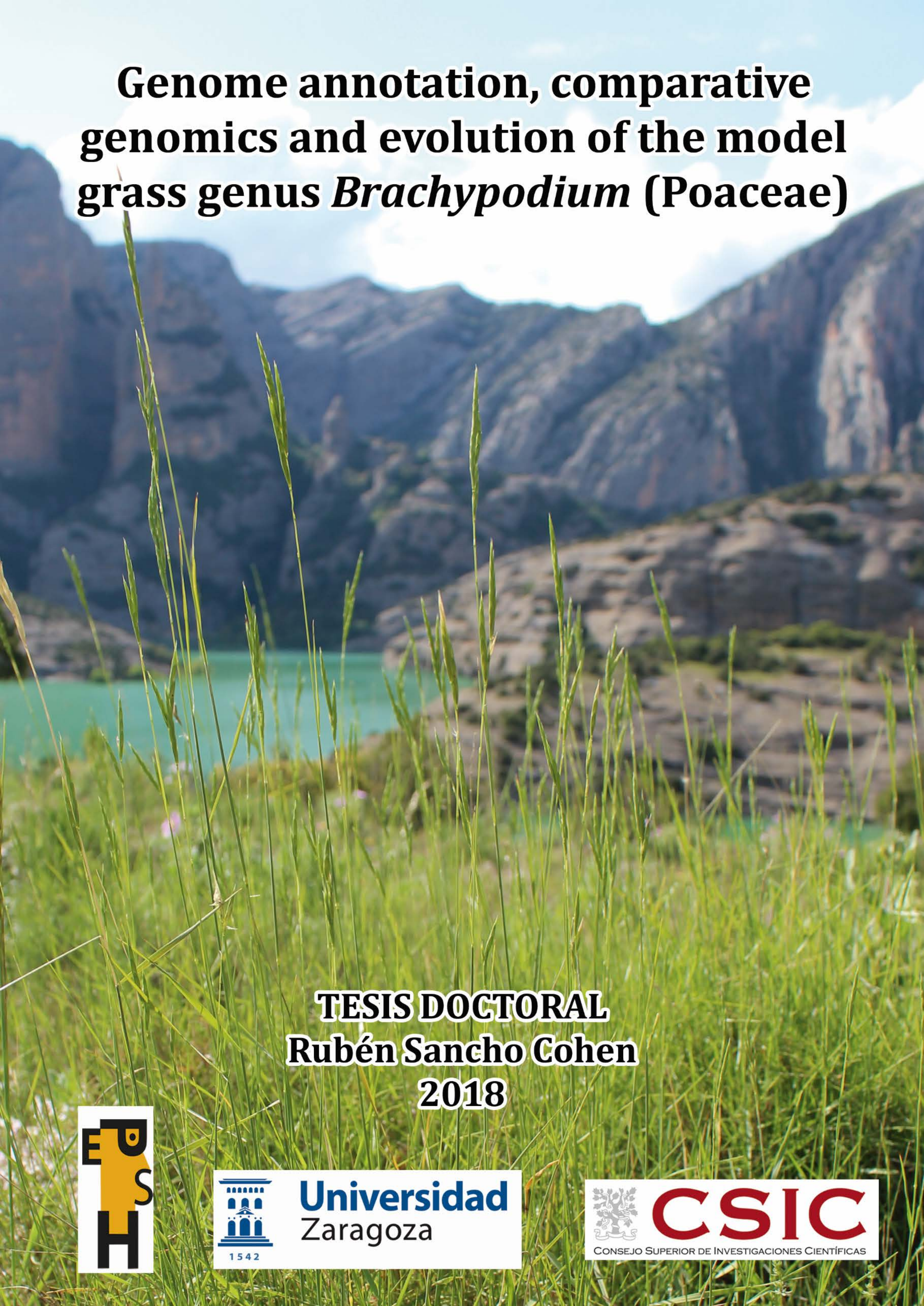
Director/es

Catalán Rodríguez, Pilar
Contreras Moreira, Bruno

UNIVERSIDAD DE ZARAGOZA

Ciencias Agrarias y del Medio Natural

2018



Genome annotation, comparative genomics and evolution of the model grass genus *Brachypodium* (Poaceae)

TESIS DOCTORAL

Rubén Sancho Cohen

2018

EDSH

Universidad Zaragoza
1542

CSIC
Consejo Superior de Investigaciones Científicas

TESIS DOCTORAL

DOCTORADO INTERNACIONAL

Genome annotation, comparative genomics and evolution of the model grass genus *Brachypodium* (Poaceae)

Autor:

Rubén Sancho Cohen

Directores:

Dra. Pilar Catalán Rodríguez

Dr. Bruno Contreras-Moreira

Huesca, Junio de 2018

Escuela Politécnica Superior de Huesca

Departamento de Ciencias Agrarias y del Medio Natural

Universidad de Zaragoza

Universidad Zaragoza

1542

Escuela Politécnica Superior de Huesca

Departamento de Ciencias Agrarias y del Medio Natural

Universidad de Zaragoza

Genome annotation, comparative genomics and evolution of the model grass genus *Brachypodium* (Poaceae)

Memoria presentada por

D. Rubén Sancho Cohen

para optar al Grado de Doctor por la

Universidad de Zaragoza

Directores de tesis:

Dra. Pilar Catalán Rodríguez

Dr. Bruno Contreras-Moreira

La Dra. Pilar Catalán Rodríguez, Catedrática de Botánica del Departamento de Ciencias Agrarias y del Medio Natural de la Universidad de Zaragoza, y el Dr. Bruno Contreras-Moreira, investigador del grupo de Biología Computacional y Estructural de la Estación Experimental de Aula Dei (CSIC), hacen constar que el trabajo recogido en la presente memoria de tesis doctoral ha sido desarrollado bajo su dirección y autorizan su presentación y defensa.

Huesca, Junio de 2018

Esta tesis doctoral ha sido desarrollada en los laboratorios de los grupos 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 Dra. Pilar Catalán Rodríguez, y del grupo de Biología Computacional y Estructural de la Estación Experimental de Aula Dei (CSIC) bajo la dirección del Dr. Bruno Contreras-Moreira. Parte de los resultados fueron obtenidos en el centro de investigación del Arnold Arboretum perteneciente a la Universidad de Harvard bajo la supervisión del Dr. David L. Des Marais.

Los trabajos de laboratorio fueron llevados a cabo en los centros de la Escuela Politécnica Superior de Huesca (Universidad de Zaragoza) y el Arnold Arboretum (Universidad de Harvard). Los recursos informáticos empleados fueron los de la Escuela Politécnica Superior de Huesca y la Estación Experimental de Aula Dei.

El desarrollo de esta Tesis se enmarca en el proyecto "Genómica comparada, biogeografía y evolución floral y adaptativa de gramíneas modelo" financiado por el Ministerio de Economía y Competitividad (Proyecto No. CGL2012-39953-C02-01), y fue posible gracias a un contrato pre-doctoral para la formación de doctores asociado a dicho proyecto. Adicionalmente, se obtuvieron dos ayudas de movilidad para desarrollar estancias breves de investigación en el Arnold Arboretum (Universidad de Harvard), financiadas por el Ministerio de Economía y Competitividad (No. EEBB-I-15-09760) y el programa Ibercaja-Cai para estancias de investigación (convocatoria 2015), respectivamente. Además se obtuvo una ayuda a la Investigación del Instituto de Estudios Altoaragoneses (Diputación Provincial de Huesca).

A mi familia

AGRADECIMIENTOS

Quisiera dar las gracias a todas las personas que me han acompañado y apoyado durante este largo, exigente y apasionante viaje.

A mi directora, la Doctora Pilar Catalán, por darme la oportunidad de realizar la tesis en el grupo de investigación Bioflora, por el esfuerzo invertido en enseñarme a conocer las plantas objeto de estudio, los diseños de los estudios, mi formación investigadora en el campo de la evolución, los análisis llevados a cabo, y las intensas horas dedicadas a las revisiones y mejoras de todos los manuscritos que constituyen esta tesis.

A mi director, el Doctor Bruno Contreras, por el tiempo dedicado a mi formación como investigador, especialmente en el campo de la biología computacional y la genómica, así como su ayuda en el desarrollo de algoritmos, análisis y revisiones de gran parte de los estudios que forman esta tesis.

A mi supervisor durante las dos estancias de investigación en el Arnold Arboretum, el Doctor David L Des Marais, por su esfuerzo y disposición para sacar adelante el trabajo, por su viaje desde Boston a Zaragoza para ayudar con la investigación y especialmente por su apoyo personal durante mis estancias, así como el trabajo invertido en las investigaciones de transcriptómica y expresión génica.

Al doctor Luis Ángel Inda por su contribución a los estudios citogenéticos que forman parte de esta tesis.

Al doctor Antonio Díaz por su contribución a los análisis de biogeografía y de filogenia basada en evolución mínima que forman parte de esta tesis.

A la doctora Diana López, por su contribución a la amplificación y secuenciación de regiones flanqueantes de las IR del genoma cloroplástico, así como por su ayuda y apoyo tanto personal como científico en los inicios de esta tesis.

Al doctor Carlos P. Cantalapiedra por su contribución a los análisis y los ensamblajes de los genomas cloroplásticos, así como su ayuda resolviendo un gran número de dudas en cuestiones bioinformáticas.

A Marisa López, por su ayuda e instrucción en diversas técnicas de laboratorio durante las primeras etapas de la tesis.

A los investigadores, profesores y personal de la Escuela Politécnica Superior de Huesca que me han ayudado en alguna de las etapas de esta tesis, y muy especialmente a los amigos que me han acompañado tanto profesional como personalmente: Ernesto, Juan, Ester, Rocío, Asun, Belén, así como los amigos de fuera (y dentro) de la universidad, Bea, Cristina, Ana y Javi.

A los investigadores, compañer@s y amig@s de la Estación Experimental de Aula Dei de los grupos de Biología Computacional y Estructural, y de Genética y Desarrollo de Materiales Vegetales con los que he tenido el placer de coincidir durante la tesis, Ana, Ernesto, Chus, Val, Arantxa y otros muchos con los que he compartido excursiones, comidas y conversaciones.

A los amig@s que me acompañaron en Boston, Federico y Diana, y muy especialmente a Juan, Nuria y Julieta que me hicieron sentir en familia cada segundo de mi estancia.

A todos los compañer@s con los que en algún momento de la tesis he tenido el placer de coincidir: María Fernanda, Valeria, Uriel, Teshome, Álvaro, Javier, David, Irene, y los que seguro me estoy olvidando también gracias.

A mi familia, especialmente a mis padres y hermano, por su apoyo incondicional y hacer posible que hoy esté escribiendo el final de este trabajo.

Por último a Erica, por hacer que uno de los momento más estresantes de mi vida, como es el finalizar una tesis, haya sido de los más divertidos y felices simplemente por pasarlo a su lado.

¡Gracias a todos!

INDEX

RESUMEN

La especie anual *Brachypodium distachyon* y otras especies del género *Brachypodium* han sido seleccionadas como plantas modelo de gramíneas y monocotiledóneas durante la última década. Su estudio ha aportado grandes avances en la compresión de los procesos biológicos, evolutivos y ecológicos, siendo especialmente relevantes por su posible traslación a los cereales templados y a gramíneas biocombustibles.

Dilucidar cuales han sido los orígenes y los eventos de divergencia, hibridación, poliploidización, especiación y aislamiento intraespecífico que han experimentado especies del género *Brachypodium*, especialmente las alopoliploides , ha supuesto un desafío debido a su disploidía y a su compleja y reticulada historia evolutiva, mostrando sucesivos eventos de introgresión y poliploidización. En el presente estudio se han desarrollado análisis tanto a nivel inter- como intra-específico para tratar de clarificar estos procesos.

La obtención de grandes cantidades de datos genómicos mediante tecnologías de secuenciación de alto rendimiento ha permitido pasar del estudio de unos pocos genes de estas especies a sus genomas completos o parciales, con un coste de recursos razonable. El análisis de *Big Data* supone un importante reto, por ello el desarrollo de algoritmos y la aplicación de herramientas bioinformáticas juega un papel determinante en su procesado. El empleo de distintos modelos evolutivos y de análisis filogenómicos han sido fundamentales para poder descifrar los intrincados procesos históricos experimentados por los linajes de estas plantas y para tratar de responder a las hipótesis de esta tesis sobre su origen, naturaleza y dinámica espacio-temporal, y su funcionalidad en tratamientos de estrés hídrico.

La combinación de estos tres factores, sistemas modelo, tecnologías de secuenciación de alto rendimiento y desarrollo de herramientas bioinformáticas, junto con los análisis de genómica comparada y filogenómicos, nos ha permitido obtener un gran conocimiento sobre la intrincada historia evolutiva y los complejos procesos biológicos que han tenido lugar en las plantas objeto de estudio.

Los estudios filogenómicos y biogeográficos llevados a cabo mediante secuencias tanto génicas como de genomas y transcriptomas, nos han posibilitado la reconstrucción y la datación de árboles de especie y de subgenomas de todas o de la mayor parte de las especies reconocidas del género, incluyendo las complejas alopoliploides. Estos datos nos han permitido inferir cómo tuvieron lugar las divergencias y las dispersiones de los linajes, y sus posteriores introgresiones en un marco geográfico y temporal. El estudio filogenómico de los plastomas de un elevado número de ecotipos de *Brachypodium distachyon* y la comparación del árbol infra-específico con el obtenido de los análisis de sus genomas nucleares nos ha permitido identificar las divergencias de los principales linajes de la especie, estructurados según sus tiempos de floración y su geografía, y el descubrimiento de posteriores introgresiones y capturas cloroplásticas entre esos linajes aislados que contrarrestan la potencial microespeciación.

El amplio uso de *Brachypodium distachyon* como planta modelo de gramíneas templadas y la amplia disponibilidad de recursos pan-genómicos de esta especie nos ha incitado a llevar a cabo estudios de redes de co-expresión génica y de expresión diferencial de genes en condiciones de sequía y de riego entre una amplia muestra geográfica de sus ecotipos. Los resultados nos han permitido identificar conjuntos de genes reguladores de rutas biológicas implicadas en las respuesta al estrés hídrico (síntesis de prolina, respuesta a la privación de agua, a la de fosfato inorgánico y de estímulo de la temperatura) que pueden ser también claves en procesos celulares de señalización y de respuesta a otros estreses.

El conjunto de los estudios de la tesis han dado lugar a un incremento en el conocimiento sobre el género *Brachypodium*, tanto a nivel evolutivo como funcional.

SUMMARY

During the last decade the annual species *Brachypodium distachyon* and other congeners have been selected as model plants for grasses and monocots. Their study has proportionated enormous advances in the knowledge of their biological, evolutionary and ecological processes, fostered by their potential translation to the temperate cereal crops and the biofuel grasses.

Untapping the origins and the divergence, hybridization, polyploidization, speciation and intraspecific isolation events experienced by the species of the genus *Brachypodium* has been a challenge due to their dysploidy and complex reticulate evolutionary history. Different allopolyploid *Brachypodium* species have shown successive introgression and genome duplications. We have developed both inter and intraspecific analyses aiming to clarify these events. The production of large amounts of data using high-throughput sequencing technologies has allowed researchers to move from the study of a few genes to complete or partial genomes with reasonable resource costs. The analysis of the Big Data is a main challenge, and for this reason the development of algorithms and the application of bioinformatic tools play an important role in the process of the data. The application of different evolutionary models and of phylogenomic analyses has been a fundamental step to deciphering the inextricable historical processes experienced by the lineages of these plants and to answering the main hypothesis of this thesis about their origin, nature and spacio-temporal dynamics, and their functionality under drought stress conditions.

The combination of those three factors, model systems, high-throughput sequencing technologies and development of bioinformatic tools, and comparative genomics and phylogenomic analyses, has allowed us to acquire a large knowledge about the intricate evolutionary history and the complex biological processes related to the plants under study.

The phylogenomic and biogeographic studies undertaken with the *Brachypodium* species using both genetic, and genomic and transcriptomic data, has facilitated the reconstruction and the dating of the species tree and the subgenomic tree of all or the main part of the species, including the complex allopolyploid taxa. The data allowed us to elucidate the divergences and dispersals of lineages and their subsequent mergings within a geographic and temporal evolutionary framework. The phylogenomic analysis of plastomes from a large number of *B. distachyon* ecotypes, and the comparison of the infraspecific plastome tree with the nuclear genome tree, allowed us to identify the main diverging lineages. They were structured according to their flowering times and geographic distribution; the discovery of latter introgressions and plastid captures between those isolated lineages counteracted their potential microspeciations.

The ample use of *Brachypodium distachyon* as model plant for temperate grasses and the vaste availability of pangenomic resources stimulated us to conduct analysis of co-expression networks and of differentially expressed genes under drought and water conditions among a large geographic sampling of its ecotypes. The results detected groups of regulatory hub genes implied in the response to the drought stress (synthesis of proline, responses to water deprivation, phosphate starvation and stimulus to temperature) that could also be key in the regulation of other signaling pathways and on the response to other stresses.

The compilation of studies developed in this thesis has contributed to increase the current knowledge on the evolution and functional responses of species of the genus *Brachypodium*.

PhD THESIS STRUCTURE

The PhD thesis is structured in four general chapters (Introduction, Materials and Methods, Objectives, Conclusions) and four specific chapters related to the research conducted during the PhD work (Chapters 1 to 4). A further section of References lists all the references mentioned in the general and specific chapters, and another section of Appendices includes supplementary information from each of the research chapters. The last section of the thesis lists the Publications obtained from the PhD research.

The order of the chapters and sections is as follows:

- **Introduction:** State of art on the genomic, evolutionary and systematics investigations in the genus *Brachypodium*.
- **Material and Methods:** General review of material and methods used in the thesis.
- **Objectives:** Main and specific objectives of the PhD work.
- Chapters of studies conducted in the thesis (each chapters contains the Summary, Introduction, Material and Methods, Results and Discussion sections):
    - **Chapter 1.** Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches.
    - **Chapter 2.** Reference-genome syntenic mapping and multigene-based phylogenomics reveal the ancestry of homeologous subgenomes in grass *Brachypodium* allopolyploids.
    - **Chapter 3.** Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes.
    - **Chapter 4.** Co-expression network features and differentially expressed genes explain drought-response patterns in the model grass *Brachypodium distachyon*.

- **Conclusions:** General conclusions of the thesis.
- **References:** This section includes all the bibliographic references cited in the thesis.
- **Appendices**: This section includes supporting information from each research chapter (supplementary methods, results, tables and figures):
  - **Appendix I:** Supporting Information of Chapter 1
  - **Appendix II:** Supporting Information of Chapter 2
  - **Appendix III:** Supporting Information of Chapter 3
  - **Appendix IV:** Supporting Information of Chapter 4
- **Publications of the PhD thesis:** This section lists the publications obtained from and contributed to by the PhD thesis.

INTRODUCTION – STATE OF THE ART

Contribution of Grasses to Earth ecosystems and human development

Grasses have played a fundamental role on human development being a main source of human nutrition, directly or indirectly as animal nutrition (Jacobs & Everett, 2000), and providing textile fibers during miles of years. The grass subfamilies with greater economic importance for human nutrition are Pooideae (e. g., *Triticum aestivum*, *T. turgidum*, wheats; *Hordeum vulgare*, barley; *Secale cereale*, rye), Oryzoideae (*Oryza sativa*, rice), and Panicoideae (*Zea mays*, maize; *Sorghum bicolor*, sorghum; *Saccharum officinarum*, sugar cane). More recently, grasses have also acquired a new important role for human development as a source of renewable biomass for the sustainable production of bioenergy and liquid biofuels in the form of cellulosic biomass, starch from crops, and sugar from cane (Bhattacharya & Knoll, 2012).

As consequence of the capital importance of this family, a large number of breeding programs have been developed to improve species such as wheat, barley, rice or maize, generating new cultivars to ameliorate traits such as yield, nutrition value, and biotic and abiotic stress tolerance (Bradshaw, 2017). Genetics and biotechnology techniques like marker-assisted selection (MAS) or transgenic technology represent major advances in plant breeding. Those technologies have allowed researches to regulate the expression of genes across the germplasm of crop species or the transference of target genes from a species into a crop (Brummer et al., 2009). Recently, a new group of grasses of the cool season genus *Brachypodium* have emerged as model systems for crops grasses (Vogel, 2016). Based on its optimal biological and genomic features and its close phylogenetic relatedness to the temperate cereals, *B. distachyon* and its close congeners have been proposed as suitable models for grasses and monocots (IBI, 2010; Catalán et al., 2014; Gordon et al., 2016; Scholthof et al., 2018).

Evolution of Poaceae family

Grasses (Poaceae) have played a crucial role on Earth since their origin in the Cretaceous-Paleocene transition (Prasad et al., 2005; Strömberg, 2011), and definitively since their expansion into all continents and almost all terrestrial ecosystems from the Oligocene onwards (Bouchenak-Khelladi et al., 2010; Pimentel et al., 2017b). Members of this family are currently part of grasslands and other grass- and graminoid-dominated habitats (e.g., savanna, open and closed shrubland, and tundra), which occur on every continent (Strömberg, 2005, 2011). Grasses occupy about 30–40 % of Earth's land surface, and account for 69 % of the world's agricultural area; grasslands cover more terrestrial area than any other single biome type (O'Mara, 2012; Blair et al., 2014).

Comparative genomics studies indicate that all the Poaceae derive from a grass ancestor that likely experienced a whole genome duplication (WGD) event between 90 to 70 Ma (Paterson et al., 2004; Salse et al., 2008; Murat et al., 2010). Evidence suggests that the ancient grass paleopolyploidization was followed by subsequent "diploidizations", involving differential losses of many duplicated heterologous copies in the subgenomes (Paterson et al., 2004) or by profound distinct genomic rearrangements (Salse et al., 2008), including successive centromeric chromosome fusions (Murat et al., 2010), along the divergent grass lineages. The return to the "diploid" state in plants is interpreted as the genomic reduction to disomic single copy genes, downsized genomes and small chromosome numbers (Leitch & Bennett, 2004; Ma & Gustafson, 2005). By contrast, new polyploidization events apparently led to the rising of mesopolyploids, originated some million years ago, and of neopolyploids, considered to have arisen during or after the Quaternary glaciations (Stebbins, 1985; Marcussen et al., 2015).

Allopolyploids account for 70% of the current grass species (Stebbins, 1949; Kellogg, 2015a). The Poaceae include approximately 12,000 species classified into 750 to 850 genera (Kellogg, 2001; Soreng *et al.* 2015). Evolutionary studies of grass representatives indicate a diverging grade of ancient Anomochlooideae, Pharoideae, and Puelioideae subtribal lineages that preceded the split of the main BOP (Bambusoideae, Oryzoideae and Pooideae) and PACMAD (Panicoideae, Aristoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clades (Clark et al., 1995; Zhang, 2000; Sánchez-Ken & Clark, 2010; Kellogg, 2015a; Soreng et al., 2017).

The increase in the rate of diversification detected in the temperate C3 Pooideae grasses, (Pimentel et al., 2017b) was correlated with the drop in global temperatures that took place in the Middle to Late Eocene and the Oligocene (Beerling & Royer, 2011). Interestingly, this increase in diversification of the 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 & Davis, 2001; Bouchenak-Khelladi et al., 2009). By contrast, diversification of tropical, mostly C4, PACMAD grasses concurred with the diversification of some mamalian herviborous lineages like Antilopienae *s.l.*, Hippotragineae and Alcelaphineae within the Bovidae in the Oligocene, despite the much older origin of the group (Late Eocene) (Bouchenak-Khelladi et al., 2009). 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 & Beerling, 2006). The diversification of the Pooideae during the Oligocene continued during the Miocene and the Pliocene (Pimentel et al., 2017b) and developed into primary temperate grasslands in both hemispheres (Bouchenak-Khelladi et al., 2009; Edwards et al., 2010; Strömberg, 2011).

Several phylogenetic studies have been carried out with the aim of deciphering the evolutionary history of Poaceae. Forty six structural characters, macro and micro-morphological, grouped as culm (2), leaf (5), spikelet (10), floret (14), fruit and embryo (9), seedling (6) characters, were defined to optimized the phylogeny of grasses (GPWG, 2001). The refinement of genomic analyses led to using genes or intergenic regions to conduct molecular phylogenetic studies. Chloroplast loci such as rbcL (Barker et al., 1995), ndhF (Clark et al., 1995), rpl16 intron (Zhang, 2000), rps4 (Nadot et al., 1994) or matK (Liang & Hilu, 1996; Hilu et al., 1999; Ge et al., 2002), nuclear loci such as phytochrome (Mathews & Sharrock, 1996; Mathews et al., 2000), or nuclear and/or plastomes loci (Guo & Ge, 2005; Saarela et al., 2017), or a combination of morphological characters, chloroplast and/or nuclear loci (Soreng & Davis, 1998; CPWG, 2001) have been widely used in phylogenetic studies of grasses. Complete chloroplast genomes have been used in several approaches (Daniell et al., 2016), including the reconstruction of both inter- and intra-specific phylogenies and comparative analyses in several grasses such as *Hordeum*, *Sorghum* and *Agrostis* (Saski et al., 2007), *Oryza sativa, Zea mays* and *Triticum aestivum* (Matsuoka et al., 2002), *Cynodon dactylon* (Huang et al., 2017b), Andropogoneae (Arthan et al., 2017), Bambusoideae (Wysocki et al., 2015) or representatives of all Poaceae (Saarela et al., 2018). The RNA sequencing technique has demonstrated to be a very useful tool for phylogenetic studies (phylo-transcriptomics), including orthology inference and gene synteny (Yang & Smith, 2014; Washburn et al., 2017). Synteny-based orthology determination is rooted in the assumption that orthologous genes will not only share sequence similarity, but will also reside in similar locations within the genomes of related species (Tang et al., 2008). Comparative transcriptomic studies focusing on phylogenies and evolution of gene expression have also been conducted in grasses (Davidson et al., 2012; Washburn et al., 2017; Zhou et al., 2017).

The grass subfamily Pooideae

The grass subfamily Pooideae comprises about one third of the grasses (*ca*. 177 genera and *ca*. 3850 species *sensu* Kellogg (2015a) or *ca*. 197 genera and *ca*. 4234 species *sensu* Soreng *et al.*, (2015), including some of the most prominent crops such as wheat, rye, oats and barley. Its phylogenetic structure has been thoroughly studied, but recent revisions on this topic have called for larger datasets to increase the robustness of the results (Grass Phylogeny Working Group, 2012; Soreng et al., 2015). Molecular phylogenies support the monophyly of the Pooideae within the Poaceae, and recover it as sister to the Bambusoideae in the BOP clade (Saarela et al., 2015).

The systematic positions of the different tribes and subtribes within the Pooideae are currently under discussion, and their evolutionary relationships are not totally resolved (Kellogg, 2015a; Soreng et al., 2015). The tribal arrangement of the Pooideae has varied widely over the last century. In the most recent classification twelve subtribes (plus the incertae sedis *Avenula* – *Homalotrichon*) belong to the Poeae-type plastid DNA clade and seven tribes to the Aveneae-type plastid DNA clade (Soreng et al., 2015), all of them classified within supertribe Poodae. Different studies focusing on some particular subtribes such as the Airinae, Loliinae, Poinae and Aveninae have suggested that further changes to the taxonomy of the supertribe Poodae may be necessary (Pimentel et al., 2017b). A supertribe Triticodae has also been proposed including three tribes: Bromeae, Triticeae (encompassing subtribes Triticinae and Hordeinae) and the recently created Littledaleeae (Soreng et al., 2015). The sister Poodae and Triticodae constitute the "core pooids" (Catalán et al., 1997), a highly speciose and recently evolved lineage formed by taxa showing some of the largest genomes of grasses due to the accumulation of transposons (Kellogg, 2015a).

The pooids show a karyotype evolutionary trend of increasing chromosome sizes and decreasing chromosome base numbers (Catalán et al., 1997) ranging from basal tribes with small chromosomes and high chromosome base numbers (Brachyelytreae=11; Lygeae=10; Nardeae=13; Phaenospermatae=12; Meliceae=10, 9, 8; Stipeae=12, 11, 10; Diarrheneae=10), through the intermediate ones of Brachypodieae (10, 9, 8, but also 5) (Catalán & Olmstead, 2000), to the large chromosomes and almost constant chromosome base number of x=7 present in the more recently evolved Triticodae + Poodae (Hsiao et al., 1995; Salse et al., 2008; Luo et al., 2009), although x=6, 5, 4, 2 occasionally occur in Aveneae (Poodae) (Catalán et al., 2016b).

Experimental and model organisms

Advances in biology have both benefited from and been predicated on model organisms (Lyons & Scholthof, 2016). Many non-model crop species became research tools because they were of economic importance (e. g., maize, rice, within the grasses). These plants have their limitations, primarily due to their intrinsic domesticated-crop genetic erosion, long seed-to-seed life cycles and the need for extensive growth facilities (Scholthof et al., 2018). In the past two decades, several experimental plants were also used as tractable genomic tools. Experimental organisms and model organisms differ, although both are essential for advances in biology (Leonelli & Ankeny, 2013). In particular, model organisms are systems with deep resources for large-scale biology, ecology, evolution, genetics, cell biology, and availability of diverse lines (wild, isogenic, strains, mutants), infrastructure (databases, seeds), and a culture of sharing, as well as expected features of a short lifecycle, easy and inexpensive cultivation, and readily manipulated in the lab with standard molecular biology techniques. In contrast, experimental organisms, are used to solve a specific question, or are interesting organisms or objects of scientific curiosity (Leonelli & Ankeny, 2013; Scholthof et al., 2018). From this, *Arabidopsis* (for dicots) and *Brachypodium* (for monocots) can be defined as model organisms for plant biology. Furthermore, *Brachypodium distachyon* and other congeners represent a singular example of a model group system for grasses in the post-genomic era of plant biology (Vogel, 2016; Scholthof et al., 2018).

Brachypodium: a model system for biological research in grasses

Brachypodium distachyon was selected as a model organism for grasses based on its suitability in extending our knowledge of grass biology, including fundamental research on plant development, plant-microbe interactions, abiotic stress, evolutionary biology, ecology research, and for the development of new tools and concepts towards improving other temperate C3 grasses, such as wheat and barley, that are crucial small grains used world-wide for food, forage, and feed, and tropical C4 grasses, such as switchgrass (*Panicum virgatum*), and *Miscanthus* spp., that are widely used as biofuel grasses (Kellogg, 2015b; Lyons & Scholthof, 2016; Vogel, 2016; Scholthof et al., 2018).

Model organisms for laboratory research have primarily been used to dissect specific aspects of host biology, such as growth, development or host-environment interactions (abiotic or biotic), following a reductionist approach (Scholthof et al., 2018). With the rise of *Brachypodium*, basic (theoretical, hypothesis-driven) and translational research problems are being solved with the most up-to-date tools of next generation sequencing (NGS), microscopy, and forward genetics that have demonstrated the viability of *Brachypodium* as a tool for grass biology. Additionally, *Brachypodium* spp. have maintained their wildness, providing incomparable resources for ecologists to study the plant in situ. This in turn, will bolster fundamental laboratory studies towards identifying and testing new hypothesis that will benefit agronomists and breeders to improve food and bioenergy-related grasses (Scholthof et al., 2018).

The annual *Brachypodium* species

For more than a century *B. distachyon* (L.) P. Beauv. *sensu lato* (Palisot de Beauvois, 1812) was considered to be the single annual representative species of the genus *Brachypodium* (Schippmann, 1991), and for more than three decades, three cytotypes of 2n=10, 20 and 30 chromosomes were recognized within the species, though they were considered to be diploid, tetraploid and hexaploid individuals of an ascendant autopolyploid series with x= 5 (Talavera, 1978). It was not until recently, however, that the accrued phenotypic, cytogenetic, and molecular phylogenetic evidence demonstrated that the three cytotypes corresponded to three independent species—two diploids, *B. distachyon* (2n=2x=10, x=5) and *B. stacei* (2n=2x=20, x=10), and their derived allotetraploid *B. hybridum* (2n=4x=30, x=10+5) (Catalán et al., 2012). Despite having twice the number of chromosomes, the genome size of *B. stacei* (0.564 pg/2C) was roughly similar to that of *B. distachyon* (0.631 pg/2C), whereas the genome size of *B. hybridum* corresponded to the sum of the two progenitor genomes (1.265 pg/2C). Molecular evolutionary data indicated that *B. stacei* was the oldest diploid lineage within the genus *Brachypodium*, splitting from the common ancestor approximately 10 Ma, followed by the divergence of the *B. distachyon* lineage (~7 Ma), which preceded the split of a clade of recent perennial lineages (core perennial clade; ~3 Ma), and that the allotetraploid *B. hybridum* species originated approximately 1 Ma (Catalán et al., 2012).

Maternally inherited plastid genes supported the recurrent origin of allotetraploid *B. hybridum* from bidirectional crosses of its parents, followed by whole genome duplication of the unfertile interspecific hybrid (López-Alvarez et al., 2012). Recently resequenced nuclear genomes and plastome-based analyses confirmed these findings, showing that most of the studied circum-Mediterranean *B. hybridum* populations were derived from a maternal *B. stacei* parent, whereas only relatively few western-Mediterranean populations were derived from a maternal *B. distachyon* parent.

The biological and genomic attributes that made *B. distachyon* an optimal grass model (small stature, short life cycle, predominantly self-pollinating, small genome size, low amount of repetitive DNA, easy to transform, phylogenetically close to the temperate cereals) are also shared by its congeners *B. stacei* and *B. hybridum* (Catalán et al., 2012, 2016b). This trio of species was proposed as a model complex for (allo)polyploidy, and for the potential application of their comparative functional genomics knowledge to polyploid wheats (Catalán et al., 2014; Gordon et al., 2016). Despite the close morphological resemblances of the three annual *Brachypodium* species, basic statistics and analysis of variance across a wide diversity of wild populations and inbred lines detected eight phenotypic traits [(stomata) guard-cell length, pollen grain length, (plant) height, second leaf width, inflorescence length, number of spikelets per inflorescence, lemma length, awn length] and 434 tentatively annotated metabolite signals that significantly discriminated *B. distachyon*, *B. stacei* and *B. hybridum* (Catalán et al., 2012; López-Álvarez et al., 2017).

These findings, coupled with the identification of five new qualitative traits, helped to characterize and separate the three species. Leaf blade color is an easily identified feature, with *B. distachyon* bright green, *B. stacei* pale green, and *B. hybridum* dark green. *Brachypodium stacei* can be distinguished from the other two species by leaf blade shape (curled vs. straight), softness (soft vs. stiff) and hairiness (densely hairy vs. scarcely hairy or glabrous)] (Catalán et al., 2016b). Intraspecific phenotypic variation was significant among populations of *B. stacei* and *B. distachyon* (López-Álvarez et al., 2017), and part of this morphological variation correlated with genetic divergence in western Mediterranean populations of the two parental species (Shiposha et al., 2016; Marques et al., 2017). Notably, disparate circum-Mediterranean populations of *B. hybridum* originated from contrasting bidirectional crosses were less differentiated phenotypically (López-Álvarez et al., 2017).

Phylogenomic studies of *B. distachyon* based on 54 resequenced ecotypes showed a main split of two intraspecific lineages characterized by their flowering-time features, i.e., extremely delayed flowering (EDF+) vs. non-extremely delayed flowering (non-EDF+) lineages, and their respective co-evolving molecular variants of genes known to regulate vernalization (e.g., VRN1, VRN2) and flowering (e.g., CO2, FTL9, FTL13, PHYC, PPD1), whereas none of those traits co-evolved with latitude (Gordon et al., 2017). Whereas the first clade contained lines distributed across the Mediterranean region, the second clade showed the divergence of two geographically constrained eastern Mediterranean (Turkey and other countries, T+) and western Mediterranean (Spain and other countries, S+) groups.

The *Brachypodium* pangenome is based on 54 whole-genome sequence assemblies of geographically diverse ecotypes, 36 of which were also analyzed at the transcriptome level (Gordon et al., 2017). From this, 61,155 total pangenome clusters were classified as core (present in all lines), softcore (present in 95-98% lines), shell (present in 5-94% lines) and cloud (present in 2-5% lines) genes, and contained nearly twice the number of genes present in any individual genome. The study showed that core genes were enriched for essential biological functions (e. g., glycolysis) and were constrained by purifying selection, whereas shell genes were enriched for potentially beneficial functions (e. g., defense, development, gene regulation), displayed higher evolutionary rates, located closer to and were more functionally affected by transposable elements, and were less syntenic with orthologous genes in other grasses (Gordon et al., 2017). Shell genes contribute substantially to phenotypic variation and influence population evolutionary history within *Brachypodium*, as demonstrated for the three phylogenetic groups detected in the study (EDF+, T+, S+), characterized by different flowering time traits and their molecular regulators, associated with different types of core and shell ingroup genes (Gordon et al., 2017).

The availability of the *Brachypodium* pangenome (https://brachypan.jgi.doe.gov/), with its annotated genomes, transcriptomes and transposons, opens new avenues to study the regulatory networks of key physiological and adaptive processes in the model plant and other target grasses.

The perennial *Brachypodium* species

Besides the three most intensively investigated annual species, the genus *Brachypodium* also contains ~17 perennial species distributed worldwide (Schippmann, 1991; Catalán et al., 2016b). The twenty recognized *Brachypodium* taxa are characterized by their typical subsessile spikelet and exclusive embryo development, seed storage proteins, polysaccharides and globulins, stem and leaf fructosans, small genome sizes and large disploidy (Catalán et al., 2016b). They belong to the monotypic tribe Brachypodieae, evolutionarily placed in an intermediate position between the ancestral basal pooids and the recently evolved clade of core pooid lineages, including the economically important Triticeae + Bromeae and Poaeae (Catalán et al., 1997).

Perennial *Brachypodium* species vary widely both in phenotype and origin; they range from the short-rhizomatose, self-fertile, American allotetraploid *B. mexicanum*, a species closely related to the oldest *B. stacei* lineage and biologically and genomically similar to the annual species, to the strong-rhizomatose, outcrossing, and recently evolved Eurasian and African diploid and allopolyploid species of the core-perennial clade, which include some of the largely distributed palaearctic species, such as diploids *B. pinnatum* and *B. sylvaticum*, together with other more restricted endemic species (Catalán et al., 2016b). Two Mediterranean high allopolyploids, *B. retusum* and *B. boissieri*, characterized by their branched woody stems and short strongly inrolled leaves, have inherited ancestral, intermediately evolved and recent genomes, whereas core perennial allotetraploids *B. phoenicoides, B. pinnatum* 4x and *B. rupestre* 4x, characterized by their non-branched stems and long flat leaves, have only inherited recently evolved genomes (Catalán *et al.*, 2016b).

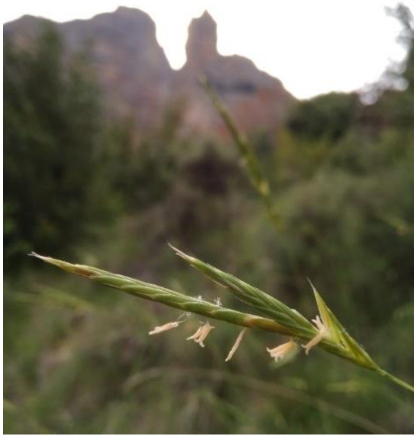


Figure 1. *B. retusum* (Huesca-Aragón-Spain). Author: E. Pérez.

The diploid *B. sylvaticum*, the best known perennial species of the genus, was recently selected as a model plant for perenniality (Gordon et al., 2016). Genomic and transcriptomic resources are available for *B. sylvaticum*, including its reference genome (*B. sylvaticum* Ain1) and a second resequenced line (*B. sylvaticum* Sin1) (see http://phytozome.jgi.doe.gov/). The plant is predominantly self-fertile (94.6%), with a small and compact genome (340 Mb) distributed in 9 chromosomes, and can be easily transformed (Steinwand et al., 2013). Though deeply nested within the core perennial clade of predominantly robust strong-rhizomatose outbreeding species, *B. sylvaticum* shows slender habit and rhizomes and selfing reproductive system; the species however, like the other perennials, is an overwintering plant, characterized by its hairy indumentum, soft leaves, nodding panicle and long awned lemma (Catalán et al., 2016b). Its distribution covers the largest native Old World geographical range of any other *Brachypodium* species, ranging from the Canary Islands (West) to Japan and New Guinea (East) and from Scandinavia and Siberia (North) to northern Africa and Malesia (South), though some of the East Asian and Malesian populations probably correspond to different microtaxa (Catalán et al., 2016b).

Disploidy is a main feature of *Brachypodium*, a genus that contains diploid species with x=10, 9, 8 and 5 chromosomes, and allopolyploid species with different combinations of chromosome base numbers (Catalán et al., 2016b). Phylogenetic and comparative chromosome painting data have been used to propose evolutionary hypotheses on descendant vs. descendant-ascendant disploidy series along the *Brachypodium* tree (Betekhtin et al., 2014) and secondary origins for the allopolyploids (Catalán et al., 2016b). The advent of the sequenced reference genomes would help to reconstruct the path of syntenic chromosome fusions that support the descendant disploidy hypothesis. Interspecific breeding barriers between *Brachypodium* species (Khan & Stace, 1999) are fully congruent with the *Brachypodium* phylogeny, and explain the reproductive isolation of the early diverging *B. stacei*-type (*B. hybridum*) and *B. mexicanum*, the crossability of the intermediately evolved *B. distachyon* with the core perennial species, and the highly fertile descendants of all attempted interspecific crosses between recently evolved core perennial taxa (Khan & Stace, 1999; Catalán et al., 2016b).

Distribution and ecology of *Brachypodium*

The ecology of the ~20 *Brachypodium* taxa varies drastically depending on their geographical distributions and adaptation to different climates and habitats. Among them, the three annual species and the perennials *B. retusum* and *B. boissieri* have adapted to xeric Mediterranean conditions, the Canarian endemic *B. arbuscula* grows in more humid places, the endemic South African *B. bolusii* and Taiwanese *B. kawakamii* thrive in alpine vegetation belts, the tropical African *B. flexum* and Malagasy *B. madagascariense* grows in the afromontane forests and the American *B. mexicanum* survives in xeric to humid neotropical habitats, the western Mediterranean *B. phoenicoides* is adapted to dry edaphically-humid places, and the predominantly Eurasian *B. pinnatum*, *B. rupestre* and *B. sylvaticum* grow in mesic to humid open grasslands and forests (Catalán et al., 2016b). Two *Brachypodium* species have been confirmed as invasive species: *B. hybridum* has successfully and predominantly colonized other Mediterranean-type eco-regions (California, South Africa, South America and southern Australia), and *B. sylvaticum* was introduced and is now spread in humid, forested regions of western North America and Australia (Catalán et al., 2016b).

Detailed ecological studies have been conducted with the three annual circum-Mediterranean species of the *B. distachyon* complex. Environmental niche modeling analysis indicated that, overall, *B. distachyon* grows in higher, cooler and wetter places, north of 33°, *B. stacei* in lower, warmer and drier places, south of 40° 30', and *B. hybridum* in places with intermediate ecological features and across latitudinal boundaries but also overlapping with those of its parents, more often with those of *B. stacei* (Lopez-Alvarez et al., 2015; Catalán et al., 2016b). It concurs with the findings that most *B. distachyon* lines require vernalization treatment to flower, whereas those of *B. stacei* and *B. hybridum* do not (Vogel et al., 2009). Additionally, *B. stacei* grows in shady habitats whereas *B. distachyon* and *B. hybridum* occur in open habitats (Catalán et al., 2016a).

Paleoenvironmental modeling data support the Mediterranean basin and adjacent areas as long-term refugia for *B. stacei* and *B. distachyon*, and some of them as potential hybrid zones which could have favored the recurrent origins of *B. hybridum* since the late Pleistocene. Niche similarity tests showed evidence of niche conservatism for *B. hybridum* and each of its parents; the allotetraploid shares niche occupancy with its progenitors but is reproductively isolated from both of them. Also, *B. hybridum* had the largest niche overlap with its parent niches, but a similar distribution range and niche breadth, indicating that the hybrid does not outcompete its parents in their native ranges (Lopez-Alvarez et al., 2015). Conversely, *B. hybridum* is the only species of the complex that has successfully colonized other non-native world regions (except for one locality in southern Australia where *B. distachyon* (2n=10) has been also found; J. Borewitz, J. Streich and D. López-Álvarez, personal communication), suggesting a greater ecological tolerance of the allotetraploid compared to the diploids that could be associated with increasing genomic and epigenomic expression, boosting diversifying selection, and with rapid shifts in physiological and adaptive traits such as photoperiod and weediness (Bakker et al., 2009; Lopez-Alvarez et al., 2015).

Field analyses have demonstrated that environmental aridity gradients in Spain affect the predominant northern and southern Mediterranean distributions of, respectively, less efficient *B. distachyon* and more efficient *B. hybridum* users of water under water-restricted growing conditions (Manzaneda et al., 2012). Under drought conditions, *B. hybridum* and *B. stacei* individuals behave as drought-escapists, maintaining higher photosynthesis and stomatal conductance and showing earlier flowering times to cope with water stress than the less adapted *B. distachyon* individuals (Manzaneda et al., 2015; Martínez et al., 2018).

Translocation experiments in admixed southern Spanish *B. distachyon* – *B. hybridum* populations have demonstrated the superior capability of the allotetraploid in colonizing densely occupied competitive habitats, and a balance of intra/interspecies competition favoring the establishment of *B. hybridum* over that of *B. distachyon* populations under natural field conditions at the rear-edge distribution of the diploid *B. distachyon* parent (Rey et al., 2017). Field analyses in southern Mediterranean Israel microsites have shown a predominant presence of allotetraploid *B. hybridum* over those of its diploid parents, especially the usually more frequent *B. stacei*, along a large-scale latitudinal range; however, the distribution of *B. hybridum* was not correlated with an aridity cline, though clustered patterns suggested that the distributions of *B. stacei* and *B. hybridum* were not random (Bareither et al., 2017). Ongoing ecogenomic studies of *B. distachyon* – *B. hybridum* populations and *B. stacei* – *B. hybridum* populations in Spain and Israel will further help to decipher the potential drivers of ecological success of parental diploid and allotetraploid populations in different microenvironments.

Genetics and genomics resources from *Brachypodium*

Brachypodium has some obvious advantages over rice (*Oryza sativa*), its closest genetic competitor as a grass model, including a smaller habit, ease of cultivation in the laboratory, and a shorter seed-to-seed life cycle. *Brachypodium* also has a smaller genome size (~272 Mbp) (IBI, 2010), while the rice genome is estimated to be ~430 Mbp (Sasaki & Antonio, 2004), although genome size may be of lesser importance with the rapid advances in NGS technologies (Scholthof et al., 2018). *Brachypodium* is also evolutionarily closer to several major cereal crops like wheat, rye and barley than rice, which makes *Brachypodium* a better suited model to study cereal biology (Draper et al., 2001; Brkljacic et al., 2011; Catalán et al., 2016b).

Brachypodium and rice, both C_3 plants, have less complex genomes, when compared to other grasses with larger genomes. For instance, comparative genomics analyses of *Brachypodium*, rice, sorghum (*Sorghum bicolor*; 730 Mbp), and goat grass (*Aegilops tauschii*; 4020 Mbp) revealed that sorghum and goat grass genes are distributed in clusters or so called "gene insulae" in the genome. The gene insulae contained an average of 3.2 genes/cluster, with non-coding regions separating the clusters. Rice and *Brachypodium*, on the other hand, have genes more uniformly distributed, with short intergenic distances. These differences highlight the complexities of genome expansion on gene distribution and spacing in grasses (Gottlieb et al., 2013), and underscores the simplicity of the *Brachypodium* genome (Scholthof et al., 2018).

The fully annotated reference genome sequence of two commonly used *Brachypodium* accessions (Bd21 and Bd21-3) are publicly available through Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). In addition to these accessions, *de novo* assemblies of 54 *B. distachyon* diverse inbred accessions were completed to enable identification of the full gamut of genes (or pan-genome) of *Brachypodium* (Gordon et al., 2017). The *Brachypodium* pangenome can be accessed from the BrachyPan website (https://brachypan.jgi.doe.gov/). Reference genomes of additional *Brachypodium* species, *B. stacei*, *B. sylvaticum* and *B. hybridum* were also recently completed and are publicly available at Phytozome. Together, these *Brachypodium* spp.

genomes are invaluable resources for grass evolutionary biology, polyploidy and speciation studies.

MATERIAL AND METHODS

Next Generation sequencing (NSG) technologies: Genomics, Transcriptomics and Genotyping

The “omic” approach is the field of research which intergrates studies of many different “omes”, including the genome (genomic), transcriptome (transcriptomic) proteome (proteomic), and metabolome (metabolomics). These approaches use high-throughput technologies enabling scientists to study genomes, transcriptomes, proteomes, and metabolomes at a huge scale (Schuster, 2008). The growing number of sequenced plant genomes has provided a large number of opportunities to study biological processes related to physiology, growth and development, and tolerance to biotic and abiotic stresses at the cellular and whole plant level using a novel systems- level approach (Agrawal et al., 2015).

In the present work we focus on genomics (study of large numbers of genes, or genomes) and transcriptomics (study of the transcriptome—the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using high-throughput methods) approaches applied to phylogeny and gene expression studies.

Next Generation sequencing

Sanger (Sanger et al., 1977) first generation sequencing (FGS) has been and still is used to characterize the genomes of several organisms including model plants as well as major crop species like rice, soybean, sorghum, maize, grape and eucalyptus (Thudi et al., 2012). Next Generation Sequencing (NGS) relies on the amplification and sequencing of single isolated DNA molecules and their analysis in a massive parallel way. Huge amounts of single-stranded DNA molecules are immobilized on solid surfaces such as glass slides or on beads, depending on the platform used (Agrawal et al., 2015). NGS technologies are referred as second-generation sequencing (SGS) technologies, utilized for *de novo* sequencing, genome re-sequencing, and whole genome and transcriptome analysis. More recent NGS technologies are referred to as third-generation sequencing (TGS) or “next-next” generation sequencing (NNGS) technologies (Thudi et al., 2012). We focus on the SGS technologies used in the present study.

Over the last decade several detailed reviews about Next Generation Sequencing have been published (Shendure & Ji, 2008; Metzker, 2010; Liu et al., 2012b; Thudi et al., 2012; Mardis, 2013; Buermans & den Dunnen, 2014; Reuter et al., 2015; Heather & Chain, 2016; Jiao & Schneeberger, 2017). Currently, the most widely adopted SGS platform is Illumina technologies (Thudi et al., 2012; Goodwin et al., 2016) and sequencing data in this study has been conducted using this platform.

Sequencing technologies include a number of steps that are grouped broadly as template preparation (sample preparation or library preparation), sequencing and data analysis.

Library preparation

The main steps to carry out RNA or DNA libraries for NGS analysis are fragmenting (physical, enzymatic or chemical methods) and/or sizing the target sequences to a desired length, converting target to double-stranded DNA, attaching oligonucleotide adapters to the ends of target fragments, and quantitating the final library product for sequencing (Head et al., 2014).

The DNA to be sequenced is used to construct a library of fragments that have synthetic DNAs (adapters) added to both ends of each fragment. Those adapters include (1) sequencing binding site, (2) indexes and (3) a pairing region to the flow cells oligos (oligonucleotides, primers).

Cluster generation and Sequencing

Two methods could be used in preparing templates for NGS reactions: clonally amplified templates originating from single DNA molecules, and single DNA molecule templates (Metzker, 2010). The Illumina platform uses a "DNA colony generation (Bridge amplification)" method and a sequencing by synthesis (SBS) technique named pyrosequencing.

Before the sequencing starts, a cluster generation step is carried out to achieve massive amplification. In bridge-PCR (Fedurco et al., 2006) the library is loaded into a flow cell where fragments are captured on a lawn of surface-bound oligos, complementary to the library adapters. Each fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing (Illumina Inc, 2017).

The subsequent pyrosequencing process (Ronaghi et al., 1998) could be performed using natural nucleotides (instead of the heavily-modified dNTPs used in the chain termination protocols), and observed in real time (instead of requiring lengthy electrophoreses). Illumina SBS technology uses a specific reversible terminator–based method that detects single bases as they are incorporated into the DNA template strands. All four reversible terminator–bounding dNTPs are present during each sequencing cycle (Illumina Inc, 2017). The massively parallel sequencing process is a stepwise reaction series that consists of nucleotide addition, detection and wash steps (Mardis, 2013).

Single-End (SE), Paired-Ends (PE) and Mate Pairs (MP): Library preparation and sequencing

Single-end (SE) and paired-end (PE) sequences are generated when only one end of the nucleic acid fragment (single read) or both ends of each library fragment (paired reads) are sequenced, respectively.

Paired reads are classified as paired-end (PE) reads or mate pairs (MP) reads (Mardis, 2013), depending of the size of the sequenced fragment (over 1 kbp and 1-20 kbp, respectively). Mapping reads to a reference genome using long distance is useful to resolve large structural rearrangements (insertions, deletions, inversions) (Van Nieuwerburgh et al., 2012) and repeated regions. PE sequences are linear fragments with two adapters, while MP fragments are circularized around a single adapter.

The final step is multiplexing, which allows large numbers of libraries to be pooled and sequenced simultaneously during a single sequencing run and then each read can be identified using a index (short DNA sequence) and sorted before the final data analysis (Illumina Inc, 2017).

NGS and Plant Genotyping

NGS technologies have provided a fast and cost-effective way to obtain large sequence sets. It has led to the development of new approaches enabling the discovery of molecular markers in a vast quantity of plant species.

The knowledge of the genotype of a plant allows us to carry out several processes as marker-assisted selection, associating phenotype with polymorphism, DNA barcoding, genetic diversity analyses, conservation genetics, and improvement of genome assemblies (Batley, 2015).

Single Nucleotide Polymorphisms (SNPs) have emerged as the most widely used genotyping markers due to their abundance in the genome and the relative ease in determining their frequency in a cost-effective and parallel way in batches of individuals (Deschamps et al., 2012). Many biological challenges can now be addressed with high accuracy. For example, identifying recombination breakpoints for linkage mapping or quantitative trait locus (QTL) mapping, locating differentiated genomic regions between populations for quantitative genetics studies, genotyping large broods for marker-assisted selection (MAS), resolving the phylogeography of tens of wild populations (Davey et al., 2011) or applying this information to plant breeding (Abe et al., 2002).

Genotyping-By-Sequencing (GBS)

Genotyping by sequencing (GBS) is a form of reduced representation sequencing using restriction enzyme (REs) digested samples. Digesting genomic DNA with a frequent cutter and high-throughput sequencing of all resulting restriction fragments is the mainstay of GBS (Patel et al., 2015).

This approach uses sequences to detect and score SNPs, therefore bypassing the entire marker assay development stage (Deschamps et al., 2012), being suitable for population studies, germplasm characterization, breeding, and trait mapping in diverse organisms (Elshire et al., 2011). Adaptors containing barcodes and common adaptors without barcodes are mixed and used in the ligation reaction. Not all adaptor-ligated fragments will be sequenced, because many fragments will not be efficiently bridge-amplified on an sequencer, either because they do not feature both a barcoded adaptor and a common adaptor, or because they are too long (>1 kb) (Davey et al., 2011).

This approach has been used in a number of studies addressing the phylogeny/phylogeography of plant species such as coffee (Hamon et al., 2017), Carex (Escudero et al., 2014) and Amaranthus (Stetter & Schmid, 2017). It has also been used to analyze plant populations (maize (Beissinger et al., 2013; Romay et al., 2013), barley and maize (Elshire et al., 2011), *Brachypodium distachyon* (Tyler et al., 2016)) and in crop breeding (He et al., 2014; Pootakham et al., 2015; Kim et al., 2016; Scheben et al., 2017).

RNA-seq: library preparation and sequencing

Next-generation cDNA sequencing (RNA-seq) makes possible to sequence complete transcriptomes. A set of RNA (total or fractionated, such as poly-A) is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without amplification, is then sequenced using NGS technologies to obtain sequences from one end (single-end sequencing) or both ends (paired-end sequencing). The reads are typically 30–400 bp, depending on the DNA sequencing technology used (reviewed by Wang *et al.*, 2009). The RNA-seq library could included all RNAs, to capture the complete transcriptome, start with ribosomal RNA (rRNA) depletion or mRNA enrichment, or capture specific RNAs as small RNA (miRNA) (Dijk et al., 2014; Head et al., 2014).

RNA-seq (cDNA) library

The main steps in preparing RNA library for NGS analysis are: (1) fragmenting and/or sizing the target sequences to a desired length, (2) converting target to double-stranded DNA, (3) attaching oligonucleotide adapters to the ends of target fragments, and (4) quantitating the final library product for sequencing (reviewed by Head *et al.*, 2014).

The protocols to RNA-seq (cDNA) library preparation can be classified into two main categories: non-stranded protocols, such as Illumina's TruSeq RNA Sample Preparation Kit, in which RNA sense and antisense strand information is lost (which could be problematic when there are overlapping genomic features), and stranded protocols, such as Illumina's TruSeq Stranded mRNA Sample Preparation Kit, in which the strand information is preserved (Griffith et al., 2015; Hou et al., 2015).

To focus on mRNA, poly-T oligonucleotides fixed to magnetic beads are added to total RNA and selectively bound to messenger RNAs. Any RNA not bound is removed during a wash step and mRNAs are eluted from the beads to use in the first step of library preparation.

3’ Tag-based sequencing

Low-cost Tag-based methods applied to RNA-seq, called TagSeq, have been developed for differencial expression studies (Meyer et al., 2011).

The 3′ RNA library contains only those RNA fragments possessing a poly-A tail and this method yields only one single-end (SE) read per transcript, avoiding the bias produced in long transcripts which are represented by more reads than shorter transcripts. The levels of expression can be estimated directly by the number of reads corresponding to a certain transcript, as a single fragment per mRNA molecule is sampled (Tandonnet & Torres, 2017).

The main caveat of these methods is that they are unable to distinguish between alternatively spliced transcripts from a single locus, or to identify polymorphism or allele-specific expression in much of a gene's coding sequence. Their main strength is the capacity of precisely measuring locus-level transcriptional differences with high replication (Lohman et al., 2016).

Differencial expression (DE) analysis and co-expression networks

The parallel advances of NGS and bioinformatics allowed researchers to apply these technologies to expression profiling (Teixeira Torres et al., 2008). When the main goal is not to obtain the assembled transcriptome but a subset of it, stranded-specific single-end approach is a valid choice (Corley et al., 2017).

RNA expression profiles among samples can be compared to identify differentially expressed genes (DEGs) with the aim of explaining phenotypic differences and short-listing candidates genes involved in responses to abiotic or biotic stresses.

Gene co-expression networks (GCN) are powerful graph-theory tools to carry out simultaneous identification and linking of thousands of genes through analyses of their expression profile from transcriptomic data (microarray and RNA-seq data) comparing different conditions as treatments, tissues or species. Genes, nodes in the terminology of graphs, are arranged in adjacency matrices that summarize their co-expression patterns. Nodes with similar expression profiles are clustered in modules. Studies of gene co-expression networks have demonstrated that modules are often constituted by genes with similar functions (Stuart et al., 2003; Wolfe et al., 2005).

Graph topological features are used to define the structure of a network and the interacctions between its nodes. Those features are defined according to Zhang & Horvath (2005), Dong & Horvath (2007) and Horvath & Dong (2008).

- Connectivity (degree): row sum of the adjacency matrix. For weighted networks, sum of connection strengths to other nodes.

$$Connectivity = \ k_i = \sum_{j \neq i} a_{ij}$$

- Scaled connectivity: the connectivity vector scaled by the highest connectivity in the network. Range 0 to 1.

$$Scaled\ connectivity = K_i = \frac{k_i}{\max(k)}$$

- Clustering coefficient: measures the cliquishness of a particular node (a node is cliquish if its neighbors know each other). Range 0 to 1.

$$ClusterCoef_i = \frac{\sum_{l \neq i} \sum_{m \neq i,l} a_{il} a_{lm} a_{mi}}{(\sum_{l \neq i} a_{il})^2 - \sum_{l \neq i} a_{il}^2}$$

- Maximum adjacency ratio (MAR): a useful measure for weighted networks to determine whether a node has high connectivity because of many weak connections (small MAR) or because of strong (but few) connections (high MAR). Range 0 to 1.

$$MAR_i = \frac{\sum_{j \neq i} a_{ij}^2}{\sum_{j \neq i} a_{ij}}$$

- Density: mean adjacency.

$$Density = \frac{\sum_i \sum_{j \neq i} a_{ij}}{n(n-1)} = \frac{mean(k)}{n-1}$$

  where *n* is the number of network nodes.

- Centralization: range 1 (star topology) to 0 (all nodes with the same connectivity).

$$Centralization = \frac{n}{n-2}\left(\frac{\max(k)}{n-1} - Density\right) \approx \frac{\max(k)}{n-1} - Density$$

- Heterogeneity: coefficient of variation of the connectivity.

$$Heterogeneity = \frac{\sqrt{variance(k)}}{mean(k)}$$

Weighted gene co-expression network analysis (WGCNA) is a systems biology method for describing the correlation patterns among genes defining a ‘soft’ thresholding that assigns a connection weight to each gene pair (Zhang & Horvath, 2005; Langfelder & Horvath, 2008). Genes that show a high number of interactions with other genes, i.e. nodes which have high connectivity (“hub” genes) within a weighted co-expression network, are thought to play an important role in organizing the behavior of biological networks (Albert et al., 2000; Carlson et al., 2006; Dong & Horvath, 2007) (fig. 1).

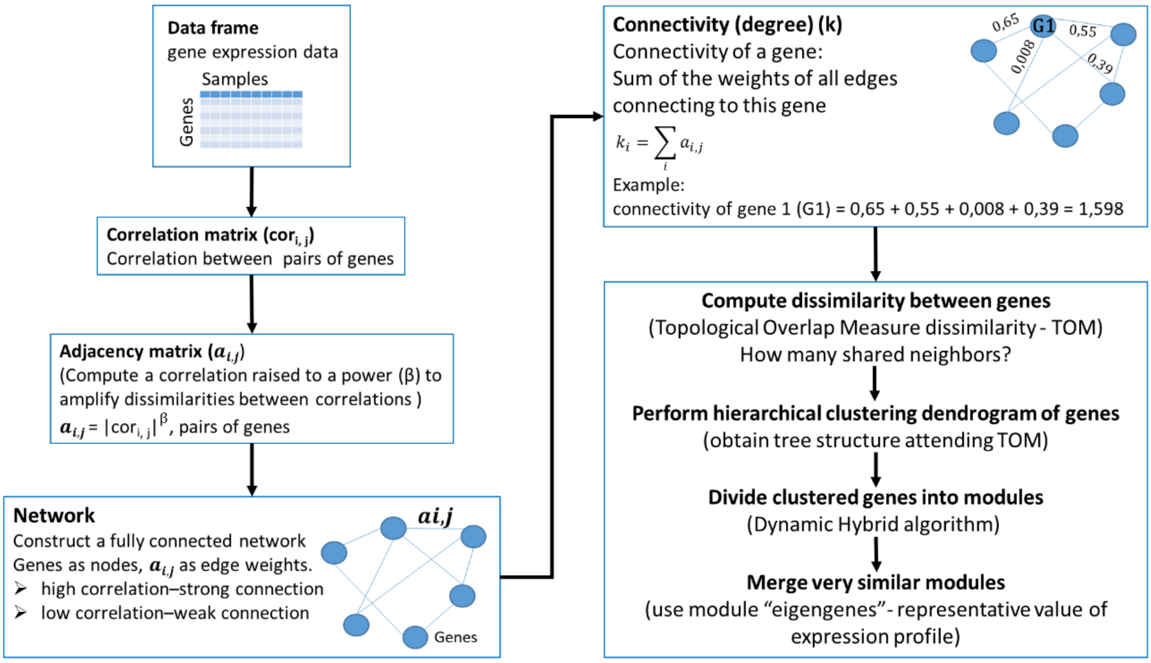

Data frame
gene expression data
Samples
Genes
Correlation matrix ($cor_{i,j}$)
Correlation between pairs of genes
Adjacency matrix ($a_{i,j}$)
(Compute a correlation raised to a power (β) to amplify dissimilarities between correlations)
$a_{i,j} = |cor_{i,j}|^{\beta}$, pairs of genes
Network
Construct a fully connected network
Genes as nodes, $a_{i,j}$ as edge weights.
➢ high correlation–strong connection
➢ low correlation–weak connection
ai,j
Genes
Connectivity (degree) (k)
Connectivity of a gene:
Sum of the weights of all edges connecting to this gene
$k_i = \sum_i a_{i,j}$
Example:
connectivity of gene 1 (G1) = 0,65 + 0,55 + 0,008 + 0,39 = 1,598
G1
0,65
0,55
0,008
0,39
Compute dissimilarity between genes
(Topological Overlap Measure dissimilarity - TOM)
How many shared neighbors?
Perform hierarchical clustering dendrogram of genes
(obtain tree structure attending TOM)
Divide clustered genes into modules
(Dynamic Hybrid algorithm)
Merge very similar modules
(use module “eigengenes”- representative value of expression profile)

Figure 1. Overview of Weighted Gene Co-expression Network Analysis

Co-expression network analyses of plants have flourished during the last decade (Aoki et al., 2007; Serin et al., 2016). These analyses combine large data bases (He & Maslov, 2016) in model plants such as *Arabidopsis* (Mao et al., 2009), grasses as rice (Childs et al., 2011) and maize (Huang et al., 2017a). In some cases even different species have been compared in the same study to identify genes linked to specialized metabolic pathways (Wisecaver et al., 2017). Those approaches also can be applied to study genes involved in adaptation to different abiotic stressor conditions as temperature or water deficiency (Des Marais et al., 2017a; Miao et al., 2017) or biotic stresses as those caused by microbial pathogens (Amrine et al., 2015).

Systematics, Phylogenetics, Phylogenomics and Biogeography

Systematics is the science or field of biology focused on the recognition of basic units in nature (taxa), the classification of those taxa in a hierarchic scheme and the placement of information about them and their classification in some broader context (Schuh, 2000).

Three main evolutionary schools have applied their criteria to the study of taxonomy: evolutionary systematics, phenetics and cladistics. Evolutionary systematics, proposed by Theodosius Dobzhansky, Ernst Mayr, G. G. Simpson, and Julian Huxley, is based on the character-state similarity of a group. Groups are designated using combinations of derived, ancestral, unique and shared characters. The phenetic approach (Sneath & Sokal, 1973) concedes that evolutionary history can not be deciphered as consequence of parallelism and reversal. Phenetic classification is based on the observation of many characteres turn to quantitative values. The UPGMA method (unweighted pair-group method with arithmetic mean) is a typical method for constructing trees in phenetic approach. UPGMA assumes the same evolutionary speed on all lineages (constant rate of mutations over time and for all lineages in the tree). The cladistic approach, developed by Willi Hennig, is the majoritarily adopted approach (though not the unique, see, for example, the evolutionary systematics approach) for systematics (Davis, 1997; Judd et al., 2008). The cladistic approach is rooted on the criterion that only shared derived characters could proportionate information about phylogeny and, therefore, on systematics.

Similar features (character states) between two species that have been inherited from common ancestors are called homologous features (homology) and those features could be informative for resolving the evolutionary relationships between organisms. By contrast, when the similar features between species could be a consequence of convergent or parallel evolutions (e. g., species with similar adaptive or genomic traits), those features are called analogous (homoplasy), and they could not be used for phylogenetic reconstruction. Furthermore, only the shared homologies, called synapomorphies, evidence that two organisms are closely related (Lipscomb, 1998).

Although phylogenetic descent relationships can be disrupted by reticulation (e.g. hybrid allopolyploid species), a point of divergence among lineages is usually reached at which phylogenetic relationships show a hierarchical structure represented by a tree with diverging branches. Nodes for every pair of elements can be identified as the most recent common ancestor (MRCA) (Davis, 1997). Monophyly has been a topic of discussion from the beginning of the cladistic school (reviewed by Vanderlaan *et al.*, 2013). In brief, monophyly is a unified criterion in hierarchical descent systems, where three key attributes occur simultaneously, common ancestry for all the members of a group (i.e., the group includes an ancestor and all of its descendants), exclusivity of kinship (i.e., every member of the group is more closely related to every other member than to any non-member), and phylogenetic nesting of such groups (i.e., if there is any overlap in the membership of two different monophyletic groups, one of the groups is completely included in the other) (Davis, 1997).

Some pitfalls as speciation events closely spaced in time (e. g., small phylogenetic signal, short internal tree branches), leading to undesirable lineage sorting events, or ancient events largely spaced in time(e. g., long terminal branches with multiple substitutions occurring at the same position), leading to disturbing long branch attraction events, contribute significantly to the difficulty of reconstructing the correct phylogenetic tree for a set of sequences (Philippe et al., 2011).

The phylogenomic method proposed by Eisen et al. (1998) is an approach to combine evolutionary and genetic information to improve functional predictions. This method is based on the assumption that gene functions change as a result of evolution, and therefore reconstructing the evolutionary history of genes should help predict the functions of uncharacterized genes (Eisen, 1998). The main steps of this approach are to infer the phylogenetic tree representing the evolutionary history of the gene of interest and its homologues (genes that descended from a common ancestor) to biologically determine functions of the various homologues that are overlaid onto the tree and the structure of the tree, and to trace the history of functional changes from the relative phylogenetic positions of genes of different functions in the tree, which is then used to predict functions of uncharacterized genes (Eisen, 1998).

The most popular phylogenomic approach is known as the ‘‘supermatrix’’ (or superalignment), consisting in concatenating numerous orthologous genes into a single supergene data set (reviewed by Philippe *et al.*, 2011).

Reconstruction methods to infer phylogenetic/phylogenomic trees

Three families of reconstruction methods can be used to infer phylogenetic/phylogenomic trees: distance-based methods, and character-based methods divided into, respectively, maximum parsimony and likelihood based methods (reviewed by Delsuc *et al.*, 2005).

Distance methods

These methods first convert the character matrix into a triangular distance matrix that represents the evolutionary distances between all pairs of species. The phylogenetic tree is inferred from the distance matrix using algorithms such as unweighted pair group method with arithmetic mean (UPGMA by Sokal & Michener (1958)), neighbour joining (NJ by Saitou & Nei (1987)) or minimum evolution (ME by Kidd & Sgaramella-Zonta (1971) and Rzhetsky & Nei (1992)).

Maximum parsimony method

Maximum parsimony method (MP; (Fitch, 1971; Farris, 1983)) is a character-based method which analyses all possible tree topologies from the given input data and chooses the optimal tree (most parsimonious tree), i. e. the tree that requires the minimum number of character changes to explain the observed data (Delsuc et al., 2005). This method is also used to infer phylogenetic networks (Kannan & Wheeler, 2012).

Likelihood methods: Maximum Likelihood and Bayesian methods

Maximum likelihood (ML, (Felsenstein, 1981) methods estimate the parameters of one or more statistical evolutionary models and assign probabilities for a group of possible phylogenetic trees. The optimal tree is considered to be the one that has the highest probability (Felsenstein, 1981; Eisen, 1998), thus the less negative log likelihood.

Bayesian methods (Rannala & Yang, 1996; Mau et al., 1999; Li et al., 2000) are based on the posterior probability of a tree defined as "the probability that the tree is correct, assuming that the model is correct" and it is calculated by numerical methods as Markov chain Monte Carlo-MCMC (Huelsenbeck et al., 2001; Huelsenbeck & Rannala, 2004).

One of the major distiction is that Maximum likelihood analyses take a long time to run, and bootstrap analyses requiere a high-performance computer. Bayesian methods estimate support for the tree from all saved posterior probability trees (Judd et al., 2008).

Distance, maximum likelihood and Bayesian methods use an evolutionary model to describe the data, whereas maximum parsimony methods do not have an explicit model. Models of evolution describe the rates of change of fixed mutations among sequences and constitute the basis of the evolutionary analysis of genetic data at the molecular level (reviewed by Arenas, 2015).

Dating the molecular phylogenetic tree

Sequence data can be used to estimate the timing of the evolutionary events and the rates of molecular evolution through the association of externally derived dates obtained from fossil or biogeographical evidence to internal nodes of the tree. This calibration system, using an external source of information, is required to convert relative into absolute divergence times.

One calibration approach is to assign dates to internal nodes representing the most recent common ancestors (MRCAs) between lineages using information from the fossil record or from dated biogeographical events. Other approaches take information about the age of the sequenced samples themselves to calibrate the phylogeny by assigning dates to the tips (sometimes also called terminal nodes) of the tree, hence the term tip dating (reviewed by Rutschmann, 2006; Rieux & Balloux, 2016). One of the most popular tools for phylogenomic dating, “Bayesian Evolutionary Analysis by Sampling Trees” (BEAST2), is conducted using a set of Bayesian tree priors (e. g., Yule model, Birth-death model, coalescent model) (Bouckaert et al., 2014).

Biogeography

Reconstructing the historical biogeography of a clade relies on our ability to infer the nodal ancestral ranges of its lineages (Bouchenak-Khelladi et al., 2010). Inferring the evolution of ancestral ranges of clades within a phylogenetic context is a major focus of historical biogeography (Ree & Smith, 2008).

Parametric biogeography integrates processes, patterns and time (Sanmartín, 2012). These methods are termed “model based” or “parametric” because they are based on statistical models of range evolution, whose parameters (“variables”) are biogeographic processes such as dispersal, range expansion, or extinction. Range evolution—i.e., the change in geographic range from ancestor to descendants—is modeled as a stochastic process that changes along the branches of the phylogenetic tree according to a probabilistic Markov-chain model. The Markov-chain model uses a matrix of transition probabilities that determines the instantaneous rate of change from one state to another. The states of the Markov process are the set of discrete geographic areas that form the distribution range of the group (e.g., A, B, and AB). The parameters of the model are biogeographic processes that change the geographic range of the species, such as range contraction (extinction, EA) or range expansion (DAB). By letting the model evolve along the branches of the phylogeny, which here represents the time since cladogenesis, we can estimate the rates (probabilities) of occurrence of the biogeographic processes (DAB, DBA, EA, EB) and infer the most probable ancestral ranges at every cladogenetic event (Buerki et al., 2011; Sanmartín, 2012).

Parametric methods can evaluate every possible ancestral area in terms of its "likelihood" (probability) of explaining the data. They integrate the uncertainty in the reconstruction of ancestral ranges in the phylogeny ("mapping uncertainty"). Parametric methods can estimate the parameters over every possible tree topology and combination of branch lengths, and therefore they can account for the uncertainty associated with the phylogenetic inference itself ("phylogenetic uncertainty"). Parametric methods provide an appropriate statistical approach to compare alternative biogeographic hypotheses or scenarios. Each scenario is formulated in terms of different parametric models, which can be compared on the basis of how well they fit the data. Because the parameters of each alternative model are biogeographic processes, one can identify the processes that best explain the biogeographic patterns by identifying the "best-fitting" model, for example, by using likelihood-based statistical tests. Parametric methods integrate into the biogeographic inference estimates of the evolutionary divergence between lineages or the time since cladogenesis, which are represented by the length of branches in the phylogeny (Sanmartín, 2012).

One of the most commonly used parametric methods is the DEC (Dispersal-Extinction-Cladogenesis) model (Ree et al., 2005; Ree & Smith, 2008). DEC is a ML-based method that allows estimating by maximum likelihood rates of range expansion (dispersal) and contraction (extinction), and range inheritance scenarios at cladogenetic events from a time-calibrated phylogeny with terminal lineage distributions. The DEC model is implemented in Lagrange (Ree & Smith, 2008). The DEC analysis requires a maximum clade credibility (MCC) chronogram (a fully resolved ultrametric tree), a matrix of current range distribution of species in operational areas, and a dispersal rate matrix between pairs of operational areas (for simple or stratified models). The analysis reconstruct the probabilities of the nodal ancestral areas in the phylogeny, allowing the inference of past biogeographic events (vicariance, dispersal, peripheral isolations, extinctions) along the nodes and the branches of the tree (Buerki et al., 2011).

Comparative genomics and phylogenomics tools

The increasing popularity of cost-efficiency of NGS and its application to comparative genomics and evolutionary analysis has been aided by the developent of a large number of tools and software packages. Bioinformatic tools play a fundamental role in the processing of *big data*. Evolutionary analytical methods are key approaches for analysing these data sets and for inferring new relevant conclusions on decisive biological and evolutionary events of organisms.

Some of the bioinformatic tools used in this work can be classifed as i) quality checking and pre-processing tools (FastQC (Andrews, 2010); Trimmomatic (Bolger et al., 2014)); ii) read mappers (DNA *vs* DNA as BWA (Li & Durbin, 2009); RNA *vs* DNA as HISAT2 (Kim et al., 2015)); iii) sequence aligners such as Mafft (Katoh & Standley, 2013), Clustal Omega (Sievers et al., 2011), or whole-genome aligner Cgaln (Nakato & Gotoh, 2010); iv) post-processing tools (SAMtools (http://samtools.sourceforge.net/) and BCFtools (http://samtools.github.io/bcftools/) (Li & Durbin, 2009; Li et al., 2009)); v) genotyping tools (GIbPSs (Hapke & Thiele, 2016); NGSEP (Perea et al., 2016)),), vi) genome assemblers (DNA: Velvet (Zerbino & Birney, 2008) and SSPACE (Boetzer et al., 2011); RNA: Trinity (Grabherr et al., 2011)); vii) pangenome clustering tools such as GET_HOMOLOGUES-EST (Contreras-Moreira et al., 2017) and viii) genome visualization tools such as IGV (Thorvaldsdóttir et al., 2013).

Several phylogenetic and phylogenomic inference softwares have been used in this study for Maximum Likelihood (RAxML (Stamatakis, 2014) and IQ-Tree (Nguyen et al., 2014)) and Bayesian (MrBayes (Ronquist & Huelsenbeck, 2003) phylogenetic reconstructions and for Bayesian (BEAST2 (Bouckaert et al., 2014) dating approaches.

The software STRUCTURE based on a Bayesian clustering approach using Markov Chain Monte Carlo (MCMC), population ancestry and allelic correlations models (Pritchard et al., 2000; Porras-Huratdo et al., 2013) was used to estimate population structure.

The curation and revision of sequences, as well as calculations of raw and patristic distances and Neighbor Joining (NJ) clustering analyses were conducted with the Geneious software (Kearse et al., 2012).

Reconstruction of species level gene genealogies in the form of haplotypic networks was performed with the TCS tool (Clement et al., 2000, 2002). Some putative plastome microrecombinations events were analysed using the recombination detection program RDP4 (Martin et al., 2015).

Sometimes it was necessary to develop custom pipelines and tools to complete the scheduled analyses. In those cases we chose to publish their codes and documentations in public repositories (i.e. GitHub: https://github.com/) for transparency and to ensure the reproducibility of the work.

OBJETIVOS

El objetivo general de la tesis doctoral ha sido utilizar especies del género *Brachypodium* (Poaceae) como plantas modelo para descifrar procesos evolutivos de especiación híbrida y alopoliploide en las gramíneas templadas mediante análisis de genómica comparada y filogenómicos, inter e intra-específicos, empleando genomas nucleares y organulares (plastoma), genes independientes y transcriptomas, así como la identificación de genes implicados en la tolerancia a estrés hídrico.

El objetivo general incluye los siguientes objetivos específicos:

- Reconstruir la filogenia y la biogeografía de las especies reconocidas de *Brachypodium* mediante el análisis evolutivo del gen nuclear copia simple GIGANTEA (GI), y de otros genes nucleares (ITS, ETS) y plastídicos (*ndh*F, *trn*LF).
- Reconstruir la filogenia y datar los orígenes de los genomas y subgenomas presentes en especies diploides y alopoliploides de *Brachypodium* empleando aproximaciones transcriptómicas y de genotipado (GBS).
- Ensamblar, anotar y analizar la evolución de los genomas organulares (plastomas) de las especies anuales de *Brachypodium* y su comparación con sus filogenias nucleares. Dilucidar la dinámica poblacional y la diversificación de sus ecotipos.
- Identificar y analizar genes funcionales implicados en la respuesta ambiental a estrés hídrico mediante el análisis de redes de co-expresión génica y de genes diferencialmente expresados en diversos ecotipos de la planta modelo *Brachypodium distachyon*.

OBJECTIVES

The main goal of the PhD thesis is to use species of the genus *Brachypodium* (Poaceae) as model plants to decipher the evolutionary processes involved in the hybrid and allopolyploid speciation events of temperate grasses. The objective was attained through comparative genomic and phylogenomic analyses at inter and intra-specific levels of the studied species, using nuclear and plastid genes, genomes and transcriptomes, and through the identification of genes involved in the tolerance to drought stress.

The main objective includes the following specific objectives:

- The reconstruction of the phylogeny and the biogeography of the recognized species of *Brachypodium* based on evolutionary analyses of the nuclear single copy gene GIGANTEA (GI), and of other nuclear (ITS, ETS) and plastid (*ndh*F, *trn*LF) genes
- The reconstruction of the phylogeny and the estimation of the times of divergence of the genomes and subgenomes present in diploid and allopolyploid species of *Brachypodium* using transcriptomic and genotyping-by-sequencing approaches (GBS).
- The assembly, annotation and evolutionary analysis of the organellar genomes (plastomes) of the annual *Brachypodium* species and their comparison to nuclear genome based phylogenies. The elucidation of the population dynamics and diversification of their ecotypes.
- The identification and the analysis of the functional genes involved in the environmental response to drought stress through the study of co-expression gene networks and of differentially expressed genes across several ecotypes of the model plant *Brachypodium distachyon*.

Chapter 1. Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches

Summary

The identification of homeologous genomes and the biogeographical analyses of highly reticulate allopolyploid-rich groups face the challenge of incorrectly inferring the genomic origins and the biogeographical patterns of the polyploids from unreliable strictly bifurcating trees. Here we reconstruct a plausible evolutionary scenario of the diverging and merging genomes inherited by the diploid and allopolyploid species and cytotypes of the model grass genus *Brachypodium*. We have identified the ancestral *Brachypodium* genomes and inferred the paleogeographical ranges for potential hybridization events that originated its allopolyploid taxa. We also constructed a comprehensive phylogeny of *Brachypodium* from five nuclear and plastid genes using Species Tree Minimum Evolution allele grafting and Species Network analysis. The divergence ages of the lineages were estimated from a consensus maximum clade credibility tree using fossil calibrations, whereas ages of origin of the diploid and allopolyploid species were inferred from coalescence Bayesian methods. The biogeographical events of the genomes were reconstructed using a stratified DEC model with three temporal windows. Our combined ME-coalescence-Bayesian approach allowed us to infer the origins and the identities of the homeologous genomes of the *Brachypodium* allopolyploids, matching the expected ploidy levels of the hybrids. To date, the current extant progenitor genomes (species) are only known for *B. hybridum*. Putative ancestral homeologous genome have been inherited by *B. mexicanum*, ancestral and recent genomes by *B. boissieri*, and only recently evolved genomes by *B. retusum* and the core perennial clade allopolyploids (*B. phoenicoides*, *B. pinnatum* 4x, *B. rupestre* 4x). We dissected the complex spatio-temporal evolution of ancestral and recent genomes and have detected successive splitting, dispersal and merging events for dysploid homeologous genomes in diverse geographical scenarios that have led to the current extant taxa. Our data support Mid-Miocene splits of the Holarctic ancestral genomes that preceded the Late Miocene origins of *Brachypodium* ancestors of the modern diploid species. Successive divergences of the annual *B. stacei*

and *B. distachyon* diploid genomes were implied to have occurred in the Mediterranean region during the Late Miocene-Pliocene. By contrast, a profusion of splits, range expansions and different genome mergings were inferred for the perennial diploid genomes in the Mediterranean and Eurasian regions, with sporadic colonizations and further mergings in other continents during the Quaternary. A reliable biogeographical scenario was obtained for the *Brachypodium* genomes and allopolyploids where homeologous genomes split from their respective diploid counterpart lineages in the same ancestral areas, showing similar or distinct dispersals. By contrast, the allopolyploid taxa remained in the same ancestral ranges after hybridization and genome doubling events. Our approach should have utility in deciphering the genomic composition and the historical biogeography of other allopolyploid-rich organismal groups, which are predominant in eukaryotes.

Introduction

Phylogenetic and biogeographical studies of highly reticulate allopolyploid plant groups have been severely hampered by the difficulty or impossibility of reconstructing bifurcated tree-like topologies from genome-mergers and genome-doubled species, which render network-like phylogenies (Jones et al., 2013; Marcussen et al., 2015). In grasses, where allopolyploids account for 70% of the current species (Stebbins, 1949; Kellogg, 2015a), comparative genomic studies support the existence of an ancient Whole Genome Duplication (WGD) event, estimated to have occurred ca. 90 Ma (Salse et al., 2008). The return to the diploid state was followed by new polyploidizations, leading to the rise of meso- and neo-polyploids, which originated in the Early-Mid Neogene and the Quaternary, respectively (Stebbins, 1985). Though the role of allopolyploidy in species diversification has been extensively debated (Soltis et al., 2014a; Soltis & Soltis, 2016), there is general agreement on the importance of this mechanism and its preeminence in some angiosperm lineages (Brysting et al., 2007; Marcussen et al., 2015). Most allopolyploids have experienced multiple recurrent origins from different parental populations (Soltis et al., 2014a). In some instances, similar directional crosses led to distinct allopolyploid grass speciation events (e. g., *Aegylops*; (Meimberg et al., 2009)), whereas in others all sorts of bidirectional crosses led to the same speciation outcome (e. g., *Brachypodium hybridum*;(López-Álvarez et al., 2017)).

Brachypodium has received considerable attention since the selection of the annual *B. distachyon* as model functional plant for temperate cereals and biofuel grasses (IBI, 2010; Mur et al., 2011) and of its three annual species as a model group for allopolyploid speciation (Catalán et al., 2014; Gordon et al., 2016). This genus, characterized by its small-size and compact genomes (Betekhtin et al., 2014), is as an ideal model for comparative genomics of monocots (Kellogg, 2015b). *Brachypodium* belongs to the monotypic tribe Brachypodieae and contains between 18 and 20 taxa (Catalán et al., 2016b) (Fig. 1). Dated phylogenies of plastid and nuclear rDNA genes support a rapid and relatively recent radiation of the genus since the Mid-Miocene, showing the early divergences of annual and short-rhizomatose lineages and the recent split of the strong-rhizomatose core perennial lineages (Catalán et al., 2012). Phylogenetic trees reconstructed from single-copy nuclear genes supported this hypothesis, but also showed homeologous copies in all of the polyploid lineages studied to date (Wolny et al., 2011; Catalán et al., 2012, 2016b).

Alternative phylogenetic methods have been proposed to reconstruct and date the species network of reticulate allopolyploid groups, including comparative statistical analysis of diploid/polyploid multiple gene tree discordances (Cai et al., 2012) and dated allopolyploid network analysis (Marcussen et al., 2015). Other authors used multilabeled gene trees (Huber et al., 2006), with auto- and allo-polyploids represented by one or more tip leaves, respectively, to estimate the relative time of origin of homeologous genomes (Estep et al., 2014). However, some of these scenarios appear to be constrained for complex groups such as *Brachypodium*, where highly divergent homeologous genomes have been observed within single allopolyploids (Catalán et al., 2016b). This, in turn, suggests that putative *Brachypodium* ancestors could have evolved in different geographic locations.

A preliminary evolutionary analysis of the *Brachypodium* taxa was performed in our previous work (Catalán et al., 2016b). We grafted the polyploid alleles into a diploid species tree using a minimum evolution criterion aiming to draft a general scenario explaining the putative origins of the polyploid species. We observed four main placements of polyploid allelic copies in basal, *stacei*, *distachyon* and core perennial clade branches, with some putative recent polyploids sharing also basal allelic copies. Nevertheless, statistical refinements are necessary to correct the excess of allelic copies grafted to different branches of the skeleton diploid species tree in order to properly infer the origins and the hybridization patterns of the homeologous genomes present in the allopolyploids.

In this study we have incorporated a statistical treatment that corrects the excess of allelic copies by fusing closely related copies located in close branches. The main objectives were to identify the genome donors of the allopolyploids and to obtain a biogeographic scenario for the known taxa of *Brachypodium*. Homeologous genomes now merged in the allopolyploids could have arrived at their current geographic locations from different ancestral ranges historically occupied by diploid or low polyploid ancestors. Therefore, we decided to adopt a novel biogeographic approach that independently handles each homeologous genome with the aim of inferring its ancestry range and its time of divergence from its closest diploid lineage. This approach allowed us to reconstruct a chronogram that included all grafted heterologous copies of a polyploid species to inform the biogeographic analysis. This strategy is conceptually different from most current biogeographic studies, where typically a single genomic copy is selected for each polyploid species (Linder & Barker, 2014; Fougére-Danezan et al., 2015).

Given these considerations, the objectives of our research were i) to incorporate statistical support for the allele grafting method to identify specific *Brachypodium* homoeologous genomes; ii) to reconstruct a robust explicit phylogenetic framework using a multigenic Species Network to disentangle the complex reticulate history of diploid and allopolyploid taxa, including all the identified genomic copies; iii) to build a dated chronogram for the multigenic allelic copies of *Brachypodium*; iv) to reconstruct the historical biogeography of its genomes using parametric dispersion-extinction-cladogenesis models, inferring the paleo-scenarios for the dispersals and merging of genomes; and v) to estimate the coalescence ages of polyploid genomes from their closest diploid relatives, identifying and dating the hybridization events that gave rise to the allopolyploid species and cytotypes.

Materials and Methods

We used the data matrices generated by Catalán *et al.* (2016), although the data processing and the statistical methods used to reconstruct the diploid species tree and the grafting of polyploid alleles into this tree have been updated and are described in detail in this study. We have included new divergence time estimations, coalescence dating analysis and biogeographic methods. A general scheme of the analyses performed in this study is shown in Fig. 2.

Sampling, DNA sequence data processing and haplotype networks

Our sampling was designed to represent the taxonomic diversity and geographic distribution of *Brachypodium* taxa (Catalán et al., 2016b) as well as the intraspecific cytotypic variability described for some perennial species (Betekhtin et al., 2014). A total of 110 ingroup samples representing the 17 recognized species plus one variety of *Brachypodium* were included (Fig. 1; Table S1 and Methods/Results S1). The outgroup species were represented by ancestral and recently evolved Pooideae (*Melica* *ciliata*, *Glyceria declinata, Secale cereale*, *S. montanum*, *Festuca arundinacea*, *F. pratensis,* and *Lolium perenne*). *Oryza sativa* (Oryzoideae) was included as external outgroup (BOP clade) and used to root the trees.

DNA sequences from three nuclear [rDNA ETS and ITS genes, and a single-copy GIGANTEA (GI) gene] and two plastid (*ndh*F, *trn*LF) genes were used to reconstruct the phylogeny of *Brachypodium*. The protocols used for DNA isolation, amplification, cloning and sequencing are described in Methods/Results S1. Five clones per sample were used for each nuclear locus in both diploid and polyploid taxa, aiming to detect all potential copies. A total of 973 *Brachypodium* sequences were aligned with sequences retrieved from Genbank (Table S1 and Methods/Results S1). The final data sets consisted of 431 sequences/682 aligned positions for ETS, 368/645 for ITS, 280/831 for GI, 95/564 for *ndh*F, and 100/941 for *trn*LF. The non-recombinant *ndh*F + *trn*LF plastid (cpDNA) sequences were concatenated into a combined 105/1505 data set. In order to discard spurious variation generated from PCR or cloning artifacts, intraspecific consensus (type) sequences were generated following Díaz-Pérez *et al.* (2014). Closely related sequences of the same species that showed a p-distance lower than 0.01 base differences per site were collapsed into a consensus type sequence using MEGA v. 5 (Tamura et al., 2011) and BIOEDIT v. 7.0.9.0 (Hall, 1999) (Methods/Results S1). The consensus types that were represented by a single clone were discarded. The haplotype networks were constructed using statistical parsimony (Clement et al., 2000) and POPART (Leigh & Bryant, 2015), with a 95% cut-off for the maximum number of mutational connections between pairs of sequences.

Diploid species tree reconstruction

A Bayesian diploid backbone species tree was constructed from consensus sequences (types) from each separate locus with *BEAST v.2.1.3 (Bouckaert et al., 2014), using *Festuca pratensis* to root the tree. All parameters were unlinked across loci to allocate different evolutionary models in the species tree estimation. Initially, we imposed nucleotide substitution models according to the selection of the best model based on the AIC criterion computed in MODELTEST v.3.4 (Posada & Crandall, 1998), and the maximum likelihood test (LSet command) computed in PAUP* v.4.0b10 (Swofford, 2003), among alternative models and a strict molecular clock model. However, convergence of the MCMC chain for the four data sets could only be achieved after imposing the simple HKY85 substitution model and a strict molecular clock model. In these searches the evolutionary rate was set to 1.0, scaling node and root heights in units of mutations per site, and assuming a Yule birth tree prior. The MCMC was run twice for 500 million steps, logging parameters every 10 thousand samples, and checking for convergence in TRACER v.1.6.0 (Rambaut et al., 2014)(Rambaut et al., 2014) with effective sample size (ESS) values above 200. Log-files were combined after discarding the first 50% of each sampling as burn-in. The posterior distribution of trees was summarized through a maximum clade credibility tree with TREEANNOTATOR v.2.1.2 and visualized with FIGTREE v.1.4.2 in the BEAST package (Bouckaert et al., 2014).

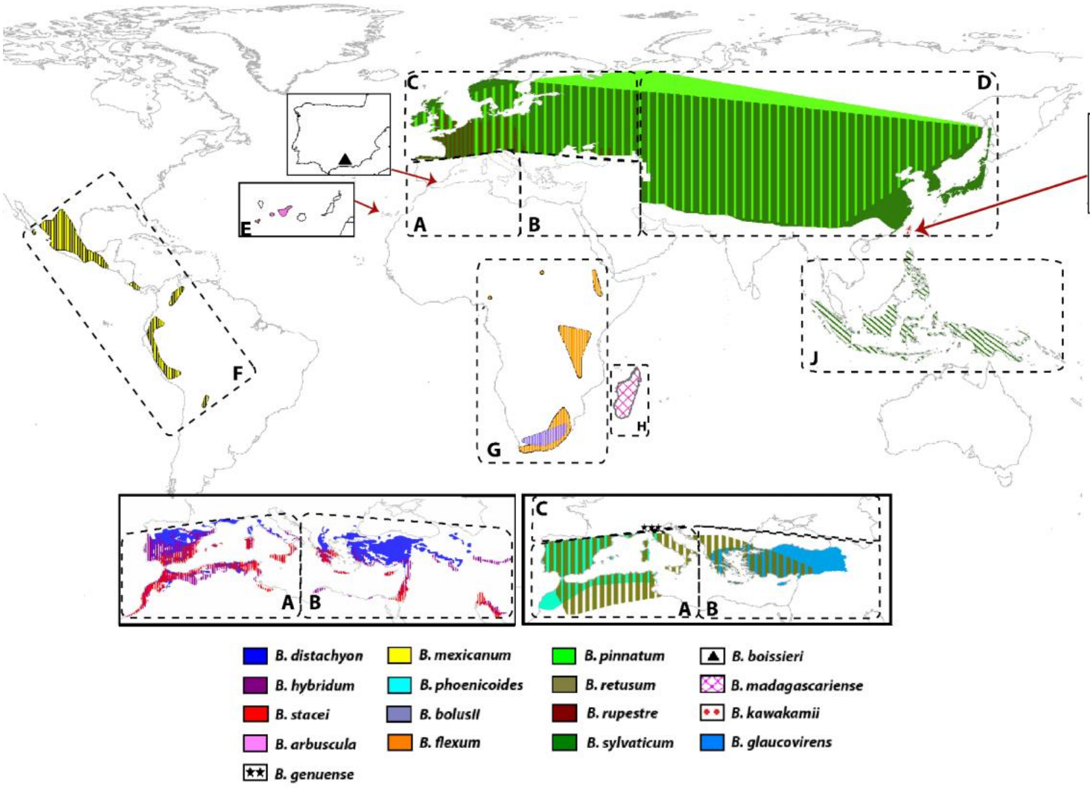


Figure 1. The worldwide geographic distribution of the 18 *Brachypodium* taxa and the boundaries of the 10 operational areas used in the biogeographic study [A) western Mediterranean; B) eastern Mediterranean + SW Asia; C) western Eurasia (from Atlantic to Urals); D) eastern Eurasia (from Urals to Pacific and eastern Asia); E) Canary Isles; F) America (from Mexico to Peru-Bolivia); G) Africa (Tropical Africa and South Africa); H) Madagascar; I) Taiwan; J) Malesia (including Papua-New Guinea)]. The species ranges colors and marks are indicated in the chart.

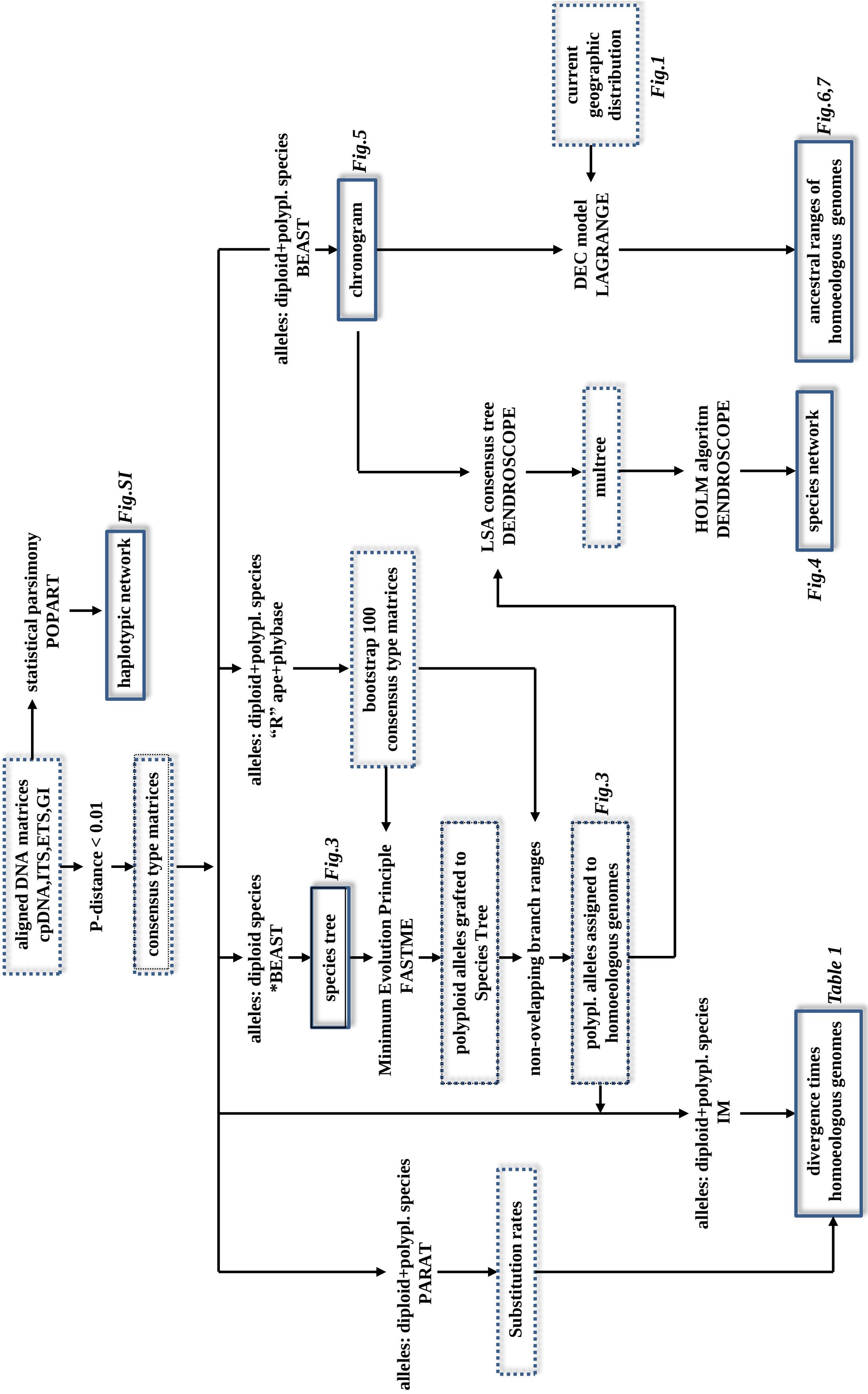


Figure 2. The general pipeline used for the statistical methods employed in this study. The boxes with solid and dashed lines represent main and secondary outputs, respectively. The software used for each aspect of the pipeline is indicated in capital letters.

Grafting polyploid alleles into the diploid species tree

A modified procedure of Cai *et al.* (2012) was used to graft individual alleles of polyploid species to specific branches of the diploid species tree using the Minimum Evolution criterion. In this analysis, all polyploid and skeleton diploid alleles (used to generate the species tree) per locus were analyzed to construct a gene tree. Different pruned gene trees were generated by pruning all polyploid alleles except one, per analysis. This excluded allele was treated as missing in the remaining gene trees of the other three loci, which were solely composed of skeleton diploid alleles. Several integrated distance matrices were constructed by averaging distances between diploid species from the four loci, but each time the process included single-locus internodal distances between the respective polyploid allele and diploid sequences. The distances were estimated by the average number of internodes between all pairs of tips from the gene trees. For diploid species, internode distances were averaged across all gene trees and all pairs of samples for each species-pair. This generated as many distance matrices as single-locus polyploid alleles were available. Distance matrices were calculated from maximum likelihood gene trees that were previously estimated through RAxML v.7.2.6 (Stamatakis, 2006), using the R-package APE (Paradis et al., 2004). The rooted species tree of all diploid *Brachypodium* taxa had 15 branches after excluding the branch leading to the outgroup. To estimate the optimal placement(s) of the polyploid allele in this tree, each polyploid allele was inserted in every potential branch, rendering 15 species trees per allele. The lengths of the trees were calculated according to the Minimum Evolution method implemented in FASTME (Desper & Gascuel, 2002), using the integrated distance matrices and fixing each of the 15 species trees per polyploid allele. A set of contiguous branches was selected as the optimal placement for each polyploid allele in the diploid tree. This set was defined as those branches whose associated tree lengths were contained in the lowermost 5% cutoff of the observed range of tree lengths. For each allele, this process was repeated 100 times from bootstrap pseudoreplicates, as indicated in Cai *et al.* (2012), giving bootstrap support for the contiguous range of branches where this allele was grafted. Non-overlapping ranges were treated as different sets of polyploid alleles. In *B. mexicanum*, two ranges partially overlapped, but each range showed a marked concentration of bootstrap placements in different branches of the tree. Each set of alleles was considered as a single putative homeologous genome. Homeologous genomes were classified depending on their topological proximity to counterpart diploid lineages in the tree.

Dating analysis

We constructed a chronogram including all *Brachypodium* polyploid and diploid alleles using BEAST v.2.1.3. For this, we assumed that the origin of polyploid alleles was circumscribed to an interval of time delimited by the parent and child nodes of the specific branch of the species tree onto which these alleles were grafted. Consequently, the topology of the diploid species tree and the minimum evolution placement of each polyploid allele were fixed in this analysis. To fix a set of polyploid alleles to a single branch of the tree, we constrained in BEAST v.2.1.3 the monophyly of a group that included these alleles, plus all of the diploid and polyploid alleles previously nested in more recent branches. To graft the polyploid alleles onto the terminal branches of the species tree, they were constrained to a monophyletic group that also included the respective diploid species. Parameters were unlinked across the four loci using an optimal GTR+GAMMA substitution model. The MCMC and posterior distribution processing and summarizing were similar to those of the diploid species tree reconstruction, except that the MCMC was run five times for 100 million steps.

The selection of tree priors were based on Bayes factors (BF) where Marginal Likelihood Estimators (MLE) were generated according to the Path-Sampling (PS) and Stepping-Stone (SS) methods as implemented in BEAST. The Uncorrelated Relaxed Clock (UCLD)-Birth-Death model was chosen over the UCLD-YULE with a PS and SS MLE of -13211.5 vs -13236.9 and -13211.6 vs -13237.0, respectively, yielding a decisive BF of 22.5 with both estimators. The Strict Clock tree prior did not reach convergence so we could not estimate BF to test them against UCLD models. MLE are highly influenced by prior distributions, but we did not detect any mismatch between simulated and theoretical prior distributions for multiple calibrated internal nodes (see below), as suggested by Heled & Drummond (2012). Moreover, "the ucld.sdev" estimate obtained from UCLD models was clearly different from zero, indicating variability of branch rates, giving an indirect support to UCLD over the Strict prior. Because there are no described fossils of *Brachypodium*, we dated the more inclusive data sets. For this, we calibrated the crown node of the BOP clade imposing a secondary calibration of 54.9 ± 5.7 Ma (normal prior distribution) according to the family-wide analysis of Bouchenak-Khelladi *et al.* (2010). A pooid epidermal phytolith fossil from the Middle Eocene (Strömberg, 2011) provided a minimum age for the crown node of Pooideae of 48.4 Ma [log-normal prior distribution mean=3.88, stdev=0.05, 95% highest posterior density (HPD) interval 44.6 to 52.58 Ma].

Divergence times of homeologous genomes and plausible ages of hybridization events

We assumed that a homeologous *Brachypodium* genome diverged from an ancestral diploid parental lineage, represented by the current diploid closest relative(s) identified in the Minimum Evolution tree. Pairwise divergence times were computed using an "Isolation-with-Migration" model according to the Bayesian method of Hey & Nielsen (2004) implemented in the program IM v.3.5. The bidirectional migration rates and population size parameters were enforced to be the same in all cases. These parameters were used to simplify the model and to maintain agreement with the recent radiation observed for the *Brachypodium* clade lineages (Catalán et al., 2012, 2016b). Population parameters were scaled by μ (the neutral mutation rate), the effective number of gene copies (*Ne*), the migration rate (*M*) and the divergence time (*T*). These parameters were estimated from the model parameters $\theta = 4Ne\mu$, $m = M/\mu$ and $t = T\mu$. The estimated IM coalescent diverging times should not be confused with the estimated *BEAST lineage diverging times; *BEAST estimates the relative divergence times of diploid genome lineages, whereas IM estimates the demographic divergence time of each homeologous genome from its diploid relative. Three simulations per pairwise divergence estimation between a homeologous genome and its counterpart diploid genome were performed with $2x10^6$ burn-in and $3x10^6$ iterations to check for convergence, in addition to ESS > 300. A total of 22x3=66 pairwise runs were performed (Table 1). Wide uniform priors were assigned in the first run to set appropriate limits for the priors of the two subsequent independent runs. There were a variable number of loci available for pairwise comparisons, ranging from one to four loci depending on the genome (Table 1). In this case, we suggest that most estimates should be taken as approximate values, despite the fact that convergence was achieved and the replicated runs generated similar values. Considering that homeologous genomes could never have originated before than their more recent genome donors, we equated the time of the putative hybridization event with the time of the origin of the most recent counterpart diploid genome.

To transform model population parameters estimates into demographic units, μ rates of the four loci were approximated through the estimation of substitution rates (K) using the program PARAT (Meyer & Haeseler, 2003). This program included an iterative procedure to estimate the topology, branch lengths and site specific substitution rates. For each pair of sequences, the neutral mutation rate was estimated as $\mu = K/2T_C$, where T_C is the coalescence time obtained from the BEAST chronogram (see above). Pairwise μ's for consensus sequences located in different clades of the chronogram tree were averaged to feed the IM analysis. Estimates of substitution rates (x10^{-9} s/s/y) generated in this study were 1.317, 1.5535, 2.4667 and 2.7064 for the GI, cpDNA, ETS and ITS loci, respectively.

Table 1. The estimated age (Ma) of homeologous genomes present in the allopolyploid *Brachypodium* species. This is inferred from the coalescent splits from their respective closest counterpart diploid genome lineages, computed through the Isolation-Migration model implemented in IM. A square box represents the age of the most recent homeologous genome in a taxon and the inferred time for the putative origin of the hybrid. The ploidy levels correspond to those indicated in Table S1. The numbers within the square brackets indicate the number of loci used for each estimation. The numbers within parentheses correspond to the homeologous genomes participating in the allopolyploids, ranging from the youngest (1) to the oldest (2) or (3). The Ancestral Areas (AAs) represent a matrix occupied by the homeologous genomes (rows) when they diverged from their respective diploid relatives (columns). The AAs of a cell represent the sum of the AAs of all parent nodes of all allelic copies assigned to a homeologous genome (see colored lineages in Figs. 6 and 7), just before the time of divergence from its diploid genome. For example, in *B. flexum* its *ARBUSCULA* (0.609 Ma), *SYLVATICUM* (0.197 Ma) and *PINNATUM* (0.024 Ma) homeologous genomes originated in BG, B and G, respectively; when *SYLVATICUM* and *PINNATUM* split, the more ancestral *ARBUSCULA* was already distributed in G, and when *PINNATUM* split *SYLVATICUM* was also distributed in G; all three ancestral homeologous genomes merged in the same area (G) giving rise to *B. flexum*. The AA codes represent: A, western Mediterranean; B, eastern Mediterranean + SW Asia; F, America; G, Africa; H, Madagascar; and I, Taiwan. The designation (*) ANCESTRAL indicates the ancestral homeologous genome without any known diploid relative. The age estimation was performed using *B. stacei* as a reference. The designation (**) IM indicates coalescent diverging times that are estimates of the demographic divergence time of each homeologous genome from its diploid relative. For example, the *STACEI* homeologous genome of *B. hybridum* might have diverged more recently from *B. stacei* than the *DISTACHYON* homeologous from *B. distachyon* (this Table), despite the BEAST species tree indicates that the *B. stacei* lineage is more ancestral than that of *B. distachyon* (Fig. 6).

Polyploid species	time**	AA		Polyploid species	time	AA		
		(2)	(1)			(3)	(2)	(1)
***B. hybridum* (4x)**				***B. boissieri* (cf. 8x)**				
(1) *STACEI* [4]	$\boxed{0.035}$		AB	(1) *SYLVATICUM* [1]	$\boxed{0.030}$			A
(2) *DISTACHYON* [2]	0.060	AB	AB	(2) *DISTACHYON* [1]	3.750		A	A
				(3) *ANCESTRAL* [3]	16.915*	A	A	A
***B. bolusii* (unknown)**								
(1) *ARBUSCULA* [2]	$\boxed{0.027}$		G	***B. flexum* (unknown)**				
(2) *SYLVATICUM* [2]	0.379	G	G	(1) *PINNATUM* [1]	$\boxed{0.024}$			G
				(2) *SYLVATICUM* [1]	0.197		B	G
B. madagascariense				(3) *ARBUSCULA* [2]	0.609	BG	G	G
(unknown)								
(1) *ARBUSCULA* [1]	$\boxed{0.390}$		H					
(2) *SYLVATICUM* [2]	0.441	AG	AG	***B. retusum* (6x)**				
				(1) *PINNATUM* [1]	$\boxed{0.036}$			A
***B. mexicanum* (4x)**				(2) *ARBUSCULA* [1]	0.037		B	B
(1) *STACEI* [2]	$\boxed{3.377}$		F	(3) *SYLVATICUM* [2]	0.466	A	B	B
(2) *ANCESTRAL* [2]	11.070	F	F					
				B. kawakamii				
				(unknown)				
***B. phoenicoides* (4x)**				(1) *PINNATUM* [1]	$\boxed{0.067}$			A
(1) *PINNATUM* [1]	$\boxed{0.048}$		A	(2) *ARBUSCULA* [1]	0.309		I	I
(2) *SYLVATICUM* [3]	0.052	A	A	(3) *SYLVATICUM* [2]	0.476	AI	AI	I

Species Network reconstruction

A species network was reconstructed from the BEAST chronogram using the HOLM algorithm (Huber et al., 2006) implemented in DENDROSCOPE v.3.2.10 (Huson & Scornavacca, 2012). This algorithm generates a phylogenetic network with a minimum number of polyploidization events, suggesting the merging pattern of homeologous genomes of a polyploid species. Alleles from the four loci grafted to different branches in the same allopolyploid species were given the same code to convert the chronogram into a multilabeled tree. To simplify the representation of the network, each homeologous genome per polyploid species was represented by a single consensus type in the multilabeled tree. Nonetheless, we observed that the polyploids *B. phoenicoides*, *B. madagascariense* and *B. kawakamii* showed two consensus types assigned to the same *SYLVATICUM* homeologous genome according to the Minimum Evolution criterion (see Results). Consequently, and aiming to correct it, we generated different alternative multilabeled trees, each time dropping one consensus type of each species from the chronogram. Then, these topologies were condensed into a single consensus tree using the Lowest Stable Ancestor algorithm implemented in DENDROSCOPE v.3.2.10. Starting from the multilabeled tree, a collection of maximal inextendible subtrees (MIS) were subdivided, identified and pruned. The resulting network contained fewer leaves than the original multilabeled tree and, in some cases, different collections of MIS. The search steps were repeated until no MIS remained (Huber et al., 2006).

Biogeographic reconstruction of *Brachypodium* genomes

We used the BEAST chronogram and a parametric Dispersal-Extinction-Cladogenesis (DEC) approach to reconstruct the ancestral range distributions and the biogeographic scenarios of the *Brachypodium* genomes. We assumed that before the hybridization, each separate genome evolved independently from each other and that after the hybridization the merged homeologous genomes (subgenomes) evolved in parallel within the same allopolyploid lineage and ancestral range (see Results). This assumption is justified by the fact that once two homeologous genomes reached the same ancestral area, they did not disperse to different areas later (see Table 1 and Results for more details). Alternative DEC models were compared through Maximum Likelihood analysis in LAGRANGE v. 20130526 (Ree & Smith, 2008). The chronogram was also used to infer global extinction and dispersal rates and range inheritance scenarios at each node.

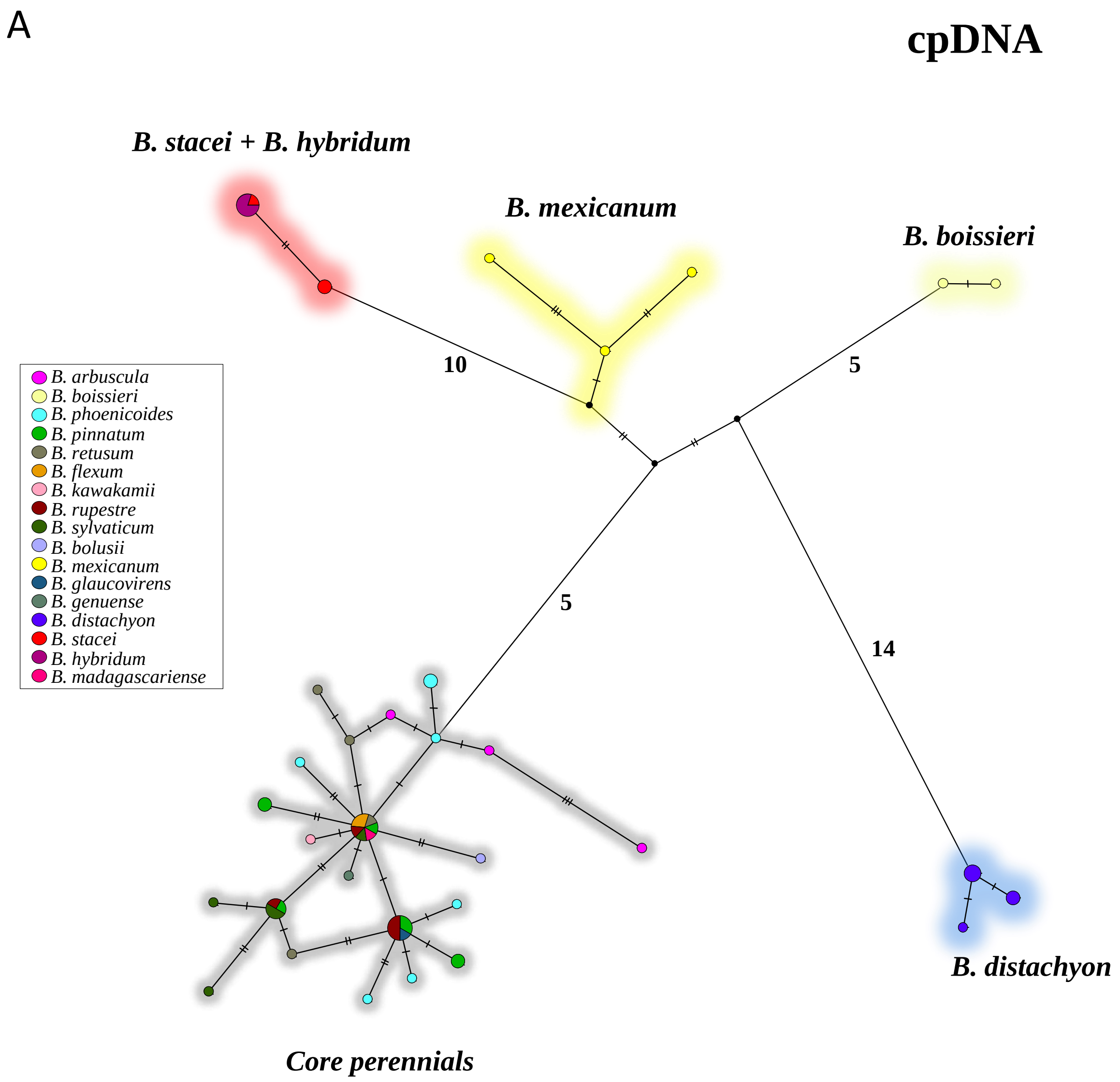
We defined 10 operational areas (OAs) for reconstructing the biogeography of the *Brachypodium* genomes (Fig. 1; Table S2). The OAs were selected according to the current distribution of taxa, but also reflected the geological history of the study area: A) western Mediterranean; B) eastern Mediterranean + SW Asia; C) western Eurasia (from the Atlantic to the Urals); D) eastern Eurasia (from the Urals to the Pacific and eastern Asia); E) Canary Islands; F) America (from Mexico to Peru-Bolivia); G) Africa (Tropical Africa and South Africa); H) Madagascar; I) Taiwan; and J) Malesia (including Papua-New Guinea). Given the relatively disjunct and isolated distribution of most current *Brachypodium* taxa, the DEC analyses were constrained to a maximum number of two areas at ancestral nodes, assuming that ancestors (and genomes) were not more widespread than their extant descendants.

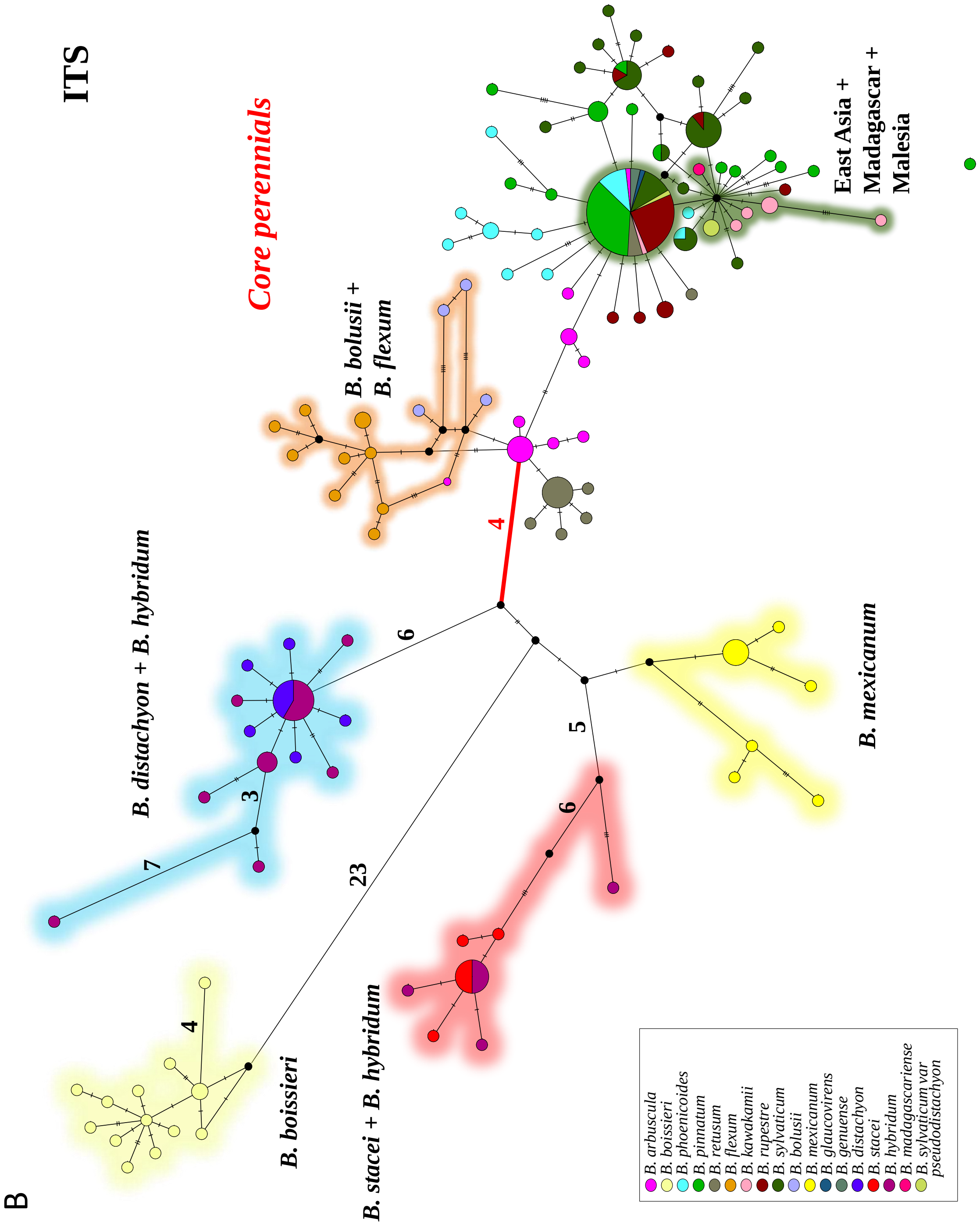
Two alternative DEC models were used to infer the biogeographical events along the branches of the *Brachypodium* chronogram, an unconstrained model (M0), where dispersal rates between all biogeographic areas were constant through time, and a constrained stratified model (M1), where the topology was divided into three temporal windows, each with a specific matrix of dispersal rates set according to paleogeographic connectivity (Table S2). Three time slices were defined: TSI, Mid-Miocene (Langhian) to Messinian (16.2-7.2 Ma); TSII, Messinian to Pleistocene (7.21-2.6 Ma); and TSIII, Quaternary (2.61-0 Ma). These time slices were used to reflect the foremost paleogeographic events of both hemispheres that could have affected the divergence of the current *Brachypodium* lineages.

Results

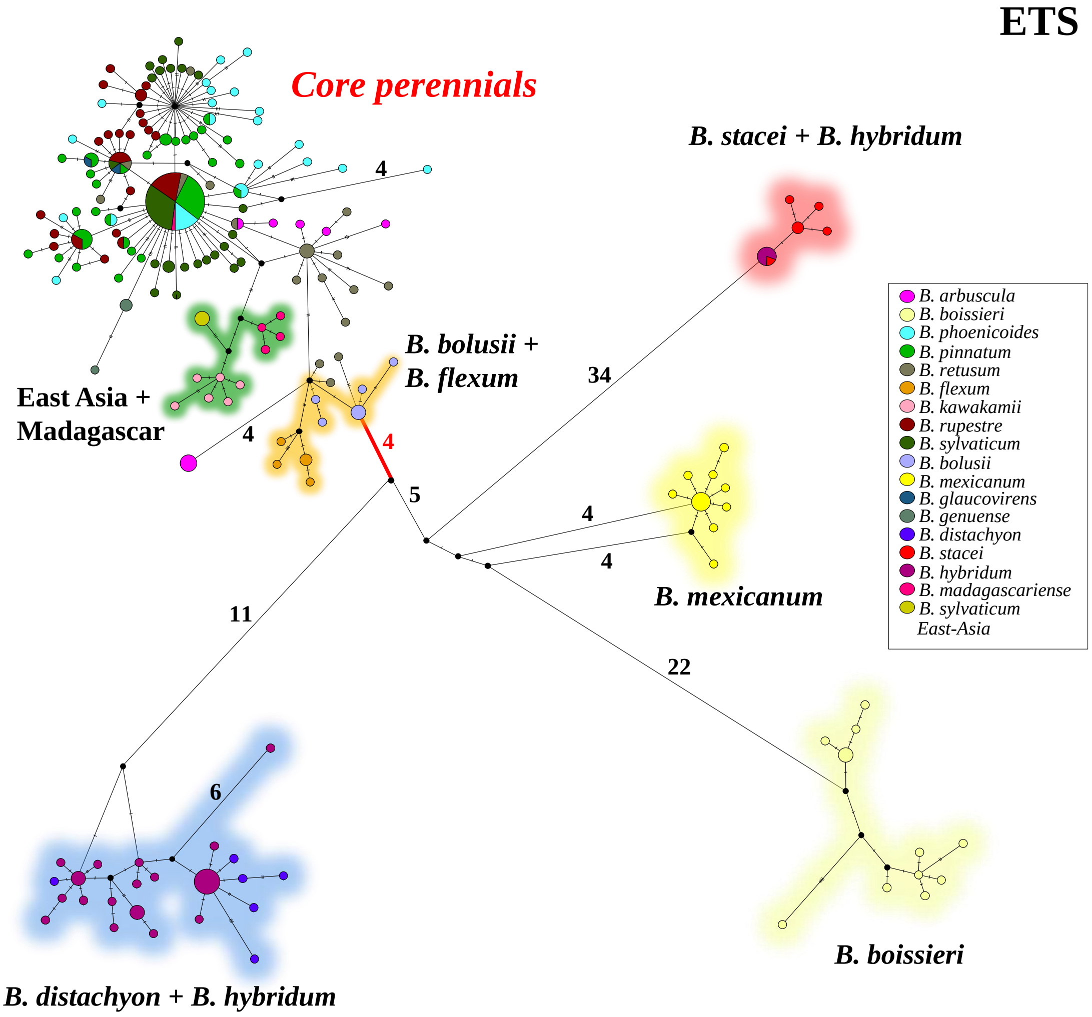
<u>The *Brachypodium* Species Tree and inference of allopolyploid homeologous genome lineages</u>

Single-locus haplotypic networks and phylogenetic trees of *Brachypodium* based on plastid, ITS, ETS and GI data were in agreement with in the earliest divergences of *B. stacei*, *B. mexicanum* and *B. distachyon* lineages, and of a more recent split of the core perennial group (Figs. 3A-D; Methods/Results S1). The ETS and ITS data also detected the early divergence of the African *B. bolusii*/*B. flexum*, the Canarian *B. arbuscula* and the Mediterranean *B. retusum* lineages within the core perennials clade, and the clustering of endemic East Asia- Madagascar [*B. sylvaticum* (China)/*B. kawakamii*, *B. madagascariense*] and East Asia-New Guinea (*B. kawakamii*/*B. sylvaticum* var. *pseudodistachyon*) haplotypes, respectively. The three nuclear genes (ETS, ITS, and GI) identified co-inherited *B. stacei*-type and *B. distachyon*-type parental copies in *B. hybridum*, and a number of co-inherited ancestral and recently evolved homeologous copies among the perennial allopolyploid species (Figs. 3B-D and Methods/Results S1).




B
ITS
Core perennials
East Asia +
Madagascar +
Malesia
B. bolusii +
B. flexum
B. distachyon + B. hybridum
B. mexicanum
B. boissieri
B. stacei + B. hybridum
4
6
5
23
3
7
6
4
B. arbuscula
B. boissieri
B. phoenicoides
B. pinnatum
B. retusum
B. flexum
B. kawakamii
B. rupestre
B. sylvaticum
B. bolusii
B. mexicanum
B. glaucovirens
B. genuense
B. distachyon
B. stacei
B. hybridum
B. madagascariense
B. sylvaticum var
pseudodistachyon

C


ETS
Core perennials
B. stacei + B. hybridum
B. bolusii +
B. flexum
East Asia +
Madagascar
B. mexicanum
B. distachyon + B. hybridum
B. boissieri
4
34
4
4
4
5
4
11
6
22
● B. arbuscula
● B. boissieri
● B. phoenicoides
● B. pinnatum
● B. retusum
● B. flexum
● B. kawakamii
● B. rupestre
● B. sylvaticum
● B. bolusii
● B. mexicanum
● B. glaucovirens
● B. genuense
● B. distachyon
● B. stacei
● B. hybridum
● B. madagascariense
● B. sylvaticum
East-Asia

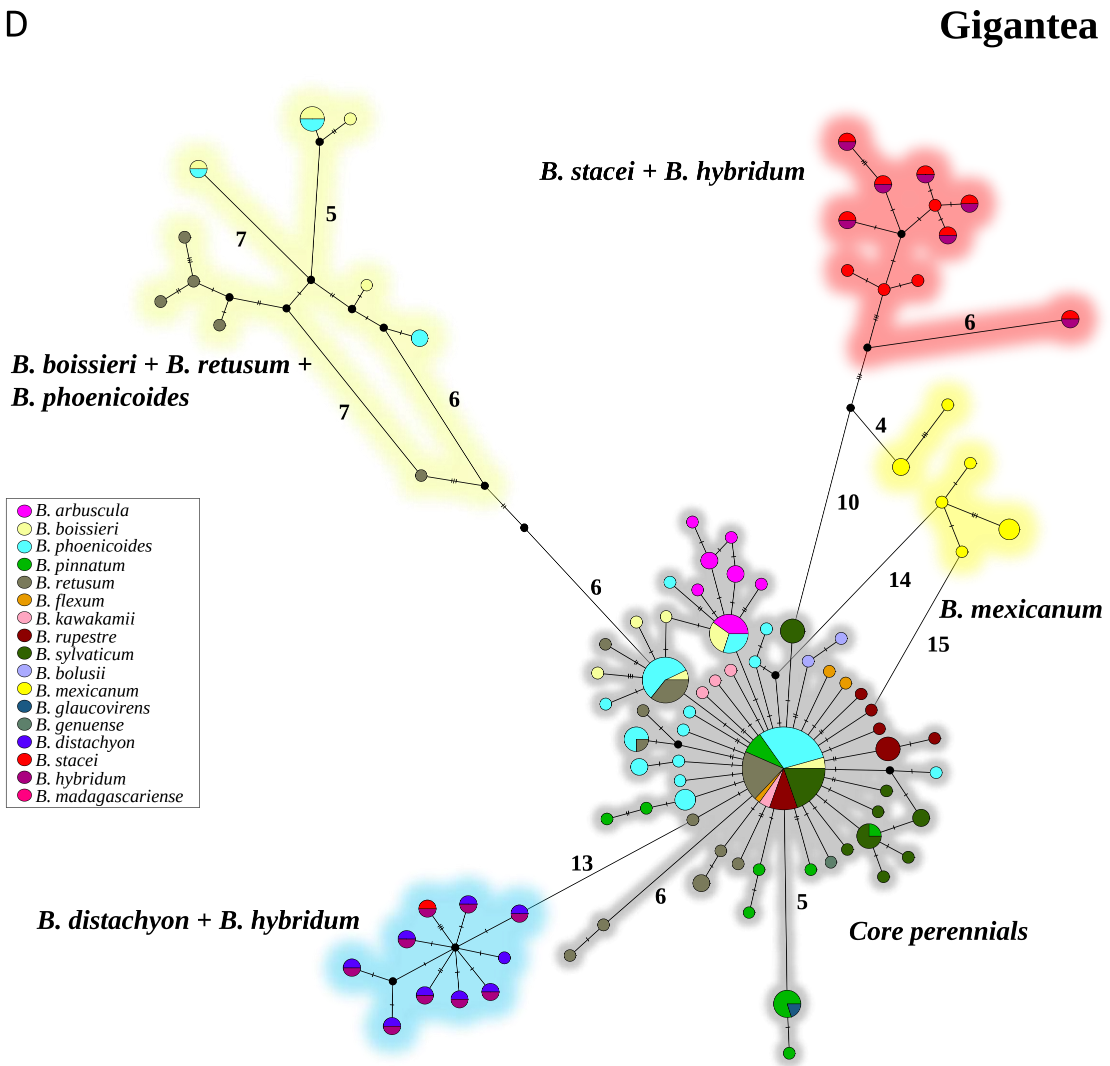


Figure 3. The statistical parsimony networks constructed with POPART for **(A)** the chloroplast (ndhF + trnLF), **(B)** the nuclear ITS, **(C)** the nuclear ETS, and **(D)** the nuclear GIGANTEA (GI) haplotypic data sets. The species colors are indicated in the charts. The size of the circles is correlated with the number of samples showing the haplotype.

Our diploid tree, which included only *Brachypodium* species of confirmed diploid nature (Fig. 4), showed the earliest divergence for the annual *B. stacei* lineage, then the annual *B. distachyon* and lastly the clade of core perennial taxa, which successively split into the *B. arbuscula*, *B. genuense*, *B. sylvaticum*, *B. glaucovirens*, and *B. pinnatum* 2x (2n=18)/*B. rupestre* 2x (2n=18) lineages.

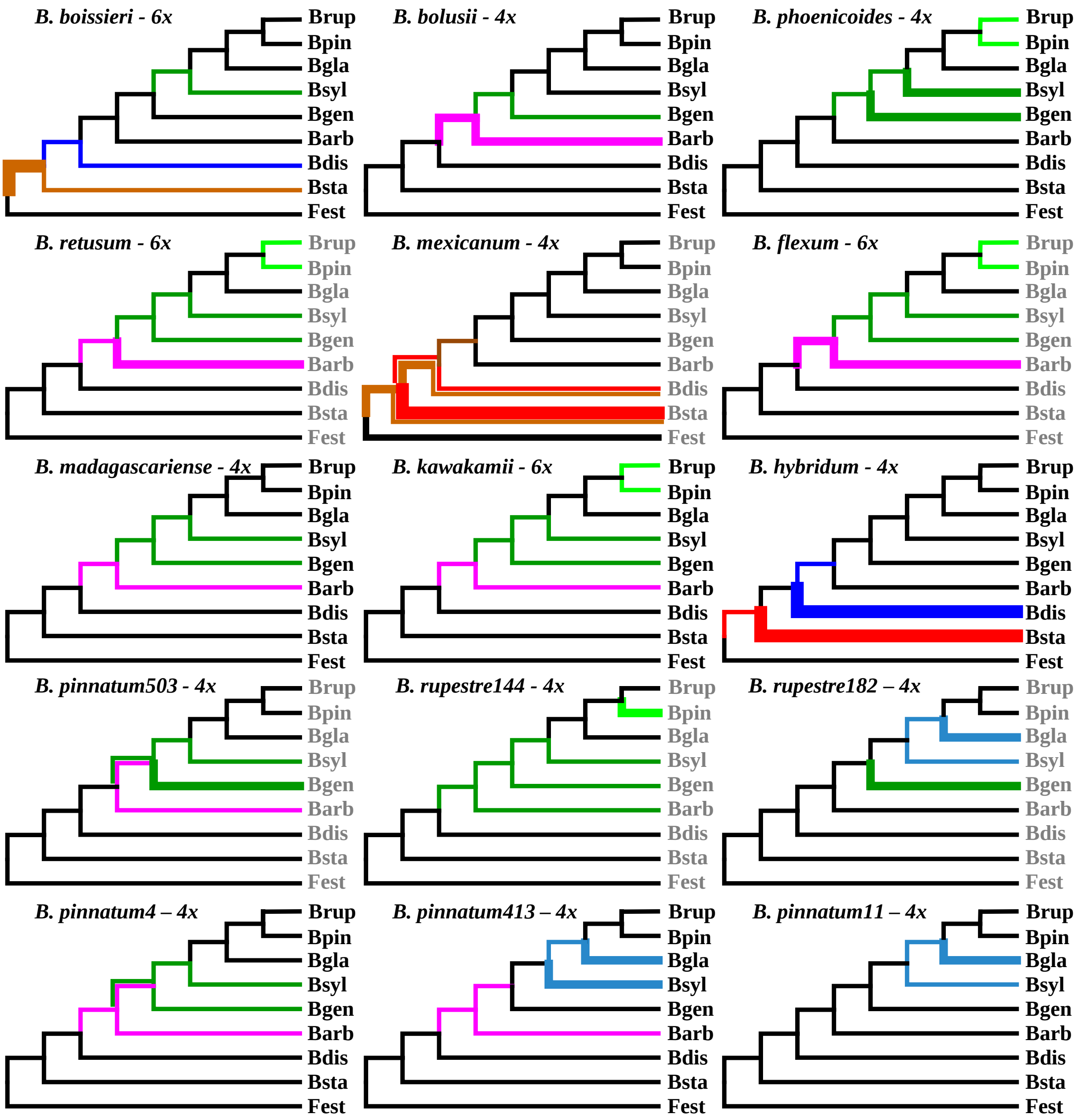


Figure 4. Minimum Evolution grafting of single-locus polyploid alleles into the *BEAST diploid species tree. The polyploid alleles of each species are grafted (in color) along the branches, according to the bootstrap pseudoreplications. The thick, medium, and thin lines represent allele placement with >75, 51-75, and <51 bootstrap support, respectively The different colors differentiate the groups of alleles associated with several homoeologous genomes (dark green, SYLVATICUM; light green, PINNATUM; purple, ARBUSCULA; dark blue, DISTACHYON; red, STACEI; brown, ANCESTRAL; and, light blue, GLAUCOVIRENS). The polyploid alleles grafted to the same branches are considered copies of the same homeologous genome. Festuca pratensis (Poeae) was used to root the tree. The color codes for the *Brachypodium* species are indicated in the chart.

The grafting of *Brachypodium* polyploid alleles, —inferred from the minimum evolution approach along the branches of the species tree—, suggested there were six homeologous genomes that could have participated in allopolyploidization events within *Brachypodium*, spanning several levels of phylogenetic depth (Figs. 4 and 5).

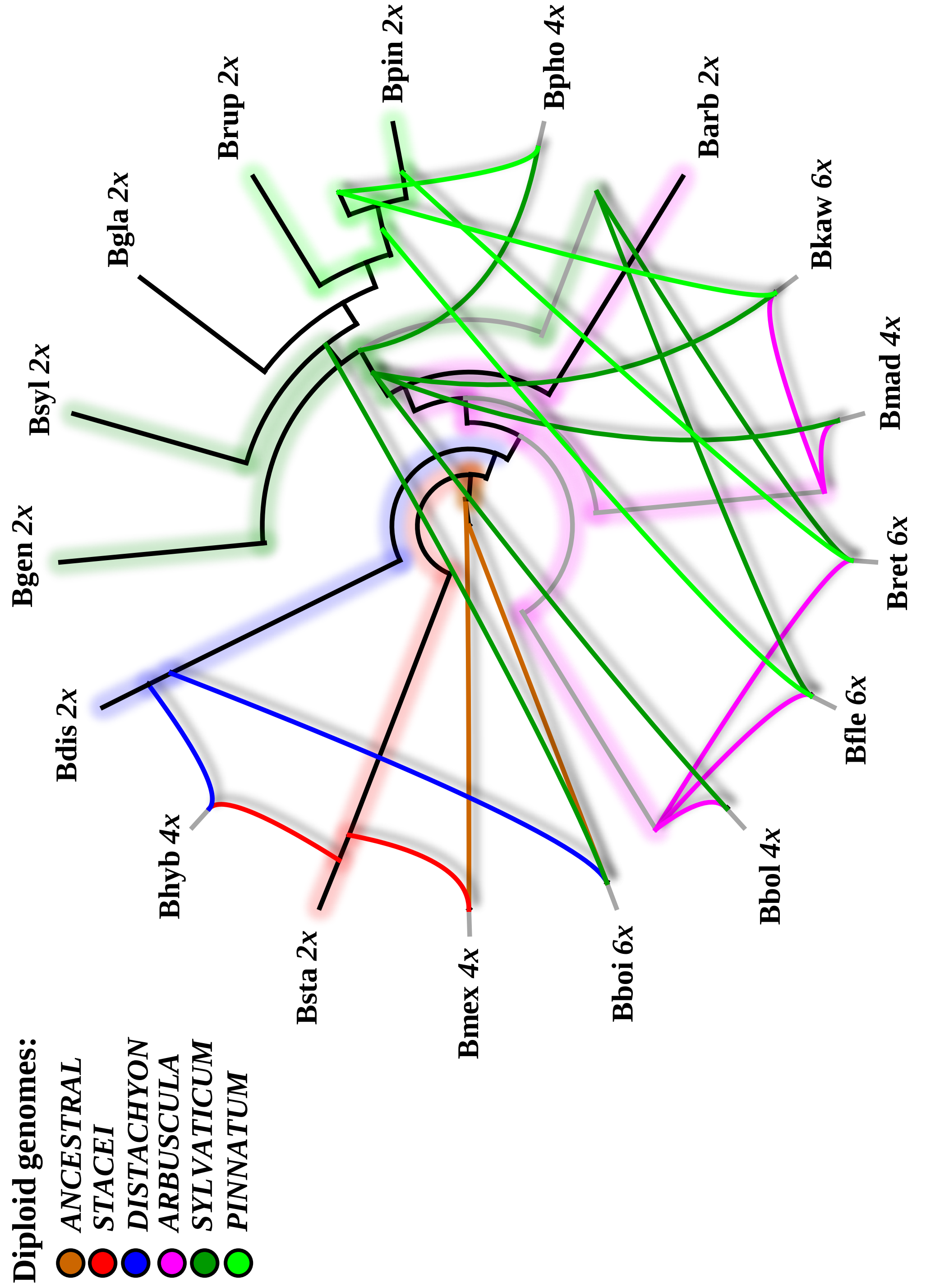


Figure 5. HOLM species network. The putative homeologous genomes are represented by colored lines diverging from specific branches. The diploid species lineages and branches generated by the HOLM algorithm that are associated with the same homoeologous genome have the same background color.

We have named core genomes all recently evolved genomes falling within the core perennial clade, and out-core genomes those showing more ancestral divergences. We also traced the sources of one of the most ancestral out-core type genomes (*ANCESTRAL*), two more recently diverged out-core diploid genomes [*STACEI* (stacei-like)] and *DISTACHYON* (distachyon-like)], one ancestral core-type genome (*ARBUSCULA*), and two recently diverged core-type diploid genomes [*SYLVATICUM* (sylvaticum-genuense-like) and *PINNATUM* (pinnatum-rupestre-like)] (Figs. 4 and 5). Both *SYLVATICUM* and *PINNATUM* were represented by polyploid alleles grafted to *B. sylvaticum* + *B. genuense* and *B. pinnatum* + *B.rupestre* terminal branches, respectively. However, we considered each of them as constituting a single genome, because they were grafted to both branches with similar though moderate-to-low bootstrap support. In addition, the *GLAUCOVIRENS* genome was represented by alleles grafted to *B. glaucovirens* + *B. sylvaticum* branches; although, in this case, strong bootstrap support was also observed for alleles grafted to the *B. glaucovirens* terminal branch.

The Minimum Evolution reconstruction placed the alleles of *B. mexicanum* in the out-core *ANCESTRAL* and *STACEI* genomes (Fig. 4). The *B. hybridum* alleles were strongly associated with two out-core terminal branches, suggesting parental *B. stacei*-like (*STACEI*) and *B. distachyon*-like (*DISTACHYON*) ancestors. The perennial species *B. boissieri* had alleles strongly related to out-core *ANCESTRAL* and *STACEI* genomes and to the recent core genome *SYLVATICUM*. Grafting allelic copies of the remaining polyploid or unknown-ploidy *Brachypodium* species was restricted to the recent stem branch and internal branches of the core perennial clade. The *ARBUSCULA, SYLVATICUM* and *PINNATUM* genomes were potentially involved in the origins of seven allopolyploid core perennial species: *B. phoenicoides*, *B. kawakamii*, *B. madagascariense*, *B. retusum*, *B. flexum* and *B. bolusii* (Figs. 4, 5; Methods/Results S1). With respect to six allotetraploid *B. pinnatum* and *B. rupestre* cytotypes (*B. pinnatum* 4, 11, 413 and 503, and *B. rupestre* 144 and 182), we observed the overall participation of the *SYLVATICUM* and *ARBUSCULA* genomes in most of them, plus two additional sources of genome ancestry associated to *GLAUCOVIRENS* in *B. pinnatum* 11 and 413 and *SYLVATICUM* in *B. pinnatum* 503 and *B. rupestre* 182, (Fig. 4). In contrast, the PINNATUM genome was found only in *B. rupestre* 144 (Figs. 4).

Divergence times and biogeography of the *Brachypodium* lineages

The consensus maximum clade credibility chronogram indicated that the *Brachypodium* lineage branched off from its stem node (S) in the Late Eocene (38.8 Ma) and the split of the crown node (CR) occurred in the Mid-Miocene (12.6 Ma) (Fig. 6). Our analyses also showed successive Late-Miocene and Early-Pliocene divergences for the basalmost currently extant *Brachypodium* genome lineages (*B. stacei*, 6.8 Ma; *B. distachyon*, 5.1 Ma). This was followed by a rapid radiation of the core perennial genome lineages from the end of the Pliocene (2.4 Ma) through the Quaternary, showing the sequential divergence of *B. arbuscula* (1.5 Ma), *B. genuense* (0.7 Ma), *B. sylvaticum* (0.6 Ma), *B. glaucovirens* (0.5 Ma), and *B. rupestre*/*B. pinnatum* lineages (0.3 Ma).

According to the coalescence-based Isolation Migration model, the American *B. mexicanum* originated by the hybridization of two out-core genomes approximately 3.3 Ma (Table 1) and the Mediterranean *B. hybridum* originated from the out-core *STACEI* and *DISTACHYON* genomes in the Quaternary (0.04 Ma; Table 1). The Mediterranean *B. retusum* and *B. boissieri*, the African *B. flexum* and the eastern-Asian *B. kawakamii* species were inferred to have resulted from the merging of three distinct genomes between 0.03 and 0.07 Ma. The allopolyploids include i) the out-core *ANCESTRAL* and *DISTACHYON* genomes in *B. boissieri*; ii) the ancestral core-type genome *ARBUSCULA* in *B. flexum*, *B. kawakamii* and *B. retusum*; and, iii) the recently evolved core-type genomes *SYLVATICUM* and *PINNATUM* in all of these species (except *PINNATUM* in *B. boissieri*) (Table 1). The mid- to late-Quaternary parental *ARBUSCULA* genome of African *B. bolusii*/*B. flexum* (0.03/0.61 Ma) and Madagascar-Eastern Asian *B. madagascariense*/*B. kawakamii* (0.39/0.31 Ma) lineages merged with other genomes, resulting in the origin of the current polyploid taxa in the late Quaternary (Table 1). The sister eastern Asian *B. sylvaticum* EA/*B. sylvaticum* var*. pseudodistachyon* diverged from the Eurasian *B. sylvaticum* lineage in the late Quaternary (0.2 Ma) (Fig.6).

The stratified DEC model (M1) of *Brachypodium* showed a better fit for the data than the unconstrained (M0) model (-ln likelihood 196.7 *vs*. 206.3, respectively; Likelihood Ratio Test (LRT)=19.2, *p* =0.001), and we will refer to this model hereafter (Fig. 7).

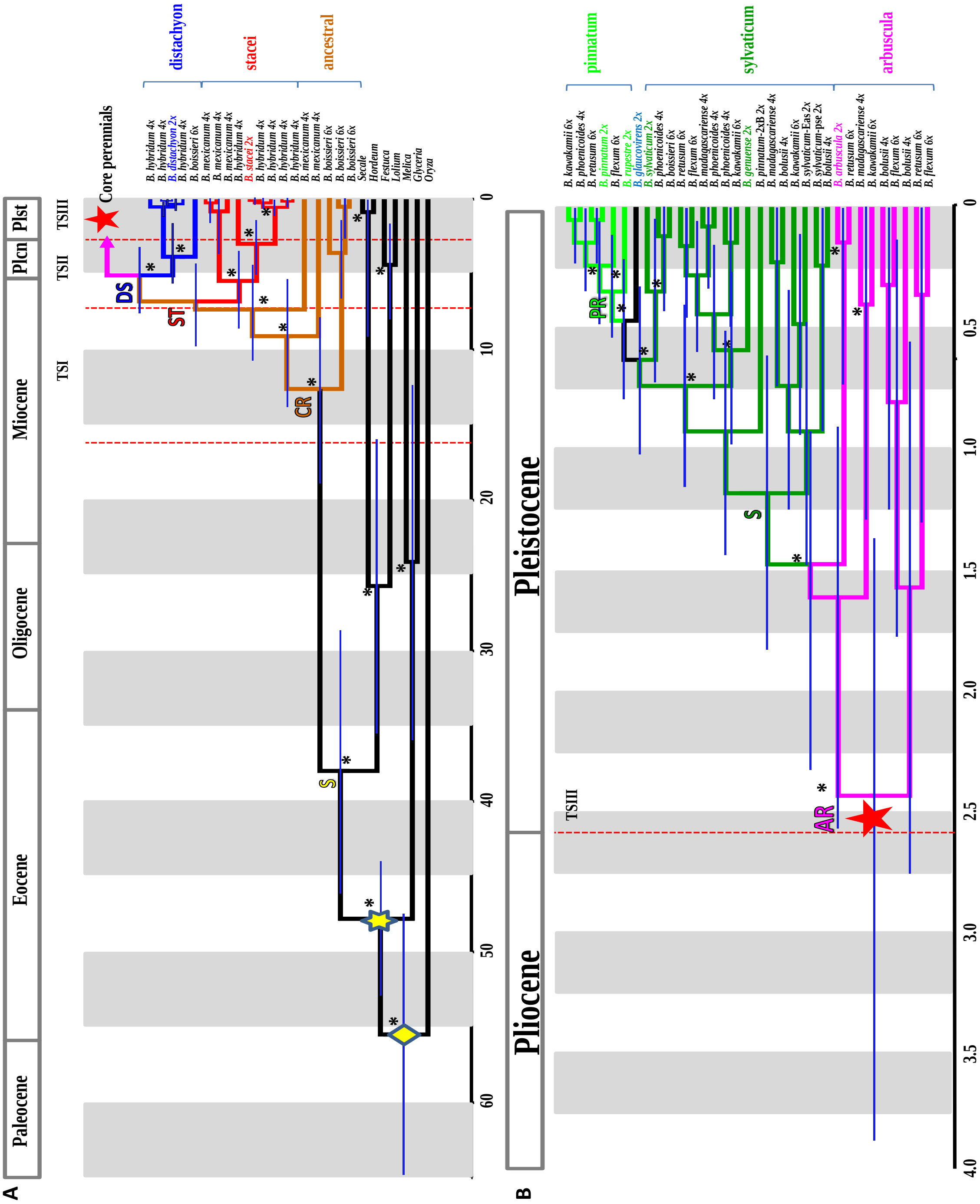


Figure 6. BEAST maximum clade credibility chronogram of *Brachypodium* and outgroup taxa based on analysis of the four studied loci. The clades are separated into **(A)**, the basalmost lineages and **(B)**, the most recently evolved core perennial clade The designations ST, DS, ARB, SG, PR correspond to nodes that define most copies associated to STACEI, DISTACHYON, ARBUSCULA, SYLVATICUM and PINNATUM genomes, respectively; and, CR (crown) represents the basalmost node of the ANCESTRAL genome. The Roman and Greek lowercase letters identify additional chronogram nodes. The right-most labels and color lines represent the allelic copies associated with homeologous genomes, following the Minimum Evolution principle. The splitting times were inferred for all genomic lineages diverging from the same species tree branch. The blue bars indicate 95% highest posterior density (HPD) intervals of nodal ages. The asterisks represent nodes with BS >80%. The diamond and star symbols indicate secondary and fossil-based calibrations imposed to the BOP and Pooideae nodal ancestors, respectively (see text). The vertical red lines are used to separate the three time slices (TSI-TSIII) used in the LAGRANGE analysis (see Fig. 7). The time scale bars below each panel represent million years ago (Ma).

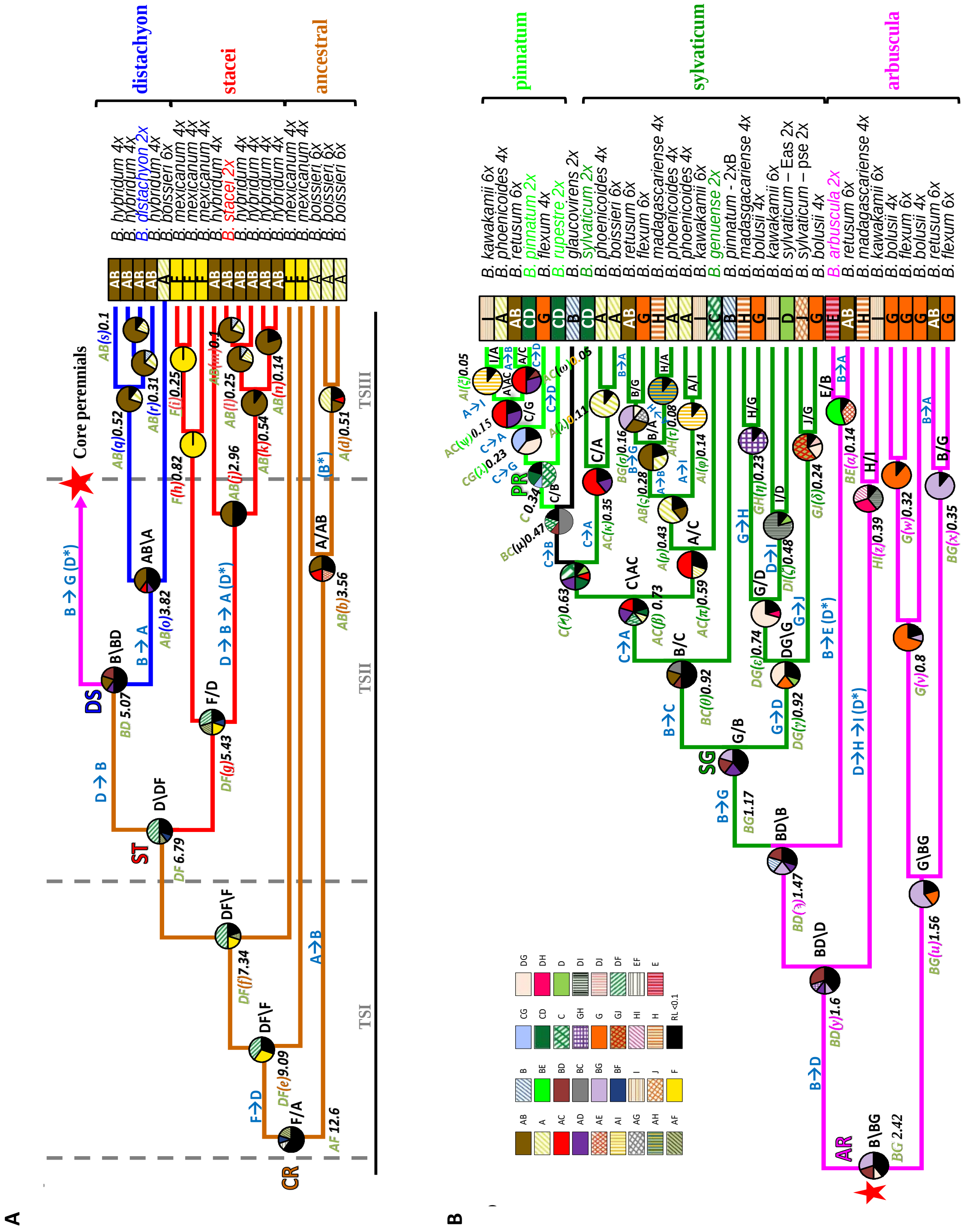


Figure 7. The estimated ancestral ranges and biogeographical events of the *Brachypodium* genomes, as inferred from LAGRANGE under the stratified M1 DEC model mapped on the BEAST maximum clade credibility tree with outgroups pruned from it. The panels represent **(A)** the basalmost lineages and **(B)** the recently evolved core perennial clade. The pie charts and numbers at the nodes indicate the relative probabilities for alternative ancestral ranges (with their color legends indicated at the inset chart), and the estimated median ages, respectively. The nodal codes (within the brackets) correspond to those indicated in Fig. 6. The vertical red lines are used to separate the three time slices (TSI-TSIII) used in the Lagrange analysis. The Operational Areas assigned to species' genomes are indicated to the right of the tree.

The global estimated dispersal rate (*dis*: 0.8314) was five times higher than the estimated extinction rate (*ext*: 0.1632) for the M1 model. The estimation of the geographic origin of the ancestral Mid-Miocene MRCA of *Brachypodium* showed considerable uncertainty (CR). The western Mediterranean and American ranges (AF) were inferred as the most likely area for it, followed by vicariance and the spread of the American genomic lineage to eastern Eurasia (DF) in the Mid-Miocene (N_e, N_f) (Figs. 7, 8). Different Mid- to Late-Miocene biogeographical events, involving the Palaeartic and Nearctic regions, were inferred to explain the ancestral distributions of the earliest diverging genome lineages of *Brachypodium* (the ancestral Mediterranean genome, *B. stacei*, *B. mexicanum*, *B. distachyon*) (nodes N_a, N_e, N_f, N_{ST}, N_g, N_{DS}; Fig. 7). The origin of the ancestor of the core perennial clade was estimated to have occurred between the Late Miocene in the eastern Eurasia-eastern Mediterranean region (N_{DS}, BD, 5.1 Ma) and the Pliocene in the eastern Mediterranean-Africa region (N_{AR}, BG, 2.42 Ma) (Figs. 7, 8). Several Quaternary Long Distance Dispersal (LDD) events had to be invoked to explain the successive colonizations of eastern Mediterranean-eastern Eurasian perennial ancestral genomes to Africa (N_{AR}, BG, 2.42 Ma), Macaronesia ($N_{ϡ}$-N_α, BD-BE, 1.47-0.14Ma), Madagascar (N_ε-N_η, DG-GH, 0.74-0.23Ma), East Asia (Nζ, DI, 0.5Ma), and Malesia (N_δ, GI, 0.24Ma), plus the parallel expansions to the western Eurasian-western Mediterranean ranges (Figs. 7, 8). Successive Quaternary LDDs involved colonization from the eastern Mediterranean to western Eurasia (N_θ, BC, 0.92Ma), western Eurasia to the western Mediterranean (N_β, AC; 0.73 Ma) and from the western to eastern Mediterranean (N_ξ, AB; 0.28 Ma) areas (Figs. 7, 8).

The western and eastern Mediterranean ranges hosted the most complex hybridization and genome doubling processes, which generated the high ploidy level *Brachypodium* allopolyploids (*B. boissieri*, *B. retusum*) (Table S1). The genomes of several recent lineages from western Eurasia (*SYLVATICUM*, *PINNATUM*) have converged with the ancestral local core lineage (*ARBUSCULA*) in *B. retusum* or with local out-core western Mediterranean genomes (*DISTACHYON*+*ANCESTRAL*) in *B. boissieri* (Figs. 7, 8). Similar patterns of genomic colonization, but involving long distance transoceanic dispersal, mostly from eastern to western Mediterranean regions (N_{AR}, N_{SG}, N_ρ), but also from eastern Eurasia (N_z, N_μ) to Africa and Madagascar, could have contributed to the presumed allopolyploids *B. bolusii*, *B. flexum* and *B. madagascariense*. In Taiwan, the putative allopolyploid *B. kawakamii* likely resulted from the merging of colonizing genomes from eastern Eurasia (N_y, Nζ) and the western Mediterranean region (N_φ, N_ς) (Figs. 7, 8).

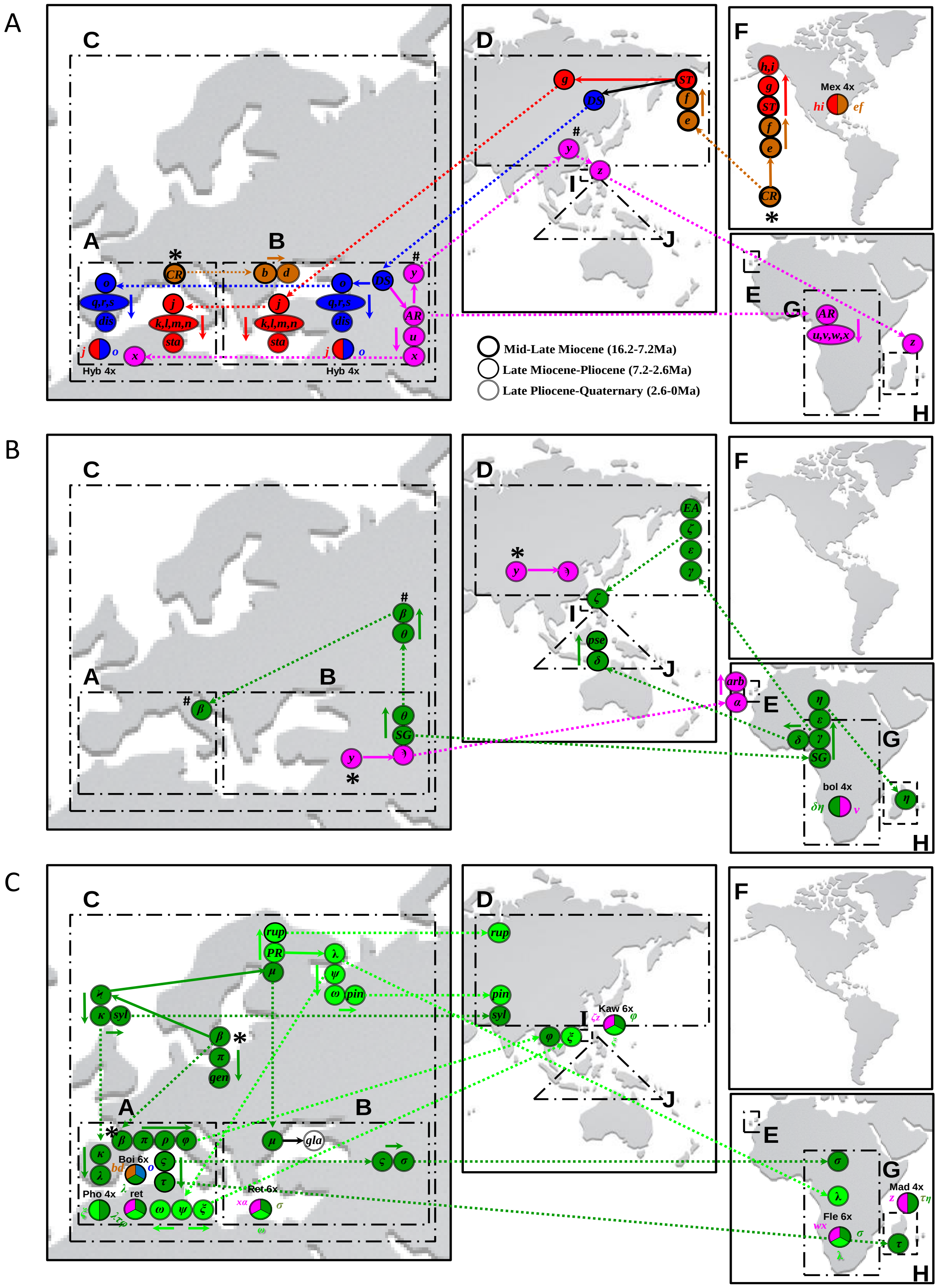


Figure 8. A map of the continents showing the ancestral areas and the dispersal and merging events of *Brachypodium* genomes, inferred under the optimal stratified M1 DEC Model (Fig. 7). Subfigures **A**, **B** and **C** show the nodes related to different sections of the BEAST maximum clade credibility tree (Fig. 7). The dashed arrows represent main dispersals between areas and the solid arrows represent the evolution of genomic lineages within the same area (phylogeny). The ancestral and recent genomes of the diploid skeleton tree and the Beast chronogram are depicted as circles that are color coded according to their respective main ancestral genome. The polyploid species are represented by circles with colored sections, representing homeologous genomes. The species abbreviations are: arb, *B. arbuscula*; boi, *B. boissieri*; bol, *B. bolusii*; dis, *B. distachyon*; EA, *B. sylvaticum* East Asia; fle, *B. flexum*; gen, *B. genuense*; gla, *B. glaucovirens*; hyb, *B. hybridum*; kaw, *B. kawakamii*; mad, *B. madagascariense*; mex, *B. mexicanum*; pho, *B. phoenicoides*; pin, *B. pinnatum*; pse, *B. sylvaticum* var. *pseudodistachyon*; ret, *B. retusum*; rup, *B. rupestre*; sta, *B. stacei*; and, syl, *B. sylvaticum*.

Discussion

A baseline phylogeny for *Brachypodium*: unravelling the evolutionary reticulate polyploid history of its model grass species

Reconstructing the evolutionary history of organismal groups where high level allopolyploids outnumber extant parental genomes is a major challenge in phylogenetic research (Brysting et al., 2007; Kamneva et al., 2017). Several studies, however, have applied alternative approaches to unravel the splits and mergings of the homeologous genomes that originated highly reticulate polyploid groups. These approaches include multilabeled genomes tree and species network dating analysis (e. g., *Cerastium,* (Brysting et al., 2007); *Viola*, (Marcussen et al., 2015)); Bayesian concordance, multilocus species tree and coalescence-based dating analysis (*Hordeum*, (Brassac & Blattner, 2015)); and multilabeled gene trees, network clustering and coalescence-based hybridization tests (*Fragaria*, (Kamneva et al., 2017)). These analyses have faced the difficulty of identifying potential “ghost genomes”—currently present only in the allopolyploids (Brysting et al., 2007; Marcussen et al., 2015)—and accounting for plausible gene copy losses and lineage sorting events (Brassac & Blattner, 2015; Kamneva et al., 2017) that could confound the recovery of all homeologous genomes.

Our study provides a comprehensive and updated phylogenetic reconstruction of the model genus *Brachypodium* with respect to previous work (Wolny & Hasterok, 2009; Catalán et al., 2012, 2016b), including the 18 currently recognized taxa that are distributed worldwide (Fig. 1, Figs. 3A-D). A statistical correction for the excess of allelic copies has allowed for the retrieval of diploid homeologous genomes participating in known allopolyploid species and cytotypes, congruent with their expected chromosome ploidy level (*B. hybridum* 4x*, B. mexicanum* 4x*, B. phoenicoides* 4x*, B. pinnatum* 4x, *B. retusum* 6x, and *B. rupestre* 4x) (Table S1, Figs. 4, 5). Our analysis retrieved only three homeologous genomes for the putative allo-octoploid *B. boissieri* (2n=42, 46; (Schippmann, 1991)). Because we did not include in the reconstruction some consensus types that were supported only by one clone, this led to the exclusion of one potential ancestral copy of *B. retusum*, which was preliminarily grafted to the ancestral branch of the species tree, suggesting an ancient genomic composition in the species similar to that of *B. boissieri*. We have provided further evidence for the potential allopolyploid nature of other karyologically unknown taxa (*B. bolusii*, *B. flexum, B. kawakamii*, *B. madagascariense*) (Fig. 4), though their ploidy levels have to be confirmed through cytogenetic data. Our Minimum Evolution analysis identified A*NCESTRAL*, a putative old ghost genome, in *B. mexicanum* and *B. boissieri* (Figs. 4). This lends support for a slightly earlier Miocene split of the crown *Brachypodium* ancestor (12.6 Ma), than was previously estimated from current extant taxa and whole plastome analyses of most ancestral annual *Brachypodium* lineages (10.1 Ma; (Sancho et al., 2018)). Evolutionary relationships have been corroborated for six poorly studied taxa (*B. bolusii*, *B. flexum*, *B. genuense*, *B. kawakamii*, *B. madagascariense*, *B. sylvaticum* var. *pseudodistachyon*), all falling within the core perennial clade (Figs. 3A-D, 4, 5). Approximately half of the species in the genus are diploids (8) and most of the remaining taxa (10) are likely allopolyploids (Figs. 3A-D, 4, 5), as determined for other model grasses, such as *Oryza* (Zhou et al., 2015).

Our Species Network reconstruction is in agreement with previous studies of the more ancestral divergences of the annual *B. stacei* and the short-rhizomatose *B. mexicanum*, and in the sister relationship of the annual *B. distachyon* and the core perennial clade (Figs. 3A-D, 4, 5). The derived allotetraploid origin of the annual *B. hybridum* from its diploid ancestors, *B. stacei* and *B. distachyon*, is supported by our loci and bootstrapping analyses (Fig. 4). This confirms that *B. hybridum* is, thus far, the only allopolyploid *Brachypodium* species with known extant diploid progenitors (Gordon et al., 2016). Our dated chronogram (Fig. 6) and IM analysis (Table 1) indicates that *B. mexicanum* could be considered a mesopolyploid, showing only ancestral out-core homeologous copies, and an estimated age of 3.37 Ma. By contrast, the core perennial allopolyploid species are neopolyploids, with estimated ages younger than 0.4 Ma. They either have homeologous copies from both ancestral out-core and recent core genomes (Table 1; Fig. 6), or only from recent core genomes, similar to the perennial relatives of rice and barley (Brassac & Blattner, 2015; Zhou et al., 2015). In general, the estimated coalescent times of origins of the core perennial *Brachypodium* allopolyploids were very recent (Table 1), although they overlap with the time divergence HPD intervals estimated for some species clades in other studies (e. g., *B. hybridum*; (Catalán et al., 2012)). The Species Network reconstruction shows two potential origins (*ANCESTRAL*, *STACEI*) for the alleles of *B. mexicanum* (Figs. 4, 5). This connection to the *STACEI* genome could explain the shared biological, morphological and genomic features of *B. mexicanum* and *B. stacei* (Catalán et al., 2016b).

The Minimum Evolution and coalescent analyses have clarified the genomic composition and recent origin of the perennial allopolyploid *B. boissieri* (*ANCESTRAL*, *DISTACHYON* and core *SYLVATICUM* genomes; 0.03 Ma), previously treated as an early split of the genus (Catalán et al., 2012), and of a similar age but different genome composition than the phenotypically close *B. retusum* (core *ARBUSCULA*, *SYLVATICUM* and *PINNATUM* genomes, 0.036 Ma) (Figs. 3D, 4, 5, Table 1). The genomic composition of *B. retusum* concurs with its allohexaploidy (Betekhtin et al., 2014; Catalán et al., 2016b). However, only three homeologous genomes have been detected in the purported allo-octoploid *B. boissieri*, suggesting a potential convergent evolution of some rDNA copies (Nieto-Feliner & Rosselló, 2007) or a loss of GI copies for the lost genome. The allotetraploid *B. phoenicoides* shows alleles associated with the recent core genomes *SYLVATICUM* and *PINNATUM* (Figs. 4, 5) and the tetraploid cytotypes of *B. pinnatum* and *B. rupestre*, alleles associated to the core species *B. glaucovirens* (*GLAUCOVIRENS* genome), but also to *SYLVATICUM*, *PINNATUM* and *ARBUSCULA* (Fig. 4). It should be emphasized that, contrary to our expectations, the *PINNATUM* genome, present in the *B. pinnatum* and *B. rupestre* diploid cytotypes, was only involved in the origin of a single allotetraploid cytotype of this group, *B. rupestre*144 (Fig. 4).

Our study has revealed the evolutionary origins of *B. bolusii*, *B. flexum*, *B. kawakamii* and *B. madagascariense* (Figs. 4, 5). These lineages show homeologous *ARBUSCULA* allelic copies grafted to the core perennial clade, indicating a putative hybrid origin from recently divergent genomes. By contrast, some of the studied loci (ITS, ETS) have identified a Malagasy-East Asian lineage composed of *B. madagascariense*, *B. kawakamii*, *B. sylvaticum* var. *pseudodistachyon* and an infraspecific *B. sylvaticum* var. *sylvaticum* East Asian lineage (Figs. 3B-C). This suggests the easternmost populations of the widespread Palaearctic *B. sylvaticum,* selected as a model grass for perenniality (Gordon et al., 2016), could belong to a separate taxon. The species network analysis did not show any clear concurrence of sequential hybridizations in the origin of high allopolyploid species (Fig. 5). However, potential low allopolyploid progenitors were presumably formed, especially when their ancestral genomes co-occurred in the same geographic area (e. g., *B. boissieri*: *DISTACHYON* and *ANCESTRAL* co-occurring in the western Mediterranean; *B. retusum*: *ARBUSCULA* and *SYLVATICUM* co-occurring in the eastern Mediterranean + SW Asia; and *B. kawakamii*: *ARBUSCULA* and *SYLVATICUM* co-occurring in Taiwan), or when they had different geographical origins but all merged in the same ancestral range (e. g., *B. flexum*: *ARBUSCULA*, *SYLVATICUM* and PINNATUM) (Fig. 8, Table 1). Our results do not support the hypothesis of the potential participation of a *B. distachyon*-like parent with x=5 chromosomes (and a perennial parent with x=9) in the origin of the 2n=28 allotetraploids *B. pinnatum* 4x, *B. rupestre* 4x and *B. phoenicoides* (Wolny & Hasterok, 2009; Betekhtin et al., 2014). However, the inferred participation of only core perennial genomes in these allotetraploids (Fig. 4) disagrees with the chromosome base numbers of x=9, 8 found among their closest current diploid species (Table S1). Plausible hypotheses for their in-core origins suggest the participation of two distinct genomes with x=9 or x=8, and their consequent chromosome fusions/losses after the genome doubling.

Historical biogeography of the *Brachypodium* genomes and taxa: a spatio-temporal scenario for successive splittings and mergings

Biogeographical reconstructions of large allopolyploid plant groups have been mostly drawn from matrilineal plastid DNA trees (e. g., *Primula*, (Guggisberg et al., 2006); *Rosa*, (Fougére-Danezan et al., 2015)) or from combined trees of reciprocally congruent nuclear and plastid gene topologies (e. g*., Cardamine*, (Carlsen et al., 2009); Loliinae, (Inda et al., 2014); Danthonioideae, (Linder & Barker, 2014)) where allopolyploids were represented by a single sequence per genotype. However, these simplistic historical reconstructions are prone to errors if the plastid or the nuclear genome donors had ancestral areas different from those of the current allopolyploids. Other studies have inferred the ancestral ranges after excluding the conflicting hybrid polyploids (e. g., *Abies*, (Xiang et al., 2015); *Tolpis*, (Gruenstaeudl et al., 2017)), which impeded the recovery of the biogeographical history of their homeologous genomes.

Our study, using the species and cytotypes of the grass genus *Brachypodium* as models, represents the first attempt to reconstruct the biogeography of ancestral genomes inherited by current diploid and allopolyploid taxa. The proposed biogeographical scenarios for the *Brachypodium* genomes and taxa fit the conceptual requirements for appropriate ancestral range reconstruction, and show i) that the splits of the allopolyploids' homeologous (sub)genomes from those of their diploid counterparts occurred in the same ancestral areas, although they could have dispersed independently (Fig. 7), and ii) that following the genome mergings, the homeologous genomes participating in the new allopolyploids had the same biogeographical patterns (Figs. 7, 8). The inferred existence of parallel evolution of homeologous genomes within the allopolyploid *Brachypodium* species might have artificially increased the global rate of dispersion estimated by LAGRANGE (*dis*: 0.8314). This is predicated on our approach that considered a dispersal event of an allopolyploid as two or three independent events, each related to a single subgenome. We contend that this was not important in *Brachypodium* because all homeologous genomes of *B. hybridum*, *B. boissieri*, *B. bolusii*, *B. retusum*, *B. mexicanum* and *B. phoenicoides* originated in the same geographic location (Table 1, Fig. 8), thus precluding these species acting as genetic sources for additional dispersions. For the remaining allopolyploids (*B. madagascariense*, *B. flexum* and *B. kawakamii*), some dispersion events were observed (Table 1, Fig. 8), but they were limited to a single genome at a time.

Our DEC M1 model has provided a biogeographical scenario for the *Brachypodium* genomes and taxa that supports the origin of their MRCA in the Holarctic region, followed by successive dispersals to Northern and Southern hemisphere ranges from the Miocene to the present (Figs. 7, 8). This parallels similar cases with other temperate grasses and angiosperms (e. g., Cardueae, (Barres et al., 2013); *Hordeum*, (Blattner, 2006); Loliinae, (Minaya et al., 2017)). Of 32 total inferred dispersals, 25 occurred in the Quaternary (TSIII), 5 in the the Pliocene (TSII) and two in the Miocene (TSI), (Fig. 7). This indicates that most *Brachypodium* genomes and species, especially those of the core perennial clade, emerged very recently. The western Mediterranean and American ranges were reconstructed as the ancestral areas with the highest marginal probabilities for the MRCA of *Brachypodium* (CR, 12.6 Ma). In the Mid-Miocene the areas were probably connected through Asia and the Bering Land Bridge, favoring the migrations of these and other xerophytic ancestors (Sanmartin et al., 2001). A Mid-Miocene vicariance (CR; A/F), coincident with a major temperature drop in the global climate (Meijer & Krijgsman, 2005), would explain the distribution of an isolated W Mediterranean genome (N_a), later inherited by the local polyploid *B. boissieri* and by the American *B. mexicanum* (Figs. 7, 8). Several connections between America and Asia through Beringia enabled genomic exchanges between the two areas (e. g. *Rosa*, (Fougére-Danezan et al., 2015)). A Mid-Late Miocene range expansion from America to Asia (N_e, 9.1 Ma; DF), followed by peripheral isolations, probably originated the *ANCESTRAL* genome of *B. mexicanum*, whereas a Late Miocene American/Asian vicariance (N_g, 5.4 Ma; F/D), followed by dispersal of the Old World lineage to the Mediterranean region in the Pliocene (N_j, 3.0 Ma; AB), likely separated the *STACEI* genome of *B. mexicanum* from that of *B. stacei* (Figs. 7, 8, Table 1).

Mediterranean migrations could have been facilitated by the closure of Mediterranean-southwestern Asian land bridges as a consequence of the Messinian salinity crisis (Krijgsman, 2002; Meulenkamp & Sissingh, 2003). Two concomitant independent Late Miocene-Pliocene LLDs from eastern to western Mediterranean ranges would explain the respective widespread AB distributions of xeromorphic *B. stacei* (N_g-N_j, 5.4-3.0 Ma) and meso-xeromorphic *B. distachyon* plus *DISTACHYON*-like genomes (II-N_o, 5.1-3.8 Ma), whereas western Mediterranean Pliocene and Quaternary peripheral isolations within the *DISTACHYON* lineage probably originated a distachyon-like genome, also inherited by the local *B. boissieri* polyploid (Figs. 7, 8, Table 1). Our data strongly support the merging of the *STACEI* (x=10) and *DISTACHYON* (x=5) diploid genomes in the derived allotetraploid (heteroploid) annual *B. hybridum* in the Mediterranean region during the Quaternary (ca. 0.05 Ma) (Figs.4-8, Table 1). This corroborates the potential existence of multiple hybridization scenarios in the region at different Pleistocene and Holocene times (Catalán et al., 2012) that could have facilitated the recurrent origin of the species (Lopez-Alvarez et al., 2015).

Multiple colonizations of Eurasia and other continents by ancestral perennial *Brachypodium* genomes (x=9, 8) were inferred to have occurred profusely in the Pliocene-Pleistocene (Fig. 7). These genomes merged with more ancestral annual-type genomes (x=10, 5), giving rise to a dysploid series of strongly-rhizomatose core perennial allopolyploid taxa (Fig. 8) (Betekhtin et al., 2014; Catalán et al., 2016b). In addition, a Late Miocene-Pliocene range expansion from the eastern Mediterranean region to Africa would explain the widespread distribution of ancestral genomes of the core perennial clade (N_{DS}-N_{AR}, 5.1-2.4Ma; BG). This migration likely occurred through the southwest Asian and Arabian platform corridor, a main migratory pathway of temperate Holarctic elements into East Africa and South Africa (Gehrke & Linder, 2009). Subsequent peripheral isolations and colonization of Asia, Madagascar and

Taiwan, concomitantly with the Quaternary climatic oscillations (Hewitt, 2000) and the recent uplifts of the high African and Central and East Asian mountains, were inferred to explain the origins of the oldest core-type *ARBUSCULA* genome. This genome was inherited from a putative polyploid African (*B. bolusii*, *B. flexum*), Malagasy (*B. madagascariense*) and Taiwanese (*B. kawakamii*) species (Figs. 7, 8). A Mid-Quaternary LDD of a perennial genome from the eastern Mediterranean region to Macaronesia (Canary Islands), followed by vicariance ($N_{ϡ}$-N_{α}, 1.47-0.14Ma), would explain the origin of the Canarian endemic *B. arbuscula*, following the emergence of these volcanic islands. New range expansions from the E Mediterranean region to Africa, and separate migrations from Africa to Asia (N_{SG}-N_{γ}, 1.17-0.92 Ma; DG) and from the Mediterranean region to Europe (N_{SG}-N_{θ}, 1.17-0.92 Ma; BC), were inferred to have caused the disjunct distributions of the ancestral genomes of the East and West Palaearctic perennial lineages (Figs. 7, 8). In the East, Late Quaternary LDDs of genomes from Africa to Madagascar (N_{ε}-N_{η}, 0.74-0.23 Ma), and from Asia to Taiwan (N_{ε}-N_{ζ}, 0.74-0.48Ma), over the respective straits, would explain the origins of newly recruited genomes, inherited by the local polyploids. The diploid *B. sylvaticum* var. *pseudodistachyon* could have originated following transoceanic colonization of an African genome in Malesia (N_{γ}-N_{δ}, 0.92-0.21 Ma), possibly facilitated by the mountain chains in New Guinea (Heads, 2006) (Figs. 7, 8). In the West, Upper Pleistocene range expansion from Europe to the Mediterranean region (N_{θ}-N_{β}, 0.92-0.73 Ma AC), and their respective Ionian-Holocene dispersals to Asia, were inferred to have been the origin of the most recent genomes of Mediterranean diploids *B. genuense* and *B. glaucovirens* and local polyploids, and of Eurosiberian *B. sylvaticum*, *B. rupestre* and *B. pinnatum* diploids. Some of the recent *SYLVATICUM* and *PINNATUM* genomes were also inferred to have migrated to Africa, Madagascar and Taiwan, contributing to the genomic dosage of the local polyploids (Figs. 7, 8). The current widespread Palearctic distribution of *B. sylvaticum* and *B. pinnatum* (Figs. 1, 8) probably resulted from recent Holocene postglacial colonizations from different Eurasian refugia, as indicated for other temperate grass lineages (Inda et al., 2014).

Chapter 2. Reference-genome syntenic mapping and multigene-based phylogenomics reveal the ancestry of homeologous subgenomes in grass *Brachypodium* allopolyploids

Summary

Phylogenomic analyses of a 505,512 RNA-seq SNP data set, mapped against the syntenic *B. distachyon*, *B. stacei* and *B. sylvaticum* reference genomes, and of 397 orthologous genes obtained from 12 *Brachypodium* taxa and ecotypes allowed us to reconstruct and date the splits of the dysploid *Brachypodium* diploid backbone species tree and of its allopolyploid sublineages. The transcriptome phylogenetic framework together with genome size (GS) data elucidated complex hybridization scenarios for the homeologous subgenomes participating in six *Brachypodium* allopolyploid species. Interspecific hybridization followed by whole genome duplication (IH+WGD) was the predominant scenario inferred for most the genome mergings, as illustrated by the recent allotetraploid *B. hybridum* (Quaternary), derived from *B. stacei*- and *B. distachyon*-type parents. Allotetraploid *B. mexicanum* emerges as the oldest polyploid species, having ancestral (A) and stacei-like (B) subgenomes (Mid-late Miocene), and the largest GS reported in the genus. The high polyploids *B. boissieri* (potential allo-octoploid) and *B. retusum* (potential allo-hexaploid) accumulate three (A, B, and intermediately evolved distachyon-type C, Miocene-Pliocene) and four (A, B, C, and recently evolved core perennial-type D, Quaternary) subgenomes, respectively. Reconciliation of their chromosomes and the inferred subgenomes requires the assumption of past chromosome fusions or genomic losses. Core perennial allotetraploids *B. rupestre* and *B. phoenicoides* show recently evolved C and D subgenomes (Quaternary); plastome data indicates that diploid *B. pinnatum* and *B. sylvaticum* could be their respective maternal parents. Pan-transcriptome analysis detected 5,202 transcript clusters across the studied *Brachypodium* samples, with a number of exclusive genes annotated in annual, perennial and ancestral *Brachypodium* lineages.

Introduction

Despite recent debate about the evolutionary fate of allopolyploids, alternatively viewed as drivers of biodiversity (Otto & Whitton, 2000) or evolutionary dead-ends (Mayrose et al., 2011), cumulative evidence suggest that hybrid polyploids could be considered true evolutionary winners in the eukaryotic kingdom (Otto, 2007; Van de Peer et al., 2009; Soltis et al., 2014b, 2016). In many groups of plants recurrent polyploidization events have led to allopolyploid species with highly dynamic genomes showing higher genetic variation than their diploid progenitors (Madlung, 2013; Soltis et al., 2014a). This is especially remarkable in seed and angiosperm plants, which are all considered descendants of paleopolyploid ancestors (Jiao et al., 2011, 2012). Allopolyploids are predominant in the grass family, accounting for 70% of the current species (Stebbins, 1949; Kellogg, 2015a). Despite genome duplication is considered generally irreversible in the short term (Marcussen et al., 2015), evidence suggests that the protograss whole genome duplication (WGD) was likely followed by subsequent diploidizations (Murat et al. 2010). These involved profound distinct genomic rearrangements, such as nested chromosome fusions, chromosome inversions and inactivation of paleocentromeres, coupled with differential losses of duplicated heterologous copies in the subgenomes along the divergent lineages. In contrast, new allopolyploidization events apparently led to the rising of grass mesopolyploids, originated some million years ago, and of grass neopolyploids, considered to have arisen during or after the Quaternary glaciations (Stebbins, 1985; Marcussen et al., 2015). These resulted in their current overwhelming representation within the grasses. Whilst the genomic origins of the recent plant neopolyploids can be traced through comparative genomics (e. g. wheats; Marcussen et al. 2014), deciphering the genomic origins of recent allopolyploids has proven challenging when the contributing parental genomes are genomically similar (Brassac & Blattner, 2015; Kamneva et al., 2017).

Brachypodium is a small genus of subfamily Pooideae (Poaceae) that contains ~20 species (3 annuals, 17 perennials) distributed worldwide (Catalán et al., 2016b). Its flagship species *B. distachyon* was selected as model system for grasses and monocots (IBI, 2010; Mur et al., 2011). Moreover, the three annual species (*B. distachyon*, *B. stacei*, *B. hybridum*) have recently been proposed as a model group to investigate grass allopolyploidy (Catalán et al., 2014; Gordon et al., 2016), and *B. sylvaticum* as a model for perennial grasses (Gordon et al., 2016). The selection of *Brachypodium* as a model genus was due to the small genomic sizes and low fraction of repetitive DNA found in all its currently sequenced genomes (*B. distachyon* (IBI, 2010); *B. stacei*, *B. hybridum*, *B. sylvaticum* http://phytozome.jgi.doe.gov/) and its genomic and evolutionarily closeness to both temperate and tropical grasses (Sancho et al., 2018). Recent plastid and nuclear phylogenetic studies of the genus (Catalán et al., 2016b) indicated that approximately half of its species are diploids and the other half are likely allopolyploids (from allotetraploids to putative allo-octoploids), suggesting that allopolyploidy has been the prevalent speciation mechanism in a large portion of the genus. These analyses also detected the early divergence of annuals and short rhizomatose (*B. mexicanum*) lineages, and a recent split of mostly strong-rhizomatose core perennials (all remaining perennial *Brachypodium* species excluding *B. mexicanum*). In contrast to core pooid cereal and forage grasses, where interspecific hybridization involved homo- or heteroploid parents with the same chromosome base number (e. g., x=7, Triticeae, Bromeae, Poeae, Aveneae), most allopolyploid *Brachypodium* species likely resulted from crosses of dysploid homo or heteroploid parents, showing different chromosome base numbers (x=10, 9, 5) (Betekhtin et al., 2014). The best-known case is the annual allotetraploid *B. hybridum* (2n=30, x=10+5), derived from the cross and subsequent genome doubling of diploid *B. stacei*-type (2n=20, x=10) and *B. distachyon*-type (2n=10, x=5) parents. The hybrid originated from bidirectional crosses approximately 1 Ma and still maintains almost intact progenitor subgenomes (López-Alvarez et al., 2012; Lopez-Alvarez et al., 2015). The recent recreation of a stable synthetic allotetraploid that phenotypically resembles the wild *B. hybridum* corroborates the natural allopolyploid origin of this annual neopolyploid species (Dinh Thi et al., 2016). In contrast, the evolutionary history of the perennial allopolyploids remains unclear (Catalán et al., 2016b).

Deciphering the evolutionary history of perennial allopolyploid *Brachypodium* species is a crucial step for understanding the genomic composition and origins of these and other grass species [e. g., robust perennial allopolyploid biofuel (*Miscanthus*, 4x; *Paspalum*, 2x-8x; *Thinopyrum* 2x-10x) and forage (*Festuca* 4x-12x) grasses (Catalán et al., 2016b)]. In this study, we develop two approaches to shed light into the reticulate phylogeny of this model genus, focusing on its allopolyploid species. First, we take advantage of three available genomes representing different evolutionary depths of the *Brachypodium* tree [early diverging *B. stace*i, intermediately evolved *B. distachyon* and recently evolved core-perennial *B. sylvaticum* lineages (Catalán et al., 2016b)] to perform a synteny-based, read mapping approach for calling homeologous SNPs from RNA-seq reads, and building genomic phylogenetic trees that would identify the homeologous genomes of the allopolyploids. Second, we assembled core nuclear and plastid transcripts of both diploid and polyploid species to build gene trees targeting labeled homeologous genes and identifying the parental genome donors of the allopolyploids. To accomplish the first task, we developed a phylogenomic workflow and analyzed two independent datasets, i) transcripts obtained by assembling a large data set of RNA-seq reads, and ii) a complementary restricted data set of genomic sequences produced by Genotyping-By-Sequencing (GBS) that was used to validate the RNA-seq based findings. To attain the second goal, we filtered core transcript isoforms from a *Brachypodium* wide pan-transcriptome and used them to build the *Brachypodium* core-genes subgenome tree. We validated the computational pipelines, estimated genome size (GS) values and reconstructed a robust phylogeny for 12 *Brachypodium* species and ecotypes, six of them allopolyploids [including two ecotypes of tetraploid *B. phoenicoides* (Bpho6 and B422) in the RNA-seq based analyses (SNPs, core transcripts), and two cytotypes of *B. pinnatum* (2x and 4x) in the GBS based analysis]. Our approaches allowed us to propose plausible hypotheses about the identities of the parental genome donors and the times of origin of the lineages participating in the studied allopolyploid species.

Materials and Methods

Plant materials and gDNA and total-RNA extractions

Thirteen individual plants of twelve *Brachypodium* species and cytotypes [one accession per species for nine species, two accessions (ecotypes) of *B. phoenicoides* and two cytotypes of *B. pinnatum* (2x, 4x)] collected in their native circum-Mediterranean, Eurasian and North American (Mexico) regions were studied (table 1).

Genomic DNA (gDNA) was extracted from fresh leaves, which were grinded with liquid nitrogen, using Wizard Genomic DNA purification kit (Promega). Sample quality, concentration and integrity were checked with BioDrop µLITE, Qubit 2.0 fluorometer, Quanti-iT dsDNA HS Assay Kit (Invitrogen), and visual inspection on 1% agarose gel, respectively.

For transcriptomic (RNA-seq) analysis, each plant was divided into four tillers , re-established and allowed to tiller, and then each new plant received one of the following abiotic stress treatments: control (watered plant every 48 h, 25ºC), soil drying stress (no water for one week), hot stress (40ºC day/25ºC nigh for 24 h), salt stress (500 mM NaCl administered in water daily for two days).

Total RNA was isolated from 50 – 200 mg of leaf tissue using the E.Z.N.A Plant RNA kit (Omega) and the RNeasy Plant Mini kit (Qiagen). PVP 2.5% w/v was added to the extraction buffer and an on-column DNase treatment was carried out following manufacturer's protocols. RNA integrity was measured with a RNA nano-chip on the Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified by BioDrop µLITE. Pooled RNAs from the four treatments were used in subsequent library preps.

Genome size and ploidy level estimations

Leaves of adult plants growing in pots were used for genome size (GS) estimation through flow cytometry. Nuclear suspensions were prepared from 200 mg of leaf sample and 200 mg of leaf internal standard. The nuclear DNA content of *B. retusum* was calculated using nuclei isolated from young leaves of *Raphanus sativus* "Saxa" (1.11 pg/2C DNA; Dolezel et al. 1998) and those of *B. arbuscula*, *B. boissieri*, *B. mexicanum*, *B. phoenicoides*, *B. pinnatum* and *B. rupestre* using *Lycopersicon esculentum* 'Stupicke' (1.96 pg/2C DNA; Dolezel et al. 1992) as standards. Leaves in 500 µl Otto I reagent (Otto, 1992) were chopped by razor blade on a Petri dish. The suspension was filtered using 50 µm pore nylon filters and 1000 µl of Otto II reagent, nuclei were stained with propidium iodide, and RNAse was added. Samples were analyzed using a CyFlow Ploidy Analyser SYSMEX. At least 5,000 nuclei were analyzed per sample. Each sample (two replicates) was analyzed three times at different days. Only measurements with coefficient of variation < 3.5% were recorded.

Ploidy levels were inferred from chromosome counts (2n) performed in the same accessions used in our study or through contrasted GS and 2n values obtained in conspecific accessions that showed similar GS values (table 1; Inda, unpub. data).

Table 1. List of *Brachypodium* species and cytotypes and outgroup taxa used in the study. Information on locality of origin, chromosome number (2n), chromosome base number (x), ploidy level, genome size (GS), life-cycle, and data set type are provided for each accession. Genome size values estimated in this work are shown in bold; superscripts indicate genome size and chromosome number values of species obtained in previous studies [1. (Suda et al., 2005); 2. (Wolny & Hasterok, 2009); 3. (Catalán et al., 2012); 4. (Johnston et al., 1999); 5. (Uozu et al., 1997); 6. (Robertson, 1981); 7. (Schippmann, 1990); 8. (Schippmann, 1991) 9. (Shi, 1991); 10. (von Bothmer et al., 1995); 11. (Vaughan, 1994)].

Taxa	Accession codes	Abbreviations in Figures	GS (pg/2C)	2n	x	Ploidy	Life-cycle	Locality	Data set
B. arbuscula Gay ex Knoche	Barb502	Barb	**0.713±0.004**	18[6]	9	2x	perennial	Spain: Canary Isles: La Gomera: Vallehermoso	RNA-seq/GBS/GS
B. arbuscula Gay ex Knoche	Barb405	Barb405	**0.7203±0.006**	18[6]	9	2x	perennial	Spain: Canary Isles: Tenerife: Teno	GS
B. arbuscula Gay ex Knoche	-	-	0.69±.0.01[1]	18[6]	9	2x	perennial	Spain: Canary Isles: Tenerife: Teno, rock crevices in the Roque El Fraile	GS (Suda et al. 2005)
B. boissieri Nym.	Bbois3	Bboi	**3.236±0.072**	**ca. 46** 42[7,8] 46[7,8]	?	6x? 8x?	perennial	Spain: Granada: Puerto de la Mora	RNA-seq/GBS/GS
B. distachyon (L.) P. Beauv.	Bdis_Bd21	Bdis	-	10[3]	5	2x	annual	Iraq: Salah ad Din, 4 km from Salahuddin	GBS
B. distachyon (L.) P. Beauv.	Bd21 (SRX1721359)	Bdis	-	10[3]	5	2x	annual	Iraq: Salah ad Din, 4 km from Salahuddin	RNA-seq data (Bettgenhaeuser et al. 2017) The Sainsbury Laboratory
B. distachyon (L.) P. Beauv.	Brachypodium distachyon v3.1, line Bd21	-	-	10[3]	5	2x	annual	Iraq: Salah ad Din, 4 km from Salahuddin	Reference Genome Brachypodium distachyon v3.1; Vogel et al. 2010
B. distachyon (L.) P. Beauv.	ABR1[2]	Bdis_ABR1	0.631±0.015[2]	10[3]	5	2x	annual	Turkey	GS (Wolny and Hasterok 2009)
B. hybridum Catalán, Joch. Müll., Hasterok & Jenkins	Bhyb_ABR113	Bhyb_ABR113	1.265[3]	30[3]	5 + 10	4x	annual	Portugal: Lisbon	GBS GS (Catalán et al. 2012)
B. hybridum Catalán, Joch. Müll., Hasterok & Jenkins	Bhyb_BdTR6g (W6 39378)	Bhyb_BdTR6g	-	30[3]	5 + 10	4x	annual	Turkey	RNA-seq
B. mexicanum (Roem. & Schult.) Link	Bmex347	Bmex	**3.774±0.033**	40[9]	?	4x?	short-perennial	Mexico: Hidalgo: Sierra de Pachuca	RNA-seq/GBS/GS
B. phoenicoides (L.) P. Beauv. ex Roem. & Schultes	Bpho_B422	Bpho B422	**1.469±0.012**	28[6]	?	4x	perennial	Slovakia: Ružomberok	RNA-seq/GBS/GS
B. phoenicoides (L.) P. Beauv. ex Roem. & Schultes	Bpho6	Bpho6	**1.443±0.019**	28[6]	?	4x	perennial	Spain: Huesca: Panzano	RNA-seq/GBS/GS
B. phoenicoides (L.) P. Beauv. ex Roem. & Schultes	PI 89817 PI 253503 PI 283196	-	1.476±0.049[2] 1.499±0.005[2] 1.497±0.009[2]	28	?	4x	perennial	Spain Spain Portugal	GS (Wolny and Hasterok 2009)

B. pinnatum (L.) P. Beauv. (diploid A)	Bpin505	Bpin-2x	**0.822±0.009**	18[6]	9	2x	perennial	Norway USDA PI 345982	RNA-seq/GBS/GS
B. pinnatum (L.) P. Beauv. (diploid A)	PI 230114 PI 345982	-	0.882±0.005[2] 0.881±0.027[2]	18	9	2x	perennial	Iran Norway	GS (Wolny and Hasterok 2009)
B. pinnatum (L.) P. Beauv. (tetraploid)	Bpin520	Bpin-4x	**1.499±0.014**	28[2,9]	?	4x	perennial	Netherlands: Scherpenzeel	GBS/GS
B. pinnatum (L.) P. Beauv. (tetraploid)	PI 4193 PI 249722 PI 251445 PI 430277	-	1.462±0.007[2] 1.547±0.064[2] 1.574±0.038[2] 1.532±0.037[2]	28	?	4x	perennial	Germany Greece Turkey Ireland	GS (Wolny and Hasterok 2009)
B. retusum (Pers.) P. Beauv.	Bret400	Bret	**1.704±0.024**	**ca. 32**	?	4x? 6x?	perennial	Spain: Huesca: Angües	RNA-seq/GBS/GS
B. retusum (Pers.) P. Beauv.	PI4195	Bret(PI4195)	2.570±0.036[2]	38[2]	?	6x	perennial	Greece	GS (Wolny and Hasterok 2009)
B. retusum (Pers.) P. Beauv.				27[8] 28[8] 31[8] 32[7,8] 36[6] 38[2] 40[7] 42[8]					
B. rupestre (Host) Roem. & Schult. (tetraploid)	Brup5	Brup	**1.469±0.037**	28	?	4x	perennial	Spain: Huesca: Jaca: Aratorés: Castiello de Jaca	RNA-seq/GBS/GS
B. rupestre (Host) Roem. & Schult. (diploid)	PI440172	Brup435	**0.820±0.005**	18[2]	9	2x	perennial	Greece	GS
B. stacei Catalán, Joch. Müll., Mur & Langdon	Bsta_ABR114	Bsta_ABR114	0.564[3]	20[3]	10	2x	annual	Spain: Balearic Isles: Formentera	GBS GS (Catalán et al. 2012)
B. stacei Catalán, Joch. Müll., Mur & Langdon	Bsta_TE4.3 (INIA-CRF: NC050363)	Bsta_TE4.3	-	20[3]	10	2x	annual	Spain: Canary Isles: La Gomera: Agulo	RNA-seq
B. stacei Catalán, Joch. Müll., Mur & Langdon	Brachypodium stacei v1.1, line ABR114	-	-	20[3]	10	2x	annual	Spain: Balearic Isles: Formentera	Reference Genome (Brachypodium stacei v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/)
B. sylvaticum (Huds.) P. Beauv.	Brachypodium sylvaticum v1.1 line Ain1	-	-	18[6]	9	2x	perennial	Tunisia	Reference Genome (Brachypodium sylvaticum v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/)
B. sylvaticum (Huds.) P. Beauv.	Bsyl_Sin1	Bsyl_Sin1	-	18	9	2x	perennial	Turkey	Genomic data (Vogel & Gordon, 2017 – unpublished data)
B. sylvaticum (Huds.) P. Beauv.	Brasy-Cor Population OR-C1	Bsyl_Cor	-	18	9	2x	perennial	USA: Oregon: Corvallis	RNA-seq data (Fox et al. 2013)
B. sylvaticum (Huds.) P. Beauv.	Brasy-Esp (F-88) (PI318962)	Bsyl_Esp	-	18	9	2x	perennial	Spain: Ávila	RNA-seq data (Fox et al. 2013)
B. sylvaticum (Huds.) P. Beauv.	Brasy-Gre (PI206546)	Bsyl_Gre	-	-	-	-	perennial	Greece: Thessaloniki	RNA-seq data (Fox et al. 2013)
B. sylvaticum (Huds.) P. Beauv.	PI237792 PI297868 PI318962 PI380758	Bsyl443 Bsyl444 Bsyl445 Bsyl446	0.0863±0.029[2] 0.873±0.008[2] 0.844±0.007[2] 0.898±0.0045[2]	18	9	2x	perennial	Spain Australia Spain: Ávila Iran: Ardebil	GS (Wolny and Hasterok 2009)

Hordeum vulgare subsp. vulgare cv. Morex	ERR247357	Hvul	11.13[4]	14[10]	7	2x	annual	-	Genomic data Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) (IPK-GATERSLEBEN)
Hordeum vulgare subsp. vulgare cv. Morex	ERR159679	Hvul					annual	-	Transcriptomic data JHI
Oryza sativa L. (Japonica group)	Nipponbare (SRX743701)	Osat	0.91[5]	24[11]	12	2x	annual	South Korea: Suwon	Genomic data SEOUL NATIONAL UNIVERSITY
Oryza sativa L. (Japonica group)	Nipponbare (SRX738077)	Osat					annual	missing	Transcriptomic data SUN YAT-SEN UNIVERSITY

Genotyping-by-sequencing (GBS) and RNA library preparation and transcript assembly

Samples were digested with PstI and the resulting fragments in the 100-400 bp size range were sequenced in a single lane on a Illumina HiSeq2000 platform, obtaining paired-end (PE) reads of 150 bp. Adapter-trimmed and demultiplexed PE reads were used in downstream analysis. Quality control of PE reads was done with FastQC 0.11.3 software (Andrews, 2010) (table S1).

cDNA library preparation of *Brachypodium* RNA samples was carried out using TruSeq Stranded mRNA Library Prep Kit (Illumina, Inc.), generating PE libraries with insert size of 300 to 600 bp. Sequencing was performed using a Illumina HiSeq2500 platform (125 bp PE sequencing). Quality control of PE reads was performed with FastQC software. Adapters and low quality reads were removed and filtered with Trimmomatic-0.32 (Bolger et al., 2014) (table S1). Transcript sequences were assembled with trinityrnaseq-r20140717 (Grabherr et al., 2011) using default parameters (table S2).

Pre-processing, concatenating and aligning reference genomes

The three *Brachypodium* reference genomes were downloaded from Phytozome (Goodstein et al., 2012). They corresponded to *Brachypodium distachyon* line Bd21 from Irak (*Brachypodium distachyon* v3.1; Vogel et al. 2010), *B. stacei* line ABR114 from Spain (*Brachypodium stacei* v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/), and *B. sylvaticum* line Ain-1 from Tunisia (*Brachypodium sylvaticum* v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/). In all three cases only complete chromosome arms were included in the analysis. The three reference genomes were concatenated into a single FASTA file (*B. distachyon* – *B. stacei* – *B. sylvaticum*) for mapping. After soft-masking genome repeats, whole-genome synteny-based alignments of the *B. stacei* and *B. sylvaticum* assemblies to *B. distachyon* were conducted with Cgaln v1.2.3 software (Nakato & Gotoh, 2010).

Read mapping, SNP calling and multiple alignments

Clean RNA-seq and GBS pair-end reads were mapped to the three concatenated reference genomes using bwa 0.7.12-r1039 (Li & Durbin, 2009) and hisat2-2.0.4 (Kim et al., 2015), respectively. Only reads with mapping score ≥ 30 were considered for downstream analyses, as recommended for polyploids (Clevenger et al., 2015).

We developed a pipeline (fig. 1) to filter, align and validate SNPs requiring a minimum read coverage of 10x. Constant sites were included in the RNA-seq data set to recover the maximum number of syntenic sites for downstream analyses. *vcf2alignment* takes as input a merged VCF file with mapped GBS or RNA-seq reads and outputs a multiple alignment with called SNPs in FASTA format, *mapcoords* extracts syntenic sites from whole-genome alignments of the three reference genomes, and *vcf2alignment_synteny* combines called SNPs and syntenic positions to make a multiple subgenome-based alignment where all sites correspond to common (syntenic) positions with respect to the master *B. distachyon* reference genome (fig. 1). Within this framework, SNPs called in a given species were split in up to three sequences in the resulting multiple alignment, each retrieved from a different reference genome. Subgenomes sequences with negligible mappings/SNPs were removed, and those from diploid species collapsed into a single sequence. The workflow is fully described in Supplementary Methods. The complete protocol is available at https://github.com/eead-csic-compbio/vcf2alignment.

Other independent approaches were tested for validation of our pipeline using previously published tools. In particular, GIbPSs v1.0.2 (Hapke and Thiele 2016) was used to analyze our GBS datasets, while NGSEP (Duitama et al., 2014; Perea et al., 2016) was tested with both GBS and RNA-seq data.

Clustering expressed genes and pan-transcriptome analyses

Transcripts assembled *de novo* from RNA-seq data, together with annotated transcripts or CDS from three accessions of *B. sylvaticum* obtained from Fox *et al*. (2013) and from three species *Brachypodium distachyon* (Bdistachyon_314_v3.1; http://phytozome.jgi.doe.gov/; (IBI, 2010)), *Oryza sativa* (Osativa_323_v7.0; http://phytozome.jgi.doe.gov/; (Ouyang et al., 2007)) and *Hordeum vulgare* (ftp://ftpmips.helmholtz-muenchen.de/plants/barley/genome_release2017/; (Mascher et al., 2017)) were clustered to define core and accessory transcripts with our software GET_HOMOLOGUES-EST (Contreras-Moreira et al., 2017). The OMCL algorithm was selected (-M), percent sequence identity threshold was calibrated to –S 80 to properly include sequences from the two outgroups (*Oryza sativa*, *Hordeum vulgare*), and an Average Nucleotide Identity matrix (-A) was produced. A pan-transcriptome matrix was generated and subsequently interrogated to identify core transcripts, expressed in all species, and also accessory sequences expressed only in some species (e. g., diploids), but not in others (e. g., polyploids). Clusters were functionally annotated with databases Pfam (Finn et al., 2016), RefSeq (Leary et al., 2016) and SwissProt (Boutet et al., 2016). Redundant and overlapping cluster sequences were collapsed with script *annotate_cluster.pl*, producing multi-copy FASTA files.

Enrichment analyses of Gene Ontology (GO) biological processes associated to sequence clusters within each target species group were carried out. Bdistachyon_314_v3.1 gene identifiers, either from sequences in the same clusters or matched by BLASTN (ncbi-blast-2.6.0+; (Camacho et al., 2009)) with at least 75% and 90% of coverage and identity, respectively, were used as input for PANTHER13.1 (Protein ANalysis THrough Evolutionary Relationships) Overrepresentation Test (http://pantherdb.org/). The *Brachypodium distachyon* GO was used as background for Fisher's Exact test with False Discovery Rate (FDR) multiple test correction (Thomas et al., 2003; Mi et al., 2013).

Phylogenomic analyses of RNA-seq, GBS and core transcripts data sets

Nucleotide alignments inferred from stacked RNA-seq and GBS SNPs, as well as from clusters of core transcripts, were analyzed to infer the phylogeny of *Brachypodium*, using *Oryza sativa* and *Hordeum vulgare* as outgroups. Constant sites were only included in the RNA-seq subgenomic and core gene based analyses attempting to recover large orthologous fragments and more syntenic sites.

Maximum Likelihood (ML) analyses of the concatenated RNA-seq and GBS data sets and the syntenically aligned RNA-seq data set were performed with IQ-TREE (Nguyen et al., 2014); the best-fit evolutionary model was automatically selected by ModelFinder (Kalyaanamoorthy et al., 2017) in terms of the Akaike Information Criterion corrected (AICc). Topological congruence among alternative tree pairs was tested through Likelihood Ratio Test (SH-aLRT), and branch support ultrafast bootstrap searches were performed with 1000 replicates (Minh et al., 2013; Chernomor et al., 2016). The resulting trees were rooted with *Oryza sativa*, except for the GIbPSs species tree, which was rooted with *B. stacei* ABR114. The phylogenomic tree of 397 core transcripts clusters (see results, Supplementary Methods) was conducted by "Partitioned analysis for multi-gene alignments" using the –spp option (Edge-proportional partition model with proportional branch lengths) of the IQ-TREE software.

Multi-labeled ML trees obtained from 397 multi-copy core clusters with IQ-tree were alternatively analyzed with the software GRAMPA (Thomas et al., 2017) aimed at confirming the ploidy level and nature of each species (diploid or polyploid, discerning between allopolyploid and autopolyploid), and inferring the plausible polyploidization events. Procedures for these analyses were done for one species at a time, fixing the particular node to search in each case as the polyploid clade (-h1). For each species, the parsimony scores of the obtained multi-labeled trees were compared to the corresponding score of the reference single-labeled species tree in order to infer the potential polyploidization events and the putative ancestral parental genomes involved in each event. The reference species tree was the consensus topology resulting from the highly-supported RNA-seq and GBS trees inferred by *vcf2alignment* and NGSEP (see above); labels were simplified by GRAMPA (--labeltree option).

Ancestral divergence ages of the *Brachypodium* homeologous subgenomes were estimated from the 397 core transcript data set with BEAST 2.5.0 (Bouckaert et al., 2014). We imposed independent site substitution models, lognormal relaxed clock and Birth-Death tree models, a broad uniform distribution prior for the uncorrelated lognormal distribution (ucld) mean (lower = 1.0E-6; upper = 0.1) and an exponential prior for ucld standard deviation (SD) to each partition. We used two calibration points, imposing normal distribution secondary age constrains for the crown nodes of the BOP clade [*Brachypodium* + *Oryza* + *Hordeum*] (normal prior mean = 51.9 Ma, SD =2.0) and the Brachypodium + core pooids clade [*Brachypodium* + *Hordeum*] (normal prior mean = 30.9 Ma, SD =2.5) following the grass-wide plastome based dating analysis of (Sancho et al., 2018). We ran 600,000,000 Markov chain Monte Carlo (MCMC) generations in BEAST with a sampling frequency of 1000 generations. The adequacy of parameters was checked using TRACER v.1.6 (http://beast.bio.ed.ac.uk/Tracer) with most parameters showing Effective Sample Size (ESS) >200. Maximum clade credibility (MCC) trees were computed after discarding 10% of the respective saved trees as burn-in.

Plastome data set and phylogenomic analysis

Plastome reads were filtered from the pool of RNA-seq data with DUK (http://duk.sourceforge.net) (Li et al., 2011a) using a reference set of 23 grass plastomes and a matching K-mer composition of K=24. Plastome reads were used to generate two datasets through de novo assembling and mapping to a reference *Brachypodium stacei* plastome, respectively (table S3a).

De-novo assembling and clustering of *B. pinnatum*-2x, *B. rupestre, B. phoenicoides* (two ecotypes), *B. mexicanum*, *B. boissieri* and *B. retusum* (see supplementary table S3b) transcripts, plus CDS sequences extracted from plastomes of *B. sylvaticum* (Sin1, assembled and annotated for this work), *B. arbuscula* (Barb502, assembled and annotated for this work), *B. distachyon* Bd21 (NC_011032; Bortiri et al. 2008), *B. stacei* ABR114 (NC_036837) and *B. hybridum* ABR113 (NC_036836), was performed with NOVOPlasty (Dierckxsens et al., 2017) and the pipeline described in Sancho et al. (2017) rendering an aligned data matrix. A total of 31 plastome core transcripts (atpA, atpF, ccsA, cemA, clpP, matk, ndhB, ndhJ, ndhK, petA, petB, petD, psaA, psaB, psaI, psbA, psbB, psbC, psbE, psbH, psbI, psbK, psbM, psbN, rbcL, rpl22, rpoA, rpoB, rps16, rps4 and rps7) were recovered from this data set, aligned and concatenated for phylogenomic analyses (fig. S7a). A validation for this approach was performed through the mapping of plastome reads of *B. distachyon* (Bd21), *B. stacei* (TE4.3), *B. hybridum* (BdTR6g), *B. arbuscula*, *B. pinnatum*-2x, *B. sylvaticum*_Esp, *B. sylvaticum*_Cor, *B. sylvaticum*_Gre, *B. rupestre, B. phoenicoides* (two ecotypes), *B. mexicanum*, *B. boissieri*, *B. retusum* and two outgroups (*Oryza sativa* and *Hordeum vulgare*) to the large *Brachypodium stacei* ABR114 plastome (NC_036837; Sancho et al. 2017) with hisat2 v2.0.5 (Kim et al., 2015). SNPs were called with the vsf2alignment and used to build a second aligned data matrix (fig. S7b; table S3a). A plastome based phylogenomic tree of *Brachypodium* was constructed with IQ-TREE using the concatenated data set of 31 core transcripts, imposing the optimal GTR+R3 substitution model selected by ModelFinder in terms of the AICc.

Results

Reference-genome syntenic mapping of RNA-seq and GBS data

A pipeline was designed to call SNP variants after mapping transcriptomic (RNA-seq) and genomic (GBS) paired-end sequence reads data obtained from the 12 *Brachypodium* species, cytotypes and ecotypes under study to three diploid reference genomes of *Brachypodium* (fig. 1; table 1; table S1). First, we mapped reads to a synthetic reference genome obtained by concatenating the genome sequences of diploid species *B. distachyon* (Bd, x=5), *B. stacei* (Bs, x=10) and *B. sylvaticum* (Bsy, x=9) (figs. 1a). Mapping statistics are provided in supplementary figure S1 and tables S4a, b, S5a, b and S6a, b. Second, we called and piled up SNPs to produce a multiple alignment containing a single sequence per accession (type I data set, fig. 1a). Third, whole genome alignments of chromosomes from the master *B. distachyon* genome and the recently assembled *B. stacei* and *B. sylvaticum* genomes that showed high collinearity between their respective chromosome complements (5 Bd and 9 Bsy chromosomes resulting from predominant centromeric chromosome fusions of 10 Bs chromosomes with minor rearrangements; see Supplementary Methods) were computed to obtain aligned sequences of syntenic genomic positions (fig. 1b). A high synteny was observed between the three *Brachypodium* reference genomes and that of *Oryza sativa* (fig. 1c). Four, we used this syntenic alignment to partition the previously aligned samples (type I data set) into up to three potential subgenomes per sample (*B. distachyon*-type, *B. stacei*-type, and *B. sylvaticum*-type homeologous genomes). In this way, we generated syntenically aligned data matrices of, respectively, RNA-seq and GBS data (type II data sets, fig. 1b), each of them comprising the three reference genomes and up to three sequences (subgenomes) per accession.

The amount of aligned syntenic SNPs per data set varied; the RNA-seq data set was overall one order of magnitude larger than the GBS data set (fig. S1; table S6).

Therefore, most downstream analysis were based on the robust RNA-seq data, using the GBS data as a validation approach of the methods and of some results. We tested our pipeline for the two types of data (transcriptomic RNA-seq and genomic GBS) data and for the two mapping strategies (type I and II) using both biological and bioinformatics controls. The biological controls included: i) two ecotypes of the allotetraploid *B. hybridum* (BdTR6g and ABR113 for RNA-seq and GBS data, respectively), which showed ~50% of reads mapped to each of its two progenitor *B. distachyon* and *B. stacei* genomes, and almost none to *B. sylvaticum* (table S5a, b), ii) three *B. sylvaticum* samples (Bsyl-Cor, Bsyl-Gre, Bsyl-Esp, RNA-seq data; table 1) that were resolved as monophyletic in the RNA-seq based trees, and iii) two *B. phoenicoides* samples (Bpho6 and B422; for both RNA-seq and GBS data) that were resolved as sister taxa. Our pipeline was further validated bioinformatically by comparing the resulting trees to those obtained with other software such as NGSEP (for RNA-seq data) and NGSEP and GIbPSs (for GBS data) (table S7), rendering in all cases congruent topologies (see Results).

As expected, the *B. distachyon*, *B. stacei* and *B. sylvaticum* RNA-seq and GBS SNPs mapped almost totally or preferentially to the chromosomes of their respective reference genomes (fig. S1, table S6), though some percentages of the *B. sylvaticum* SNPs also mapped to the *B. distachyon* (17.2-22.2% RNA-seq; 11.6% GBS (Bsyl-Sin1, table 1) and *B. stacei* (8.3-10.0% RNA-seq; 5.4% GBS) chromosomes (table S6). SNPs from *Brachypodium* species, cytotypes and ecotypes of the core perennial clade (*B. arbuscula*, *B. phoenicoides*, *B. pinnatum* 2x and 4x, *B. rupestre* 4x) mapped mainly to *B. sylvaticum* chromosomes (≥66.1% RNA-seq; ≥73.5% GBS) and less frequently to *B. distachyon* (20.2-23.6% RNA-seq; 15.8-18.0% GBS) or *B. stacei* (8.8-10.3% RNA-seq; 6.6-8.3% GBS) chromosomes (table S6). *B. mexicanum* and *B. boissieri* SNPs mapped similarly to each of the three reference genomes (~30%) for RNA-seq data and only slightly more to *B. sylvaticum* chromosomes (41.0-45.0%) for GBS data. Most *B. retusum* SNPs mapped to the *B. sylvaticum* genome (53.6% RNA-seq; 58.5% GBS), with smaller fractions of them mapping to the *B. distachyon* (26.8%; 23.4%) and *B. stacei* (19.6%; 18.1%) chromosomes, respectively (table S6). The GBS data set was too small to accurately retrieve homeologous subgenomes of allopolyploids and only the RNA-seq dataset was used for further phylogenomic analysis based on type II data. The pipeline steps are described in more detail in Supplementary Methods.

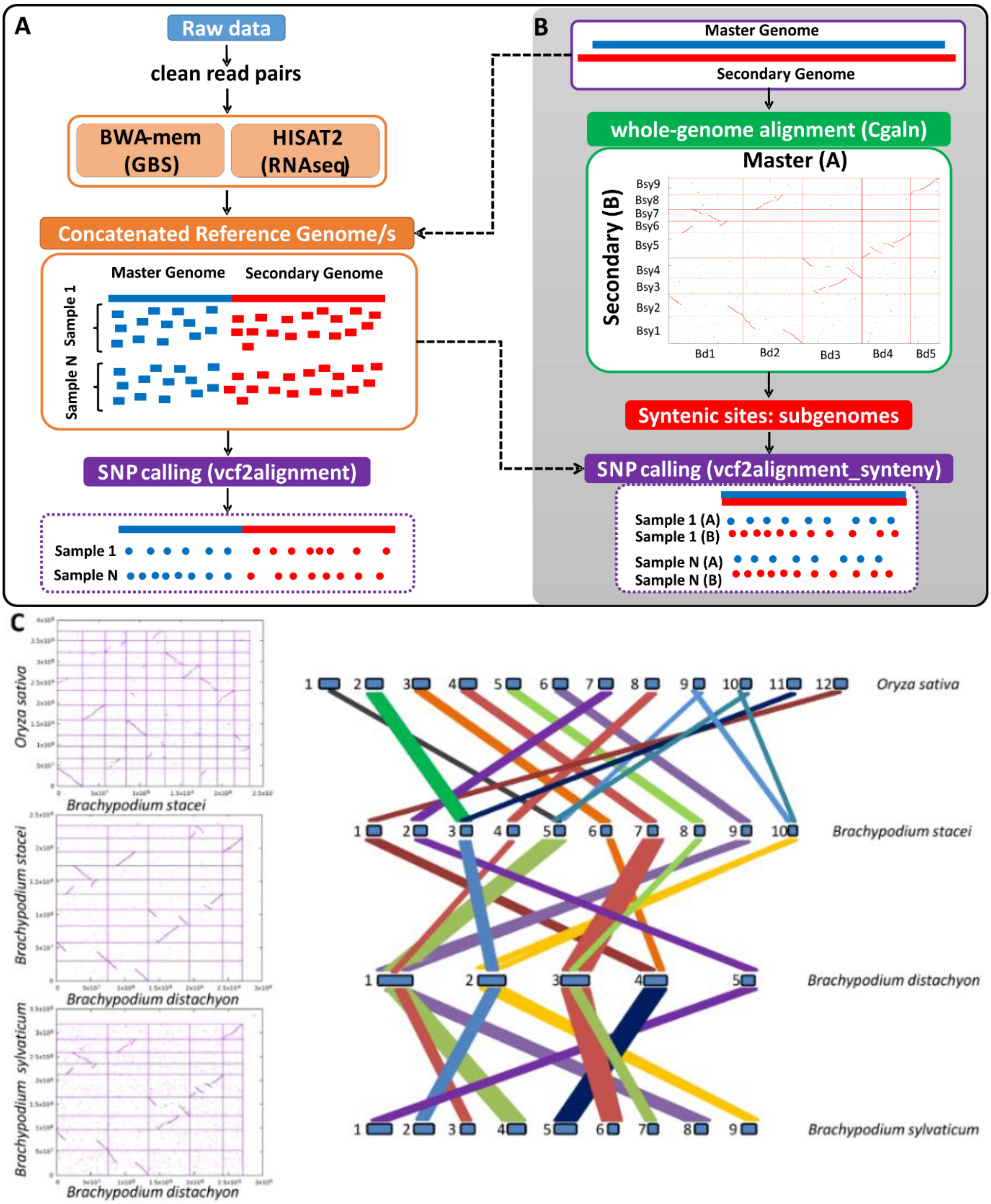


Figure 1. Pipeline used for reference-genome syntenic mapping and alignment of *Brachypodium* RNA-seq and GBS data. **(A)** Mapping of RNA-seq or GBS reads of *Brachypodium* species, cytotypes and ecotypes to the three concatenated *B. distachyon* - *B. stacei* -*B. sylvaticum* reference genomes and SNP calling with the *vcf2alignment* tool. **(B)** Whole genome syntenic alignment of the secondary *B. stacei* (chromosomes Bs1 to Bs10) and *B. sylvaticum* (chromosomes Bsy1 to Bsy9) reference genomes to the master *B. distachyon* (chromosomes Bs1 to Bd5) reference genome with Cgaln and syntenic alignment of the *Brachypodium* species and cytotypes SNPs (from A) to the genome data matrix with the *vcf2alignment_synteny* tool **(C)** Syntenic alignment of the *Oryza sativa*, *B. stacei* and *B. sylvaticum* genomes against the *B. distachyon* genome.

Nuclear core transcripts: allelic assignation to allopolyploid subgenomes

A second pipeline was designed to cluster sequences from expressed genes, produce multiple alignments and infer gene phylogenies. As illustrated in fig. 2a, core and accessory transcripts were called for, respectively, phylogenetic reconstruction of and comparative expression among *Brachypodium* taxa, thus defining a pan-transcriptome. In subsequent steps, independent core transcripts were aligned and trees were computed (fig. 2b).

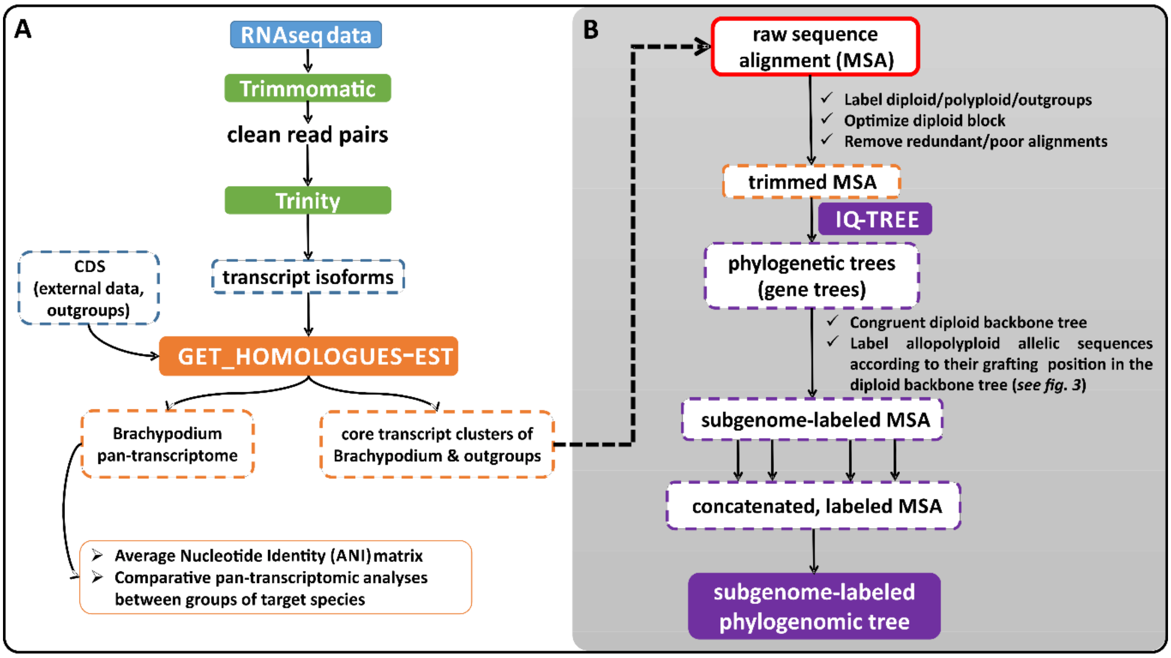
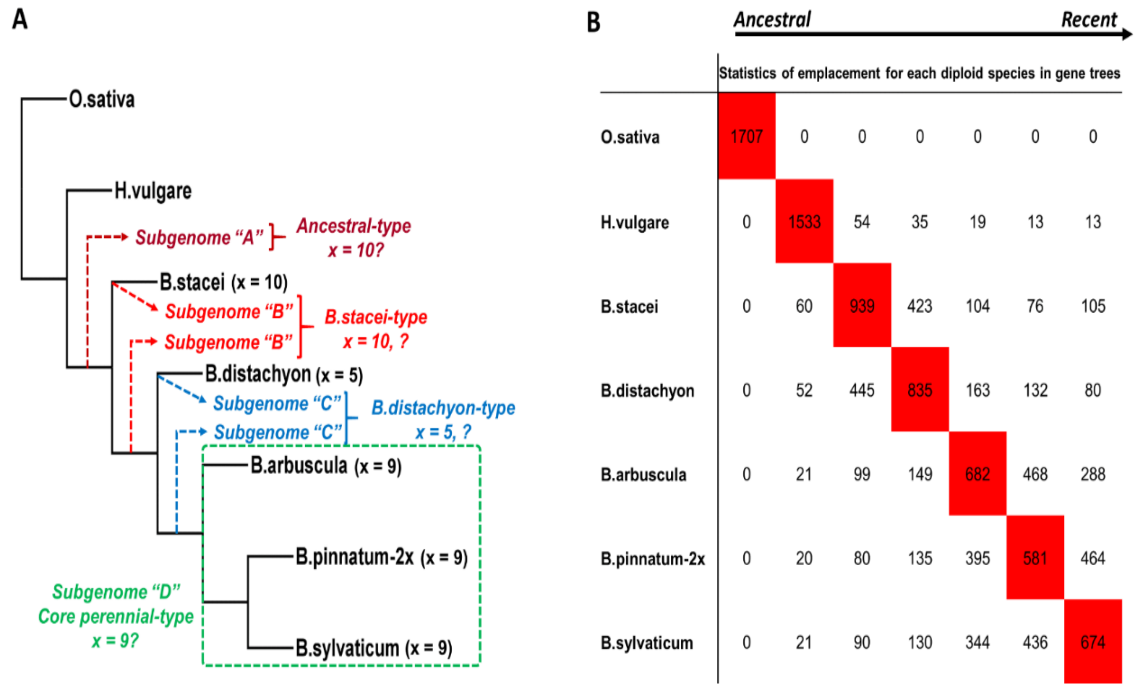

A
RNAseq data
Trimmomatic
clean read pairs
Trinity
CDS (external data, outgroups)
transcript isoforms
GET_HOMOLOGUES-EST
Brachypodium pan-transcriptome
core transcript clusters of Brachypodium & outgroups
➢ Average Nucleotide Identity (ANI) matrix
➢ Comparative pan-transcriptomic analyses between groups of target species
B
raw sequence alignment (MSA)
✓ Label diploid/polyploid/outgroups
✓ Optimize diploid block
✓ Remove redundant/poor alignments
trimmed MSA
IQ-TREE
phylogenetic trees (gene trees)
✓ Congruent diploid backbone tree
✓ Label allopolyploid allelic sequences according to their grafting position in the diploid backbone tree (*see fig. 3*)
subgenome-labeled MSA
concatenated, labeled MSA
subgenome-labeled phylogenomic tree

Figure 2. Pipeline used for phylogenomics analyses using the core transcript data set. **(A)** Filtering, assembling and transcript analysis for phylogenomic inference (core transcripts) and *Brachypodium* pangenome (all transcripts). **(B)** Workflow of core transcript and phylogenetic trees filtering for phylogenomic reconstructions of diploid genomes and allopolyploid homeologous genomes (subgenomes) guided by allelic grafting positions in the topology of the consensus diploid backbone tree.

After selecting core gene trees showing a congruent diploid backbone tree (see below), allelic sequences from *Brachypodium* allopolyploids were labeled according to their grafting position as sister or as closest ancestral or descendant branches of the backbone tree diploid lineages, thus representing homeologous copies from their respective putative homeologous genomes (fig. 3a, b). The multilabeled multiple sequences alignments (MSA) from those genes were combined as separate partitions into a data set and used to compute maximum likelihood (ML) and Bayesian trees and to estimate dates of divergence of diploid genomes and of allopolyploid homeologous subgenomes.



Statistics of emplacement for each diploid species in gene trees

O.sativa	1707	0	0	0	0	0	0
H.vulgare	0	1533	54	35	19	13	13
B.stacei	0	60	939	423	104	76	105
B.distachyon	0	52	445	835	163	132	80
B.arbuscula	0	21	99	149	682	468	288
B.pinnatum-2x	0	20	80	135	395	581	464
B.sylvaticum	0	21	90	130	344	436	674

Figure 3. Statistics of topological placement of each *Brachypodium* and outgroup (*Oryza*, *Hordeum*) diploid lineages in the consensus diploid backbone tree based on 1,707 core transcripts. Values indicate the number of genes that support the topological placement of a specific diploid lineage in the consensus backbone tree **(A)**; *Brachypodium* and outgroup (*Oryza*, *Hordeum*) consensus diploid backbone tree (black branches) based on 397 common core genes showing the four potential grafting position of the allopolyploids' allelic copies corresponding to their putative homeologous genomes according to the subgenome-type criterion [A: ancestral-type (brown), B: stacei-type (red), C: distachyon-type (blue), D: core perennial-type (green); see Results and table S8] **(B)**.

After assembling RNA-seq reads, between 72 and 160 thousand transcript isoforms were obtained with median lengths ranging between 414 to 555 bp (table S2). Transcripts from all *Brachypodium* species, cytotypes and ecotypes, plus coding DNA sequences (CDS) from *Brachypodium distachyon* (Bdistachyon-314-v.3.1; http://phytozome.jgi.doe.gov/), and from *Oryza sativa* and *Hordeum vulgare* outgroups, were compared, producing a total number of 5,202 clusters. A subset of 3,324 clusters contained sequences from all *Brachypodium* species plus outgroups, and were consequently annotated as core clusters. Core MSAs were computed, partially aligned sequences removed and alignments with missing diploid backbone tree lineages sequences discarded, yielding in total 1,786 complete clusters of core transcripts. Phylogenetic trees were subsequently estimated for each of these MSAs, obtaining 1,707 curated gene trees. The nesting frequencies of the diploid lineages across the trees were recorded and used to select the most congruent diploid backbone tree (fig. 3a). Overall, we recovered 397 clusters and trees showing that diploid topology. Allopolyploid allelic sequences in those trees were labeled with codes A to D (corresponding to potential homeologous genomes A to D) according to their relative position in the diploid backbone tree (fig. 3b, table S8). More details of this protocol are provided in Supplementary Methods.

Genome size analysis and ploidy level inference

Genomes size (GS) values of new *Brachypodium* accessions were analyzed by flow cytometry and contrasted with GS values and chromosome count values obtained in previous studies (table 1). Diploid *Brachypodium* species of the core perennial clade showed values ranging between 0.713±0.004 pg/2C (*B. arbuscula*) and 0.822±0.009 pg/2C (*B. pinnatum*-2x), that corresponded to chromosome numbers of 2n=18. Genome sizes of tetraploid core perennial clade species (*B. rupestre*-4x, *B. pinnatum*-4x and *B. phoenicoides*) were approximately constant, ranging between 1.4 and 1.5 pg/2C, corresponding to chromosome numbers of 2n=28. The putative hexaploid *B. retusum* showed a GS value of 1.704±0.024 pg/2C, that corresponded to a chromosome number of 2n=32 (Inda, unpubl. data; Schippmann 1991), and the putative octoploid *B. boissieri* a high value of 3.236±0.072 pg/2C, that corresponded to a chromosome number of 2n=ca. 46 (Inda, unpubl. data; Schippmann 1991). The short-rhizomatose perennial and putative tetraploid *B. mexicanum* showed the highest GS value known within *Brachypodium*, 3.774±0.033 pg/2C (table 1). It corresponded to the same *B. mexicanum* accession that showed a chromosome number of 2n=40 (Shi et al., 1993).

Phylogenomics based on reference-genome synteny mapping: the *Brachypodium* species tree and subgenome tree

The *Brachypodium* species and subgenomes trees were computed aiming to unravel the evolutionary history of its lineages both at the species and genomic levels. First, we performed separate phylogenomic analyses with RNA-seq and GBS SNP data mapped onto the three concatenated reference genomes (type I data, fig. 1a), using a single aligned sequence per accession and targeting the *Brachypodium* species tree. Maximum likelihood (ML) trees were constructed with the IQTREE software after inferring best-fit evolutionary models (table S7). Overall, strongly supported and congruent topologies were obtained from both data sets (figs. S2; S3, S4; Supplemental Results); however, only the more widely sequenced RNA-seq based topology will be further explained. The RNA-seq *Brachypodium* species tree showed the successive early splits of annual diploid *B. stacei* and *B. distachyon* lineages and an intermediate position of the allotetraploid *B. hybridum* between these two parental lines. It was followed by the successive divergences of intermediate evolved perennial polyploid *B. mexicanum* and *B. boissieri/B. retusum* lineages, and by the recent split of the core perennial clade lineages in which the early divergence of diploid *B. arbuscula* was followed by that of the sister clades of diploids *B. pinnatum*-2x/*B. sylvaticum* and tetraploids *B. rupestre*-4x/*B. phoenicoides* (see fig. S2a). This topology was validated by re-analyzing the RNA-seq data with an independent methodology (NGSEP; fig. S3).

Second, we conducted phylogenomic analyses using the RNA-seq data (type II data (fig. 1b), searching for the *Brachypodium* subgenome tree where the putative homeologous lineages present in the allopolyploid species, cytotypes and ecotypes could be identified. The multi-labeled syntenic RNA-seq data matrix, consisting of 28,563,327 aligned sites (505,512 of them informative) and 24 sequences, was used to build a ML tree with IQ-TREE imposing the best-fit GTR+R4 model (selected by AICc) and using *Oryza sativa* as root (table S7). Subgenomic sequences of diploid or allopolyploids species forming monophyletic clades (dashed boxes in fig. 4a) were collapsed into single consensus sequences (fig.4b).

Up to three putative homeologous subgenome sequences were obtained for some allopolyploid species, one per reference genome. The resulting *Brachypodium* subgenome tree (fig. 4) showed that the two subgenome sequences of allotetraploid *B. hybridum* were resolved as sister to, respectively, its parental *B. stacei* and *B. distachyon* lineages, whereas the three subgenome sequences of the allopolyploids *B. mexicanum*, *B. boissier*i and *B. retusum* were resolved in basal and sub-basal evolutionary positions (*B. mexicanum, B. boissieri*) and in basal, intermediate and recently evolved evolutionary positions (*B. retusum*). Homeologous *B. mexicanum, B. boissieri* and *B. retusum* sequences mapped to *B. stacei* aligned close to this lineage, whereas those mapped onto *B. distachyon* were placed in an intermediate position between the *B. stacei* and *B. distachyon* splits. In contrast, those mapped to *B. sylvaticum* were sister to this lineage (*B. retusum*) or were placed more ancestrally between the *B. stacei* - *B. distachyon* lineages (*B. mexicanum, B. boissieri*). Two of the three homeologous sequences retrieved for allotetraploids *B. rupestre* and *B. phoenicoides* (Bpho6, B422) (Bsta and Bdis types) were sister groups of a clade nested before the split of the core perennial clade, whereas the third type of homelogous sequences (Bsyl-type) were separately nested (B. phoenicoides-Bsyl/B. retusum-Bsyl intermediate between *B.*

pinnatum and *B. sylvaticum*; B. rupestre-Bsyl sister to *B. sylvaticum*) within the core perennial clade (fig. 4b).

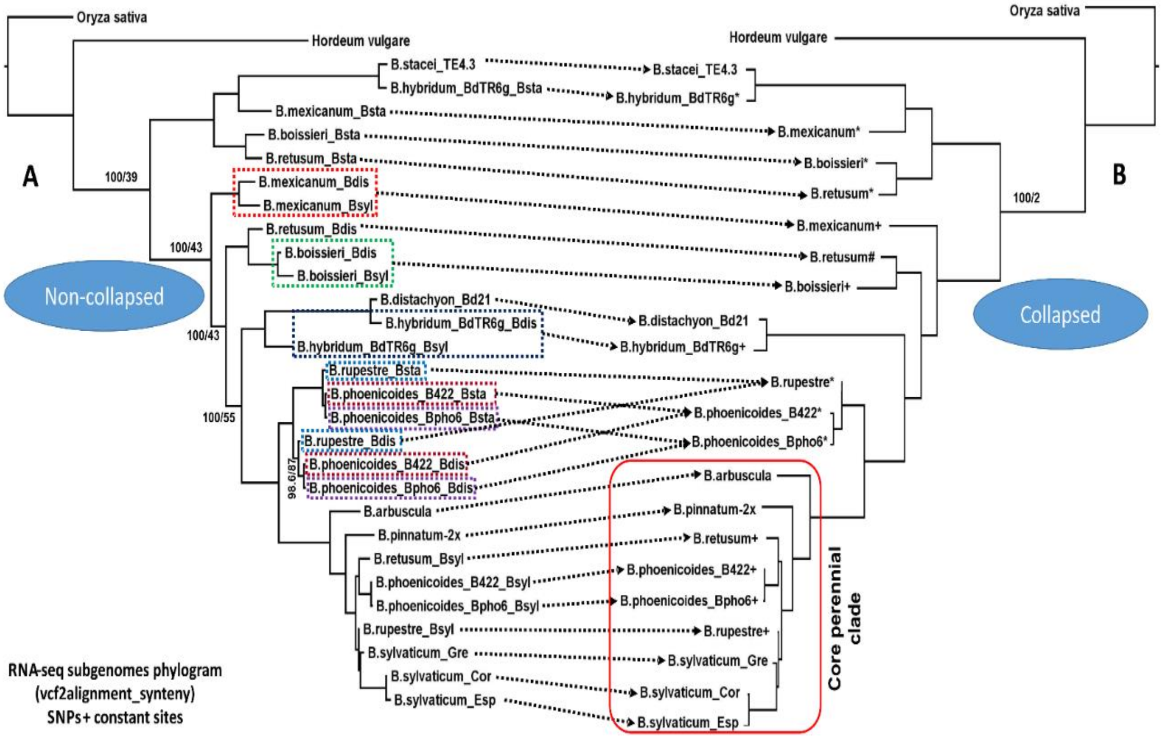


Figure 4. *Brachypodium* maximum likelihood subgenomic tree based on RNA-seq SNPs from diploid and allopolyploid accessions mapped and syntenically aligned to the three *Brachypodium* reference genomes (*B. distachyon*: Bdis; *B. stacei*: *Bsta*; *B. sylvaticum*: Bsyl) using *vcf2alignment_synteny*, IQTREE topologies showing non-collapsed (A) and collapsed (B) monophyletic subgenomic clades of allopolyploid *Brachypodium* species. Asterisk (*), hash (#) and plus (+) symbols indicate the inferred ancestral, intermediate evolved and recently evolved subgenomes of each allopolyploid species and sample. *Oryza sativa* was used to root the trees. SH-aLRT/UltraFast Bootstrap supports (<99) values are shown on branches.

A partial validation of the *Brachypodium* RNA-seq subgenome tree was conducted with Gene-tree Reconciliation Algorithm with MUL-trees for Polyploid Analysis (GRAMPA) (Thomas et al., 2017), which recovers a maximum of two homeologous genomes per sample, using 3,173 transcript clusters. In the course of these analyses, all studied *Brachypodium* polyploids were consistently reported as allopolyploids (fig. S5). The best parsimony trees of allopolyploids were congruent with our previous approach, recovering the ancestral subgenomes of *B. mexicanum* (fig. S5a) and *B. boissieri* (fig. S5b), the intermediately and recently evolved subgenomes of *B. retusum* (fig. S5c), the two subgenomes of *B. hybridum* (fig. S5d), sister to each of its *B. stacei* and *B. distachyon* parental lineages, and the two recently evolved subgenomes of core perennial allotetraploids *B. rupestre* and *B. phoenicoides* (fig. S5e, f, g).

Phylogenomics based on 397 nuclear core expressed genes: the *Brachypodium* nuclear gene tree

A *Brachypodium* ML nuclear gene tree was computed from the 397 individual gene trees that were congruent with the diploid backbone topology, including all detected allopolyploids' alleles coded according to the subgenome criterion (A to D), as described above and in fig. 3a, b, table S8 and Supplementary methods. It was clear from the overall statistics that some homeologous allelic copies ("subgenomes") of allopolyploids were found in many gene trees, while others could only be observed marginally (table S8). In order to reduce the effect of potential artefacts, allopolyploid allelic (subgenomic) copies found in less than 15% of the gene trees were removed from downstream analyses, and the remaining copies were used to infer the parental genome lineages of the allopolyploids. The *Brachypodium* gene tree computed with IQTREE (fig. 5) confirmed the hybrid origin of allotetraploid *B. hybridum*; 53% and 44% of its core genes (or core allelic copies) were found to be sister to its parental *B. stacei* (B subgenome) and *B. distachyon* (C subgenome) lineages, respectively.

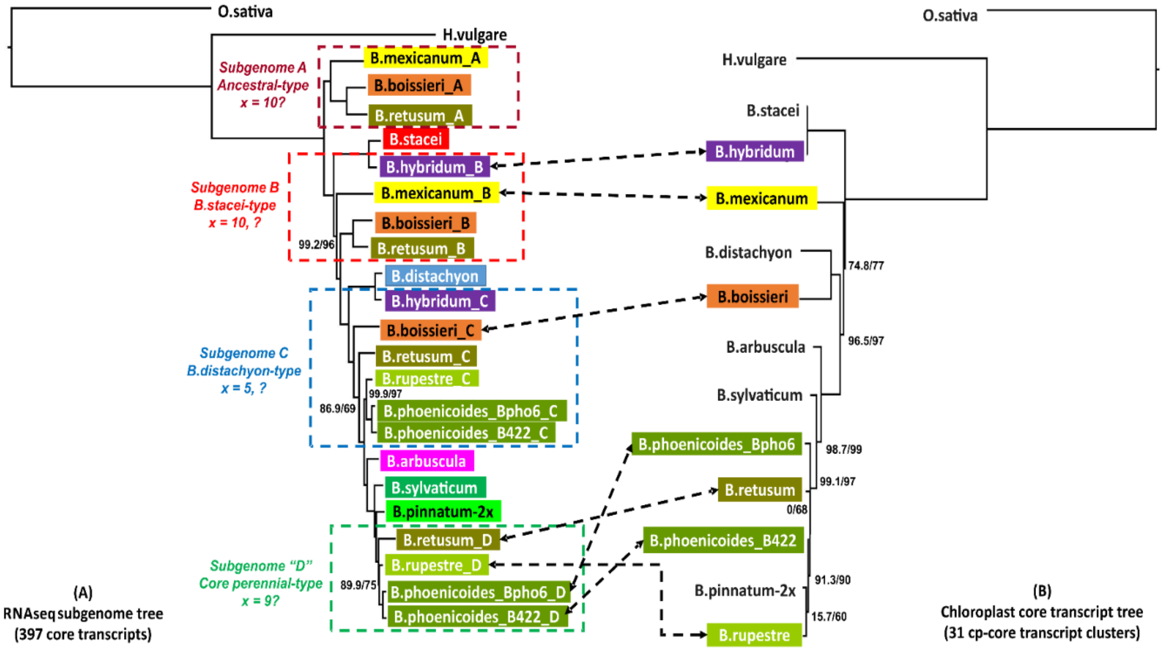


Figure 5. *Brachypodium* gene trees based on maximum likelihood analysis of 397 independent nuclear core genes **(A)**, following the allopolyploid allelic copy grafting to diploid backbone tree branches procedure (subgenomic classification criterion) described in Results, fig. 3 and table S8, and of 31 plastome core gene **(B).** The IQTREE nuclear topology shows the inferred ancestral-type (A), stacei-type (B), distachyon-type (C) and core perennial-type (D) homeologous genomes of the studied allopolyploid species and samples, and the IQTREE plastome topology the subgenomic lineages that acted as maternal genome donors of the studied allopolyploid accessions (dashed lines). *Oryza sativa* was used to root the trees. SH-aLRT/UltraFast Bootstrap supports (<99) values are shown on branches.

The most expressed core genes of *B. mexicanum* indicated that this allopolyploid inherited only ancestral subgenomes (A, B) (fig. 5); the participation of a third putative subgenome C based on 34 core genes was rejected based on their low frequency (table S8). Three and four subgenomes were respectively detected for *B. boissieri* and *B. retusum* in the IQTREE topology. Both species shared ancestral subgenomes A and B and intermediately evolved subgenome C; *B. retusum* also presented a recently evolved subgenome D (fig. 5). Core genes from subgenome A were more frequent in *B. boissieri* and those from subgenomes C and D in *B. retusum* (table S8). Fifteen core genes of subgenome D were also detected in *B. boissieri* (fig. S6a-f, table S8), although they were discarded from analysis due to their low frequency. Two subgenomes, C and D, were found in the core perennial allopolyploids *B. rupestre* and *B. phoenicoides* (ecotypes Bpho6, B422) (fig. 5); core genes from the recentmost D subgenome were the most expressed in these accessions (table S8).

A BEAST maximum clade credibility (MCC) tree of 397 nuclear core genes yielded the same *Brachypodium* gene topology (fig. 6) than that of the IQ-TREE (fig. 5).

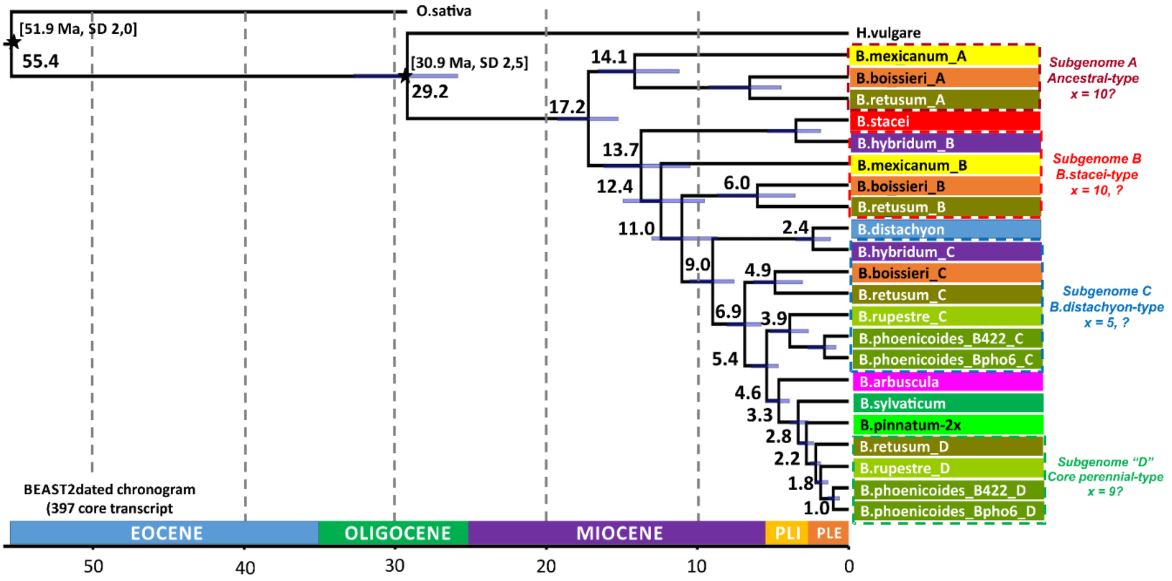


Fig. 6. *Brachypodium* BEAST2 maximum clade credibility (MCC) dated chronograms of 397 independent nuclear core genes (with allopolyploid allelic copies classified as subgenomes types "A, B, C and D", see fig. 5) (A) and 31 plastome core genes (B) showing estimated nodal divergence times (meadians, in Mya) and 95% highest posterior density (HPD) intervals (bars). Stars indicate secondary nodal calibration priors (means ± SD, in Mya) for the crown nodes of the BOP [*Oryza* + *Brachypodium* + *Hordeum*] and *Brachypodium* + core pooids [*Brachypodium* + *Hordeum*] clades. Accessions codes correspond to those indicated in table 1.

The splits of the *Brachypodium* stem and crown nodes were estimated to have occurred in the Mid-Oligocene (29.2 Ma) and Early-Miocene (17.2 Ma), respectively (fig. 6). Mid-late Miocene ages were estimated for the successive splits of *B. stacei* (13.7 Ma) and *B.*

distachyon (9.0 Ma) lineages, and Early-late Pliocene ages for those of *B. arbuscula* (core perennial clade) (4.6 Ma), and *B. sylvaticum*/*B. pinnatum*-2x (2.8 Ma) (fig. 6). The split of the ancestral subgenome A lineage was inferred to have occurred in the Mid-Miocene (14.1 Ma), predating the split of the oldest extant diploid (*B. stacei*) lineage. The more ancestral *B. mexicanum* subgenome B lineage was estimated to have split in the Mid-Miocene (12,4 Ma) whereas those inherited by *B. boissieri* and *B. retusum* diverged more recently (11 Ma). The subgenome C lineage inherited by *B. boissieri* and *B. retusum* was inferred to be more ancestral (6.9 Ma) than that inherited by *B. rupestre* and *B. phoenicoides* (5.4 Ma). The split of subgenome D lineage of *B. retusum*, *B. rupestre* and *B. phoenicoides* was dated to the Pliocene-Pleistocene transition (2.2 Ma). The origin of the recenmost *B. distachyon*-type parental lineage of *B. hybridum* (ABR113) was estimated to have occurred in the Pleistocene (2.4 Ma) (fig. 6).

Phylogenomics based on 31 plastome expressed genes: the *Brachypodium* plastome tree

Thirty one core plastome transcripts were assembled de novo from filtered RNA-seq reads obtained for the *Brachypodium* accessions under study and from additional genome data (table 1), and were concatenated and used to build a ML *Brachypodium* plastome tree with IQ-TREE (fig. S7a). A validation approach for this topology was conducted using filtered SNPs from the RNA-seq plastome reads mapped to the reference plastome of *B. stacei* (fig. S7b) (see Material and Methods below). The plastome tree was contrasted to the *Brachypodium* nuclear gene tree and used to infer the maternal genome donors of the studied allopolyploid accessions (fig. 5b).

The two plastome trees were highly congruent to each other (fig. S7) and to the nuclear core gene tree (fig. 5). The 31 core plastome gene tree showed the successive moderate to well supported divergences of *B. stacei* (and *B. hybridum* with stacei-type plastome), *B. mexicanum*, *B. distachyon/B. boissieri, B. arbuscula, B. sylvaticum, B. phoenicoides* (Bpho6, B422), *B. retusum,* and *B. pinnatum/B. rupestre* lineages (fig. 5b, S7). The SNP plastome tree showed a congruent topology with the plastome gene-based tree but with swapped positions for *B. retusum* (sister to *B. sylvaticum*) and *B. phoenicoides* B422 (sister to *B. pinnatum*) (fig. 5b). The topological comparisons between the 397 nuclear core gene tree and the 31 plastome core gene tree indicated that the maternal genome donors of *B. mexicanum* and *B. boissieri* were their respective subgenomes B, and of *B. retusum*, *B. rupestre* and *B. phoenicoides* their respective subgenomes D (fig. 5, fig. S8).

Analysis of the *Brachypodium* pan-transcriptome

The complete collection of *Brachypodium* clusters of expressed genes was further analyzed within a pan-genome context. First, a list of 3,324 core genes found to be expressed in all species and outgroups were systematically compared in order to compute a matrix of Average Nucleotide Identities (ANI), summarizing the average gene identity between any pair of species under study. Gene sequence identity among *Brachypodium* species was on average over 94%. A heat-map and hierarchical clustering of species based on this data showed the highest identities among allotetraploids and diploid core perennials (fig. S8). Two sister groups, annuals + *B. mexicanum* + *B. boissieri* (ANI1) *versus B. retusum* + core perennial species (ANI2), were detected showing high intragroup and low intergroup sequence identity. These results are congruent with the previously described phylogenies (figs. 4, 5).

Second, a larger group of 5,202 transcript clusters, comprising both core and accessory genes (*sensu* Contreras-Moreira et al. 2017), were used to compile a presence-absence pan-transcriptome matrix of *Brachypodium*. Interrogation of this matrix identified exclusive gene clusters found to be core in a subset of species and absent in the remaining (table S9 for all transcript clusters which could be annotated by sequence similarity). For instance, there were 14 gene clusters expressed in the ANI1 group which were not observed in ANI2, including a putative MYB transcription factor (table S9a). The reverse comparison yielded 52 transcript clusters exclusive of ANI2 species, comprising, among others, disease resistance genes and a cell wall transporter (table S9b). We also observed 30 gene clusters expressed in all perennials and absent in annuals, including two NB-LRS resistance genes, a GLABRA homeobox transcription factor associated to maintaining floral identity and a G-type lectin S-receptor-like serine/threonine protein kinase (table S9c). In addition, 49 gene clusters were found to be expressed in annuals but absent in all perennials, including a beta subunit of RNA polymerase or a potential gene encoding a CCCH domain (table S9d). When comparing polyploids and diploids, it was found that all gene clusters expressed in diploids were also present in polyploids; however, there were 14 transcript clusters found in polyploids but not expressed in diploids, including a putative amino acid permease, a plastid aspartokinase and an ATP-dependent 6-phosphofructokinase (table S9e). In order to identify unique ancestral gene copies within clusters, *B. mexicanum*, *B. boissieri* and *B. retusum* were compared to the remaining species. A total of 143 putative ancestral transcript clusters were reported in these old allopolyploids (table S9f). In contrast, only 8 expressed gene clusters were found to be missing in these species and present in the remaining species, including a putative universal stress protein (table S9g).

Enrichment analyses was also carried out with all private sequence clusters. Only one statistically significant (False Discovery Rate (FDR)<0.05) GO biological process, corresponding to transcription by RNA polymerase I, was associated to the set of genes expressed in annuals but absent in perennials.

Discussion

Contrasting evolutionary histories of the *Brachypodium* lineages, discovering homeologous subgenomes in allopolyploid species

Deciphering the origins of plant allopolyploids face the challenge of accurately capturing the parental subgenomes contributing to these hybrid genome doubling species and their divergence times (Levin, 2013; Bombarely et al., 2014; Soltis et al., 2016). Approaches using coalescent-based analyses of multi-labeled trees and networks of a variable range of nuclear genes have been hampered by homeolog loss and incomplete lineage sorting (ILS) (Marcussen et al., 2015; Thomas et al., 2017). Targeted NGS methods have provided large amounts of nuclear genomic or exomic data (Buggs et al., 2012; Kamneva et al., 2017), however the deconvolution of the hybrid subgenomes is still controversial, especially in the predominant absence of known extant parents and of whole genome sequence data for the studied species. Recovery of potential subgenomes has been accomplished through a combination of reference-based mapping (to a unique reference genome) and *de novo* assembly, obtaining phasing haplotypes of allopolyploid individuals in *Hordeum* (Brassac & Blattner, 2015). Syntenic read/SNP mapping to the respective parental diploid genomes allowed the separation of both subgenomes in three allotetraploid *Glycine* species (Bombarely et al., 2014) and two allotetraploid *Arabidopsis* species (Novikova et al., 2016) of know hybrid origin. Our multiple reference-genome syntenic mapping approach upgrades the last method allowing the detection of unknown parental genomes in wide ploidy-level *Brachypodium* species (4x-8x) that have had different dysploid ancestral origins. This strategy, developed with SNPs from a large transcriptomic data set, allowed us to uncover all potential homeologous subgenomes (2) of allotetraploids *B. hybridum*, *B. mexicanum*, *B. rupestre*-4x and *B. phoenicoides* and of putative allohexaploid *B. retusum* (3), and up to two out of the four potential subgenomes of putative allo-octoploid *B. boissieri* (fig. 4). Our gene-based phylogenetic approach refines previous methods (e. g. Bombarely et al. 2014) through the filtering of gene trees congruent with the strongly supported diploid backbone tree and of most frequent (≥15%) allopolyploid homeologous alleles grafted to its successively divergent nodal-branch groups (A-D) (fig. 3a, b, fig. 5a). Our nuclear subgenome tree detected the same number of potential homeologous subgenomes than the syntenic SNP tree for the allotetraploids (2), increasing the number of potential subgenomes for *B. boissieri* (3) and *B. retusum* (4) (fig. 5a). Furthermore, our nuclear subgenome tree found a hypothetical ancestral genome (A) only detected in the oldest allopolyploids (*B. mexicanum*, *B. boissieri*, *B. retusum*), similar to that retrieved using few cloned nuclear genes (Catalán et al. 2016; Díaz-Pérez et al. 2018), but using a large representation of 397 core expressed genes (fig. 5a). The strong evidence for the existence of this ancestral diploid A genome (private to *B. mexicanum*, *B. boissieri* and *B. retusum*), that could not be detected by the constrained syntenic SNP mapping approach where the oldest reference genome was that of current diploid *B. stacei* (fig. 1,4), supports an earlier split of the extremely isolated *Brachypodium* lineage [17.2 Ma, *Brachypodium* crown node, 14.1 Ma ancestral genome A crown node, (fig. 6a)] than previous estimates (Sancho et al. 2017, and references therein) but with overlapping HDP intervals. The inheritance of this ancestral (presumably extinct or unsampled) diploid genome in the current *Brachypodium* allopolyploid species would have contributed to increasing the diversification rates of the genus, as in other pooid allopolyploids (Pimentel et al., 2017b).

The *Brachypodium* nuclear and plastome phylogenies (fig. 4, 5, 6) and the GS ploidy level analysis (table 1) provide an optimal framework for the reconstruction of the evolutionary history of the studied *Brachypodium* allopolyploids. Our whole-genome synteny approach (fig. 1b, Cgaln) has inferred a solid descendent dysploidy evolutionary scenario of nested chromosome fusions with occasional reshuffling (Robertsonian translocations, inversions) of the 10 oldest *B. stacei* chromosomes into the independently evolved 5 chromosomes of *B. distachyon* and 9 chromosomes of *B. sylvaticum* (figs. 1c, 3a), paralleling that proposed for the intermediate ancestral karyotype of grasses (x=12) into those of modern *Oryza sativa* (x=12) and *B. distachyon* (x=5) (Murat et al., 2010). The use of the model *B. hybridum* allotetraploid, a species that experienced interspecific hybridization followed by WGD (Catalán et al., 2014; Dinh Thi et al., 2016), as control, reinforces the value of our approaches. All the nuclear data sets (RNA-seq, GBS) and analytical methods assayed have detected equally likely participations of its *B. stacei*-type (Bs) and *B. distachyon*-type (Bd) parental genomes and the negligible presence of the *B. sylvaticum*-type (Bsy) genome (table S5, S6) in the species. Further, its nuclear homeologous genomes (SNPs, genes) are the only allopolyploid subgenomes studied resolved as sister to their respective extant parental diploid lineages (figs. 4, 5). Our dating analysis supports the recent origin of this hybrid species (2.4 Ma for its Bd lineage; fig. 6) that presumably spanned the Quaternary (Catalán et al., 2012, 2016b). Our plastome gene tree detected the stacei-type maternal donor of the studied BdTR6g line (fig. 6b, S7) though our previous analysis demonstrated that *B. hybridum* originated recurrently in its native circum-Mediterranean region and from bidirectional crosses (López-Alvarez et al., 2012; Catalán et al., 2016b).

The successfully tested syntenic evolutionary framework has also contributed to elucidate the origins of other allopolyploid *Brachypodium* species (fig. 7). Three major routes have been proposed for explaining the cytological mechanisms that might cause the mergin of parental genomes and the production of new allopolyploid species: i) the fusion of reduced (n) female and male gametes with heterologous genomes followed by WGD of the interspecific sterile F1 hybrid, leading to the restoration of fertility in the amphidiploid allopolyploid; ii) the fusion of unreduced (2n) gametes with putative homeologous genomes, via homoploid or heteroploid hybridization, resulting in a fertile segmental (or non-segmental) allopolyploid; and iii) the “triploid bridge”-type route, which involves the formation of a semi-fertile F1 individual resulting from the fusion of reduced (n) and unreduced (2n) gametes with homeologous genomes that sporadically produces unreduced 3n gametes (Ramsey & Schemske, 1998; Matsuoka et al., 2014).

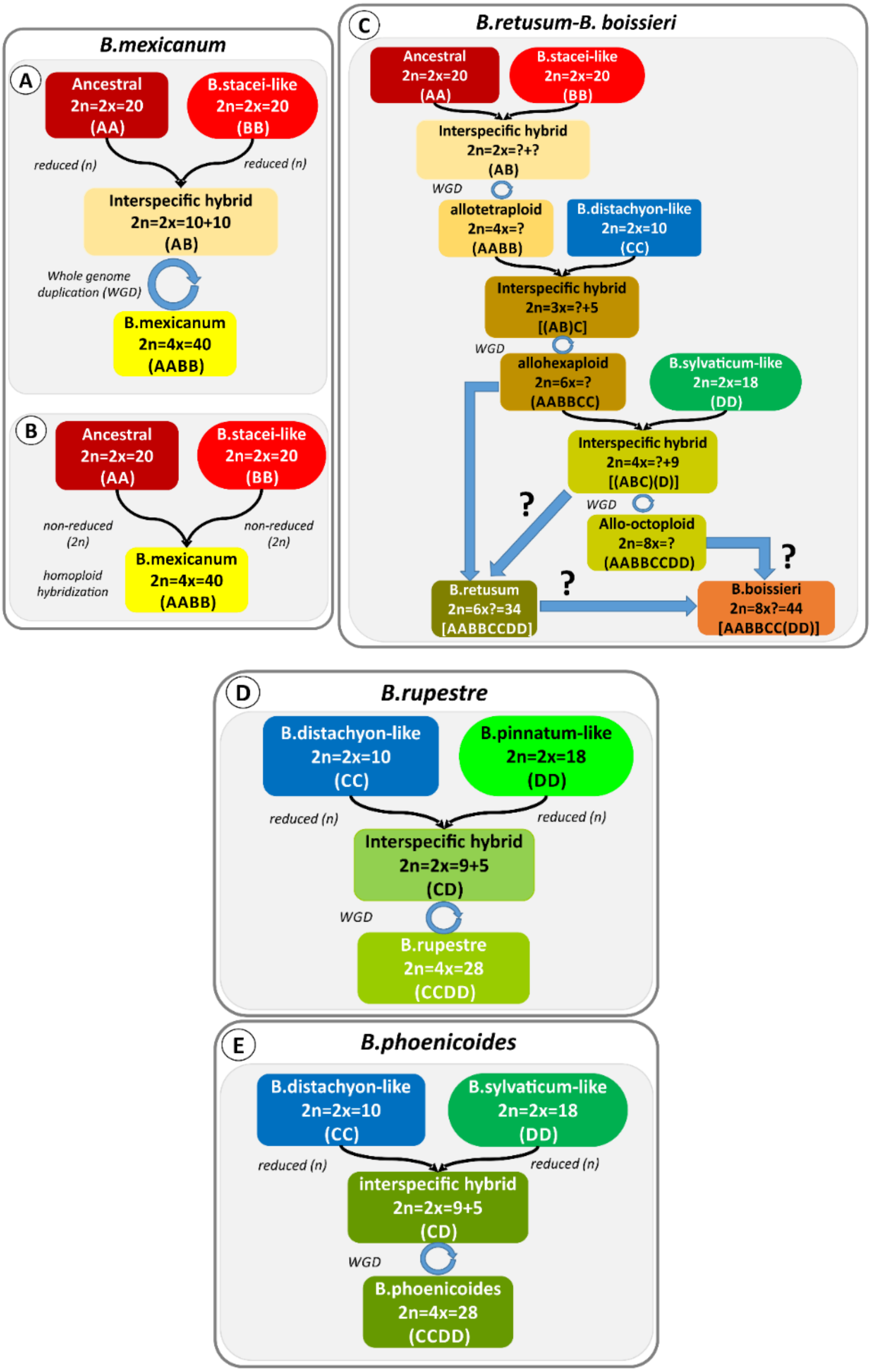


Figure 7. Diagrams representing putative origins for the studied *Brachypodium* allopolyploid lineages based on RNA-seq phylogenomic data and GS data. **(A)** and **(B)** *B. mexicanum*; **(C)** *B. retusum*; *B. boissieri*; **(D)** *B. rupestre*; **(E)** *B. phoenicoides.*

These gametes could cross with reduced (n) parental gametes to form fertile segmental allotetraploid individuals or infertile descendants that would become fertile allo-octoploids after WGD. Of the three mechanisms, the interspecific hybridization of heterologous genomes followed by WGD (IH+WDG) emerges as the commonest route of allopolyploid synthesis in grasses for both paleo and neo-allopolyploids (Murat et al., 2010; Kellogg, 2015b), though the existence of segmental allopolyploidy has been also proposed for some lineages (e. g., maize, Gaut and Doebley 1997); bread wheat, Marcussen et al. 2015). As illustrated by the model *B. hybridum* species, the artificial creation of a fertile synthetic allotetraploid could only be accomplished via WGD of the unfertile interspecific F1 hybrid (Dinh Thi et al., 2016), paralleling what is assumed to have occurred in nature (Catalán et al., 2016b). The possession of largely divergent heterologous and dysploid parental genomes by the majority of the studied *Brachypodium* allopolyploids (fig.5; (Catalán et al., 2016b) lends support to the preferential IH+WDG hypothesis to explain their origins (fig.7).

Our nuclear gene tree points to *B. mexicanum* as the most ancestral extant allopolyploid *Brachypodium* species, resulting from the merging of the two oldest ancestral (A) and stacei-like (B) subgenomes (fig. 5, 7a, b), that probably occurred from the Mid-late Miocene onwards (12.4 Ma for the split of its recenmost B genome, fig. 6a). It is also supported by the more constrained RNA-seq subgenomic tree (fig. 4). The GS value obtained for *B. mexicanum* (3.774 pg/2C; table 1) indicates a weighted genome, that considerably exceeds the small genome sizes of most *Brachypodium* species (Betekhtin et al., 2014; Catalán et al., 2016b), and approaches those of other cool-season grasses (Plant DNA C-values Database; http://data.kew.org/cvalues/). Though first considered to be an octoploid with a putative chromosome base number of x=5 (Shi et al., 1993), the possession of ancestral genomes (A and B) with putative ancestral x=10 chromosome base numbers suggests that *B. mexicanum* would be an allotetraploid (fig. 7, table 1; Díaz-Pérez et al. 2018). Our plastome gene tree indicates that its stacei-like B parent was the maternal genome donor (fig. 6b; S7) of the studied *B. mexicanum* line. According to this hypothesis, two alternative evolutionary scenarios could be inferred for the origin of *B. mexicanum*: the IH+WGD scenario (fig. 7a) would require the merging of reduced heterologous A and B genomes followed by genome doubling, whereas the segmental allopolyploidy scenario (fig. 7b) would demand the merging of unreduced AA and BB genomes via homoploid hybridization. The large GS of *B.*

mexicanum, not found in other species with the A or B subgenomes, could have been acquired after its genome doubling (IH+WGD scenario). The ongoing sequencing of the *B. mexicanum* genome would help to clarify the genomic composition and origin of this species (Des Marais et al. unpub. data).

The reconstruction of the origins of the high-ploidy level *B. boissieri* and *B. retusum* allopolyploids is more complex due to the apparent incongruence between their inferred chromosome numbers (table 1) and potential subgenomes (fig. 4, 5). These two phenotypically close Mediterranean species show large sets of phylogenetically divergent heterologous subgenomes. The nuclear gene tree identifies up to three (A, B, C) and four (A, B, C, D) subgenomes in *B. boissieri* and *B. retusum*, respectively (fig. 5), and a similar resolution but with less subgenomes in the SNP subgenomic tree (fig. 4). The estimated GS value of *B. boissieri* (3.236 pg/2C) is the second highest value found in the genus and fits a chromosome number of near 2n=46, whereas that of *B. retusum* (1.704 pg/2C) fits a 2n=32, corroborating one of the chromosome values reported by (Schippmann, 1991) for this species (table 1). Our predicted chromosome numbers suggest that *B. boissieri* could be an allo-octoploid and *B. retusum* an allohexaploid but with reduced genome size. It is intriguing, however, the retrieval of more subgenomes in *B. retusum* than in *B. boissieri* in both the nuclear core gene tree and the syntenic subgenomic tree (fig. 4, 5). A scarce number of recently evolved D subgenome genes are also expressed in *B. boissieri* (table S8, fig. S6), suggesting a potential biased reduced expression of genes from this subgenome, their potential lost or pseudogenization (Panchy et al., 2016), or diverging evolution from older subgenomes. The two species share similar ancestral A and B and intermediately evolved distachyon-type C subgenomes, indicating a possible common ancestry. A succession of two consecutive IH+WDGs resulting in a putative AABBCC allohexaploid (fig. 7c) represents the most likely scenario for the origin of their most recent ancestor. A further IH+WDG involving this ancestor and a recently evolved sylvaticum-type D genome would have originated a putative AABBCCDD allo-octoploid (fig. 7c). Conciliating the putative chromosome base numbers of the supposedly dysploid (A, B, x=10), C (x=5) and D (x=9) subgenomes and the predicted chromosome numbers of the studied *B. boissieri* and *B. retusum* accessions requires assuming the existence of different chromosome fusions in both species and/or ample genomic losses in the studied *B. retusum* accession. Our plastome gene tree has identified the distachyon-type C subgenome and the core perennial-type D subgenome as the respective maternal genome donors of the studied *B. boissieri* and *B. retusum* accessions (fig. 5b, S7). The current evidences suggest a more recent Quaternary origin (from 2.2 Ma on, the split of its D subgenome) of *B. retusum* and an older Pliocene origin (from 4.9 Ma on, the split of its C subgenome) of *B. boissieri* (fig. 6); however, the elucidation of their respective evolutionary scenarios remains still elusive.

Deciphering the origins of the recently evolved core perennial clade allotetraploids *B. phoenicoides* and *B. rupestre*-4x was more straightforward due to the matching of their detected subgenomes and the estimated GS and chromosome number values (figs. 5a, table 1). Similar subgenomes C and D were detected in *B. rupestre* and in the two analysed *B. phoenicoides* samples (Bpho6, B422) in the nuclear gene tree (fig. 5a), which correspond to the same homeologous genomes detected in the SNP subgenomic tree (fig. 4). Noticeably, the distachyon-type C subgenomes found in *B. rupestre* and *B. phoenicoides* are more recent (5.4 Ma, stem node, 3.9 Ma crown node, fig. 6) than those found in *B. boissieri* and *B. retusum* (6.9 Ma and 4.9 Ma, respectively, fig. 6) and are sister to the core perennial clade (4.6 Ma, crown node, fig. 6), suggesting than these parental C subgenomes could be more core-type than distachyon-type. The GS values obtained for the tetraploids *B. rupestre*-4x (1.469 pg/2C) and *B. phoenicoides* (1.443 pg/2C Bpho6; 1.469 pg/2C, Bpho_B422) (table 1) are similar to those recoded for *B. phoenicoides* by other authors (Wolny & Hasterok, 2009). They fit a chromosome number of 2n=28, corroborating previous findings. GS values of core perennial diploid *B. pinnatum*-2x (0.822 pg/2C) and of tetraploid *B. pinnatum*-4x (1.499 pg/2C) (only studied with GBS data) fit chromosome numbers of 2n=18 and 2n=28, respectively (table 1), agreeing also with previous records (Wolny & Hasterok, 2009). Our plastome gene tree identifies *B. pinnatum*-2x as the maternal parent of *B. rupestre*-4x and two alternative maternal parents for *B. phoenicoides*, *B. sylvaticum* (phylogenetically close to Bpho6) and *B. pinnatum*-2x (phylogenetically close to Bpho_B422) (figs. 5b, S7). Our dated chronogram indicates that the two allotetraploids originated very recently in the Quaternary (from 1.8 Ma on, crown node of their D subgenomes; from 1.0 Ma on, crown node of the *B. phoenicoides* D subgenomes) (fig. 6). According to the above evidences, two similar IH+WGD evolutionary scenarios are proposed for the respective origins of *B. rupestre* (fig. 7d) and *B. phoenicoides* (fig. 7e). The two species share a close paternal C subgenome (phylogenetically divergent from current *B. distachyon* genome) and also close but probably distinct D maternal subgenomes; *B. pinnatum*-2x emerges as the maternal parent of the studied *B. rupestre*-4x and *B. sylvaticum* as the potential maternal parent of some but not all the studied *B. phoenicoides* accessions (fig. 6). Alternatively, *B. rupestre*-4x and *B. phoenicoides* could also have originated though heteroploid hybridizations of their respective non-reduced CC and DD genomes, but this scenario seems less likely.

Allopolyploid evolution in *Brachypodium* fits the Darlington's rule, which proposes that allopolyploids should form between reproductively isolated species rather than between reproductively compatible diploids that tend to form homoploid hybrids (Darlington, 1937; Bombarely et al., 2014). The large chromosomal differences observed in its dysploid series (fig. 3a) supports this model, which is further corroborated through the production of the synthetic *B. hybridum* allotetraploid from largely divergent parental species (Dinh Thi et al., 2016). Nonetheless, closely related core perennial species could cross and produce fertile descendants (Khan & Stace, 1999), suggesting that homo- or heteroploid hybridizations could also be ongoing evolutionary drivers of current diversification within *Brachypodium*. Recent cytogenetic studies based on *B. distachyon*-type Bd2 and Bd3 chromosomal karyotypes (Idziak et al., 2014) exclude *B. distachyon* as potential parent of Eurasian core perennial allopolyploids, whereas others based on centromeric composition and structure of annual and perennial *Brachypodium* species (Li et al., 2018) have featured two main types of centromeres, proposing a phylogeny that links *B. stacei* with *B. pinnatum*-2x and *B. pinnatum*-4x and *B. distachyon* with *B. sylvaticum* and *B. phoenicoides*. Our results contradict both proposals. The syntenic evolutionary framework identifies the series of nested chromosome fusions and additional reorganizations experienced by the Bs, Bd and Bsy genomes and reconstruct their evolution (figs. 1c, 3a); the extreme chromosomal reduction experienced by *B. distachyon* apparently occurred independently in its lineage, though its genome is highly collinear with the *B. stacei* and *B. sylvaticum* genomes (fig 1c). Our evolutionary scenarios support the participation of a distachyon-type C genome (although divergent from the current *B. distachyon* genome) in the core perennial allopolyploids *B. rupestre* and *B. phoenicoides* (figs. 4-7). Our robust diploid backbone tree, based on a large number of RNA-seq SNPs and 397 core expressed genes, covering all chromosomes of the studied *Brachypodium* species, *Hordeum* and *Oryza*, reconstruct the early successive divergences of *B. stacei* and *B.*

distachyon, and then those the close and most recently evolved *B. arbuscula* and the sister *B. sylvaticum* and *B. pinnatum*-2x (figs.3-6). Transposon rich centromeric regions could be prone to highly dynamic burst and extinctions (Feschotte & Pritham, 2007), thus being incongruous markers for phylogenetic reconstruction. On the other hand, our solid subgenomic and gene based phylogenies (figs. 4-6) could be used as suitable evolutionary frameworks to map karyotypic changes on them (Acosta et al., 2015; Baltisberger & Hörandl, 2016).

A *Brachypodium* pan-transcriptome draft: insights into the evolution of private core gene groups

Recently published pan-genomes of model plant species have revealed the substantial genomic diversity of populations that contain genes outnumbering those found in any single individual, as demonstrated in the analysis of the pan-genome of the flagship *B. distachyon* species (Gordon et al., 2017) and of the *Oryza sativa*-*O. rufipogon* species complex (Zhao et al., 2018). The pan-genome differentiates core and accessory genes according to their complete or incomplete presence across genotypes and their potential implication on mostly essential or conditionally beneficial functions, respectively (Gordon et al., 2017). Here, we have extended this approach at supra-specific level and have constructed a pan-transcriptome draft for *Brachypodium*, using 5,202 expressed gene clusters found in the pooled RNA-seq libraries of the studied *Brachypodium* samples (table S9). Whereas highly conserved core genes are suitable for phylogenetic reconstruction, as used in this study (figs. 5, 6), presence/absence of *Brachypodium* accessory genes in specific *Brachypodium* groups may draw further insights into the evolution of these taxa and their genomes and their functions. Differences in gene content in plant polyploids could be due to neofunctionalization, subfunctionalization, paralogue interference, subgenome dominance or fractionation (gene loss) of the duplicated genes present in more than one subgenome (Cheng et al., 2018). However, gene content differences have also been found among congeneric diploid species (Zhao et al., 2018) and among accessions of the same diploid species, like the accessory genes of *B. distachyon* (Gordon et al., 2017). Accessory genes that usually display faster evolutionary rates than core genes and that contribute to phenotypic and potentially adaptive variation, become core for certain biological, evolutionary or ecological groups. Here we have extended our pan-transcriptome survey across the phylogeny of *Brachypodium*.

Average Nucleotide Identity hierarchical clustering of core transcripts separates two clear evolutionary groups; group ANI1 comprising the more ancestral *B. mexicanum*, *B. boissieri* and annual lineages, and group ANI2 including *B. retusum* and the core perennial lineages (fig. S8). Inspection of the pan-transcriptome discovered 14 gene clusters private to group ANI1 and 52 gene clusters private to group ANI2. Among the former a MYB transcription factor was identified which is not expressed in the core perennial species; the latter include a transporter protein which has been associated to secondary cell wall formation in *Arabidopsis thaliana* (Ranocha et al., 2010).

Further interrogation of the pan-transcriptome detected the highest number of private gene clusters (143) present in *B. mexicanum*, *B. boissieri* and *B. retusum* (table S9f). Their exclusive presence in these old allopolyploids suggests that they could have been inherited via the ancestral A subgenomes (fig. 5). Most of these gene clusters appear to encode proteins involved in general metabolic/physiological processes with no clear functional enrichment. The second group in gene content differences is that of perennials and annuals (tables S9c, S9d). Among the 30 gene clusters private to *Brachypodium* perennials a GLABRA homeobox leucine zipper stands out, as similar proteins have been shown to control floral identity in *A. thaliana* (Kamata et al., 2013). Among the 49 private to annuals there are interesting candidate genes coding for a protein containing a CCCH Zn finger domain, considered to be involved in processing mRNA in developmental processes (Peng et al., 2012) and a DNA methylation factor. Enrichment analyses of transcript clusters private to annual *Brachypodium* species recovered the transcription of a RNA polymerase I as a statistically significant biological process. This group also included an annotated RNA polymerase beta subunit (table S9d). DNA-directed RNA polymerases (RNAPs) were related to lineage-specific duplication in plant families. The number of genes encoding for RNAP subunits are relatively constant in animals, fungi and algae; however they vary in land plants, showing independent duplications and diversification events in different lineages (Wang & Ma, 2015). The annotated RNA polymerase and the transcription of RNA polymerase I process found only in annual species of *Brachypodium* could indicate differences between copies or expression levels of RNAPs between annuals and perennial species, or the loss of retained ancestral copies in more recently evolved perennial species.

A few number of private expressed gene clusters were exclusive of polyploids (14) and none of diploids (table S9e). This finding corroborates that virtually all core genes present in the diploid genomes are also present in the diploid-like subgenomes (A to D) inherited by the hybrid allopolyploids (fig. 5), whereas only a few number of gene clusters may have arisen by polyploidy per se. Our *Brachypodium* pan-transcriptome draft of pooled leaf transcripts under hydric, salt, temperature stresses and control treatments has shed some light on the differentially expressed gene contents across lineages, life-cycle and ploidy-level groups. The evolutionary fate of *Brachypodium* genomes and genes should be accomplished within a broad *Brachypodium* gene atlas framework.

Chapter 3. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes

Summary

Few pan-genomic studies have been conducted in plants, and none of them have focused on the intra-specific diversity and evolution of their plastid genomes. We address this issue in *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum*, for which a large genomic data set has been compiled. We analyze inter- and intra-specific plastid comparative genomics and phylogenomic relationships within a family-wide framework.

Major indel differences were detected between *Brachypodium* plastomes. Within *B. distachyon*, we detected two main lineages, an majoritarily Extremely Delayed Flowering (EDF+) clade and a majoritarily Spanish (S+) –Turkish (T+) clade, plus nine chloroplast capture and two plastid DNA (ptDNA) introgression and micro-recombination events. Early Oligocene (30.9 millions of years ago (Ma)) and Late Miocene (10.1 Ma) divergence times were inferred for the respective stem and crown nodes of *Brachypodium* and a very recent Mid-Pleistocene (0.9 Ma) time for the *B. distachyon* split.

Flowering time variation is a main factor driving rapid intra-specific divergence in *B. distachyon*, though it is counterbalanced by repeated introgression between previously isolated lineages. Swapping of plastomes between the three different genomic groups, EDF+, T+, S+, likely resulted from random backcrossing followed by stabilization through selection pressure.

Introduction

Plastid DNA (ptDNA) has been widely used in inter- and intra-specific phylogenetic analyses in multiple species and populations of plants (Waters et al., 2012; Ma et al., 2014; Middleton et al., 2014; Wysocki et al., 2015). Phylogenetic dating of monocots and eudicots has also been based on ptDNA (Chaw et al., 2004). Comparative genomics of whole plastid genomes has provided a way to detect and investigate genetic variation across seed plants (Jansen & Ruhlman, 2012). The proliferation of Whole Genome Sequencing (WGS), which typically includes a substantial amount of plastid sequence, has provided large data sets which can be utilized to assemble and analyze plastomes (Nock et al., 2011).

Brachypodium is a small genus in the family Poaceae that contains approximately 20 species (17 perennial and 3 annual) distributed worldwide (Schippmann, 1991; Catalán & Olmstead, 2000; Catalán et al., 2012, 2016a,b). The three annuals include two diploids [*B. distachyon* (2n=2x=10; x=5), *B. stacei* (2n=2x=20; x=10)] and their derived allotetraploid [*B. hybridum* (2n=4x=30; x=5+10)]. These three species had previously been considered cytotypes of *B. distachyon* (Catalán et al., 2012). In addition to the large, overlapping distribution in their native circum-Mediterranean region (Catalán et al., 2012, 2016a; López-Alvarez et al., 2012; Lopez-Alvarez et al., 2015), *B. hybridum* has naturalized extensively around the world.

The evolutionary relationship between *Brachypodium* and other grasses has been thoroughly studied (Catalán et al., 1997; Catalán & Olmstead, 2000; Döring et al., 2007). Most recent phylogenetic analyses place *Brachypodium* in an intermediate position within the Pooideae clade (Minaya et al., 2015; Soreng et al., 2015; Catalán et al., 2016a,b). By contrast, only a few studies of intra-specific variation have been conducted in the genus *Brachypodium*, primarily focusing on *B. distachyon* (e. g., Filiz et al., 2009; Vogel et al., 2009; Mur et al., 2011; Tyler et al., 2016).

Brachypodium distachyon has been selected as a model plant for temperate cereals and biofuel grasses (IBI, 2010; Mur et al., 2011; Catalán et al., 2014; Vogel, 2016). Additionally, the B. distachyon complex has been proposed as a model system for grass polyploid speciation (Catalán et al., 2014; Dinh Thi et al., 2016). Nuclear and plastid genomes of the Bd21 ecotype of B. distachyon have been sequenced, assembled and annotated. The nuclear genome is 272 Mbp in size (IBI, 2010) and contains 31,694 protein-coding loci. The current plastid genome reference (NC_011032.1) is 135,199 base pairs (bp) long and encodes 133 genes (Bortiri et al., 2008).

In parallel with the creation of the nuclear pan-genome of *B. distachyon* from 53 diverse lines (Gordon et al., 2017), and the genome sequencing of its close congeners *B. stacei* and *B. hybridum* (*Brachypodium stace*i v1.1 DOE-JGI, http://phytozome.jgi.doe.gov/ and *B. hybridum* early access available through Phytozome), we isolated ptDNA sequences from WGS paired-end reads to assemble the corresponding plastomes. Our aim was to compile a large plastome data set and investigate the evolutionary relationships of the annual *Brachypodium* species within the grass phylogenetic framework. The specific objectives of this study were to: (1) assemble, annotate and compare 57 plastomes of *B. distachyon*, *B. stacei* and *B. hybridum*; (2) reconstruct and date the divergences within the *Brachypodium* lineages and a family-wide plastome phylogeny, (3) infer the genealogical relationships within the studied accessions of *B. distachyon* and compare them with the nuclear genome genealogy, and (4) investigate the potential existence of plastid introgression and recombination in *B. distachyon* ecotypes known to hold nuclear introgressions.

Materials and Methods

Plant materials

Brachypodium distachyon, B. stacei and B. hybridum ecotypes used in this work are inbred lines derived from our own collections (Vogel et al., 2009; Mur et al., 2011; Catalán et al., 2012) and from the National Plant Germplasm System (NPGS) and Brachyomics collections (USDA and ABER lines; Vogel *et al.*, 2006; Garvin, 2007; Garvin *et al.*, 2008). Most ecotypes were originally collected in Spain, Turkey and Iraq (Table S1, Fig. 1) (Vogel & Hill, 2008; Filiz et al., 2009; Mur et al., 2011). Available plastome data from the main grass lineages were retrieved from GenBank (Table S2). Flowering time data were obtained from (Gordon et al., 2017). Briefly, flowering time was measured as the number of days elapsed from the end of vernalization to inflorescence heading, in the growth chamber, and assigned to flowering time classes following Ream *et al*. (2014, see Table S3).

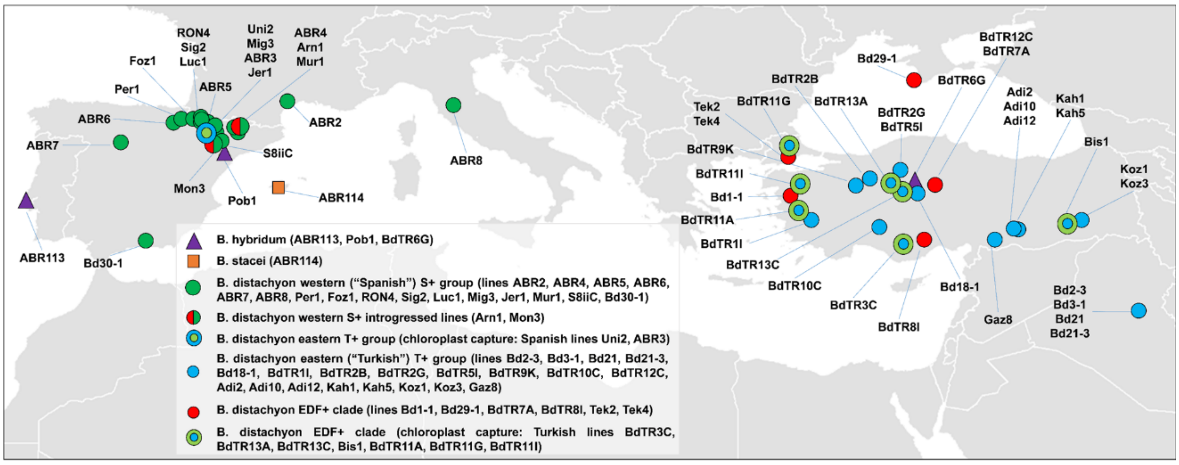

RON4
Sig2
Luc1
Uni2
Mig3
ABR3
Jer1
ABR4
Arn1
Mur1
Foz1
Per1
ABR5
ABR6
ABR7
ABR2
S8iiC
ABR8
Mon3
Pob1
ABR114
ABR113
Bd30-1
▲ B. hybridum (ABR113, Pob1, BdTR6G)
■ B. stacei (ABR114)
● B. distachyon western ("Spanish") S+ group (lines ABR2, ABR4, ABR5, ABR6, ABR7, ABR8, Per1, Foz1, RON4, Sig2, Luc1, Mig3, Jer1, Mur1, S8iiC, Bd30-1)
◐ B. distachyon western S+ introgressed lines (Arn1, Mon3)
◎ B. distachyon eastern T+ group (chloroplast capture: Spanish lines Uni2, ABR3)
● B. distachyon eastern ("Turkish") T+ group (lines Bd2-3, Bd3-1, Bd21, Bd21-3, Bd18-1, BdTR1I, BdTR2B, BdTR2G, BdTR5I, BdTR9K, BdTR10C, BdTR12C, Adi2, Adi10, Adi12, Kah1, Kah5, Koz1, Koz3, Gaz8)
● B. distachyon EDF+ clade (lines Bd1-1, Bd29-1, BdTR7A, BdTR8I, Tek2, Tek4)
◎ B. distachyon EDF+ clade (chloroplast capture: Turkish lines BdTR3C, BdTR13A, BdTR13C, Bis1, BdTR11A, BdTR11G, BdTR11I)
BdTR12C
BdTR7A
Bd29-1
BdTR2B
BdTR6G
Tek2
Tek4
BdTR11G
BdTR13A
BdTR2G
BdTR5I
Adi2
Adi10
Adi12
Kah1
Kah5
Bis1
BdTR9K
BdTR11I
Bd1-1
BdTR11A
BdTR1I
BdTR13C
BdTR10C
BdTR3C
BdTR8I
Bd18-1
Gaz8
Koz1
Koz3
Bd2-3
Bd3-1
Bd21
Bd21-3

Figure 1. Native circum-Mediterranean geographic distributions of the *B. distachyon*, *B. hybridum* and *B. stacei* ecotypes used in the plastome evolutionary and genomic analyses. Symbol and color codes for accessions are indicated in the chart. Accession numbers correspond to those indicated in Table S1.

Plastid DNA automated assembly, annotation and validation

Illumina paired-end and mate-pair libraries from 53 *B. distachyon*, 1 *B. stacei* and 3 *B. hybridum* accessions were produced from total genomic DNA, isolated as described previously (Peterson et al., 2000), randomly sheared, and filtered to target fragments sizes of 250 bp and 4 kbp, using Covaris LE220 (Covaris) and HydroShear (Genomic Solutions), respectively. The KAPA-Illumina library creation (KAPA Biosystems) and TruSeq v3 paired-end cluster kits were used for library construction. Sequencing was performed at the Joint Genome Institute on the Illumina HiSeq2000 sequencer, yielding reads of 76, 100 and 150 bp length.

We developed a pipeline, available at https://github.com/eead-csic-compbio/chloroplast_assembly_protocol, for the assembly and annotation of plastid genomes (Methods S1, Table S4, Fig. S1). Briefly, plastid reads were extracted from WGS data using DUK (http://duk.sourceforge.net), followed by quality control and error correction, with FastQC v.0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), Trimmomatic v.0.32 (Bolger *et al*., 2014) and Musket v.1.0.6 (Liu et al., 2013). Then, pass-filtered reads were assembled with Velvet v.1.2.07 (Zerbino, 2010), SSpace Basic v.2.0 (Boetzer et al., 2011), and GapFiller v.1.11 (Boetzer and Pirovano 2012; Nadalin *et al*., 2012).

This pipeline can be used to perform both *de novo* and reference-guided assemblies. Both strategies were performed with 55 out of 57 accessions; in most cases (46, see

Table S5) the reference-guided approach produced fewer and longer contigs than *de novo* assemblies. Other parameters affecting assembly outcome were optimized, such as k-mer size or the number of input reads. Assembly errors were corrected with SEQuel v.1.0.2 (Ronen et al., 2012), and by visual inspection of read mappings using IGV v2.3.8 (Thorvaldsdóttir *et al*., 2013).

Gene annotation was performed exhaustively for a single plastome of each species, and then transferred with custom scripts to the remaining plastid assemblies. The ptDNA genomes were compared with Organellar-Genome DRAW web version (Lohse et al., 2013) and Circos v.0.69 (Krzywinski et al., 2009). Typical plant plastomes show four main regions: large single-copy (LSC), first inverted-repeat (IRa), short single-copy (SSC), and second inverted-repeat (IRb), as sorted in the current Bd21 accession (NC_011032.1). Junctions between IR-LSC, LSC-IR, IR-SSC and SSC-IR regions, as well as main structural variations of *B. stacei* and *B. hybridum* plastomes were confirmed by polymerase chain reaction (PCR) amplification and Sanger sequencing (Table S6). The annotated plastomes of *B. distachyon*, *B. stacei* and *B. hybridum* ecotypes were deposited at ENA (European Nucleotide Archive) with accession numbers LT222229 - 30 and LT558582-LT558636.

Intra-specific genealogy, haplotypic network, and genomic diversity and structure analyses

Plastomes from the 53 *B. distachyon* accessions (Table S1) were aligned using MAFFT v.7.031b (Katoh & Standley, 2013); poorly aligned regions were removed with trimAl v.1.2rev59 (Capella-Gutiérrez *et al*., 2009) using option automated1, which excludes columns after heuristically computing appropriate gap and similarity thresholds. However, most robust gaps were included in the final aligned data set and used in the phylogenetic Maximum-Likelihood (ML), Bayesian inference (BI) and dating Bayesian evolutionary analysis (BEAST) approaches. The second inverted repeat region (IRb) accumulated most ambiguous nucleotides in our assemblies, probably due to biases in the pipeline (see histogram in Fig. 2). Considering that both repeats are essentially redundant in plastids, only IRa was included in subsequent phylogenetic analyses (Nock et al., 2011; Middleton et al., 2014; Saarela et al., 2015). Alignments were revised and manually curated using Geneious v.8.1.4 (Kearse et al., 2012).

Maximum-Likelihood (ML) and Bayesian inference (BI) phylogenomic analyses were performed with RAxML v.8.1.17 (Stamatakis, 2014) and MrBayes v.3.2.4 (Ronquist *et al*., 2011; Ronquist and Huelsenbeck 2003), respectively. The generalized time-reversible plus gamma distribution plus proportion of invariant sitessubstitution model (GTR+G+I), selected by JModelTest v.2.1.7 based on the Akaike Information Criterion (Guindon & Gascuel, 2003; Darriba et al., 2012), was imposed in the searches. In the ML search we computed 20 starting trees from 20 distinct randomized Maximum Parsimony (MP) trees and 1000 bootstrap replicates. In the BI search, two sets of four chains were run for 2 million generations, sampling trees and parameters every 100th generation. A 50% majority rule consensus tree was computed discarding the first 25% saved trees as 'burn-in'. All trees were mid-point rooted.

Haplotypic network analysis was conducted with the 53 *B. distachyon* plastome alignment after removing IRb and columns with missing data (Ns), both including and excluding indels. Statistic parsimony analysis was performed with TCS v1.21 (Clement et al., 2000), setting a maximum connection of 1000 steps. Haplotype polymorphism and genetic diversity statistics of the plastome data set, such as the number of segregating sites (S) and haplotypes (h), the haplotype diversity index (Hd), and the number of shared mutations (shm) and the average number of nucleotide differences (d) among the three intra-specific genetic groups retrieved from the phylogenomic analysis (see Results) were calculated with DnaSP v.5 (Librado & Rozas, 2009).

Bayesian genomic clustering analysis was performed to infer the structure of the data, using a *B. distachyon* ptDNA data matrix of 298 mapped polymorphic positions, and to assign accessions' plastomes to the inferred groups using Structure v.2.3.4 (Pritchard et al., 2000). The program was run for a number of potential genomic groups (K) from 1 to 6, imposing ancestral admixture and correlated allele frequencies priors. Ten independent runs with 100,000 burn-in steps, followed by 1,000,000 generations were computed for each *K*. The number of genetic clusters was estimated using Structure Harvester (Earl & vonHoldt, 2012), which identifies the optimal *K* based both on the posterior probability of the data for a given K and the ΔK (Evanno et al., 2005). The potential existence of inter-plastome recombination in two introgressed ecotypes (see Results) was further assessed through visual inspection of the mapped polymorphic alignments and through the recombination detection methods implemented in RDP4 v.4.56 (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, LARD, 3SEQ (Martin et al., 2015) and in OrgConv v.1.1 (Hao 2010), using default settings in all cases.

Phylogenetic and molecular dating analyses

A grass plastome alignment was built including all *B. distachyon,* one *B. stacei* and one *B. hybridum* ecotypes (55 accessions; Table S1) plus the plastomes of 90 grasses (Table S2). ML analysis was performed with RAxML following the same steps indicated above. Pairwise Tamura-Nei (TN) raw genetic distances and pairwise TN patristic (RAxML-tree) distances were computed between all pairs of grass entries using MEGA v.7.0.14 (Kumar et al., 2016) and Geneious (Kearse et al., 2012), respectively.

Divergence time estimations of the *Brachypodium* lineages were calculated within a family-wide dated phylogeny using a Bayesian nested dating partitioned approach (Pokorny et al., 2011; Mairal et al., 2015) in BEAST v1.8.2 (Drummond et al., 2012). Because there are no known fossil records of *Brachypodium*, a high-level more inclusive grass data set (93 samples = 90 grass species + 1 *B. distachyon* + 1 *B. stacei* + 1 *B. hybridum* accessions, 110,370 bp length, 22,489 polymorphic positions) was used to estimate divergence times within the *B. distachyon* ingroup (53 samples, 110,370 bp length, 415 polymorphic positions). The grass tree was rooted with the ancestral species *Anomochloa marantoidea*. The estimated ages were drawn from deep-time calibrations imposed in the Poaceae partition and were used to constrain the molecular clock rate of the linked *B. distachyon* population-level data set and to calibrate the divergence time of its crown node. We estimated divergence times among the Poaceae lineage imposing GTR+G+I, lognormal relaxed clock and Yule tree models, a broad uniform distribution prior for uncorrelated lognormal distribution (ucld) mean (lower =1.0E-6; upper = 0.1) and a default exponential prior for ucld standard deviation. Calibrations were drawn from the compilation of grass fossils of Strömberg (2011) and from fossil-rich dating analyses of the grass family (Bouchenak-Khelladi *et al.*, 2010; Christin *et al.*, 2014). In order to accommodate uncertainties in the fossil records and fossil-based calibrations, we incorporated into the divergence time analysis normal distribution priors with mean and standard deviation values of the normal distribution set for upper and lower dates of the geological period of the fossil, or the estimated divergence ages of the calibrated tree node, representing 5% and 95% quantiles of the distribution**.** We used two calibration points, imposing secondary age constrains for the crown nodes of Poaceae (normal prior mean = 90.0 Ma, SD = 1.0) and of the BOP (Bambusoideae, Oryzoideae, Pooideae) + PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae, Danthonioideae) clade (normal prior mean = 55.0 Ma, SD = 0.5), covering the age ranges of their respective fossil records and nodal age estimates. For the intra-specific *B. distachyon* data set we imposed a coalescent constant-size tree model. We ran 1,000,000,000 Markov Chain Monte Carlo (MCMC) generations in BEAST with a sampling frequency of 1,000 generations after a burn-in period of 1%. The adequacy of parameters was checked using Tracer v1.6 (http://beast.bio.ed.ac.uk/Tracer), noting effective sample size (ESS) values > 200. Maximum clade credibility (MCC) trees were computed for the Poaceae and for the *B. distachyon* data sets after discarding 1% of the respective saved trees as burn-in.

Results

Structure, gene content and sequence in *B. distachyon*, *B. stacei* and *B. hybridum* plastomes

Assemblies were obtained for 57 plastomes. Forty-one contained ≤ 10 contigs, with an average longest contig length of 84 kbp and 176x depth coverage (Table S5). After scaffolding, 45 assemblies had ≤ 4 scaffolds with a mean plastome length of 124.5 kbp. Missing data ranged from 0 to 6%, with most plastomes (38) showing ≤ 0.1%. Most of the missing sequence was located in the IRb region which was difficult to assemble because of its redundancy. The resulting *Brachypodium* plastomes were highly conserved in terms of synteny and gene number. Plastome lengths varied from 134,991 to 135,214 bp in *B. distachyon*, and between 136,326 and 136,330 bp in *B. stacei* and *B. hybridum* (Table S5).

Reference accession *B. distachyon* Bd21 (NC_011032.1; Bortiri *et al.*, 2008; 2010 – direct submission) and the *B. distachyon* Bd21 control (Bd21C, assembled and annotated in the current study) showed some differences [10 single nucleotide polymorphisms (SNPs) and 19 indels; Table S7a)]. These polymorphisms had read depth coverage ranging from 219 – 16,750 and were also confirmed in several of the other *B. distachyon* accessions (see Table S7a). While most of these polymorphisms lay in intergenic regions, some were located in protein coding genes such as *psb*A (1 synonymous (Syn) mutation), *psb*K (1 non-synonymous (NSyn) mutation), *rpo*C2 (1 Syn and 1 NSyn), *psa*A (1 Syn), and also in one copy of the rRNA 16S locus.

Brachypodium distachyon plastomes showed the same gene arrangement and number (133) as Bd21C (Table S7a, b). In particular, they contained 76 protein coding genes, 7 of which were duplicated genes, 20 non-redundant tRNAs (out of a total 38), 4 rRNAs in both inverted repeats, 4 pseudogenes (*trn*I, *rps*12a, *trn*T and *trn*I) and 2 hypothetical open reading frames (*ycf)*. Several polymorphisms, mostly non-synonymous, were detected in comparison to several grass plastomes. The most polymorphic loci were *rpo*C2 (70 SNPs), *ndh*F (59 SNPs), *rpo*B (31 SNPs) and *mat*K (30 SNPs), suggesting a significant correlation between SNP frequency and gene length ($R^2 = 0.68$, $p < 2.2e\text{-}16$; Table S7b).

Brachypodium stacei and *B. hybridum* accessions showed the same overall plastid genomic features as the *B. distachyon* accessions, with two exceptions (Fig. 2). They both contained a 1,161bp insertion between *psa*I and *rbc*L in the Large Single-Copy (LSC) region. This insertion was confirmed by read mapping (Fig. S2a, b), and it was also detected in homologous regions of several grasses (Table S7c). It corresponds to a coding sequence (CDS) fragment annotated as pseudogene *rpl*23 (Table S7d).

The *B. stacei* and *B. hybridum* plastomes also contained a deletion of an *rps*19 copy between *psb*A and *trn*H in the IRb repeat, which was confirmed through PCR amplification and Sanger sequencing (Fig. S2c; Methods S1). The presence of these indels in the plastid genomes of the three *B. hybridum* accessions suggests that they were inherited from *B. stacei*-type maternal parents. Six polymorphisms were detected between the *B. hybridum* and *B. stacei* plastomes (Table S7e). These polymorphisms were located in intergenic regions, except for a Syn substitution in *psb*T (ecotype BdTR6G, *B. hybridum*) and a NSyn mutation in one copy of *rpl*23 (ecotype ABR113, *B. hybridum*).

Furthermore, a conceptual RNA-edited translation (U to C) was inferred in the ndhB gene of all the *B. hybridum* accessions and *B. stacei*, as well as in the ndhK gene of the *B. distachyon* Gaz8 accession.

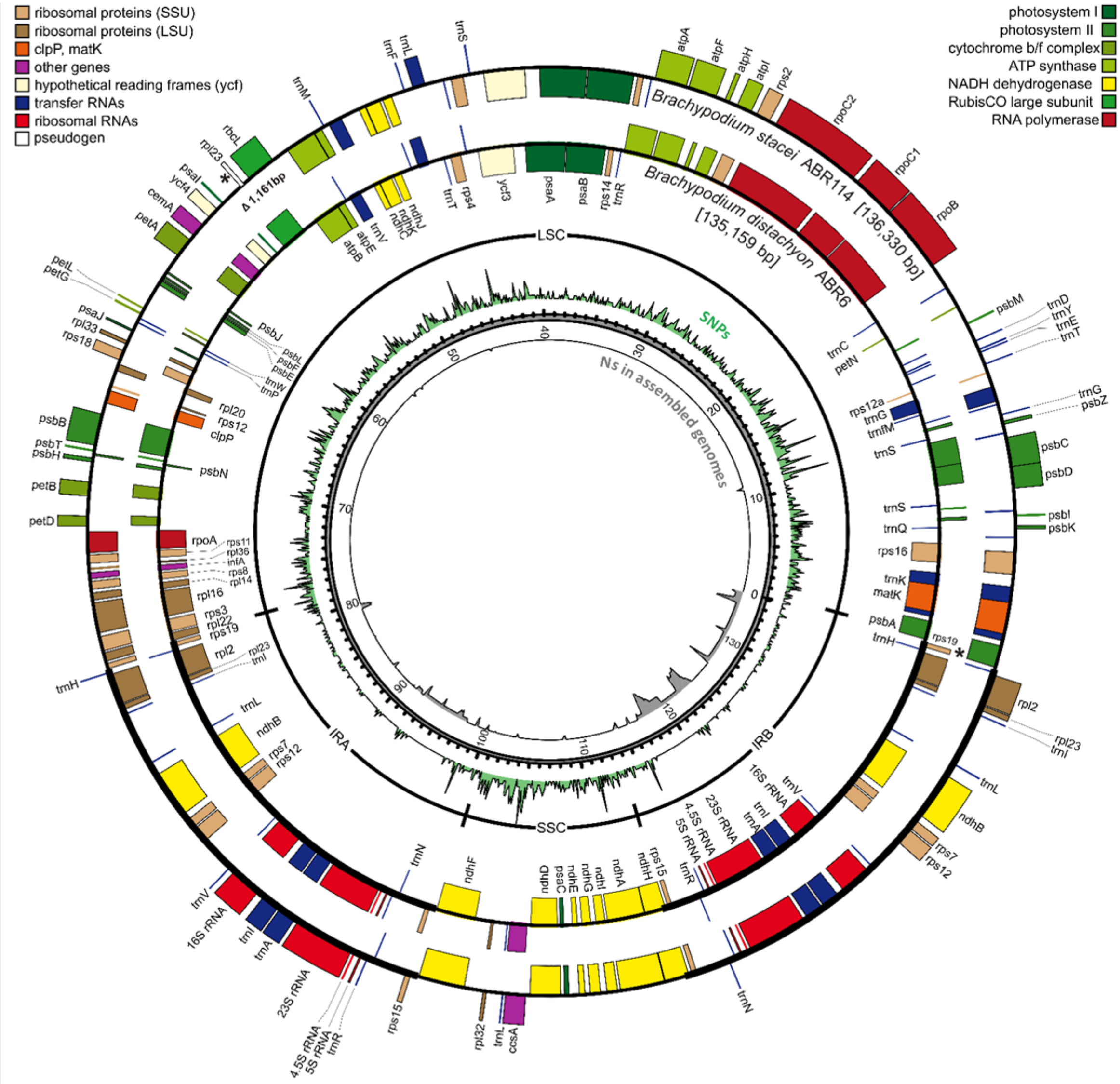


Figure 2. Plastome maps of *B. distachyon* ABR6 (inner circle) and *B. stacei* ABR114 (outer circle). A 1,161 bp insertion is shown in the *B. stacei* map (Δ, see upper-left quadrant), as well as a deletion of *rps*19 locus (*, see lower-right quadrant). Smaller inner circles and tracks correspond respectively to a map of plastome regions (LSC, SSC, IRA and IRB), a histogram of observed SNPs across all 57 aligned plastomes, and a histogram of undetermined nucleotides, marked as N characters in the alignments.

Genealogy, haplotypic groups and diversity of *B. distachyon* plastomes

BEAST (Fig. 3a), ML (Fig. S3a) and BI (Fig. S3b) analyses detected two main diverging lineages within *B. distachyon* that were structured phenotypically (Fig. 3a – Plastome tree, Table S3). One of them corresponded to an EDF+ clade, and the second to a S+T+ clade of remaining accessions, which showed a mixture of flowering phenotypes (Fig. 3a – Plastome tree, Table S3). The second clade was divided by further geographical substructure into a paraphyletic Western group ("Spanish" group – S+), including almost all ecotypes from Spain, France and Italy, and a monophyletic Eastern group ("Turkish" group – T+), including ecotypes from Turkey and Iraq, plus two Spanish accessions (ABR3, Uni2). While the divergences of the main lineages and sublineages had high bootstrap support (BS) and posterior probability support (PPS), the support of some internal branches of the S+ group was low (Figs. 3a – Plastome tree, S3a, b).

Haplotypic network analyses detected 36 or 32 distinct ptDNA haplotypes, including or excluding indels, respectively (Table S8). A set of 298 nucleotide polymorphic sites extracted from the full *B. distachyon* plastome alignment confirmed the occurrence of 32 distinct ptDNA haplotypes; 6 haplotypes were shared by different accessions (H1: 13; H2: 2; H3: 3; H4: 4; H5: 3; H6: 2) and 26 haplotypes were unique (Table S8). The TCS analysis clustered the 32 haplotypes into six groups (Fig. 3b), matching the structure observed in the genealogical ptDNA tree (Fig. 3a – Plastome tree). The haplotypic network was fully resolved except for one internal loop. The EDF+ haplotypes were separated from the cluster of S+ group and T+ group haplotypes by 59 and 74 step mutations, respectively. Within the EDF+ group there were two highly isolated clusters separated by 57 steps, one including only Turkish accessions (BdTR7A, H3, H5) and the second including Turkish and eastern European accessions (H4, Bd1-1, Bd29-1). The isolated Spanish Arn1 + Mon3 accessions of the S+T+ group showed an internal loop connecting its haplotypes with those of the EDF+ group (70 steps) and those of the remaining accessions of the S+T+ group (61 steps). Within the core S+T+ group, haplotypes clustered into four relatively close clusters, three of them including only accessions from the West (Spain, France and Italy), and the fourth cluster including mostly accessions from the East (Turkey, Iraq, plus Uni2 and ABR3) (Fig. 3b).

Plastome genomic diversity was variable within *B. distachyon* accessions (number of segregating sites (S) = 298, haplotypes (h) = 32, haplotypes diversity index (Hd) = 0.933), and especially within the S+ (S = 137, h = 17, Hd = 0.993) and EDF+ (S = 107, h = 6, Hd = 0.846) groups (Table 1a). Our analyses indicated that the T+ group was less variable (S =12, h = 9, Hd = 0.658) than the others. Diversity θπ values were not significantly different among groups. The S+ and T+ groups showed the lowest average number of nucleotide differences (d = 33.970), reflecting their close genomic affinities. In contrast, the EDF+ group showed the highest nucleotide differences to any of them (EDF+ – S+, d = 112.632; EDF+ – T+, d = 112.790) though it also shared 6 polymorphisms with the S+ group (EDF+ – S+, shm = 6) (Table 1b).

Table 1. (a). Chloroplast haplotype diversity analysis of *B. distachyon* ecotypes and genomic groups (EDF+, S+, T+). Group size and chloroplast haplotype diversity parameters. **(b)**. Pairwise estimates of the number of shared mutation (above diagonal) and the average number of nucleotide differences (below diagonal) between genomic groups.

(a)

Genomic groups	N	S	h	Hd	θπ
EDF+	13	107	6	0.846	12.780 (3.872 – 31.128)
S+	18	137	17	0.993	12.388 (3.804 – 30.837)
T+	22	12	9	0.658	12.683 (3.784 – 28.087)
B. distachyon (all ecotypes)	53	298	32	0.933	12.442 (4.218 – 28.245)

(b)

shm ***d***	**EDF+**	**S+**	**T+**
EDF+	---	6	0
S+	112.632	---	0
T+	112.790	33.970	---

When the *B. distachyon* plastome genealogy was compared to a SNP-based nuclear pan-genome genealogy generated in our parallel study (Fig. 3a – Nuclear tree, Gordon *et al.*, 2017, in press), the plastome tree revealed eleven cases of potential chloroplast capture and introgression. Seven cases (BdTR11A, BdTR11I, BdTR11G, BdTR13A, BdTR13C, BdTR3C, Bis1), corresponded to nuclear T+ ecotypes nested within the plastid EDF+ clade, two cases (ABR3, Uni2) to nuclear S+ ecotypes nested within the plastid T+ group, and two cases (Arn1, Mon3) to introgressed nuclear EDF+ ecotypes nested (and introgressed) within the plastid S+T+ clade (Fig. 3).

All these cases suggest the existence of gene flow between the most diverged *B. distachyon* lineages. The STRUCTURE search further confirmed the potential 'admixed' nature of the Arn1 and Mon3 plastomes. The Bayesian structure analysis selected two optimal plastome groups respect to second order rate of change of the log probability of data between successive K values for a particular K (ΔK), the best ΔK = 2 corresponded to the EDF+ and S+T+ clades, with individual haplotypes showing high percentages of membership (>95%) to their respective groups except the Arn1 and Mon3 haplotypes that showed similar percentages (40-60%) to both groups (Fig. 3a – plastome structure; Table S9).

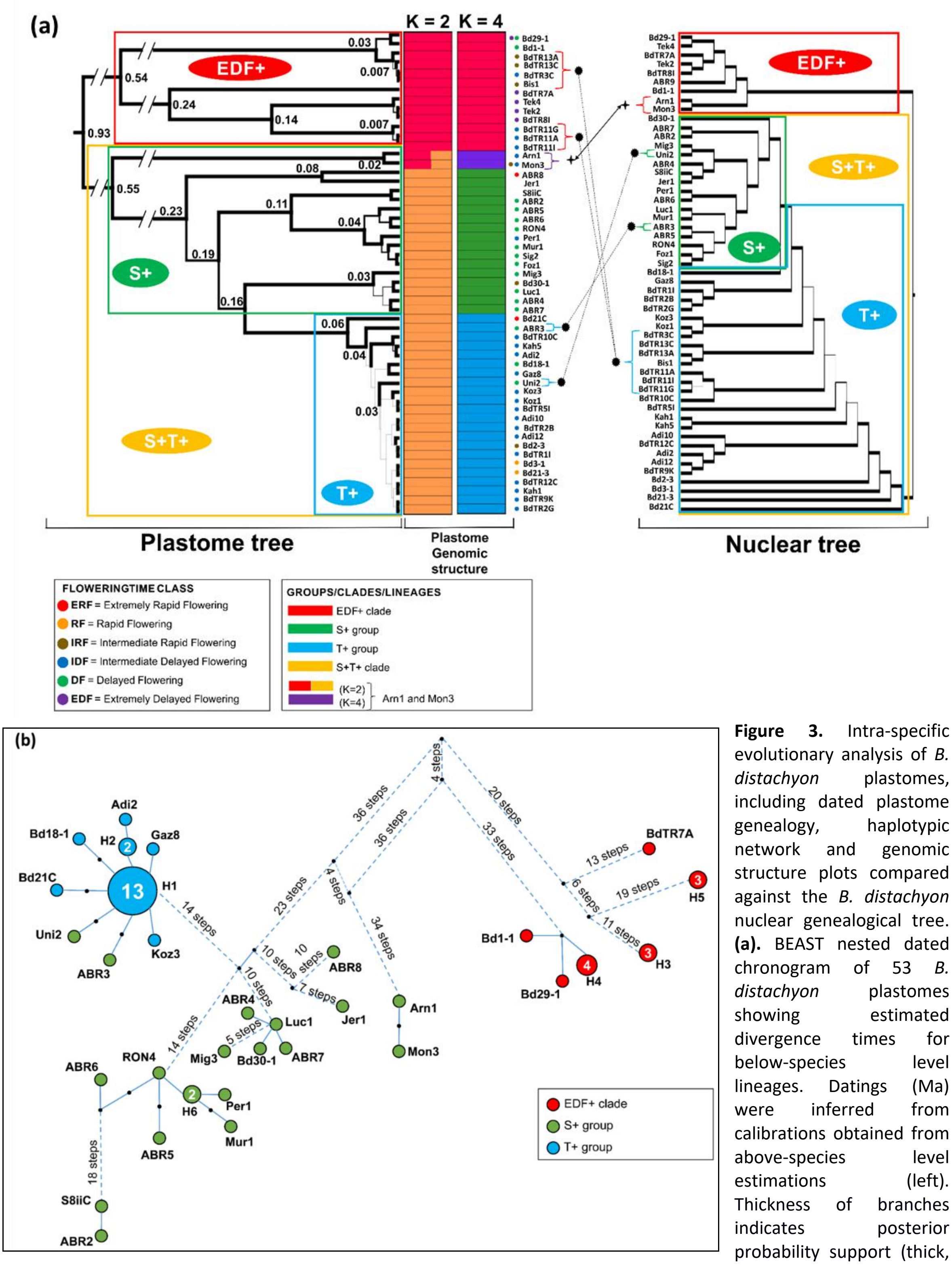


Figure 3. Intra-specific evolutionary analysis of *B. distachyon* plastomes, including dated plastome genealogy, haplotypic network and genomic structure plots compared against the *B. distachyon* nuclear genealogical tree. **(a).** BEAST nested dated chronogram of 53 *B. distachyon* plastomes showing estimated divergence times for below-species level lineages. Datings (Ma) were inferred from calibrations obtained from above-species level estimations (left). Thickness of branches indicates posterior probability support (thick, 0.95-1; intermediate, 0.90-0.94.; thin, <0.90). Genomic structure plots showing percentages of membership of plastomes' profiles to K=2 and K=4 genomic groups (center). Chloroplast capture and introgression events detected through topological contrast of the plastome and the nuclear trees (nuclear DNA (nDNA) tree from Gordon SP *et al*. 2017, in press) (right). Discontinuous and continuous lines mark potential chloroplast capture events and introgression events, respectively. Colour codes for flowering time class groups and phylogenetic groups are indicated in the respective charts. Flowering time class groups are classified according to Ream *et al*. (2014) (see Table S3) **(b).** Haplotypic statistical parsimony network constructed with the *B. distachyon* plastomes using TCS. Dots represent mutation steps; number of mutation steps are indicated on branches. Color codes for clusters are indicated in the chart.

The next optimal grouping was for ΔK = 4; in this partition EDF+, S+ and T+ haplotypes clustered separately and the Arn1 and Mon3 haplotypes formed an independent group (all memberships >95%). None of the recombination methods assayed in RDP4 and OrgConv detected significant recombination in our data set; however, visual inspection of the polymorphic data matrix detected potential micro-recombination events in Arn1 and Mon3 (Fig. S4). Both haplotypes showed a large part of their sequences (polymorphic positions 1 - 225) similar to S+T+ sequences, and a small part of them (polymorphic positions 226 - 230) similar to EDF+ sequences. Polymorphic positions 1 - 237, 238 - 245 and 246 - 298 were located in the LSC, IR and SSC regions, respectively (Figs. 2, S4).

Plastid phylogenomics and divergence time estimations of Poaceae and *B. distachyon* lineages

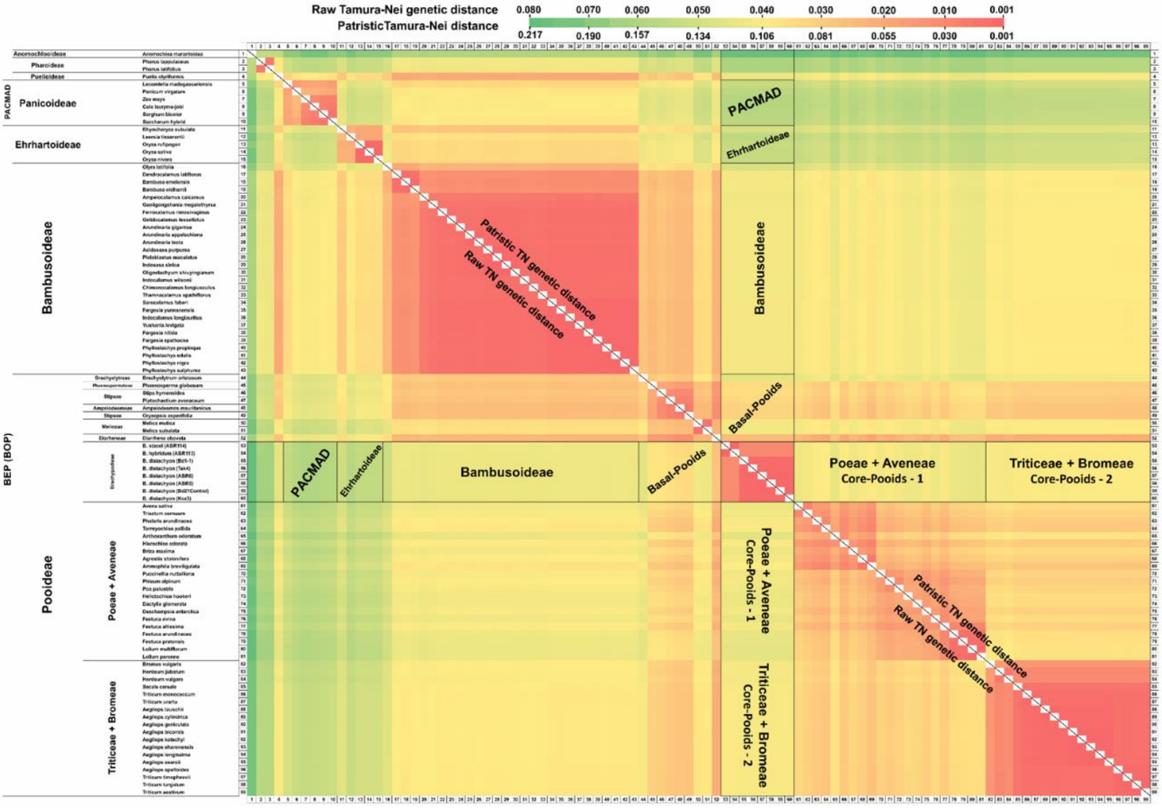


Figure 4. Color-coded matrices of pairwise Tamura-Nei (TN) genetic distances between the plastome sequences of 99 Poaceae species and 3 *Brachypodium* (*B. distachyon*, *B. stacei*, *B. hybridum*) species. Below diagonal: pairwise raw TN genetic distances; above diagonal: pairwise phylogenetically-based patristic TN genetic distances (computed on the RAxML tree, see Fig. S5b). Color-associated distance values are indicated in the chart.

ML (Fig. S5a, b) and BI (Fig. S5c, d) phylogenomic analysis of the grass plastome data set (Table S2) placed the monophyletic *Brachypodium* lineage in an intermediate and strongly supported diverging position within the Pooideae clade. *Brachypodium* was resolved as sister to the recently evolved core pooid clade, whereas the close Diarrheneae (*Diarrhena*) lineage was sister to the *Brachypodium* + core clade. Relationships among successively diverging basal Pooideae (Brachyelytreae, Phaenospematae, Meliceae, Stipeae) and BOP (Bambusoideae, Oryzoideae) and PACMAD (six Panicoideae species) lineages were congruent with previous studies; most bifurcations in the topology showed strong BS and PPS values. Within *Brachypodium*, the *B. stacei* clade (formed by *B. stacei* and the stacei-like *B. hybridum* plastomes) was resolved as sister to the *B. distachyon* clade. The latter lineage showed the divergence of the strongly supported EDF+ and S+T+ clades (Figs. S5a, c).

Both plastome raw pairwise genetic distances and pairwise patristic (RAxML tree) distances (Table S10, Fig. 4) supported the intermediate evolutionary position of *Brachypodium* within the Pooideae clade (Fig. S5a, b, c, d). Moreover, Tamura-Nei (raw) genetic and patristic distances indicated a closer relationship of Brachypodieae to more ancestral basal pooid lineages (e. g., smaller genetic /patristic distances to Stipeae and Phaenospermatae than to recently evolved core pooid lineages (Triticodae, Poodae) (Table S10, Fig. 4). They also revealed its closest relatedness to its evolutionarily nearest relative Diarrheneae. Distances of Brachypodieae to some Poodae lineages (e. g., Loliinae, Anthoxanthiinae) were similar to those observed to less related (e. g., Bambusoideae, Oryzeae (*Rhynchorhiza*), or even much less related Puelioideae (*Puelia*) lineages (Table S10, Fig. 4).

The BEAST ptDNA maximum clade credibility (MCC) tree yielded the same topology of Poaceae (Figs. 5, S6a) as that of the ML and BI trees (Figs. S5a, b, c, d). The dating analysis inferred intermediate Early Oligocene divergence times for the stem nodes of the Diarrheneae (31.9 Ma) and Brachypodieae (30.9 Ma) lineages, and divergence ages ranging from the more ancestral Mid-Late Eocene splits of the basal pooids (Brachyelytreae, 44.2 Ma; Phaenospermatae, 38.4 Ma; Meliceae, 36.7 Ma; Stipeae, 35.3 Ma) to the recent Late Oligocene-Early Miocene splits of the core pooids (crown, 27.8 Ma; Poodae, 23.9 Ma; Triticodae, 17.6 Ma) lineages. A Mid-late Miocene age (10.1 Ma) was estimated for the *B. stacei* / *B. distachyon* split and a recent Mid-Pleistocene age

(0.9 Ma) for the split of the most recent common ancestor (MRCA) of *B. distachyon* (Figs. 5, S6a). According to our nested dating analysis, intra-specific divergences within *B. distachyon* occurred very recently, during the last half million years (e. g., EDF+ and S+T+ splits, 0.55 Ma; Figs. 3a – Plastome tree, S6b).

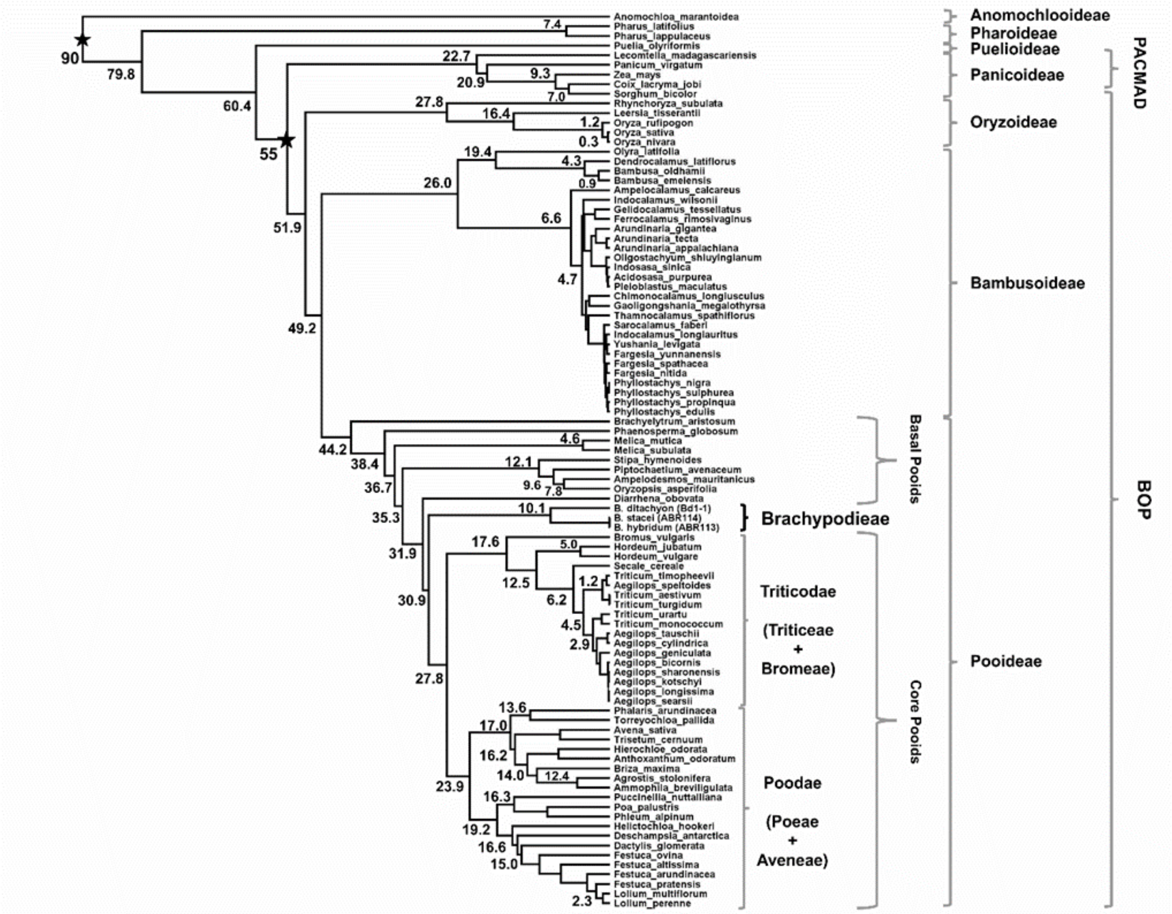


Figure 5. BEAST nested dated chronogram of 93 grass plastomes showing estimated divergence times and posterior probability support values for above-species level lineages. Stars indicate nodal calibration priors (ages) for the Poaceae and BOP+PACMAD clades. Line thickness indicates posterior probability support, which was greater than 0.97 in all branches.

Discussion

The plastid genomes of *Brachypodium*

Our study allowed us to construct the first large-scale intra-specific plastome analysis of a grass for the model species *B. distachyon* and a comparative genomics analysis with its close congeners *B. stacei* and *B. hybridum* (Fig. 2; Table S5). We detected two main indels between *B. distachyon* and *B. stacei/B. hybridum* plastomes (Fig. S2), and no structural changes but a total of 415 polymorphisms (298 without indels) among the 53 *B. distachyon* ecotypes (Tables S7a, b). A 1,161 bp insert and the deletion of one copy of the *rps*19 gene, discovered in both the *B. stacei* and *B. hybridum* ecotypes, indicates that the former is likely the maternal diploid plastome donor of the *B. hybridum* accession used in this study, which is consistent with previous findings reporting *B. stacei* as the maternal progenitor of most, though not all, wild *B. hybridum* populations (López-Alvarez et al., 2012). The scarce number of polymorphisms (6) found in the *B. hybridum* as compared to the *B. stacei* plastome (Table S7e) indicates either that the *B. hybridum* plastome has remained almost intact since the formation of *B. hybridum* or that there has been continuous gene flow from *B. stacei* into *B. hybridum* (e. g., in Pleistocene-Holocene times, after the dated split of *B. distachyon* parent; Figs. 3a, S6b).

The 1,161 bp insert found in the *B. stacei/B. hybridum* plastomes contains a *rpl*23 pseudogene of 225 bp located around position 56,335 bp (Table S7c; Figs. 2, S2a, b). The presence of a *rpl*23 pseudogene in this region has been reported in several monocots and in a large number of grasses, with insert sizes ranging from 40 – 243 bp (Morris & Duvall, 2010), whereas other authors have detected a functional *rpl*23 copy in *Agrostis stolonifera* (NC_008591) and *Sorghum bicolor* (NC_008602) (Saski et al., 2007). In this study, all the assessed *B. distachyon* plastomes lack the insert and show two annotated *rpl*23 functional copies and no pseudogene, whereas the *B. stacei/B. hybridum* plastomes have also two functional *rpl*23 copies plus the *rbc*L - *psa*I insert *rpl*23 pseudogene (Table S7c, Fig. 2a, b).

In monocots, the *trn*H-*rps*19 cluster is located near the junctions of LSC and the two inverted repeats (Borsch and Quandt 2009 and references therein). Wang *et al.* (2008) described three types of IR-LSC junctions based on the organization of their flanking genes in several monocots and dicots. While the studied *B. distachyon* plastomes fit the type III class typical of monocots (*trn*H-*rps*19 clusters contain the *rps*19 gene in both

IRs), the *B. stacei/B. hybridum* plastomes show a single *rps*19 copy near the *rpl*22 functional LSC flanking gene, and the lack of the second *rps*19 copy (Fig. S2c), fitting best the type I junction model. The type I class is mostly found in basal angiosperms, Magnoliids and Eudicots (Wang et al., 2008). Thus the *rbc*L - *psa*I insert *rpl*23 pseudogene and the *trn*H-*rps*19 type I cluster constitute landmarks of the more ancestral *B. stacei* chloroplast genome.

Flowering time divergence, chloroplast capture and introgression in B. distachyon plastomes

Our genealogical and haplotypic network analyses have detected a main split of two intra-specific *B. distachyon* lineages (EDF+ *vs* S+T+) that are not primarily connected with geography but with flowering time phenotypic traits, though the second clade is further separated into two geographically disjunct western (S+) and eastern (T+) circum-Mediterranean groups (Figs. 3a – Plastome tree, S3a, b, Table S3). Though our geographic sampling is biased towards Spain, Turkey and Iraq, these regions span the entire native distribution area of *B. distachyon* (López-Alvarez *et al*., 2012, 2015), and our results are comparable with those obtained by Tyler *et al*. (2016) using nuclear SNPs from genotyping-by-sequencing (GBS) data. Haplotypic divergence data confirm the isolation of the EDF+ clade from the S+ and T+ genomic groups and similar haplotypic diversity values of EDF+ and S+ (Table 1a, b). Intra-specific evolutionary studies of organisms tend to recover the spatio-temporal divergence of populations, that are usually associated with a geographical distribution, detecting a typical isolation-by-distance (IBD) pattern (Wright, 1943; Jenkins et al., 2010). However, long distance dispersal events and biological and ecological traits have influenced the population structure in *B. distachyon* (Vogel *et al.*, 2009; Mur *et al.*, 2011; López-Alvarez *et al.*, 2012; Tyler *et al.*, 2016). Here, we have detected a strong influence of flowering time in the ancestral divergence of the *B. distachyon* EDF+ and S+T+ lineages, as several EDF+ lines (BdTR7A, BdTR8I, Tek2, Tek4) flower considerably later than the S+T+ lines (Fig. 3a – Plastome tree, Table S3). Our parallel nuclear pan-genome study of *B. distachyon* has also recovered a main EDF+ clade, including all the extremely delayed flowering (EDF) lines of our plastome clade (Fig. 3a – Nuclear tree), and recent population genetic studies of *B. distachyon* based on GBS data (Tyler et al., 2016) have also found it. Thus, flowering time is a main biological factor controlling the divergence of the major annual *B. distachyon* clades since the late Pleistocene (0.9-0.55 Ma) (Figs. 3a – Plastome and Nuclear trees, S6b). Flowering time has been extensively studied in temperate cereals (barley, wheat), which have winter and spring races governed by vernalization and photoperiod requirements analogous to the delayed and rapid flowering phenotypes observed in *B. distachyon* (Vogel & Bragg, 2009; Schwartz et al., 2010; Colton-Gagnon et al., 2014; Ream et al., 2014; Woods et al., 2014). Although inflorescence heading-date phenotypic data in this work come from growth chamber experiments (Gordon et al., 2017), they parallel the outcomes observed in field experiments (e. g., variation in flowering time was detected between winter-annual and spring-annual wild accessions of *B. distachyon*; Manzaneda *et al*., 2015, and Manzaneda AJ, pers. comm.). Our study highlights the evolutionary importance of flowering time in driving intra-species divergence.

It could be expected that flowering time isolation would create a barrier to gene flow, which might ultimately lead to (micro) speciation (Silvertown *et al.*, 2005; Lowry *et al.*, 2008; Noirot *et al.*, 2016). However, our study has demonstrated that it is not the case in *B. distachyon*, where frequent introgressions have apparently occurred between the EDF+ and S+T+ clades during the last half million years (Figs. 3a, S6b). Topological comparison between the plastome and nuclear trees (Figs. 3a) indicated that seven Turkish accessions (BdTR11A, BdTR11I, BdTR11G, BdTR13A, BdTR13C, BdTR3C, Bis1) that are deeply and strongly nested within the eastern group of the S+T+ clade in the nuclear tree are, however, deeply and strongly nested within the eastern group of EDF+ clade in the plastome tree and network. Similarly, two Spanish accessions (ABR3, Uni2) deeply nested within the western group of the S+T+ clade in the nuclear tree are instead nested within the eastern group of S+T+ clade in the plastome tree, though with low support (Figs. 3a, b, S3a, b). Moreover, two Spanish accessions (Arn1, Mon3) which are part of the EDF+ clade in the nuclear tree, are nested within the S+T+ clade in the plastome tree, and form a loop with an EDF+ subgroup in the plastome haplotypic network (Figs. 3a, b, S3a, b). Interestingly, genomic structure analyses indicated considerable introgression signals in the Arn1 and Mon3 nuclear and plastid genomes, whereas the seven Turkish accessions and the two Spanish accessions do not show introgression evidences to the other genetic group in their chloroplast or nuclear genomes (Figs. 3a – plastome genomic structure, S4). These results support the occurrence of two different introgression events. An early introgression of a S+T+

Spanish lineage with a member of the EDF+ clade could have originated the admixed ancestor of the Arn1/Mon3 lineage that kept most of its maternal S+T+ plastome but 2/3 of its paternal nuclear EDF+ genome over generations(Gordon et al., 2017). According to our dating analysis, this introgression likely occurred in Ionian-Upper Pleistocene times (0.55 – 0.02 Ma) (Figs. 3a, S6b). By contrast, more recent late Pleistocene-Holocene (0.025 – 0.007 Ma) introgressions between geographically close Turkish EDF+ and S+T+ lines likely resulted in the seven lines that show chloroplast capture for their intact EDF+ plastomes in combination with their intact paternal nuclear S+T+ genomes, the later probably originated through repeated back-crossing to paternal S+T+ individuals (Figs. 3a, S4, S6b). A similar late Pleistocene-Holocene scenario of introgressions and repeated back-crossing, though between geographically distant S+ and T+ lines, probably resulted in the two Spanish lines that show chloroplast capture for their intact T+ maternal plastomes and their paternal nuclear S+ genomes (Figs. 3a, S4). These observations support previous evidences of long distance dispersal of eastern *B. distachyon* seeds to the West across the Mediterranean basin (cf. López-Álvarez *et al*., 2012, 2015). Additionally, Uni2 shows a significantly smaller inbreeding coefficient (F_{is} = 0.48) than the remaining highly selfed *B. distachyon* accessions (median F_{is} = 0.88), (Gordon et al., 2017), suggesting than the reduced F_{is} might be reflective of recent potential inter-population crosses.

Our analyses also point towards the potential existence of heteroplasmic recombination in the Arn1 and Mon3 plastomes (Fig. 3a – plastome structure; Table S9). Also, visual inspection of the polymorphic data matrix identified a large proportion of their plastomes as S+T+-type and a smaller proportion of them (e. g., micro-recombinations) as EDF+-type (Fig. S4). Natural chloroplast heteroplasmy originated from biparentally inherited chloroplasts is infrequent in angiosperms (but see Mogensen 1996). While plastid inheritance is considered to be mostly maternal (Jansen & Ruhlman, 2012), evidences of ptDNA biparental inheritance and of introgression have been documented in flowering plants (Mason *et al*. 1994; Mason-Gamer *et al*. 1995; Mogensen 1996), including potential low levels of sexual organelle recombination (Greiner *et al.* 2015). For instance, heteroplasmy and potential inter or intra-specific recombination have been detected in the plastomes of the highly hybridogenous genus *Citrus* (Carbonell-Caballero et al., 2015). Also, inter-specific chloroplast recombination was observed after somatic cell fusion in *Nicotiana*

(Medgyesy et al., 1985). Our study reports the first case of potential intra-specific recombination between different plastome types in these two introgressed *B. distachyon* accessions.

Evolutionary placement of a model genus for both temperate and tropical grasses

The phylogenomic analysis of 145 grass plastomes allowed us to infer the phylogenetic placement of *Brachypodium* and to calculate its genetic and patristic distances to other grass lineages (Table S10; Figs. 4, 5, S5a, b, c, d, S6a). The intermediate nesting of *Brachypodium* within the Pooideae clade and the relationships of the other Poaceae lineages agree with previous studies based on nuclear or plastid genes (Bouchenak-Khelladi et al., 2008; Schneider et al., 2011; Hochbach et al., 2015; Soreng et al., 2015) or whole plastome sequences (Saarela et al., 2015). The sister but non-inclusive relationship of *Brachypodium* to the core pooid clade [Triticodae (Triticeae+Bromeae)/Poodae (Poeae+Aveneae)], originally proposed by Davis and Soreng (1993), was abandoned in favor of the inclusion of *Brachypodium* within the 'core pooids', a non-taxonomic but independently evolved natural group, in some recent analyses (Davis & Soreng, 2007; Saarela et al., 2015; Soreng et al., 2015). Our ML and BI analyses support the sister relationship proposed by Davis and Soreng (Figs. S5a, b, c, d) as well as divergence times intermediate between those of the basal ancestral pooids and the recently evolved core pooids (Fig. 5, S6a). Additionally, our pairwise ptDNA genetic and patristic distances have further confirmed that *Brachypodium* is closer to some basal pooid lineages than to the core pooid lineages (Table S10; Fig. 4), corroborating similar results based on nuclear single copy genes (Minaya et al., 2015). Also, our genetic and phylogenetically-based patristic data indicate that *Brachypodium* is similarly close to some core pooid groups than to more distant Oryzoideae and Puelioideae lineages. The evolutionary placement of *Brachypodium* in the Poaceae supports its utility as model system for the monocots as has been recently manifested in functional genomic studies of regulation of vernalization and flowering time. *B. distachyon* shows either seasonal response to flowering mechanisms close to those of core pooid grasses adapted to cold and temperate climates (Fjellheim et al., 2014), and new flowering repressor vernalization genes shared with basal pooids, other tropical and subtropical grasses and less related

Musaceae and Arecaceae (Woods et al., 2016). Under the sampling in this study, the isolated and 'bridging' intermediate position of *Brachypodium* within the Pooideae support its value as a model genus for many types of grasses, particularly for bioenergy crops (Brkljacic *et al*., 2011) from different grass subfamilies (e. g., *Miscanthus*, *Paspalum* (Panicoideae), *Thinopyrum* (Pooideae).

Our estimated divergence times for the main Poaceae lineages (Oryzoideae, 52 Ma; Bambusoideae 49 Ma; Pooideae, 44 Ma) (Figs. 5, S6a) are in agreement with those calculated by Bouchenak-Khelladi *et al*. (2010) and Christin *et al*. (2014) but slightly older than those estimated by Wu and Ge (2012). Our results support early Oligocene (32 Ma) and late Miocene (10 Ma) splits for the respective stem and crown nodes of *Brachypodium*, which are also slightly older than those calculated by Catalán *et al*. (2012), though the highest posterior density (HPD) range intervals overlap in both studies. The relatively old divergence inferred for the annual *B. stacei* and *B. distachyon* lineages in the late Miocene contrasts with the very recent burst of the intra-specific *B. distachyon* lineages. The estimated time of the late radiation (0.9 Ma) is in agreement with the estimated age of *B. hybridum* (~1 Ma; cf. Catalán *et al*., 2012), the allotetraploid derivative of crosses between *B. stacei* and *B. distachyon*. Thus the two complementary dating analyses fit a Mid Pleistocene scenario for the almost contemporary origins of both parent and hybrid species.

Chapter 4. Co-expression network features and differentially expressed genes explain drought-response patterns in the model grass *Brachypodium distachyon*

Summary

Gene co-expression networks have been used to gain insights into gene regulation patterns and to detect interactions between stress and development signaling pathways. We developed weighted co-expression networks from leaf transcriptome data for drought response in the purple false brome *Brachypodium distachyon* and investigated network topology and differential expression of genes putatively involved in adaptation to this stressor. Co-expression analysis united drought response genes into 38 modules covering 628 hub genes (820 hub transcripts), and water response genes into 30 modules, covering 839 hub genes (1,072 hub transcripts). Pan-genome occupancy analysis showed that most drought and water network genes were core genes, present in all ecotypes, though a fraction of the hub genes were shell genes, only present in some ecotypes.

Two exclusive drought response modules included genes enriched for cellular processes including regulating proline synthesis, response to water deprivation and phosphate starvation and temperature stimulus, indicative of their potential regulation role in other stress responses. The most differentially expressed genes were overexpressed in the drought condition and a majority of them have only been found in the drought exclusive modules. A cis-regulating ABF1 motif, corresponding to an ABA inducible leucine zipper activator, was found upstream of drought exclusive genes.

Introduction

Among other environmental stresses, drought is a critical factor determining plant growth, development and survival (Bohnert et al., 1995). Plants are capable to cope and acclimate to drought stress through the reprogramming of their physiological, growth and flowering time processes (Chaves et al., 2003). Drought response also involves changes in the regulation of transcription, gene expression, epigenetic plasticity and metabolome (Fisher et al., 2016; Miao et al., 2017). Although drought stress responses and tolerance mechanisms have been investigate in a number of crops and wild species (Li & Cui, 2014), plants exhibit distinct stress response mechanisms owing to different evolutionary and adaptive processes, controlled by complex regulatory networks (Shinozaki & Yamaguchi-Shinozaki, 2007; Joshi et al., 2016). Drought-responsive gene regulation networks have been investigated in model plant systems and model organisms, such as *Arabidopsis*, maize and rice (Hayano-Kanashiro et al., 2009; Nakashima et al., 2009, 2014; Janiak et al., 2015; Borah et al., 2017). However, beyond the different responses of tolerant and sensitives genotypes to drought stress, caused by different sets of genes (e. g., maize, Mao et al. 2015), ultimate goals aim to identify signaling pathways that program regulatory networks of responses to the stressor across genotypes, making a system level study (Pereira, 2016).

Construction of gene co-expression networks (GCN) from drought-induced transcriptome profiles has been used to identify large groups of co-regulated genes in maize (Miao et al., 2017) and to infer unknown gene functions in *Arabidopsis* networks (He & Maslov, 2016). Sets of genes, defined as nodes, with similar expression profiles are clustered in modules applying graph clustering algorithms (Mao et al., 2009). Clusters with similar overall expression (modules) are often constituted by genes with similar functions (Stuart et al., 2003; Wolfe et al., 2005). Weighted gene co-expression networks (WGCN) establish correlation patterns among genes through a threshold that assigns a connection weight to each gene pair (Zhang & Horvath, 2005; Langfelder & Horvath, 2008). High connectivity “hub” nodes (genes) that show a high number of interactions with other genes within a weighted co-expression network are thought to play an important role in regulating the cellular processes (Albert et al., 2000; Carlson et al., 2006; Dong & Horvath, 2007). By contrast, peripheral genes regulate genotype x environmental (GxE) interactions, possibly reflecting their small effect size and reduced deleterious pleiotropy (Des Marais et al., 2017a). Co-expression analysis in drought response in maize detected hub genes that were crosstalk transcription factors for drought stress and developmental signaling pathways (Miao et al., 2017). Network topologies of genes involved in, respectively, cold and drought response in *Arabidopsis* showed significantly more central and more peripheral positions that genes not involved in those responses (Des Marais et al., 2017a). The peripheral expressed genotype x environment (GxE) drought response genes of *Arabidopsis* are considered to be governed by selection by changing only a small number of traits (Des Marais et al., 2017a). However, the topological positions of drought response genes in the monocot *Oryza sativa* co-expression network were different; some genes were peripheral but a large portion of them were critical components (hub genes) of the network (Miao et al., 2017).

Brachypodium is a small genus of the subfamily Pooideae (Poaceae) that contains ~20 species (3 annuals, 17 perennials) distributed worldwide (Catalán et al., 2016b). The annual diploid species *Brachypodium distachyon* was selected as model plant for temperate cereals and biofuel grasses (Vogel et al., 2010; Mur et al., 2011; Catalán et al., 2014). Despite the limited knowledge about the interaction of this plant with the abiotic environment, recent studies have demonstrated the utility of *B. distachyon* and its close congeners for elucidating the evolution and ecology of plant-abiotic interactions, focusing especially on responses to soil drying, aridity and water use strategy (Manzaneda et al., 2012, 2015; Des Marais & Juenger, 2016; Des Marais et al., 2017b; Martínez et al., 2018). A study of natural variation in drought responses of *B. distachyon* genotypes showed that phenotypic data and metabolomic profiling discriminated drought-tolerant and drought-sensitive genotypes (Fisher et al., 2016). The analysis of the interactive effects of water limitation and high temperature on the physiological responses and fitness of 35 *B. distachyon* ecotypes found GxE interactions for several traits (e. g. proline) and strong associations between phenology, biomass and water use efficiency (WUE) with parameters describing climate of origin (Des Marais et al., 2017b). Comparative field studies of mesic *B. distachyon* with aridic *B. stacei* and *B. hybridum* highlighted the contrasting physiological responses of *B. distachyon* (low WUE and proline contents) with respect to its congeners in dry environments (Manzaneda et al., 2012, 2015; Martínez et al., 2018). Despite these advances, no investigations have been developed to date on the interactions of co-expressed genes and drought response in the model plant *B. distachyon*.

In this study we analysed the osmotic stress responses of 33 *B. distachyon* ecotypes under two water conditions, drought (restriction water) and water (control). We used weighted gene co-expression network (WGCN) analysis, differentially expressed (DE) genes, and pan-genome approaches to elucidate the main genes ("hub" genes) and paths involved in drought stressor response, aiming to dissect connection mechanisms between drought stress and development signaling pathways in *B. distachyon*.

Materials and Methods

Plant material, experimental design, total-RNA extraction and 3' cDNA tag libraries preparation

We selected 33 diploid natural accessions of *Brachypodium distachyon* (L.) P. Beauv. (table S1) studied previously by Des Marais, Lasky, et al. (2017) for hydric and temperature stresses. These were inbred for more than five generations (Vogel et al., 2006, 2009; Filiz et al., 2009) and represent a large geographic and ecological diversity of their native populations across the Mediterranean region (). Whole genome data was available for all the studied lines (Gordon et al., 2017)

A total of 264 individual plants from the 33 ecotypes were grown under two abiotic greenhouse controlled conditions, restriction of water (drought, D) and watering (water, W, control). We sampled three biological replicates and three different harvests per ecotype [33 ecotypes x 4 replicates x 2 treatments (D and W)]. Water (W) plants were watered to field-capacity every second day with fresh water, whereas drought (D) plants were hand watered daily by pipette such that the soil water was reduced by no more than 5% each day (fig. 1). For each plant, the two youngest, fully-expanded, leaves of the tallest tiller were excised with a razor blade at the base of the lamina and flash-frozen on liquid nitrogen. Tissue was ground to a fine powder under liquid nitrogen using a Mixer Mill MM 300 (Retsch GmbH). RNA was extracted using the Sigma Spectrum Total Plant RNA kit, including on-column DNase treatment, following the manufacturer's protocol.

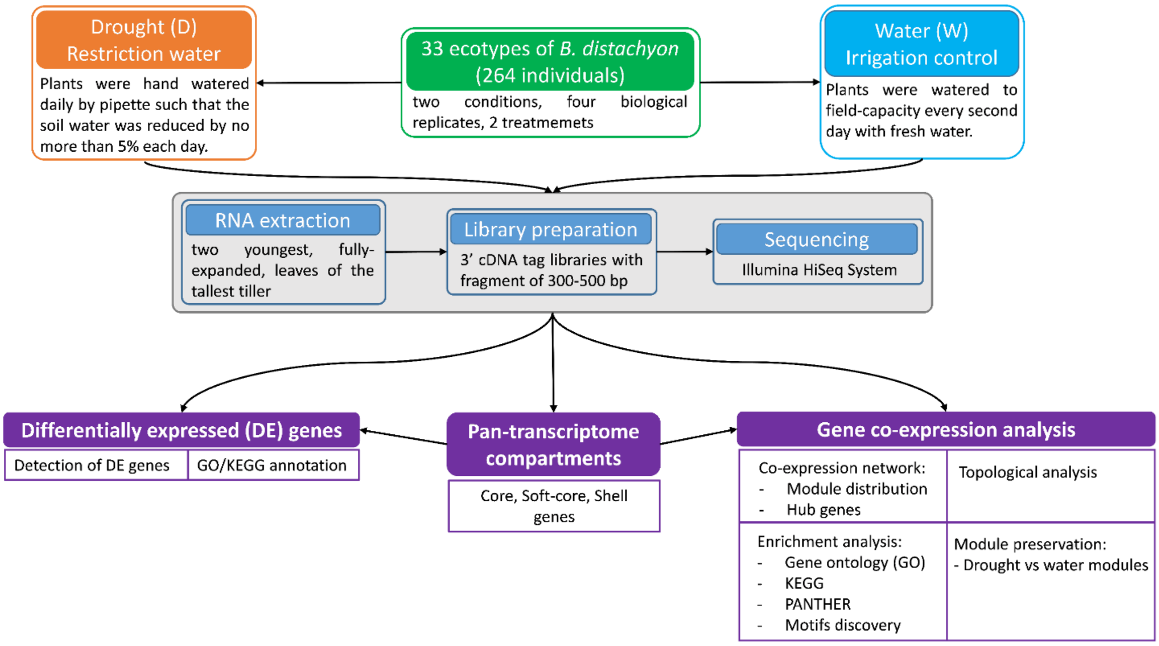

Drought (D)
Restriction water
Plants were hand watered daily by pipette such that the soil water was reduced by no more than 5% each day.
33 ecotypes of B. distachyon (264 individuals)
two conditions, four biological replicates, 2 treatmemets
Water (W)
Irrigation control
Plants were watered to field-capacity every second day with fresh water.
RNA extraction
two youngest, fully-expanded, leaves of the tallest tiller
Library preparation
3' cDNA tag libraries with fragment of 300-500 bp
Sequencing
Illumina HiSeq System
Differentially expressed (DE) genes
Detection of DE genes
GO/KEGG annotation
Pan-transcriptome compartments
Core, Soft-core, Shell genes
Gene co-expression analysis
Co-expression network:
- Module distribution
- Hub genes
Topological analysis
Enrichment analysis:
- Gene ontology (GO)
- KEGG
- PANTHER
- Motifs discovery
Module preservation:
- Drought vs water modules

Figure 1. Summary of the experimental design and analyses performed in the co-expression study of genes across 33 ecotypes of the model grass *Brachypodium distachyon* under Drought (D) and Water (W) conditions.

We used a RNA-Seq library protocol (3' cDNA tag libraries with fragment of 300-500 bp) preparation for sequencing on the Illumina HiSeq platform adapted from Meyer et al. (2011). The 3′ RNA-seq method yields only one sequence per transcript, avoiding the bias produced with long transcripts which are represented by more reads than shorter transcripts (Tandonnet & Torres, 2017).

Pre-processing of sequences, quantifying abundances of transcripts, normalizing and analysing of batch effects

Sequencing was carried out using an Illumina HiSeq2500 platform (100 bp Single-end (SE) sequencing). Quality control of SE reads was performed with FastQC software. Adapters and low quality reads were removed and filtered with Trimmomatic-0.32 (Bolger et al., 2014). Total numbers of raw and filtered SE reads for each accession and treatment are shown in table S2.

Quantifying the abundances of transcripts from RNA-seq data was done with Kallisto v0.43.1 tool (Bray et al., 2016). To accommodate the library preparation and sequencing protocols (3' tag from fragments of 300-500 bp), pseudoalignments of RNA-seq data were carried out using as references 500 bp from the 3' tails of the *B. distachyon* 314 v3.1 transcriptome (IBI 2010; http://phytozome.jgi.doe.gov/). We applied an estimated average fragment lengths of 100 bp (the approximate read length after trimming) and standard deviations of fragment length of 20. Estimated numbers of transcript per million (TPM) were recorded.

Exploratory analysis of the data set and the subsequent filtering and normalization of transcripts abundance steps between samples, and the *in silico* technical replicate step (bootstrap values computed with Kallisto), were conducted with the Sleuth tool (Pimentel et al., 2017a). A total of 16,386 targets (transcripts) and overlapping drought and water density curves (fig. S1) were recovered after the Sleuth process. This program was also used for batch-correction of data and of differentially expressed genes. To account for library preparation batch effects, date of library preparation was included as a covariate with condition variable in the full model. The reduced model only included date of libraries preparation.

Weighted gene co-expression network (WGCN) analysis of normalized transcripts abundance

Co-expression networks for the drought and water data sets were carried out using the transcript per million (TPM) estimates and the R package WGCNA (Langfelder and Horvath 2008) . We analysed the 16,386 transcripts that were filtered and normalized for 127 and 124 drought and water individual plant samples, respectively. After the removal of some putative outliers, we ended with 121 drought and 108 water samples (individual plants) that were used for network construction.

The same parameters were fitted to the drought and the water data sets to construct their respective co-expression networks. The *BlockwiseModules* function was used to perform automatic network construction and module detection on the large expression data set of 16,386 transcripts. Parameters for co-expression network construction were fitted checking different values. We chose the Pearson correlation and signed hybrid network type, the soft thresholding power 4 (high scale free, R^2>0.85), a relatively large minimum module size of 30, and a medium sensitivity (deepSplit = 2) for the cluster splitting. The topological overlap matrix (TOM) was generated using the TOMtype signed approach. Module clustering was performed with function *cutreeDynamic* and the Partitioning Around Medoids (PAM) option activated. Module merging was conducted with mergeCutHeight set to 0.25.

Transcripts (isoforms) and genes counts were calculated. Isoforms counts included all transcripts identified (e.g. Bradi1g1234.1; Bradi1g1234.2; Bradi1g1234.3) and genes counts only included different genes expressed, thus different isoforms from the same gene computed only once to gene counts (e. g. Bradi1g1234.1 and Bradi1g1234.2 count two isoforms but one gene, Bradi1g1234).

Analysis of topological features in drought and water networks and their modules

Topological features such as Connectivity, Scaled Connectivity, Clustering Coefficient, (Maximum Adjacency Ratio (MAR), Density, Centralization, and Heterogeneity were computed to compare the drought and the water networks and each of their respective modules based on an adjacency matrix calculated with the *fundamentalNetworkConcepts* function (https://www.rdocumentation.org/packages/WGCNA/versions/1.63/topics/fundamentalNetworkConcepts) of the WGCNA package. The topological features were defined according to (Zhang & Horvath, 2005; Dong & Horvath, 2007; Horvath & Dong, 2008).

Boxplots of parameters values per networks and per modules were computed using ggplot2 (Wickham, 2009) and the summary statistics of the topological features using the summary function in the R software.

In order to compare these results with WGCN analyses using all isoforms, a new round of gene co-expression networks were constructed using only the primary transcripts, encoded as Bradi[xxxx].1 (totaling 9,875 transcripts without alternative splice variants), from the filtered and normalized data set.

Detection of highly connected genes (hub genes) within co-expression networks

Three representative measures of modules, module eigengene (ME), intramodular connectivity (k_{IM}) and eigengene-based connectivity (k_{ME}) or its equivalent module membership (MM) were calculated using the WGCNA package. Briefly, ME is defined as the first principal component of a given module and can be considered a representative of the gene expression profiles within the module. k_{MI} measures how connected, or co-expressed, a given gene is with respect to the genes of a particular module. Thus, intra-modular connectivity is also the connectivity in the subnetwork defined by the module. MM is the correlation of gene expression profile with the module eigengene (ME) of a given module. MM values close to 1 or -1 indicate genes highly connected to the module. The sign of module membership indicates a positive or a negative relationship between a gene and the eigengene of the module (Langfelder & Horvath, 2010). Genes with absolute MM value over 0.8 were considered "hub genes". Correlation between MM transformed by a power of $\beta = 4$ and k_{IM} were also calculated.

Pan-genome analyses: occupancy of clustered, hub and DE genes across ecotypes

We clustered genes from the complete genome of each of the 33 studied *B. distachyon* ecotypes (Gordon et al. 2017) to define core, soft-core and shell genes with the software GET_HOMOLOGUES-EST (Contreras-Moreira et al., 2017). The search was performed selecting the OMCL algorithm (-M) and the percentage of the sequence identity threshold was calibrated to –S 98. A pan-genome matrix, with non-redundant genes within each pan-genome compartment, was generated discarding duplicated genes in downstream occupancy analyses.

This matrix was subsequently interrogated to identify core genes expressed in all 33 ecotypes, soft-core genes expressed in 32 and 31 ecotypes, and shell genes expressed only in 30 or less. Occupancy (H) was defined as the number of ecotypes that showed a particular expressed gene.

Enrichment analyses and GO/KEGG annotation of clustered genes

Enrichment analyses using the Gene ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) tools and the annotation of the *B. distachyon* 314 v.3.1 reference genome (http://phytozome.jgi.doe.gov/; IBI 2010) were performed to cluster the genes of the modules, differentiating between "hub" genes, genes of unpreserved modules and differentially expressed (DE) genes. GO and KEGG analyses were based on AmiGO 2 (http://amigo2.berkeleybop.org/amigo/landing) (The Gene Ontology Consortium, 2000, 2017; Carbon et al., 2009) and KEGG (http://www.genome.jp/kegg/kegg1.html) (Kanehisa & Goto, 2000; Kanehisa et al., 2016, 2017) databases and tools.

Gene lists were tested for functional enrichments with the PANTHER (Protein ANalysis THrough Evolutionary Relationships) Overrepresentation Test (http://pantherdb.org/). Test were conducted on both data sets, all isoforms and primary transcripts, and on both conditions, dry and water, with PANTHER13.1 using the *Brachypodium distachyon* GO Ontology database and applying a Fisher's Exact test with False Discovery Rate (FDR) multiple test correction (Thomas et al., 2003; Mi et al., 2013).

Analysis of drought *vs* water modular structure preservation

Permutation tests were performed to check for preservation of the module topology in the drought (discovery data) and the water (test data) networks by running 10,000 permutations using the modulePreservation function of the NetRep software (Gibson, 2016; Ritchie et al., 2016) with null=”all” (include all nodes) for RNA-seq data.

All test statistics (Module coherence, Average node contribution, Concordance of node contributions, Density of correlation structure, Concordance of correlation structure, Average edge weight and Concordance of weighted degree) were checked; therefore, a module was considered preserved if all the statistics have a permutation test P-value < 0.01. Searching for modules that could play a role in drought response, we focused on drought modules that were unpreserved in the water network (P-value > 0.01 for some statistics). The default alternative hypotheses "greater" tested if each module preservation statistic was larger than expected by chance comparing to random subsets of the same size.

Annotations and discovery of DNA motifs upstream of genes in unpreserved drought modules

Unpreserved modules of the drought network were further analysed with the objective of discovering DNA motifs likely involved in the control of drought responses. Analysis of putative regulatory DNA motifs was carried out using a protocol based on RSAT::Plants (Contreras-Moreira et al., 2016). Briefly, this approach allowed us to discover DNA motifs in upstream sequences of co-expressed genes, to estimate the significance (oligos and dyad tests), and to match the sequences of the discovered motifs to signatures of experimentally described transcription factors.

First, upstream sequences of modules (regulons) and 50 negative controls of equal size were extracted from the Bd21v3.1 reference genome, then peak-motifs analysis was used to discover exceptional motifs (Nguyen et al., 2018), and, finally, GO enrichment was computed. The analyses generated a report with links to similar, curated motifs in the data base footprintDB using the normalized correlation score (Ncor) (http://floresta.eead.csic.es/footprintdb/; Sebastian and Contreras-Moreira 2014), where black bars corresponded to co-expressed regulons and grey bars to negative controls.

Transcript sequences from all unpreserved drought modules were annotated by sequence similarity to proteins from UniProtKB database (The UniProt Consortium, 2017).

Analyses of differentially expressed (DE) genes

In order to determine how many transcripts and genes were differentially expressed between the two treatments (D vs W), the two data sets were analysed through the sleuth_result function. This function computes likelihood ratio tests (lrt) for null and alternative models, attending to the full and reduced fitted models. A threshold of significance level of q-value ≤ 1E-6 was fixed to detect DE transcripts.

Results

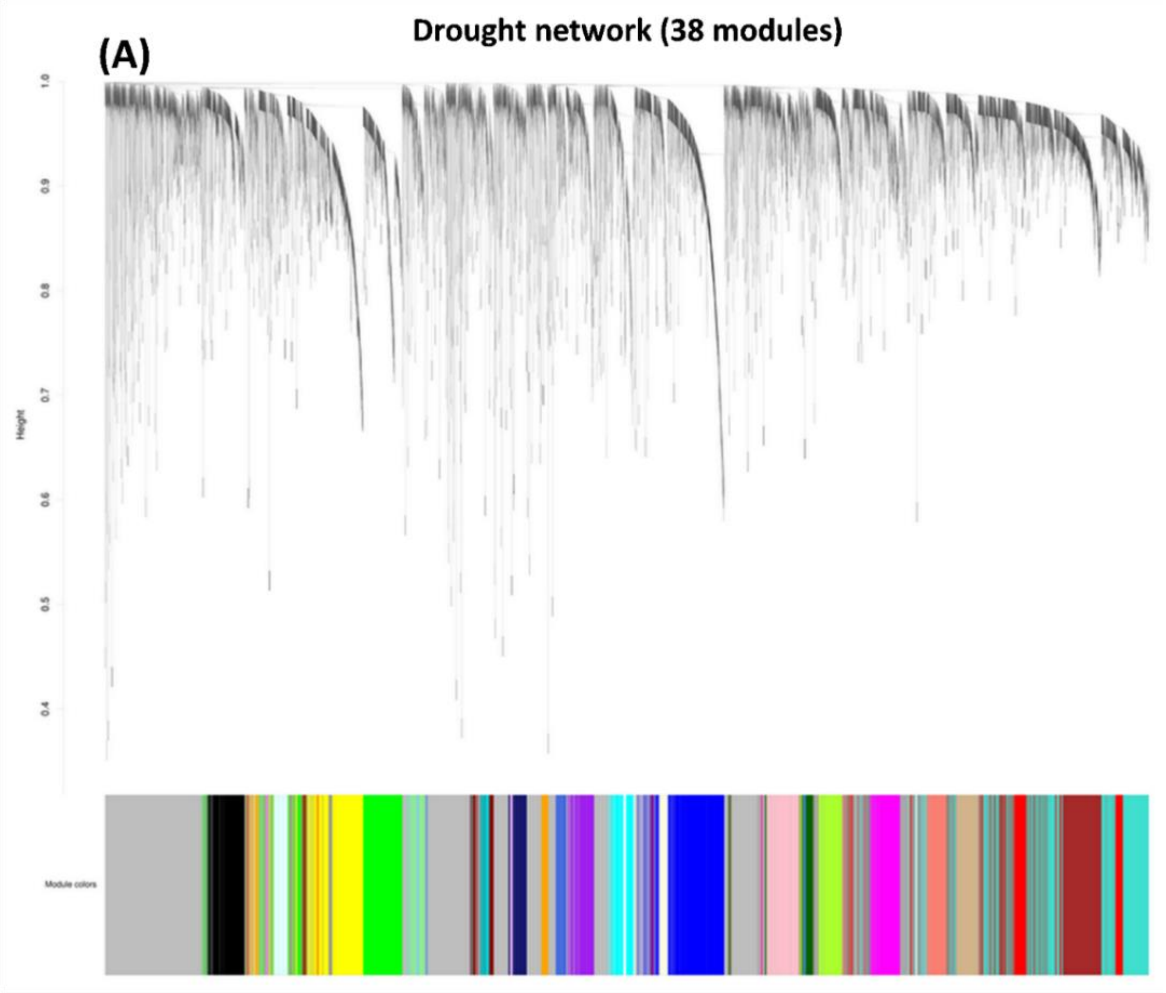
Modular distribution of gene co-expression networks

Analysis of the drought co-expression network identified 38 co-expression modules containing a total of 11,642 transcripts (min = 31, max = 1,645 transcripts) per module, corresponding to 9,072 genes (min = 16, max = 1,199 genes) per module (Fig. 2a; S2a). A total of 4,762 transcripts (3,599 genes) were not clustered in any module (fig. 2a; table 1). The water co-expression network showed 30 co-expression modules containing a total of 13,621 transcripts (min = 35, max = 2,149 transcripts) per module, corresponding to 10,587 genes (min = 23, max = 1,695 genes) per module. A total of 2,765 transcripts (2,119 genes) were not clustered in any module (fig. 2b; S2b; table 1).

Table 1. Number (#) and percentage (%) of detected transcripts and genes clustered into modules in the drought and the water networks. ID: numerical identifier of modules. Colors of modules correspond to those indicated in figs. 2 and S2

		Drought				**Water**			
modules		**transcripts**		**genes**		**transcripts**		**genes**	
ID	**color**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
0	**grey (non-clustered)**	4,762	29.1	3,599	28.4	2,765	16.9	2,119	16.7
1	**turquoise**	1,645	10.0	1,199	9.5	2,149	13.1	1,695	13.3
2	**blue**	982	6.0	783	6.2	1,470	9.0	1,170	9.2
3	**brown**	897	5.5	751	5.9	1,407	8.6	994	7.8
4	**yellow**	846	5.2	625	4.9	1,200	7.3	953	7.5
5	**green**	798	4.9	654	5.2	1,087	6.6	882	6.9
6	**red**	682	4.2	513	4.0	744	4.5	608	4.8
7	**black**	557	3.4	473	3.7	709	4.3	506	4.0
8	**pink**	494	3.0	386	3.0	704	4.3	572	4.5
9	**magenta**	465	2.8	360	2.8	636	3.9	490	3.9
10	**purple**	368	2.2	289	2.3	426	2.6	336	2.6
11	**greenyellow**	358	2.2	309	2.4	346	2.1	244	1.9
12	**tan**	354	2.2	308	2.4	330	2.0	268	2.1
13	**salmon**	303	1.8	241	1.9	322	2.0	268	2.1
14	**cyan**	302	1.8	237	1.9	263	1.6	206	1.6
15	**midnightblue**	246	1.5	188	1.5	227	1.4	163	1.3
16	**lightcyan**	229	1.4	153	1.2	221	1.3	186	1.5
17	**grey60**	212	1.3	139	1.1	192	1.2	120	0.9
18	**lightgreen**	198	1.2	169	1.3	181	1.1	152	1.2
19	**lightyellow**	169	1.0	145	1.1	128	0.8	94	0.7
20	**royalblue**	159	1.0	116	0.9	123	0.8	101	0.8
21	**darkred**	158	1.0	121	1.0	122	0.7	87	0.7
22	**darkgreen**	146	0.9	104	0.8	104	0.6	81	0.6
23	**darkturquoise**	141	0.9	110	0.9	104	0.6	85	0.7
24	**darkgrey**	125	0.8	101	0.8	95	0.6	82	0.6
25	**orange**	106	0.6	72	0.6	75	0.5	58	0.5
26	**darkorange**	93	0.6	71	0.6	63	0.4	44	0.3
27	**white**	79	0.5	74	0.6	62	0.4	53	0.4
28	**skyblue**	71	0.4	55	0.4	52	0.3	33	0.3
29	**saddlebrown**	68	0.4	55	0.4	44	0.3	33	0.3
30	**steelblue**	49	0.3	33	0.3	35	0.2	23	0.2
31	**paleturquoise**	45	0.3	35	0.3	-	-	-	-
32	**violet**	45	0.3	30	0.2	-	-	-	-
33	**darkolivegreen**	43	0.3	34	0.3	-	-	-	-
34	**darkmagenta**	43	0.3	38	0.3	-	-	-	-
35	**sienna3**	40	0.2	32	0.3	-	-	-	-
36	**yellowgreen**	39	0.2	28	0.2	-	-	-	-
37	**skyblue3**	38	0.2	25	0.2	-	-	-	-
38	**plum1**	31	0.2	16	0.1	-	-	-	-

The modular distribution of drought and water co-expression networks showed differences both in the number and the size of the modules. Thus, in the drought network 29.1% of the transcripts (28.4% of the genes) were not clustered within any module (grey or "zero" module identification). The largest drought module contained 10% of the transcripts (9.5% of the genes) whereas 12 modules clustered over 50% of the transcripts and genes, and 20 modules, 52.6% of total, clustered ≤ 1% of transcripts and genes (fig. 2a; S2a; table 1). By contrast, in water network 16.9% of the transcripts (16.7% of the genes) were not clustered within any module (grey or "zero" module identification), the largest module contained 13.1% of the transcripts (13.3% of the genes), 7 modules clustered over 50% of the transcripts and genes, and 12 modules, 40% of the total, clustered ≤ 1% of the transcripts and genes (fig. 2b; S2b; table 1).



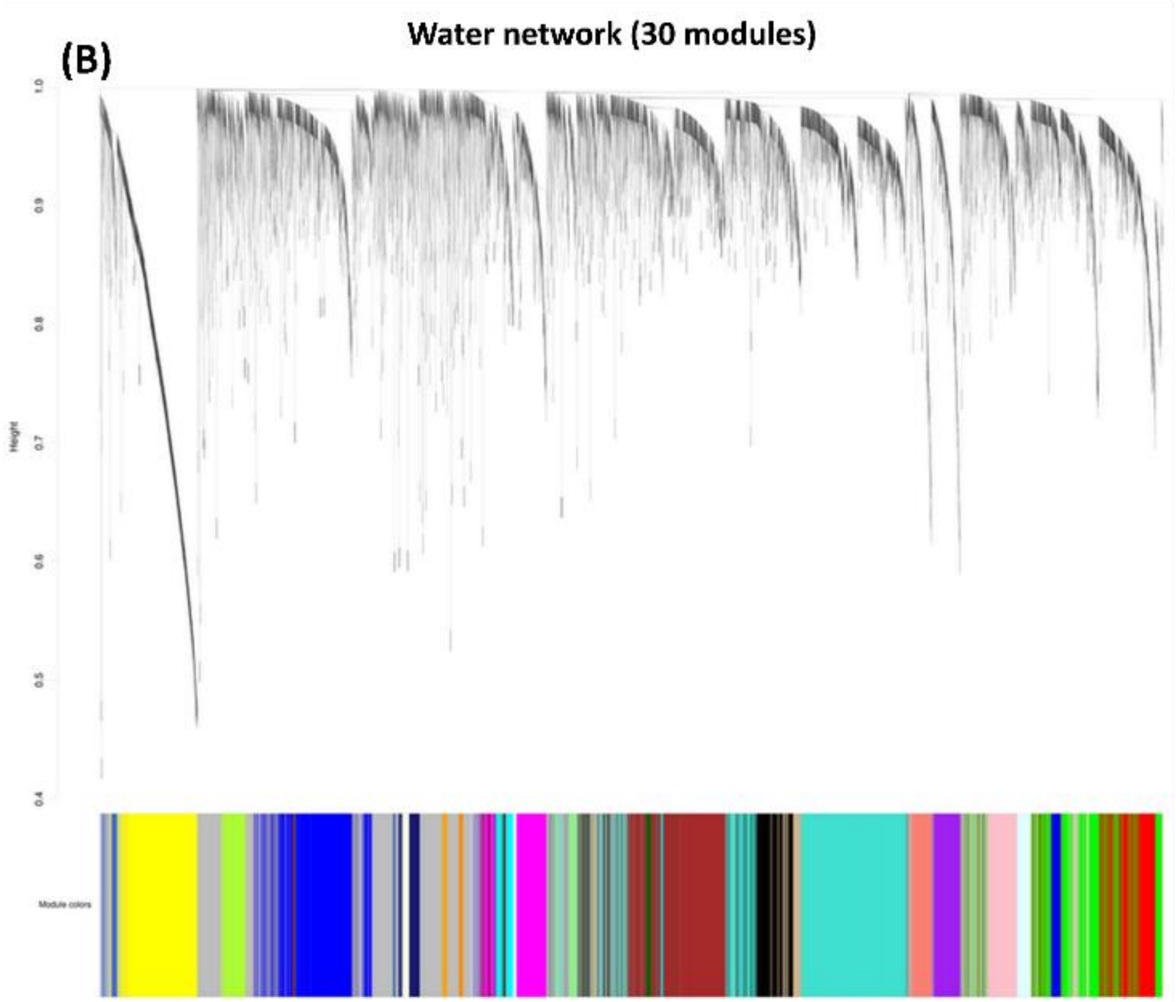


Figure 2. Clustering dendrograms with dissimilarity based on topological overlap, together with assigned module colors from weighted gene co-expression networks (WGCN) of the drought (A, 38 modules) and water (B, 30 modules) transcript analysis of *Brachypodium distachyon* accessions. Modules in A and B are color coded.

Correlation between modules in the co-expression networks

Relationships between modules within each network were established using module eigengene (ME) clustering, fixing a measure to quantify the co-expression dissimilarity of entire modules (fig. 3a, b). Modules with positive correlations greater than 0.75 (thus dissimilarities under 0.25 of height measure) were merged. The correlation between MEs was schematized to show modules with high positive (>0.70) and negative (-0.70 to -0.99) correlation between modules (fig. 3c, d) in the drought and water networks after the merging modules step.

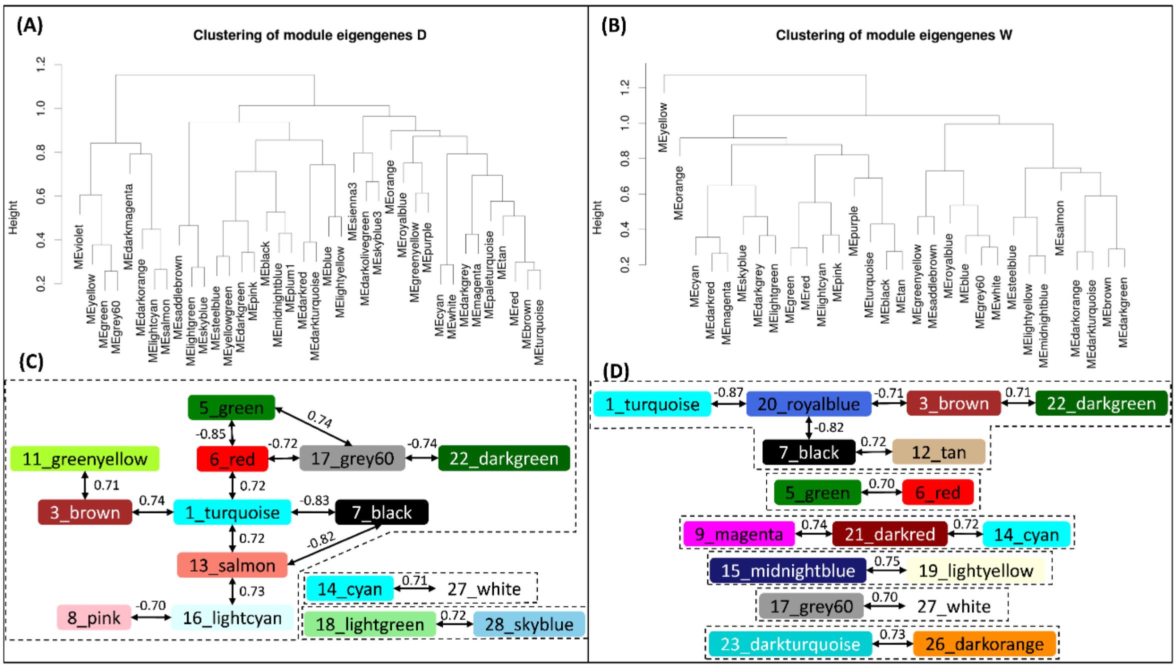


Figure 3. Dendrograms showing clustering of module eigengenes (ME) and summarized network correlations of MEs in the drought **(A, C)** and water **(B, D)** gene networks of the studied *Brachypodium distachyon* accessions, respectively. Module color codes correspond to those indicated in Fig. 2. Numerical and color identities of modules are shown.

Intra- and inter-modular connectivity was determined according to the difference between intra-modular connectivity (k_{IM}) and the connectivity out of each module (k_{out}), and computed as the difference (k_{diff}) between total connectivity and intra-modular connectivity. k_{diff} values were calculated for all transcripts within each module. Negative k_{diff} values for a transcript indicated that connectivity out the module was higher than intra-modular connectivity. High percentages of transcripts with negative k_{diff} values were recovered for each module in both drought and water networks (table S3). Two drought modules, blue (id: 2) and yellow (id: 4), showed less than 50% of transcripts with negative k_{diff}. The rest of modules showed values between 68.3 and 100% of transcripts with negative k_{diff}. The water network was found to have one module, yellow (id: 4), with 40.1% and the remaining modules with 63.7-100% of transcripts with negative k_{diff} values, respectively. These percentages indicated a high inter-modular connectivity in the two networks. High positive linear correlations between module membership (MM) transformed by a power of β = 4 and k_{IM} were recovered in both drought (fig. S3a) and water (fig. S3b) networks, thus validating the criterion of high MM (>0.8) for selecting hub genes.

Topological features of drought and water WGCN

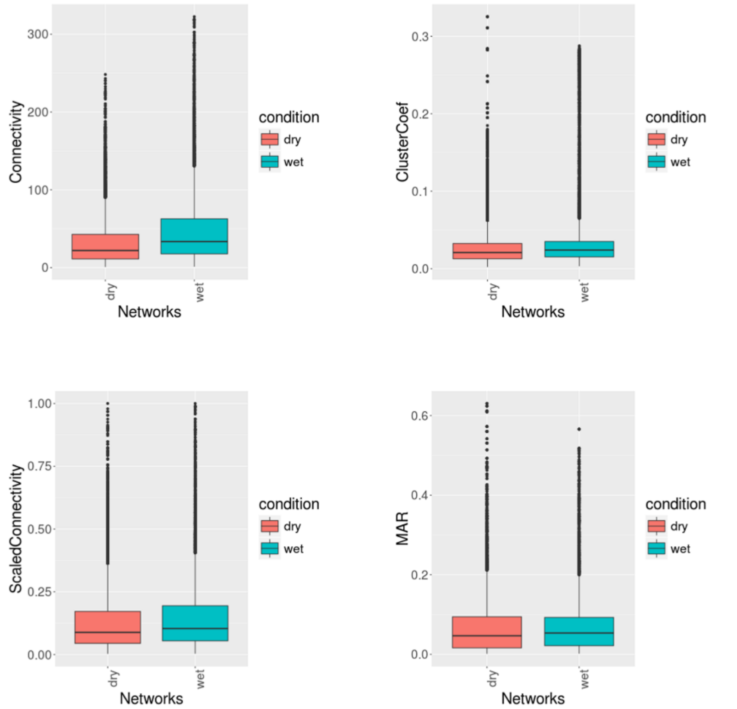


Figure 4. Boxplots of topological statistics parameters (Connectivity, Cluster coefficient, MAR (Maximum Adjacency Ratio), scaled connectivity) of drought (red) and water (blue) networks

Network topological features were compared between the drought and the water networks (fig. 4; table S4). The water network showed higher values (minimum and maximum range, quartiles, median and mean) for the studied parameters (connectivity, scaled connectivity, clustering coefficient, MAR) than the drought network. A relative high heterogeneity was observed in the two networks, indicating the presence of hub genes (fig. 4; table S4). Low centralization of 0.013 and 0.017 in drought and water network respectively, pointed out that all the nodes showed a very similar connectivity in both networks.

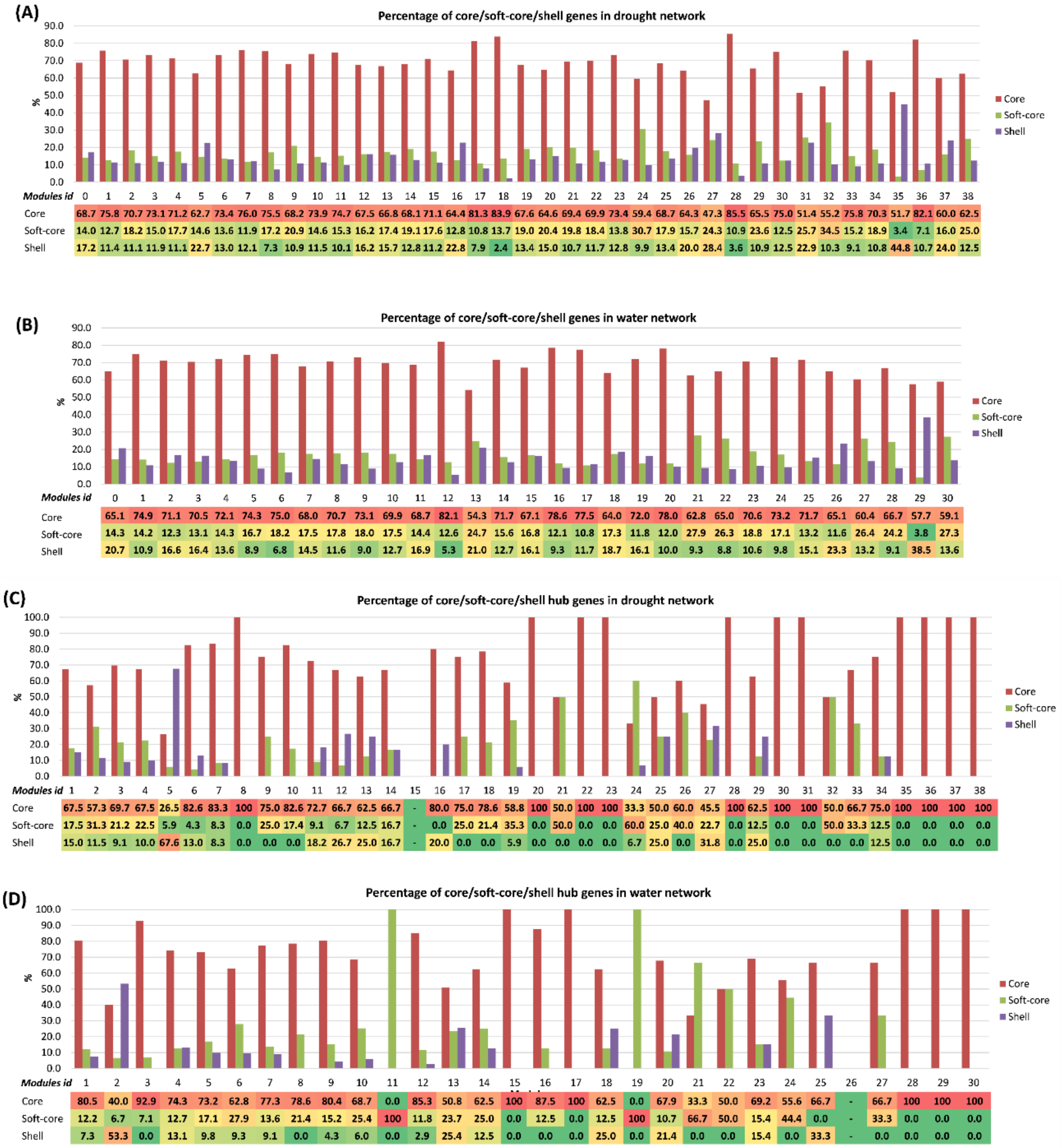
Hub genes of drought and water WGCN

Genes with absolute MM values above 0.8 were considered the most connected node “hubs”. Those nodes (transcripts/genes) were detected in both the drought and the water network (table S5). All modules showed hub nodes except modules “15” and “26” of the drought and the water networks, respectively. The modules that accumulated most hubs nodes were modules “2” and “4” with 111 and 110 hub transcripts, respectively, in the drought network and modules “4” and red “6” with 322 and 100 hub transcripts, respectively, in the water network.

Pan-genome analyses: occupancy of all clustered and hub genes

Clustered genes (assigned to modules), non-clustered genes (unassigned to modules, “zero/grey” module) and hub nodes (genes) were matched to the non-redundant pan-genome matrix to check for their occupancy (H) across the ecotypes, thus assessing the number of ecotypes that present a target gene.

Occupancy of clustered and non-clustered genes showed a clear predominance in all the ecotypes. Between 47.3 - 85.5% and 54.3 – 82.1% of these core genes were detected in the drought and the water networks, respectively (fig. 5a, b; table S6a). Hub genes were also predominantly core genes though showing a wide range of percentages in the drought (26.5 – 100%) and water (33.3 – 100%) networks. All hub genes in modules "11" and "19" of the water network were shell genes (fig. 5c, d; table S6b).



Modules id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Core	68.7	75.8	70.7	73.1	71.2	62.7	73.4	76.0	75.5	68.2	73.9	74.7	67.5	66.8	68.1	71.1	64.4	81.3	83.9	67.6	64.6	69.4	69.9	73.4	59.4	68.7	64.3	47.3	85.5	65.5	75.0	51.4	55.2	75.8	70.3	51.7	82.1	60.0	62.5
Soft-core	14.0	12.7	18.2	15.0	17.7	14.6	13.6	11.9	17.2	20.9	14.6	15.3	16.2	17.4	19.1	17.6	12.8	10.8	13.7	19.0	20.4	19.8	18.4	13.8	30.7	17.9	15.7	24.3	10.9	23.6	12.5	25.7	34.5	15.2	18.9	3.4	7.1	16.0	25.0
Shell	17.2	11.4	11.1	11.9	11.1	22.7	13.0	12.1	7.3	10.9	11.5	10.1	16.2	15.7	12.8	11.2	22.8	7.9	2.4	13.4	15.0	10.7	11.7	12.8	9.9	13.4	20.0	28.4	3.6	10.9	12.5	22.9	10.3	9.1	10.8	44.8	10.7	24.0	12.5

Modules id	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Core	65.1	74.9	71.1	70.5	72.1	74.3	75.0	68.0	70.7	73.1	69.9	68.7	82.1	54.3	71.7	67.1	78.6	77.5	64.0	72.0	78.0	62.8	65.0	70.6	73.2	71.7	65.1	60.4	66.7	57.7	59.1
Soft-core	14.3	14.2	12.3	13.1	14.3	16.7	18.2	17.5	17.8	18.0	17.5	14.4	12.6	24.7	15.6	16.8	12.1	10.8	17.3	11.8	12.0	27.9	26.3	18.8	17.1	13.2	11.6	26.4	24.2	3.8	27.3
Shell	20.7	10.9	16.6	16.4	13.6	8.9	6.8	14.5	11.6	9.0	12.7	16.9	5.3	21.0	12.7	16.1	9.3	11.7	18.7	16.1	10.0	9.3	8.8	10.6	9.8	15.1	23.3	13.2	9.1	38.5	13.6

Modules id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Core	67.5	57.3	69.7	67.5	26.5	82.6	83.3	100	75.0	82.6	72.7	66.7	62.5	66.7	-	80.0	75.0	78.6	58.8	100	50.0	100	100	33.3	50.0	60.0	45.5	100	62.5	100	100	50.0	66.7	75.0	100	100	100	100
Soft-core	17.5	31.3	21.2	22.5	5.9	4.3	8.3	0.0	25.0	17.4	9.1	6.7	12.5	16.7	-	0.0	25.0	21.4	35.3	0.0	50.0	0.0	0.0	60.0	25.0	40.0	22.7	0.0	12.5	0.0	0.0	50.0	33.3	12.5	0.0	0.0	0.0	0.0
Shell	15.0	11.5	9.1	10.0	67.6	13.0	8.3	0.0	0.0	0.0	18.2	26.7	25.0	16.7	-	20.0	0.0	0.0	5.9	0.0	0.0	0.0	0.0	6.7	25.0	0.0	31.8	0.0	25.0	0.0	0.0	0.0	0.0	12.5	0.0	0.0	0.0	0.0

Modules id	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Core	80.5	40.0	92.9	74.3	73.2	62.8	77.3	78.6	80.4	68.7	0.0	85.3	50.8	62.5	100	87.5	100	62.5	0.0	67.9	33.3	50.0	69.2	55.6	66.7	-	66.7	100	100	100
Soft-core	12.2	6.7	7.1	12.7	17.1	27.9	13.6	21.4	15.2	25.4	100	11.8	23.7	25.0	0.0	12.5	0.0	12.5	100	10.7	66.7	50.0	15.4	44.4	0.0	-	33.3	0.0	0.0	0.0
Shell	7.3	53.3	0.0	13.1	9.8	9.3	9.1	0.0	4.3	6.0	0.0	2.9	25.4	12.5	0.0	0.0	0.0	25.0	0.0	21.4	0.0	0.0	15.4	0.0	33.3	-	0.0	0.0	0.0	0.0

Figure 5. Histograms and heatmaps of all clustered genes **(A, B)** and hub genes **(C, D)** detected in the 38 and 30 modules of the drought **(A, C)** and water **(B, D)** gene networks obtained in the studied *Brachypodium distachyon* accessions. Percentages and number of genes shown in the histograms and heatmaps correspond to their respective distributions (occupancies) in gene compartments of the *B. distachyon* pan-transcriptome (core genes: found in all 33 accessions, soft-core: in 32 or 31 accessions, shell genes: in 30 or less accessions).

Enrichment analysis of genes assigned to modules

Biological processes regulated by the genes included in each module were determined for the drought and the water networks. We detected 16 drought and water modules containing genes likely involved in the regulation of several biological processes (table S7). Many of those biological processes and molecular activities were regulated by genes included in modules of both drought and water networks; however, Gene Ontology terms related to the regulation of biological processes involved in temperature and hydric stress responses, such as response to water deprivation, to heat and to phosphate starvation, were associated to genes contained in drought modules (fig. 6).

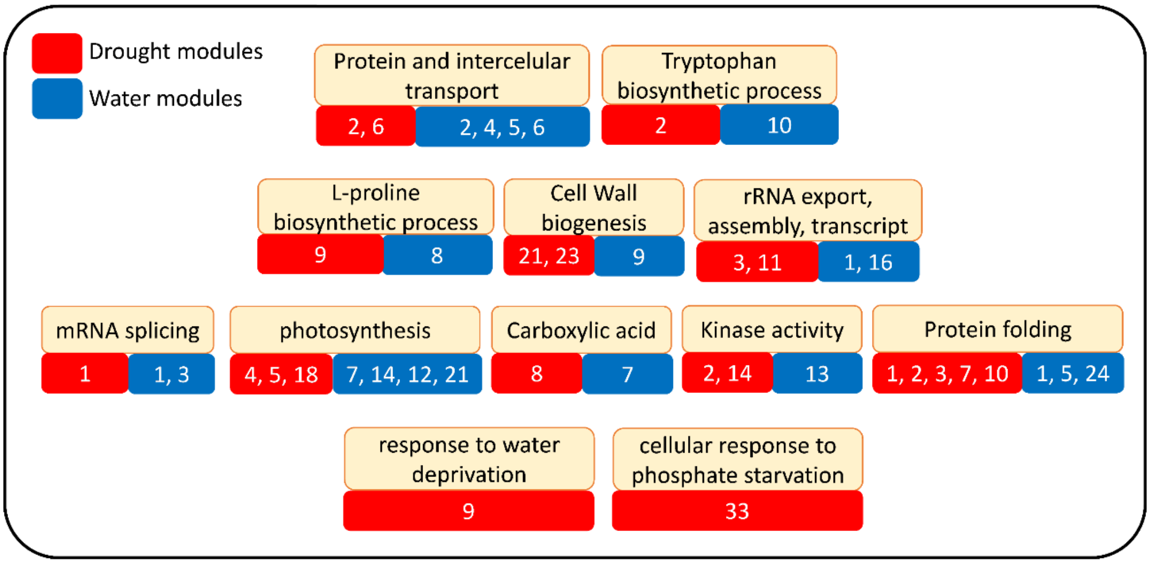


Figure 6. Biological processes significantly enriched in genes of the drought (red) and water (blue) network modules.

Unpreserved modules: exclusive drought modules absent in water network

The analysis of preservation of network modules across the drought and water datasets was focused on detecting drought modules that were absent in the water network. All drought modules were preserved in water modules (P-values < 0.01) except five modules, “9”, “15”, “22”, “30” and “33”, that showed P-values > 0.01 for some statistics (table S8). Those five modules are composed of 465 isoforms with 11 hub transcripts from 8 hub genes in module “9”, 246 isoforms without hub transcripts in module “15”, 146 isoforms with 8 hub transcripts and 1 hub genes in module “22”, 49 isoforms with 9 hub nodes and 1 hub genes in module “30” and 43 isoforms with 3 hub nodes and 3 hub genes in module “33”. Modules identified as “9” and “15” shared isoforms from the same two genes, Bradi1g20313 and Bradi2g52317 (KEGG/ec 1.8.1.8, protein-disulfide reductase).

Statistically enriched biological processes were detected in modules "9" (table 2a) and "33" (table 2b). Genes in module "9" were involved in abiotic stress responses and L-proline biosynthesis, and those in module "33" in external stimulus and response to starvation and nutrient levels.

Table 2. Enrichment analysis of non-preserved drought modules. **(A)** Drought module #9; **(B)** Drought module #33. GO (Gene Ontology).

(A)	**Ref-list**	**Drought module # 9 list (338 genes)**					
GO complete biological process	**#**	**#**	**expected**	**Fold Enrichment**	**+/-**	**raw P value**	**FDR**
L-proline biosynthetic process	3	3	.04	78.38	+	3.85E-05	3.38E-02
proline biosynthetic process	3	3	.04	78.38	+	3.85E-05	2.70E-02
proline metabolic process	4	3	.05	58.78	+	6.68E-05	3.91E-02
cold acclimation	10	4	.13	31.35	+	2.24E-05	2.62E-02
response to cold	16	5	.20	24.49	+	5.31E-06	1.86E-02
response to temperature stimulus	33	5	.42	11.88	+	1.10E-04	4.29E-02
response to water deprivation	16	4	.20	19.59	+	1.02E-04	4.48E-02
response to water	16	4	.20	19.59	+	1.02E-04	5.12E-02
macromolecule modification	3282	17	41.87	.41	-	8.30E-06	1.46E-02

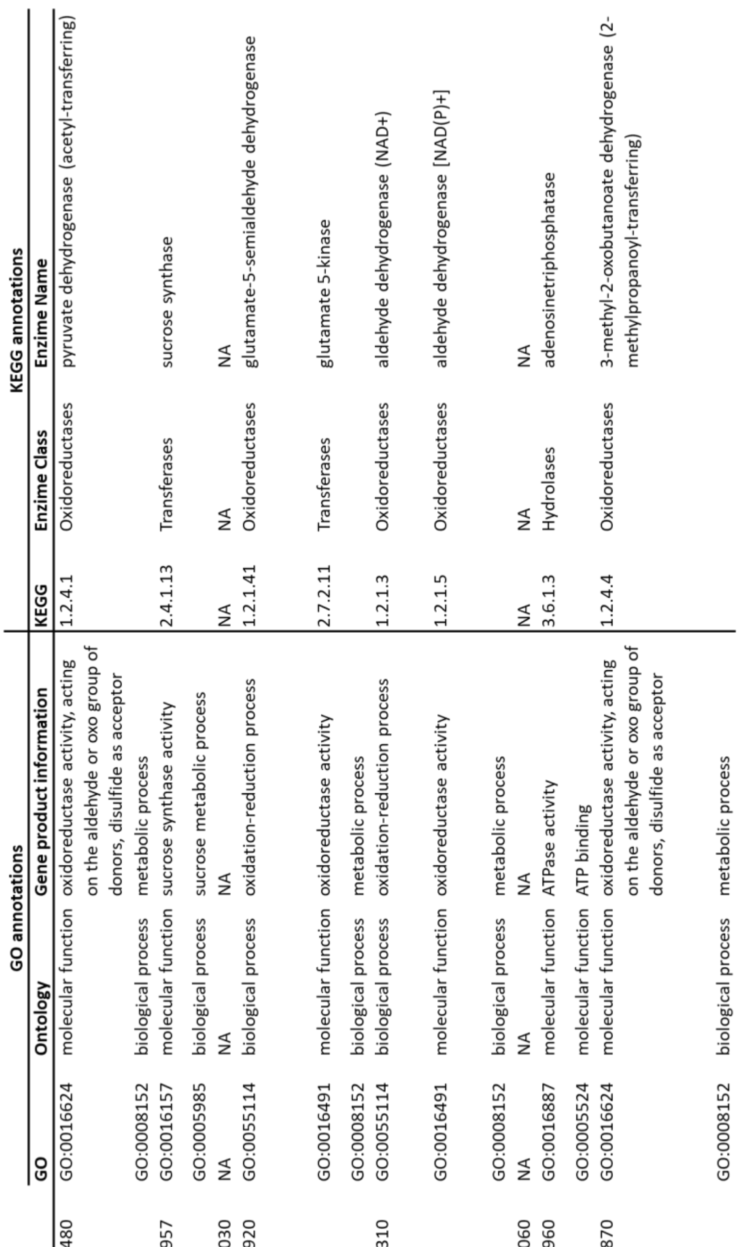
(B)	**Ref-list**	**Drought module # 33 list (30 genes)**					
GO complete biological process	**#**	**#**	**expected**	**Fold Enrichment**	**+/-**	**raw P value**	**FDR**
cellular response to phosphate starvation	5	3	.01	> 100	+	7.29E-08	2.56E-04
cellular response to starvation	14	3	.02	> 100	+	8.79E-07	1.54E-03
cellular response to nutrient levels	15	3	.02	> 100	+	1.05E-06	1.23E-03
cellular response to extracellular stimulus	18	3	.02	> 100	+	1.71E-06	1.20E-03
cellular response to external stimulus	20	3	.02	> 100	+	2.28E-06	1.14E-03
response to external stimulus	80	3	.09	33.12	+	1.13E-04	4.40E-02
response to extracellular stimulus	22	3	.02	> 100	+	2.95E-06	1.30E-03
response to nutrient levels	19	3	.02	> 100	+	1.98E-06	1.16E-03
response to starvation	16	3	.02	> 100	+	1.25E-06	1.10E-03

Ref-list (26,492 genes of *Brachypodium distachyon* from EsembleGenome source); # (number of genes in the reference (Ref)/uploaded (drought module) list that map to this particular annotation data category); expected (number of genes you would expect in your list for this category, based on the reference list); (Fold Enrichment); genes observed in the uploaded list with respect to the expected genes (number of genes in your list divided by the expected number of genes). If it is greater than 1, it indicates that the category is overrepresented in your experiment. Conversely, the category is underrepresented if it is less than 1; (+/-)A plus/minus sign indicates over/under-representation of this category in your experiment (you observed more/less genes than expected based on the reference list for this category); raw p-value as determined by Fisher's exact test. This is the probability that the number of genes you observed in this category occurred by chance (randomly), as determined by your reference list. False Discovery Rate (FDR) as calculated by the Benjamini-Hochberg procedure. By default a critical value of 0.05 is used to filter results, therefore all results shown are valid for an overall FDR<0.05 even if the FDR for an individual comparison is greater than that value.

In-depth enrichment analysis was carried out to obtain the GO and KEGG assignments of hub genes pertaining to drought modules (table 3a, b, c). GO and KEGG assignments were done for hub genes from all modules except for those of modules "22" (GO, KEGG) and "30" (KEGG). The main protein enzymatic types related to the hub genes of those modules were oxidoreductases, followed by transferases and hydrolases (table 3).

Table 3. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations of hub genes from exclusive drought modules. **(A)** module #9; **(B)** module #30; **(C)** module #33. NA (not available information. Ontology (biological process, cellular component or molecular function).

(A)



genes	GO annotations			KEGG annotations		
	GO	**Ontology**	**Gene product information**	**KEGG**	**Enzime Class**	**Enzime Name**
Bradi1g44480	GO:0016624	molecular function	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	1.2.4.1	Oxidoreductases	pyruvate dehydrogenase (acetyl-transferring)
	GO:0008152	biological process	metabolic process			
Bradi1g62957	GO:0016157	molecular function	sucrose synthase activity	2.4.1.13	Transferases	sucrose synthase
	GO:0005985	biological process	sucrose metabolic process			
Bradi2g45030	NA	NA	NA	NA	NA	NA
Bradi2g54920	GO:0055114	biological process	oxidation-reduction process	1.2.1.41	Oxidoreductases	glutamate-5-semialdehyde dehydrogenase
	GO:0016491	molecular function	oxidoreductase activity	2.7.2.11	Transferases	glutamate 5-kinase
	GO:0008152	biological process	metabolic process			
Bradi4g31310	GO:0055114	biological process	oxidation-reduction process	1.2.1.3	Oxidoreductases	aldehyde dehydrogenase (NAD+)
	GO:0016491	molecular function	oxidoreductase activity	1.2.1.5	Oxidoreductases	aldehyde dehydrogenase [NAD(P)+]
	GO:0008152	biological process	metabolic process			
Bradi4g36060	NA	NA	NA	NA	NA	NA
Bradi4g38960	GO:0016887	molecular function	ATPase activity	3.6.1.3	Hydrolases	adenosinetriphosphatase
	GO:0005524	molecular function	ATP binding			
Bradi4g40870	GO:0016624	molecular function	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	1.2.4.4	Oxidoreductases	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
	GO:0008152	biological process	metabolic process			

(B)

genes	GO annotations			KEGG annotations		
	GO	**Ontology**	**Gene product information**	**KEGG**	**Enzime Class**	**Enzime Name**
Bradi1g34100	GO:0008536	molecular function	Ran GTPase binding	NA	NA	NA
	GO:0006886	biological process	intracellular protein transport			

(C)

genes	GO annotations			KEGG annotations		
	GO	**Ontology**	**Gene product information**	**KEGG**	**Enzime Class**	**Enzime Name**
Bradi2g48420	GO:0016791	molecular function	phosphatase activity	3.1.3.2	Hydrolases	acid phosphatase
	GO:0008152	biological process	metabolic process			
Bradi3g10730	NA	NA	NA	NA	NA	NA
Bradi4g38850	NA	NA	NA	NA	NA	NA

DNA motif discovery analysis was carried out with the upstream sequences of genes clustered within exclusive drought modules using a custom protocol based on oligo-analysis and dyad-analysis RSAT tools (Contreras-Moreira et al., 2016) (fig. S4; table 4). Two of those modules (9 and 33) were found to contain statistically significant motifs. Upstream sequences of module 9 were enriched in CACGTG sites typical of bHLH (Ncor=0.756) transcription factors and of ABF1 (Ncor=0.716), an ABA (abscisic acid) inducible transcriptional activator. Moreover, promoter sequences of module 33 contained motifs similar to those of HD-ZIP (PDF2, Ncor=0.838) and Myb-like transcription factors (table 4).

Table 4. DNA motifs exclusive of genes in drought modules 9, 15, 22 30 and 33. Ncor (normalized correlation score) TF (transcription factors). Hyphen indicates unknown gene function

module id	Motif/TF name	Ncor	Description	consensus	Binding TFs
9	bHLH34	0.756	-	gtgrwwgrCACGTGycarcwygw	Helix-loop-helix DNA-binding domain
	ABF1	0.716	ABA (abscisic acid) inducible transcriptional activator	GtcgwsgKGACACGTGGCacgasr	Basic region leucine zipper, bZIP transcription factor
	bHLH31	0.662	-	swkgwgrCACGTGbswwwcwc	Helix-loop-helix DNA-binding domain
15	At2g38090	0.605	-	mCTTTCGTrtkt	Myb-like DNA-binding domain, Myb-like DNA-binding domain
	At2g41690	0.546	-	mCGAas	HSF-type DNA-binding
	AT5G22990	0.482	-	wCGwcwTCGwyttcr	T07817
22	ATGRP2B	0.624	-	kTTTTwTT	('Cold-shock' DNA-binding domain, Zinc knuckle
	TRANSCRIPTION_INITIATION_FACTOR_TFIID-1	0.608	Crystal structure of the A (-31) Adenovirus major late promoter TATA box variant bound to wild-type TBP (Arabidopsis thaliana TBP isoform 2). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution.	AATAAaAg	Transcription factor TFIID (or TATA-binding protein, TBP)
	AHL25	0.604	-	waWwWwAwTw	(Domain of unknown function (DUF296)
30	Myb-23	0.518	-	dwTkAACGGATWc	Myb-like DNA-binding domain, Myb-like DNA-binding domain
	PUCHI.DAP	0.457	-	CCGyyd	AP2 domain
	AT5G56840.ampDAP	0.448	-	ATCCAww	Myb-like DNA-binding domain
33	PDF2	0.838	-	aaarGAATATTCsaw	Homeobox domain, START domain
	AT5G29000	0.836	-	aarGAATATTCbaww	Myb-like DNA-binding domain, MYB-CC type transfactor, LHEQLE motif
	At5g29000	0.835	-	aarGAATATTCbaww	Myb-like DNA-binding domain, MYB-CC type transfactor, LHEQLE motif

Proteins encoded by genes of unpreserved modules of the drought network were further annotated (table 5). Two transcription factors (TFs) were identified in module "9" (table 5a, Bradi1g21400 and Bradi4g06867) and annotated as "Transcription elongation factor SPT4 homolog" and "MADS-box transcription factor 31", respectively. The Bradi1g21400 gene was found to be over-expressed in the drought condition whereas Bradi4g06867 was poorly expressed in both drought and water conditions.

Table 5. Characterized proteins of unpreserved drought modules. **(A)** module #9; **(B)** module #15; **(C)** module #22; **(D)** module #30; **(E)** module #33. Gene identity (id) in *B. distachyon* Bd21 v.3.1. UniProtKB data base.

(A)

module #9

gene id	protein
Bradi1g03800	NEDD8-activating enzyme E1 regulatory subunit
Bradi1g12117	Plasma membrane ATPase, EC 3.6.3.6
Bradi1g21400	Transcription elongation factor SPT4 homolog
Bradi1g29800	Catalase, EC 1.11.1.6
Bradi1g30220	Probable magnesium transporter
Bradi1g30527	Purple acid phosphatase, EC 3.1.3.2
Bradi1g40150	Reticulon-like protein
Bradi1g43780	Potassium transporter
Bradi1g44480	Pyruvate dehydrogenase E1 component subunit alpha, EC 1.2.4.1
Bradi1g45090	Glycosyltransferase, EC 2.4.1.-
Bradi1g54305	Methyltransferase, EC 2.1.1.-
Bradi1g62957	Sucrose synthase, EC 2.4.1.13
Bradi1g69160	Branched-chain-amino-acid aminotransferase, EC 2.6.1.42
Bradi1g74120	DNA topoisomerase, EC 5.99.1.2
Bradi2g04480	Phospholipase D, EC 3.1.4.4
Bradi2g04760	Glycosyltransferase
Bradi2g06627	Mannosyltransferase, EC 2.4.1.-
Bradi2g12204	Peroxidase, EC 1.11.1.7
Bradi2g12216	Peroxidase, EC 1.11.1.7
Bradi2g18970	Non-specific serine/threonine protein kinase, EC 2.7.11.1
Bradi2g23507	Delta-1-pyrroline-5-carboxylate synthase [Includes: Glutamate 5-kinase, GK, EC 2.7.2.11 (Gamma-glutamyl kinase) ; Gamma-glutamyl phosphate reductase, GPR, EC 1.2.1.41 (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase)]
Bradi2g33380	Phosphotransferase, EC 2.7.1.-
Bradi2g42180	Protein arginine methyltransferase NDUFAF7, EC 2.1.1.320
Bradi2g47840	Probable magnesium transporter
Bradi2g49160	Reticulon-like protein
Bradi2g51600	Purple acid phosphatase, EC 3.1.3.2
Bradi2g54920	Delta-1-pyrroline-5-carboxylate synthase [Includes: Glutamate 5-kinase, GK, EC 2.7.2.11 (Gamma-glutamyl kinase) ; Gamma-glutamyl phosphate reductase, GPR, EC 1.2.1.41 (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase)]
Bradi2g60730	Pyrroline-5-carboxylate reductase, EC 1.5.1.2
Bradi3g01320	Carboxypeptidase, EC 3.4.16.-
Bradi3g09850	Reticulon-like protein
Bradi3g18600	Xyloglucan endotransglucosylase/hydrolase, EC 2.4.1.207
Bradi3g35590	Trehalose 6-phosphate phosphatase, EC 3.1.3.12
Bradi3g37830	Glutamate decarboxylase, EC 4.1.1.15
Bradi3g58830	Annexin
Bradi4g00517	Plasma membrane ATPase, EC 3.6.3.6

Bradi4g06867	MADS-box transcription factor 31
Bradi4g15200	Protein ROOT HAIR DEFECTIVE 3 homolog, EC 3.6.5.- (Protein SEY1 homolog)
Bradi4g16460	Protein yippee-like
Bradi4g27240	Serine/threonine-protein phosphatase, EC 3.1.3.16
Bradi4g29640	Endoglucanase, EC 3.2.1.4
Bradi4g29920	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex, EC 2.3.1.-
Bradi4g35540	ATPase LOC100839148, EC 3.6.-.- (Arsenical pump-driving ATPase) (Arsenite-stimulated ATPase)
Bradi4g40090	Diacylglycerol kinase, DAG kinase, EC 2.7.1.107
Bradi5g12710	Peroxidase, EC 1.11.1.7
Bradi5g12982	ATP-dependent 6-phosphofructokinase, ATP-PFK, Phosphofructokinase, EC 2.7.1.11 (Phosphohexokinase)

(B)

module #15

gene id	protein
Bradi1g03700	60S ribosomal protein L36
Bradi1g06540	Ferredoxin--NADP reductase, chloroplastic, FNR, EC 1.18.1.2
Bradi1g22980	Formin-like protein
Bradi1g32792	Protein kish
Bradi1g35620	Potassium transporter
Bradi1g42270	Formate dehydrogenase, mitochondrial, FDH, EC 1.17.1.9 (NAD-dependent formate dehydrogenase)
Bradi1g43140	Vesicle transport protein
Bradi1g67700	Imidazole glycerol phosphate synthase hisHF [Includes: Glutamine amidotransferase, EC 2.4.2.-; Cyclase, EC 4.1.3.-]
Bradi1g67760	Beta-galactosidase, EC 3.2.1.23
Bradi3g33047	Flavin-containing monooxygenase, EC 1.-.-.-
Bradi3g59520	Auxin efflux carrier component
Bradi4g38380	Arogenate dehydratase, EC 4.2.1.91
Bradi4g44870	Ferritin, EC 1.16.3.1
Bradi5g14580	Endoglucanase, EC 3.2.1.4

(C)

module #22

gene id	protein
Bradi1g07160	Tubulin alpha chain
Bradi1g15590	Hexosyltransferase, EC 2.4.1.-
Bradi1g48220	Proteasome subunit alpha type, EC 3.4.25.1
Bradi1g78460	Phosphatidylserine decarboxylase proenzyme 1, mitochondrial, EC 4.1.1.65 [Cleaved into: Phosphatidylserine decarboxylase 1 beta chain; Phosphatidylserine decarboxylase 1 alpha chain]
Bradi2g10910	Coatomer subunit beta (Beta-coat protein)
Bradi2g44350	Mitogen-activated protein kinase, EC 2.7.11.24
Bradi2g60450	Phosphotransferase, EC 2.7.1.-

Bradi3g04717	Peroxidase, EC 1.11.1.7
Bradi3g52767	Tubby-like F-box protein
Bradi4g26410	Non-specific serine/threonine protein kinase, EC 2.7.11.1
Bradi4g26877	Clathrin heavy chain
Bradi4g34370	Peptidylprolyl isomerase, EC 5.2.1.8
Bradi4g35156	Proteasome subunit beta, EC 3.4.25.1
Bradi4g38460	Fatty acyl-CoA reductase, EC 1.2.1.84
Bradi5g03220	Ubiquitin-fold modifier-conjugating enzyme 1
Bradi5g08620	Carboxypeptidase, EC 3.4.16.-
Bradi5g26480	Cysteine protease, EC 3.4.22.-

(D)

module #30	
gene id	**protein**
Bradi1g20490	Defective in cullin neddylation protein
Bradi1g67960	V-type proton ATPase subunit a
Bradi3g08060	Histone deacetylase, EC 3.5.1.98

(E)

module #33	
gene id	**protein**
Bradi1g32087	Glyceraldehyde-3-phosphate dehydrogenase, EC 1.2.1.-
Bradi2g10990	Purple acid phosphatase, EC 3.1.3.2
Bradi4g27570	Glucose-1-phosphate adenylyltransferase, EC 2.7.7.27 (ADP-glucose pyrophosphorylase)

Comparative analyses between “all isoforms” and “primary transcript” WGCN analyses

To compare the previous D and W co-expressions networks with co-expression network from all 16,386 isoforms, we filtered the data set to extract and analyse only the primary transcripts (non-preprocessed mRNA WGCN analysis using the same parameters described above was conducted with 9,875 primary transcripts from the drought and the water data sets.

WGCN analysis using primary transcript drought data identified 25 co-expression modules containing a total of 7,321 primary transcripts (min = 31, max = 1,941 per module). A total of 2,554 primary transcripts were not clustered in any module whereas 505 (min = 1, max = 88) hub nodes were identified in the network (table S9a).

The water network showed 24 modules containing a total of 8,063 transcripts (min = 33, max = 913). A total of 1,812 primary transcripts were not clustered in any module whereas 721 (min = 2, max = 181) hub nodes were identified in the network (table S9b).

Primary transcripts assigned to modules that showed a negative k_{diff} were counted to analyse the intra and inter-modular connectivity relationships. We recovered a high inter-modular connectivity with more than 50% of primary transcript with negative k_{diff} in most modules from both the drought and the water data sets. Three drought modules, “1”, “2” and “4”, showed a high intra-modular connectivity with 14.9%, 42.4% and 32.9% of primary transcript with negative k_{diff} values. Three modules from the water data set also showed less than 50% of its primary transcript with negative k_{diff} with percentages of 39.2%, 4.4% and 48.7% in the “1”, “4” and “12” modules, respectively (table S9a, b).

Enrichment analyses to identify the statistically significant biological process regulated by these genes were performed (table S10). The regulated biological processes detected were similar to those retrieved previously with some exceptions. Three modules of water network, “9”, “12” and “19”, were involved in salt and heat stress response, whereas four modules of the drought network, “10”, “11”, “12” and “25”, were involved in heat, cold and stress responses.

A permutation test was computed with NetRep aiming to detect unpreserved drought modules in the water network. Three exclusive drought modules, “10”, “19” and “25” (table S11) were detected and two of them, “10” and “25”, showed statistically significant regulation of biological process involved in abiotic stress responses (table S10).

Unpreserved drought network modules of primary transcripts and “all transcripts/isoforms” data were compared and significant matches were detected. The module “10” of primary transcripts showed 169 transcripts and 5 hub nodes matching module “9 of “all transcripts”, whereas module “25” of primary transcripts showed 20 transcripts and 3 hub nodes matching module “33” of “all transcripts”. These analyses supported the preservation of these exclusive modules in the drought network and their relationships with abiotic stress responses.

Analyses of differentially expressed genes under hydric stress conditions

Most differentially expressed genes and their isoforms, summing 4,941 DE isoforms corresponding to 3,489 DE genes (q-value ≤ 1E-6), were detected by sleuth. Those DE isoforms and genes were compared to unassigned and assigned modules of both the drought and the water networks. A total of 4,229 and 4,406 DE transcripts matched exclusively drought and water modules respectively (table S12). An additional comparison was computed to detect how many DE genes (e.g. isoforms from the same gene) matched multiple modules, recovering 499 and 533 exclusive? DE genes for the drought and water networks, respectively. After comparing the medians of 4,941 DE isoforms between drought and water conditions, 1,350 isoforms were down-regulated and 3,591 over-expressed in the drought condition.

Similar DE counts of hub transcripts were recovered in the drought (423) and water (405) networks, and the same counts of DE hub genes (330) in both conditions, but they showed only 111 matching genes.

Occupancy analysis was computed comparing the DE genes to the pan-genome matrix. A total of 3,768 DE genes matched non-redundant genes in the pan-genome showing an occupancy distribution of 2,683 core, 705 soft-core and 380 shell genes.

In-depth analysis was performed to identify GO biological process, module co-expression matching and occupancy in the highest differentially expressed top-50 isoforms (fig. S5; table S13). Box plots of the highest top-50 DE transcripts (fig. S5) showed that all transcripts were over-expressed in dry condition except Bradi2g51480.1, involved in photosynthesis regulation. Three isoforms, Bradi1g37410.1, Bradi3g43870.1 and Bradi3g51200.1, were related to response to stress and water stimulus (table S13, bold entries) and all of them matched genes in the drought exclusive module "9" which clustered 31 out of 50 isoforms from the most differentially expressed genes. These isoforms did not correspond to hub genes. Pan-genome analyses showed that 23 DE genes were core, 14 soft-core, 11 shell with 30-31 occupancy, 1 shell with 24 occupancy and 1 shell with 7 occupancy.

Discussion

Weighted co-expression network analysis detects genes involved in drought stress response in cool season grass *Brachypodium distachyon*

Large scale transcriptome data sets have been used to construct co-expression networks for gene and gene regulation discovery in model plant systems and crops (Aoki et al., 2007, 2016; Masalia et al., 2017; Miao et al., 2017). The co-expression network approach further allows testing hypotheses on gene functions, from their connections with other functionally known genes classified in the same modules (Mochida et al., 2011), and on links between signalling pathways and phenotypic response to environmental stress (Des Marais et al., 2012). Gene networks operate in different biological contexts; an important proportion of the genetic interactions within a network have been demonstrated to be condition-specific (He & Maslov, 2016). Our system-level approach allowed us to construct a drought-responsive gene co-expression network from leaf tissue transcriptome profiles of *B. distachyon* accessions and to identify modules of putatively co-regulated genes within it (figs. 2, 3). Drought response mechanisms consist in extremely complex interactions of several metabolic processes, as manifested in thoroughly investigated crop grasses (e. g., barley, (Mochida et al., 2011); rice, (Yu et al., 2017); maize, (Miao et al., 2017)); however there is still a considerable gap in the knowledge of relationships between drought response genes and developmental signaling (Miao et al., 2017). Our comparative drought *versus* water study case discriminates modules and genes exclusively co-expressed under the drought conditions, shedding light into the specific pathways driving the generation of major transcriptional response involved in drought abiotic stress.

The *B. distachyon* drought (D) and water (W) WGCNs constructed from isoforms show similar topological features and no differences for all the parameters, though the connectivity of the water network was slightly higher than that of the drought network (fig. 4; table S4). By contrast, the number of assigned modules is higher in the drought (38) than in the water (30) network, despite the higher number of expressed transcripts of the latter (figs. 2, 3; table 1). Values of transcript k_{diff} were negative by more than 50% in all assigned D and W modules except for one case (table S3), indicating a high inter-modular connectivity in both networks. The number of hub genes was overall low in most modules of the D, especially in unpreserved drought modules, but also of the W networks, except for few exceptions (table S5).. Cardinally, the analysis of unpreserved drought modules in the water network provided cues to identify genes involved in the drought response in *B. distachyon*.

Two out of the five drought modules unpreserved in the water network ("9" and "33") have shown a statistically significant regulation of biological processes (table 2). Genes from module "9" are involved in L-proline biosynthesis, responses to water deprivation and temperature stimulus, and macromolecule modification, whereas those from module "33" are related to cellular response to phosphate starvation (Fig. 6, table 2). Differences in proline accumulation have been observed in *B. distachyon* and other close congeners as consequence of responses or adaptation to the dry environment. Leaf free proline abundance showed a significant strong effect of GxE (temperature/water) interactions in green house controlled experiments of *B. distachyon* (Des Marais et al., 2017b). Overall, the plants accumulated more proline in response to drought although heat and restricted water availability enhanced this response. Field experiments have found a significant role of high leaf proline and low water content traits in the response to soil water restriction conditions in the drought-escapers *B. stacei* and *B. hybridum* species but not in the dehydration avoider *B. distachyon* species (Martínez et al., 2018). Furthermore, Fisher et al. (2016) showed that drought-induced proline was significantly elevated in drought-tolerant or intermediate *B. distachyon* ecotypes but not in susceptible ecotypes. Our findings indicate that genes involved in the proline synthesis, temperature stimulus and water deprivation of module 9 (table 2) regulate the drought response pathways in *B. distachyon*. These genes could be also involved in other related signaling pathways, such as heat response (Des Marais et al., 2017b) and flowering time and development (Mattioli et al., 2009; Martínez et al., 2018). Inorganic phosphate is an essential nutrient for plant growth; plants have evolved biological mechanisms to efficiently mobilize and uptake phosphate in deficiency conditions (Yuan & Liu, 2008). Phosphate starvation signaling is affected by abiotic stresses, like drought and salt (Baek et al., 2017). Our analysis has corroborated the role played by module 33 phosphate starvation genes in the response to drought stress in *B. distachyon* (table 2; table 3). These genes could be also involved in the regulation of other crosstalk abiotic stress responses and signaling pathways, like those including phytohormones, ABA, sugars and photosynthesis found in *Arabidopsis* (Rubio et al., 2009; Baek et al., 2017). Screening analysis of genes of drought modules 9 and 33 indicates they encode for essential cellular proteins (table 5).

Drought and water co-expression networks of *B. distachyon* constructed from primary transcripts (tables S9, S10) mimic those based on isoforms. Primary transcripts drought modules unpreserved in the W network (table S11) significantly regulated the same biological processes related to proline synthesis and responses to temperature stimulus, water deprivation and phosphate starvation than in the isoform case. These findings indicate that similar co-expression and regulation mechanisms of drought response and other interconnected signaling pathways are maintained before and after the transcript maturation in *B. distachyon*. Nonetheless, the more detailed co-expression analysis using all isoforms data set shows more unpreserved exclusive drought modules and genes suspected of being involved in hydric stress responses than analyses performed with primary transcript data set alone. Moreover, the presence of isoforms from the same genes assigned to different modules indicate specificity of isoforms, not just genes, in stimulus responses. Cantalapiedra et al. (2017) studied the gene expression responses to drought and heat stress in barley detecting some cases of several isoforms associated to a single gene differentially affected by these treatments.

<u>Occupancy, differential expression and cis-regulation of drought response genes in *B. distachyon*</u>

The *B. distachyon* drought and water co-expression networks were further investigated for occupancy, over-expression and cis-regulation of total and hub genes of D and W modules across genotypes (tables 4, S6, S12, S13; fig. 5, 6).

Occupancy analysis of genes in the modules of the drought and water networks showed that the highest percentages of total genes were core genes in both cases, and therefore present in all 33 accessions, though a few modules (35 in D, 29 in W) had also relatively high percentages of shell genes (fig. 5a, b). However, inspection of hub genes occupancy indicated that despite the overall predominance of core genes in most modules, several of them had similar (21, 32 in D; 22 in W) or higher (5, 24 in D; 2 in W) percentages of shell genes than core genes or even were formed exclusively by shell (soft-core) genes (11, 19 in W) (fig. 5c, d). These results indicate that highly connected hubs of some drought response modules are only expressed in a subsample of the accessions, pointing to the exclusivity of central pathway genes in certain individuals. It agrees with conclusions from a large pan-genome study of *B. distachyon* suggesting that shell genes may confer conditionally beneficial functions to particular phenotypes, like drought tolerance (Gordon et al., 2017).

Most differentially expressed (DE) genes were assigned to modules in both drought and water networks. (table S12), though 562 and 417 DE genes were not assigned to any module in the drought and the water network, respectively. The most significant DE genes were mainly over-expressed in drought condition. Thus, 3,591 isoforms of a total of 4,941 were over-expressed in drought. The top 50 most differentially expressed transcripts were up-regulated in the drought condition (fig. S5), except for one gene (Bradi2g51480.1) encoding a photosynthesis regulator. Three of the top 50 DE genes encode drought stress and water stimulus proteins and all of them pertain to drought exclusive module 9 (Bradi1g37410, Bradi3g43870 and Bradi3g51200). This module also encompassed 31 out of the 50 most overexpressed genes (table S13). Noticeably, a few of these most up-regulated genes were hub genes in drought network. Regarding their occupancy, 2,683, 705 and 380 DE genes were core, soft-core and shell genes, and with respect to the top-50 DE genes, less than half (23) were core genes and 27 were soft-core plus shell genes, corroborating their ecotype-specific expression.

Mapping genes ids to UniProtKB database together with DNA motifs analysis detected characterized proteins of unpreserved drought modules, especially those of 9 module that corresponded to transcriptional factors SPT4, MADS-box TF31, bHLH, MYB and bZIP involved in drought (table 5). bHLH has been previously noted as a TF involved in multiple signal pathways in adaptation to drought (Castilhos et al., 2014; Mun et al., 2017). Mun et al. (2017) detected MYB and bHLH TFs up-regulated, and bZIP and MADS box TFs down-regulated in *Populus davidiana* under drought stress conditions. The basic helix-loop-helix (bHLH) superfamily of TFs has been studied in *Brachypodium distachyon* identifying 146 bHLH genes distributed in 5 chromosomes (Niu et al., 2017).

Our search for cis-regulation motifs identified a drought module 9 motif ABF1 that corresponded to an ABA inducible activator binding to leucine zipper bZIP (fig. S4, table 4). The ABA hormone regulates the plant water levels and promotes stomatal closure, thus leading to drought resistance (Christmann & Grill, 2018). A mobile CLE25 protein activated by drought that induces the synthesis of the ABA precursor could be the triggering factor causing drought resistance in plants (Takahashi et al., 2018). ABA is also a key factor of other crosstalk signaling pathways also linked to drought response, like phosphate starvation (Baek et al., 2017). It has been shown that constitutive ABA content is higher in *B. distachyon* than in its annual warm-adapted congeners and that under drought stress *B. distachyon* decreases stomatal conductance and keeps relatively high water content levels, thus avoiding dehydration (Manzaneda et al., 2015; Martínez et al., 2018). Our discovered cis-regulation ABF1 motif in drought treated *B. distachyon* accessions confirms the importance of the ABA-mediated response to drought condition.

CONCLUSIONES GENERALES

1-Los análisis evolutivos y biogeográficos de los 20 taxones reconocidos del género *Brachypodium* empleando cinco genes (tres nucleares y dos plastídicos) indican que aproximadamente la mitad de las especies son diploides y la otra mitad alopoliploides. El análisis de evolución mínima de "injerto" de alelos alopoliploides en las ramas del árbol diploide recobra los linajes homeólogos (subgenomas) de los alopoliploides. Las sucesivas divergencias de los linajes de las diploides anuales (*B. stacei, B. distachyon*) tuvieron lugar durante el Mioceno tardío-Plioceno en la cuenca Mediterránea, mientras que las de los linajes de las diploides perennes (*B. arbuscula, B. genuense, B. sylvaticum, B. glaucovirens, B. pinnatum*-2x y *B. rupestre*-2x) ocurrieron durante el Cuaternario en las regiones Mediterránea y Euroasiática, con colonizaciones esporádicas de otros continentes. Las respectivas divergencias de los linajes homeólogos de los alopoliploides tuvieron lugar en distintos tiempos evolutivos. Nuestro escenario biogeográfico apoya la existencia de dispersiones a larga distancia únicamente en los linajes diploides, mientras que todos los eventos de hibridación y duplicación genómica ocurrieron dentro de las áreas ancestrales progenitoras más recientes, sin posteriores expansiones de área.

2- Los análisis filogenómicos mediante datos de RNA-seq y GBS han identificado a *B. mexicanum* como la especie alopoliploide más antigua mostrando subgenomas de tipo ancestral (A) y materno de tipo stacei (B) (Mioceno medio-tardío). Los alopoliploides de elevado nivel de ploidía, *B. boissieri* y *B. retusum*, muestran tres y cuatro subgenomas respectivamente. Ambas especies presentan A y B así como el subgenoma intermedio tipo distachyon (C) (Mioceno-Plioceno) (heredado maternalmente en *B. boissieri*). *B. retusum* también presenta un subgenoma materno tipo *core perennial* recientemente evolucionado (D) (Cuaternario). Los alotetraploides del clado *core perennial B. rupestre* y *B. phoenicoides* muestran únicamente subgenomas recientemente evolucionados tipo C y D (Cuaternario), siendo los diploides perennes *B. pinnatum* y *B. sylvaticum* sus respectivos progenitores maternos. El reciente alopoliploide *B. hybridum* se formó repetidamente y mediante cruzamientos bidireccionales durante el Cuaternario y es el único alopoliploide del que se conocen ambos progenitores diploides actuales, *B. distachyon* y *B. stacei.*

3- Los análisis pan-transcriptómicos de 5202 conjuntos de tránscritos del género *Brachypodium* muestran genes expresados exclusivamente en los grupos de especies perennes (30), anuales (49), poliploides (14), alopoliploides más antiguos (143), especies ancestrales (14) y especies recientemente evolucionadas (52). Los tránscritos exclusivos de los alopoliploides antiguos podrían estar asociados con su genoma ancestral tipo A. Los tránscritos anotados como subunidad ARN polimerasa, encontrados únicamente en todas las especies anuales de *Brachypodium*, podrían indicar la existencia de diferencias en los niveles de expresión de las ARN polimerasas entre las especies anuales y perennes, o la pérdida de copias ancestrales en las especies perennes más recientemente evolucionadas.

4- Los análisis pan-genómicos de los plastomas de 53 ecotipos de *B. distachyon,* 3 de *B. hybridum* y 1 de *B. stacei* han detectado una inserción (1161 pb) y una deleción en una de las copias del gen *rps*19 que diferencian a los plastomas de *B. stacei* y *B. hybridum* con respecto a los de *B. distachyon*, sin que se haya observado variación en el contenido génico entre los plastomas de *B. distachyon*.

5- El árbol filogenómico de los plastomas de *B. distachyon* muestra la divergencia de dos linajes principales, correspondientes a los clados *Extremely Delayed Flowering* (EDF+) y *Spanish* (S+) – *Turkish* (T+), sugiriendo que el tiempo de floración es un factor decisivo en la divergencia intra-específica de *B. distachyon*. La comparación topológica entre las filogenias nucleares y plastídicas de esta especie revela nueve eventos de captura cloroplástica y dos de introgresión y micro-recombinación entre esos clados, apoyando la existencia de flujo génico entre linajes previamente aislados. Los intercambios de plastomas entre los tres grupos, EDF+, T+, S+, probablemente hayan sido el resultado de retro-cruzamientos aleatorios seguidos de estabilización por presión selectiva.

6- Los análisis mediante redes ponderadas de co-expresión génica llevados a cabo en 33 ecotipos de *B. distachyon* bajo condiciones de sequía y riego identificaron cinco módulos exclusivos de la red de sequía, incluyendo 465 isoformas y 11 genes altamente interconectados (*hubs*). El análisis seleccionó genes candidatos y factores de transcripción (bHLH, ABF1, MADS box) potencialmente implicados en la regulación de la respuesta a sequía, tales como la síntesis de prolina y las respuestas a carencias de agua o fosfato, así como a estímulos por temperatura. Los análisis de expresión diferencial de genes en los ecotipos han detectado 4941 tránscritos, de los cuales dos terceras partes están sobre-expresados en las plantas en condiciones de sequía con respecto a las sometidas a condiciones de riego. Los análisis pan-transcriptómicos muestran que la mayoría de los genes expresados en ambas condiciones son genes del *core*, presentes en todos los ecotipos estudiados, mientras que una fracción de los genes *hub* corresponden a genes *soft-core* y *shell*, encontrados únicamente en algunos ecotipos.

GENERAL CONCLUSIONS

1- Evolutionary and biogeographic analyses of the 20 recognized taxa of *Brachypodium* based on five genes (three nuclear and two plastidic) indicate that approximately half of the species are diploids and half are allopolyploids. Allopolyploid allelic grafting in the branches of the diploid skeleton tree, using a minimum evolution criterion, recovers the homeologous lineages (subgenomes) present in the allopolyploids. The successive divergences of the annual diploid lineages (*B. stacei*, *B. distachyon*) took place during the Late Miocene-Pliocene in the Mediterranean region, and those of the core perennial diploid lineages (*B. arbuscula, B. genuense, B. sylvaticum, B. glaucovirens, B. pinnatum*-2x and *B. rupestre*-2x) during the Quaternary in the Mediterranean and Eurasian regions with sporadic colonizations of other continents. The respective splits of the allopolyploids' homeologous lineages span different evolutionary depths. Our biogeographic scenario supports the occurrence of long distance dispersals only in diploid lineages, while all the hybridizations and genome doublings events occurred within the recentmost parental ancestral ranges without further range expansion.

2- Phylogenomic analyses of RNA-seq and GBS data identify *B. mexicanum* as the oldest allopolyploid species showing both ancestral-type (A) and maternal stacei-type (B) (Mid-late Miocene) subgenomes. The high ploidy level allopolyploids *B. boissieri* and *B. retusum* show three and four subgenomes respectively. Both species have A and B plus an intermediately evolved distachyon-type subgenome (C) (Miocene-Pliocene; maternally inherited in *B. boissieri*); *B. retusum* also shows a recently evolved core perennial-type maternal subgenome (D) (Quaternary). Core perennial allotetraploids *B. rupestre* and *B. phoenicoides* only present recently evolved C and D subgenomes (Quaternary); perennial diploids *B. pinnatum* and *B. sylvaticum* are resolved as their respective maternal parents. The recent allotetraploid *B. hybridum* originated repeatedly during the Quaternary from bidirectional crosses; the species is the only allopolyploid with known current diploid *B. stacei* and *B. distachyon* parents.

3- Pan-transcriptomic analysis of 5,202 transcript clusters of *Brachypodium* shows privately expressed genes in perennial (30 genes), annual (49), polyploids (14), old allopolyploids (143), and main ancestral (14) and recently evolved (52) groups. Exclusive transcripts of the old allopolyploids could be associated with the ancestral genome A. The transcripts annotated as RNA polymerase subunit, found only in all annual species of *Brachypodium*, could indicate differences in expression levels of RNAPs between annuals and perennial species, or the loss of ancestral copies in the more recently evolved perennial species.

4- Pan-genomic plastome analysis across 53 *B. distachyon*, 3 *B. hybridum* and 1 *B. stacei* accessions detects a major insertion (1,161 bp) and a rps19 gene copy deletion as distinctive arrangements of the *B. stacei* and *B. hybridum* plastomes with respect to the *B. distachyon* plastomes, and no variation in plastome gene content within *B. distachyon*.

5- The *B. distachyon* plastome tree shows the split of two main lineages, an Extremely Delayed Flowering (EDF+) clade and a Spanish (S+) – Turkish (T+) clade, indicating that flowering time is a main factor driving intraspecific divergence in this species. Topological comparison between the *B. distachyon* plastome and nuclear trees reveals nine chloroplast capture and two introgression and micro-recombination events across the main clades, supporting the existence of gene flow between the isolated lineages. Swapping of plastomes between the three different genomic groups, EDF+, T+, S+, likely resulted from random backcrossing followed by stabilization through selection pressure.

6- Weighted gene co-expression network analysis conducted in 33 B*. distachyon* ecotypes under drought and water conditions identifies five exclusive drought modules, including 465 isoforms and 11 highly interconnected (hub) genes. The analysis detects candidate genes and transcriptional factors (bHLH, ABF1, MADS box) potentially involved in the regulation of drought response, like proline synthesis and responses to water deprivation, phosphate starvation and temperature stimulus. Differential gene expression analysis yields 4,941 transcripts, of which two-thirds are over-expressed in dry with respect to water conditions. Pan-transcriptome analysis shows that most genes expressed in both conditions are core genes, present in all ecotypes studied, though a fraction of the hub genes corresponds to soft-core and shell genes, only found in some ecotypes.

REFERENCES

Abe, F., Saito, K., Miura, K., & Toriyama, K. (2002) A single nucleotide polymorphism in the alternative oxidase gene among rice varieties differing in low temperature tolerance. *FEBS Letters*, **527**, 181–185.

Acosta, M., Moscone, E., & Cocucci, A. (2015) Using chromosomal data in the phylogenetic and molecular dating framework: Karyotype evolution and diversification in *Nierembergia* (Solanaceae) influenced by historical changes in sea level. *Plant Biology*, **18**, 514–526.

Agrawal, P.K., Babu, B.K., & Saini, N. (2015) Omics of Model Plants. *PlantOmics: The Omics of Plant Science* (ed. by D. Barh, M.S. Khan, and E. Davies), pp. 1–32. Springer, New Delhi, India.

Albert, R., Jeong, H., & Barabási, A.-L. (2000) Error and Attack Tolerance of Complex Networks. *Nature*, **406**, 378–382.

Amrine, K.C.H., Blanco-Ulate, B., & Cantu, D. (2015) Discovery of Core Biotic Stress Responsive Genes in *Arabidopsis* by Weighted Gene Co-Expression Network Analysis. *PLoS ONE*, **10**, e0118731.

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. .

Aoki, K., Ogata, Y., & Shibata, D. (2007) Approaches for Extracting Practical Information from Gene Co-expression Networks in Plant Biology. *Plant Cell Physiology*, **48**, 381–390.

Aoki, Y., Okamura, Y., Tadaka, S., Kinoshita, K., & Obayashi, T. (2016) ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. *Plants and Cell Physiology*, **57**, e5(1–9).

Arenas, M. (2015) Trends in substitution models of molecular evolution. *Frontiers in Genetics*, **6**, 1–9.

Arthan, W., McKain, M.R., Traiperm, P., Welker, C.A.D., Teisher, J.K., & Kellogg, E.A. (2017) Phylogenomics of Andropogoneae (Panicoideae: Poaceae) of Mainland Southeast Asia. *Systematic Botany*, **42**, 418–431.

Baek, D., Chun, H.J., Yun, D.-J., & Kim, M.C. (2017) Cross-talk between Phosphate Starvation and Other Environmental Stress Signaling Pathways in Plants.

Molecules and Cells, **40**, 697–705.

Bakker, E.G., Montgomery, B., Nguyen, T., Eide, K., Chang, J., Mockler, T.C., Liston, A., Seabloom, E.W., & Borer, E.T. (2009) Strong population structure characterizes weediness gene evolution in the invasive grass species *Brachypodium distachyon*. *Molecular Ecology*, **18**, 2588–2601.

Baltisberger, M. & Hörandl, E. (2016) Karyotype evolution supports the molecular phylogeny in the genus *Ranunculus* (Ranunculaceae). *Journal of PPEES Sources*, **18**, 1–14.

Bareither, N., Scheffel, A., & Metz, J. (2017) Distribution of polyploid plants in the common annual *Brachypodium distachyon* (s.l.) in Israel is not linearly correlated with aridity. *Israel Journal of Plant Sciences*, 1–10.

Barker, N.P., Linder, H.P., & Harley, E.H. (1995) Polyphyly of Arundinoideae (Poaceae): Evidence from rbcL Sequence Data. *Systematic Botany*, **20**, 423–435.

Barres, L., Sanmartín, 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.

Batley, J. (2015) *Plant Genotyping. Methods and Protocols.* Springer Science+Business Media, New York.

Beerling, D.J. & Royer, D.L. (2011) Convergent Cenozoic CO_2 history. *Nature Geoscience*, **4**, 418–420.

Beissinger, T.M., Hirsch, C.N., Sekhon, R.S., Foerster, J.M., Johnson, J.M., Muttoni, G., Vaillancourt, B., Buell, C.R., Kaeppler, S.M., & Leon, N. De (2013) Marker Density and Read Depth for Genotyping Populations Using Genotyping-by-Sequencing. *Genetics*, **193**, 1073–1081.

Betekhtin, A., Jenkins, G., & Hasterok, R. (2014) Reconstructing the Evolution of *Brachypodium* Genomes Using Comparative Chromosome Painting. *PLoS ONE*, **9**, e115108.

Bhattacharya, A. & Knoll, J.E. (2012) Conventional and molecular breeding for improvement of biofuel crops. Past, present, and future. *Handbook of Bioenergy Crop Plants* (ed. by C. Kole, C.P. Joshi, and D.R. Shonnard), pp. 874. CRC Press, Boca

Raton, FL.

Blair, J., Nippert, J., & Briggs, J. (2014) Grassland Ecology. *Ecology and the Environment, The Plant Sciences 8* (ed. by R. Monson), pp. 389–423. Springer Science+Business Media, New York, NY.

Blattner, F.R. (2006) Multiple intercontinental dispersals shaped the distribution area of *Hordeum* (Poaceae). *New Phytologist*, **169**, 603–614.

Boetzer, M., Henkel, C. V., Jansen, H.J., Butler, D., & Pirovano, W. (2011) Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, **27**, 578–579.

Boetzer, M. & Pirovano, W. (2012) Toward almost closed genomes with GapFiller. *Genome biology*, **13**, R56.

Bohnert, H.J., Nelson, D.E., & Jensenayb, R.G. (1995) Adaptations to Environmental Stresses. *The Plant Cell*, **7**, 1099–1111.

Bolger, A.M., Lohse, M., & Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Bombarely, A., Coate, J.E., & Doyle, J.J. (2014) Mining transcriptomic data to study the origins and evolution of a plant allopolyploid complex. *PeerJ*, **2**, e391.

Borah, P., Sharma, E., Kaur, A., Chandel, G., Mohapatra, T., Kapoor, S., & Khurana, J. (2017) Analysis of drought-responsive signalling network in two contrasting rice cultivars using transcriptome-based approach. *Scientific Report*, **42131**, 1–21.

Borsch, T. & Quandt, D. (2009) Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution*, **282**, 169–199.

Bortiri, E., Coleman-Derr, D., Lazo, G.R., Anderson, O.D., & Gu, Y.Q. (2008) The complete chloroplast genome sequence of Brachypodium distachyon: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Research Notes*, **1**, 61.

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., Ní 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., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., & Drummond, A.J. (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PloS Computational Biology*, **10**, e1003537.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A.J., Poux, S., Bougueleret, L., & Xenarios, I. (2016) UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. *Plant Bioinformatics: Methods and Protocols* (ed. by D. Edwards), pp. 23–54. Springer New York, New York, NY.

Bradshaw, J.E. (2017) Plant breeding: past, present and future. *Euphytica*, **213**, 1–12.

Brassac, J. & Blattner, F.R. (2015) Species-Level Phylogeny and Polyploid Relationships in Hordeum (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci. *Systematic Biology*, **64**, 792–808.

Bray, N.L., Pimentel, H., Melsted, P., & Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, **34**, 525–527.

Bremer, K. (2002) Gondwanan Evolution of the Grass Alliance of Families (Poales). *Evolution*, **56**, 1374–1387.

Brkljacic, J., Grotewold, E., Scholl, R., et al. (2011) *Brachypodium* as a Model for the Grasses: Today and the Future. *Plant Physiology*, **157**, 3–13.

Brummer, E.C., Bouton, J.H., Casler, M.D., McCaslin, M.H., & Waldron, B.L. (2009) Grasses and Legumes: Genetics and Plant Breeding. *Grassland: Quietness and Strength for a New American Agriculture* (ed. by W.F. Wedin and S.L. Fales), pp. 157–171. ASA, CSSA,SSSA, Madison, WI.

Brysting, A.K., Oxelman, B., Huber, K.T., Moulton, V., & Brochmann, C. (2007) Untangling Complex Histories of Genome Mergings in High Polyploids. *Systematic Biology*, **56**, 467–476.

Buerki, S., Forest, F., Alvarez, N., Nylander, J.A.A., Arrigo, N., & Sanmartín, I. (2011) 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.

Buermans, H.P.J. & den Dunnen, J.T. (2014) Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, **1842**, 1932–1941.

Buggs, R.J.A., Renny-Byfield, S., Chester, M., Jordon-Thaden, I.E., Viccini, L.F., Chamala, S., Leitch, A.R., Schnable, P.S., Bradley Barbazuk, W., Soltis, P.S., & Soltis, D.E. (2012) Next-generation sequencing and genome evolution in allopolyploids. *American Journal of Botany*, **99**, 372–382.

Byars, S.G., Parsons, Y., & Hoffmann, A.A. (2009) Effect of altitude on the genetic structure of an Alpine grass , Poa hiemata. *Annals of Botany*, **103**, 885–899.

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L.A., & Spooner, D.M. (2012) Single copy nuclear gene analysis of polyploidy in wild potatoes (*Solanum* section Petota). *BMC evolutionary biology*, **12**, 1–16.

Camacho, C. (2013) BLAST+. https://www.ncbi.nlm.nih.gov/books/NBK131777/.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T.L. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**, 1–9.

Cantalapiedra, C.P., García-pereira, M.J., Gracia, M.P., Igartua, E., Casas, A.M., & Contreras-Moreira, B. (2017) Large Differences in Gene Expression Responses to Drought and Heat Stress between Elite Barley Cultivar Scarlett and a Spanish Landrace. *Frontiers in plant science*, **8**, 647.

Capella-Gutiérrez, S., Silla-Martínez, J.M., & Gabaldón, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Hub, A., & Presence, W. (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., & Dopazo, J. (2015) A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Molecular biology and evolution*, **32**, msv082–.

Carlsen, T., Bleeker, W., Hurka, H., Elven, R., & Brochmann, C. (2009) Biogeography and phylogeny of *Cardamine* (Brassicaceae). *Annals of the Missouri Botanical Garden*, **96**, 215–236.

Carlson, M.R.J., Zhang, B., Fang, Z., Mischel, P.S., Horvath, S., & Nelson, S.F. (2006) Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. *BMC genomics*, **7**, 1–15.

Castilhos, G., Lazzarotto, F., Spagnolo-Fonini, L., Helena, M., & Margis-Pinheiro, M. (2014) Possible roles of basic helix-loop-helix transcription factors in adaptation to drought. *Plant Science*, **223**, 1–7.

Catalán, P., Chalhoub, B., Chochois, V., Garvin, D.F., Hasterok, R., Manzaneda, A.J., Mur, L.A.J., Pecchioni, N., Rasmussen, S.K., Vogel, J.P., & Voxeur, A. (2014) Update on the genomics and basic biology of *Brachypodium*. *Trends in Plant Science*, **19**, 414–418.

Catalán, 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.

Catalán, P., López-Alvarez, D., Bellosta, C., & Villar, L. (2016a) Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon, B. stacei*, and *B. hybridum* (Poaceae). *Anales del Jardín Botánico de Madrid*, **73**, 1–14.

Catalán, P., López-Alvarez, D., Díaz-Pérez, A., Sancho, R., & López-Herranz, M.L. (2016b) Phylogeny and Evolution of the Genus Brachypodium. *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models* (ed. by J.P. Vogel), pp. 9–38. Springer.

Catalán, 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.

Catalán, P. & Olmstead, R.G. (2000) Phylogenetic reconstruction of the genus *Brachypodium* P. Beauv. (Poaceae) from combined sequences of chloroplast ndhF gene and nuclear ITS. *Plant Systematics and Evolution*, **220**, 1–19.

Chaves, M.M., Maroco, J.P., & Pereira, J.S. (2003) Understanding plant responses to drought — from genes to whole plant. *Functional Plant Biology*, **30**, 239–264.

Chaw, S.M., Chang, C.C., Chen, H.L., & Li, W.H. (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution*, **58**, 424–441.

Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., & Wang, X. (2018) Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants*, **4**, 258–268.

Chernomor, O., von Haeseler, A., & Minh, B.Q. (2016) Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology*, **65**, 997–1008.

Childs, K.L., Davidson, R.M., & Buell, C.R. (2011) Gene Coexpression Network Analysis as a Source of Functional Annotation for Rice Genes. *PLoS ONE*, **6**, e22196.

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.

Christmann, A. & Grill, E. (2018) Peptide signal alerts plants to drought. *Nature*, **556**, 178–179.

Clark, L.G., Zhang, W., & Wendel, J.F. (1995) A Phylogeny of the Grass Family (Poaceae) Based on ndhF Sequence. *Systematic Botany*, **20**, 436–460.

Clement, M., Posada, D., & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Clement, M., Snell, Q., Walker, P., Posada, D., & Crandall, K. (2002) TCS: Estimating Gene Genealogies. *Proceeding 16th International Parallel Distributed Processing Symposium*, 184.

Clevenger, J., Chavarro, C., Pearl, S.A., Ozias-akins, P., & Jackson, S.A. (2015) Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. *Molecular Plant*, **8**, 831–846.

Colton-Gagnon, K., Ali-Benali, M.A., Mayer, B.F., Dionne, R., Bertrand, A., Do Carmo, S., & Charron, J.B. (2014) Comparative analysis of the cold acclimation and freezing tolerance capacities of seven diploid *Brachypodium distachyon* accessions. *Annals of Botany*, **113**, 681–693.

Contreras-Moreira, B., Cantalapiedra, C.P., García-Pereira, M.J., Gordon, S.P., Vogel, J.P., Igartua, E., Casas, A.M., & Vinuesa, P. (2017) Analysis of Plant Pan-Genomes and Transcriptomes with GET _ HOMOLOGUES-EST, a Clustering Solution for Sequences of the Same Species. *Frontiers in Genetics*, **8**, 1–16.

Contreras-Moreira, B., Castro-Mondragon, J.A., Rioualen, C., Cantalapiedra, C.P., & van Helden, J. (2016) RSAT::Plants: Motif Discovery Within Clusters of Upstream Sequences in Plant Genomes. *Plant synthetic promoters* (ed. by R. Hehl), pp. 279–295. Springer Science+Business Media, New York.

Corley, S.M., MacKenzie, K.L., Beverdam, A., Roddam, L.F., & Wilkins, M.R. (2017) Differentially expressed genes from RNA-Seq and functional enrichment results are affected by the choice of single-end versus paired-end reads and stranded versus non-stranded protocols. *BMC Genomics*, **18**, 1–13.

CPWG (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden*, **88**, 373–457.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., Mcvean, G., Durbin, R., Project, G., & Vcf, T. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*, **17**, 134.

Darlington, C.D. (1937) *Recent advances in cytology.* Blakiston's Son & Co, Philadelphia.

Darriba, D., Taboada, G.L., Doallo, R., & Posada, D. (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., & Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews. Genetics*, **12**, 499–510.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S., Jiang, N., & Buell, C.R. (2012) Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *The Plant Journal*, **71**, 492–502.

Davis, J.I. (1997) Evolution, Evidence, and the Role of Species Concepts in Phylogenetics. *Systematic Botany*, **22**, 373–403.

Davis, J.I. & Soreng, R.J. (1993) Phylogenetic structure in the grass family (Poaceae) as inferred from chloroplast DNA restriction site variation. *American Journal of Botany*, **80**, 1444–1454.

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: A Journal of Systematics and Evolutionary Botany*, **23**, 335–348.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005) Phylogenomics and the reconstruction of the tree of life. *Nature reviews. Genetics*, **6**, 361–375.

Deschamps, S., Llaca, V., & May, G.D. (2012) Genotyping-by-Sequencing in Plants. *Biology*, **1**, 460–483.

Des Marais, D.L., Guerrero, R.F., Lasky, J.R., & Scarpino, S. V (2017a) Topological features of gene regulatory networks predict patterns of natural diversity in environmental response. *Proceedings of the Royal Society B: Biological Sciences*, **284**, .

Des Marais, D.L. & Juenger, T.E. (2016) Brachypodium and the Abiotic Environment. *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models, volume 18* (ed. by J.P. Vogel), pp. 291–311. Springer, Switzerland.

Des Marais, D.L., Lasky, J.R., Verslues, P.E., Chang, T.Z., & Juenger, T.E. (2017b) Interactive effects of water limitation and elevated temperature on the physiology, development and fitness of diverse accessions of *Brachypodium distachyon*. *New Phytologist*, **214**, 132–144.

Des Marais, D.L., Mckay, J.K., Richards, J.H., Sen, S., Wayne, T., & Juenger, T.E. (2012) Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. *The plant Cell*, **24**, 893–914.

Desper, R. & Gascuel, O. (2002) Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology*, **9**, 687–705.

Díaz-Pérez, A.J., Sharifi-Tehrani, M., Inda, L.A., & Catalán, P. (2014) Molecular Phylogenetics and Evolution 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.

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic acids research*, **45**, e18.

Dijk, E.L. Van, Jaszczyszyn, Y., & Thermes, C. (2014) Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research*, **322**, 12–20.

Dinh Thi, V.H., Coriton, O., Le Clainche, I., Arnaud, D., Gordon, S.P., Linc, G., Catalán, P., Hasterok, R., Vogel, J.P., Jahier, J., & Chalhoub, B. (2016) Recreating Stable *Brachypodium hybridum* Allotetraploids by Uniting the Divergent Genomes of B. distachyon and B. stacei. *PloS one*, **11**, e0167171.

Dolezel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysákt, M.A., Nardi, L., & Obermayer, R. (1998) Plant Genome Size Estimation by Flow Cytometry: Inter-laboratory Comparison*. *Annals of Botany*, **82**, 17–26.

Dolezel, J., Sgorbati, S., & Lucretti, S. (1992) Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiologia Plantarum*, **85**, 625–631.

Dong, J. & Horvath, S. (2007) Understanding network concepts in modules. *BMC Systems Biology*, **1**, 1–20.

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

Draper, J., Mur, L.A.J., Jenkins, G., Ghosh-Biswas, G.C., Bablak, P., Hasterok, R., & Routledge, A.P.M. (2001) *Brachypodium distachyon*. A New Model System for Functional Genomics in Grasses. *Plant physiology*, **127**, 1539–1555.

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**, 1969–1973.

Duitama, J., Quintero, J.C., Cruz, D.F., Quintero, C., Hubmann, G., Foulquié-Moreno, M.R., Verstrepen, K.J., Thevelein, J.M., & Tohme, J. (2014) An integrated framework for discovery and genotyping of genomic variants from high- throughput sequencing. *Nucleic Acids Research*, **42**, e44.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edwards, E.J., Osborne, C.P., Stromberg, C.A.E., et al. (2010) The Origins of C4 Grasslands: Integrating Evolutionary and Ecosystem Science. *Science*, **328**, 587–591.

Eisen, J.A. (1998) Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. *Genome Research*, **8**, 163–167.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., & Mitchell, S.E. (2011) A Robust , Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, **6**, e19379.

Escudero, M., Eaton, D.A.R., Hahn, M., & Hipp, A.L. (2014) Molecular Phylogenetics and Evolution Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Estep, M.C., Mckain, M.R., Vela Díaz, D., Zhong, J., Hodge, J.G., Hodkinson, T.R., Layton, D.J., Malcomber, S.T., Pasquet, R., & Kellogg, E.A. (2014) Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 15149–15154.

Evanno, G., Regnaut, S., & Goudet, J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, **14**, 2611–2620.

Farris, J.S. (1983) The Logical Basis of Phylogenetic Analysis. *Advances in Cladistics* (ed. by N. Platnick and V. Funk), pp. 1–36. Columbia University Press, New York.

Fedurco, M., Romieu, A., Williams, S., Lawrence, I., & Turcatti, G. (2006) BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*, **34**, .

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368–376.

Feschotte, C. & Pritham, E.J. (2007) DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics*, **41**, 331–368.

Filiz, E., Ozdemir, B.S., Budak, F., Vogel, J.P., Tuna, M., & Budak, H. (2009) Molecular, morphological, and cytological analysis of diverse *Brachypodium distachyon* inbred lines. *Genome*, **52**, 876–890.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-vegas, A., Salazar, G.A., Tate, J., & Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, **44**, 279–285.

Fisher, L.H., Han, J., Corke, F.M., Akinyemi, A., Didion, T., Nielsen, K.K., Doonan, J.H., Mur, L.A., & Bosch, M. (2016) Linking Dynamic Phenotyping with Metabolite Analysis to Study Natural Variation in Drought Responses of Brachypodium distachyon. *Frontiers in plant science*, **7**, 1–15.

Fitch, W.M. (1971) Toward Defining the Course of Evolution: Minimum Change for a Specific Tree Topology. *Systematic Zoology*, **20**, 406–416.

Fjellheim, S., Boden, S., & Trevaskis, B. (2014) The role of seasonal flowering responses in adaptation of grasses to temperate climates. **5**, 1–15.

Fougére-Danezan, M., Joly, S., Bruneau, A., Gao, X., & Zhang, L. (2015) Phylogeny and biogeography of wild roses with specific attention to polyploids. *Annals of Botany*, **115**, 275–291.

Garvin, D.F. (2007) *Brachypodium distachyon*: A New Model System for Structural and Functional Analysis of Grass Genomes. *Model Plants and Crop Improvement* (ed. by R.K. Varshney and R.M.D. Koebner), pp. 109–123.

Garvin, D.F., Gu, Y.Q., Hasterok, R., Hazen, S.P., Jenkins, G., Mockler, T.C., Mur, L.A.J., & Vogel, J.P. (2008) Development of genetic and genomic research resources for Brachypodium distachyon, a new model system for grass crop research. *Crop Science*, S69–S84.

Gaut, B.S. & Doebley, J.F. (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6809–6814.

Ge, S., Li, A., Lu, B.R., Zhang, S.Z., & Hong, D.Y. (2002) A phylogeny of the rice tribe Oryzeae (Poaceae) based on matK sequence data. *American Journal of Botany*, **89**, 1967–1972.

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.

Gibson, G. (2016) On the Evaluation of Module Preservation. *Cell Systems*, **3**, 17–19.

Gladman, S. & Seemann, T. (2012) VelvetOptimiser. http://www.vicbioinformatics.com/software.velvetoptimiser.shtml.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, **40**, 1178–1186.

Goodwin, S., Mcpherson, J.D., & McCombie, W.R. (2016) Coming of age: ten years of next- generation sequencing technologies. *Nature Publishing Group*, **17**, 333–351.

Gordon, S.P., Contreras-Moreira, B., Woods, D.P., et al. (2017) Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, **8**, 1-13.

Gordon, S.P., Liu, L., & Vogel, J.P. (2016) The Genus *Brachypodium* as a Model for Perenniality and Polyploidy. *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models* (ed. by J. Vogel), pp. 313–326. Springer, Switzerland.

Gottlieb, A., Müller, H.G., Massa, A.N., Wanjugi, H., Deal, K.R., You, F.M., Xu, X., Gu, Y.Q., Luo, M.C., Anderson, O.D., Chan, A.P., Rabinowicz, P., Devos, K.M., & Dvorak, J. (2013) Insular Organization of Gene Space in Grass Genomes. *PLoS ONE*, **8**, e54101.

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.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Trinity: reconstructing a full-length transcrptome without a genome from RNA-Seq data. *Nature Biotechnology*, **29**, 644–652.

Grass Phylogeny Working Group (2012) New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytologist*, **193**, 304–312.

Griffith, M., Walker, J.R., Spies, N.C., Ainscough, B.J., & Griffith, O.L. (2015) Informatics for RNA Sequencing: A Web Resource for Analysis on the Cloud. *PLoS Computational Biology*, **11**, 1–20.

Gruenstaeudl, M., Carstens, B.C., Santos-Guerra, A., & Jansen, R.K. (2017) Statistical hybrid detection and the inference of ancestral distribution areas. *Biological Journal of the Linnean Society*, **121**, 133–149.

Guggisberg, A., Mansion, G., Kelso, S., & Conti, E. (2006) Evolution of biogeographic patterns , ploidy levels , and breeding systems in a diploid – polyploid species complex of *Primula*. *New Phytologist*, **171**, 617–632.

Guindon, S. & Gascuel, O. (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, **52**, 696–704.

Guo, Y.L. & Ge, S. (2005) Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *American Journal of Botany*, **92**, 1548–1558.

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.

Hamon, P., Grover, C.E., Davis, A.P., Rakotomalala, J., Nathalie, E., Albert, V.A., Sreenath, H.L., Stoffelen, P., Mitchell, S.E., Hamon, S., Kochko, A. De, Crouzillat, D., Rigoreau, M., Sumirat, U., Akaffou, S., & Guyot, R. (2017) Molecular Phylogenetics and Evolution Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species. GBS coffee phylogeny and the evolution of caffeine content. *Molecular Phylogenetics and Evolution*, **109**, 351–361.

Hapke, A. & Thiele, D. (2016) GIbPSs: a toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular Ecology Resources*, **16**, 979–990.

Hayano-Kanashiro, C., Calderón-Vázquez, C., Ibarra-Laclette, E., Herrera-Estrella, L., & Simpson, J. (2009) Analysis of Gene Expression and Physiological Responses in Three Mexican Maize Landraces under Drought Stress and Recovery Irrigation. *PLoS ONE*, **4**, e7531.

He, F. & Maslov, S. (2016) Pan- and core- network analysis of co-expression genes in a model plant. *Scientific Report*, **6**, 1–11.

He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H., & Li, Z. (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in plant science*, **5**, 1–8.

Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., R, S.D., & Phillip, O. (2014) Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques*, **56**, 61.

Heather, J.M. & Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics*, **107**, 1–8.

Heled, J. & Drummond, A.J. (2012) Calibrated Tree Priors for Relaxed Phylogenetics and Divergence Time Estimation. *Systematic Biology*, **61**, 138–149.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey, J. & Nielsen, R. (2004) Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. *Genetics*, **167**, 747–760.

Hilu, K.W., Alice, L.A., & Liang, H. (1999) Phylogeny of Poaceae inferred from matK sequences. *Annals of the Missouri Botanical Garden*, **86**, 835–851.

Hochbach, A., Schneider, J., & Röser, M. (2015) A multi-locus analysis of phylogenetic relationships within grass subfamily Pooideae (Poaceae) inferred from sequences of nuclear single copy gene regions compared with plastid DNA. *Molecular Phylogenetics and Evolution*, **87**, 14–27.

Horvath, S. & Dong, J. (2008) Geometric Interpretation of Gene Coexpression Network Analysis. *PloS Computational Biology*, **4**, e1000117.

Hou, Z., Jiang, P., Swanson, S.A., Elwell, A.L., Nguyen, B.K.S., Bolin, J.M., Stewart, R., & Thomson, J.A. (2015) A cost-effective RNA sequencing protocol for large-scale gene expression studies. *Scientific Reports*, **5**, 1–5.

Hsiao, C., Chatterton, N.J., Asay, K.H., & Jensen, K.B. (1995) Molecular phylogeny of the Pooideae (Poaceae) based on nuclear rDNA (ITS) sequences. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, **90**, 389–398.

Huang, J., Vendramin, S., Shi, L., & McGinnis, K.M. (2017a) Construction and Optimization of a Large Gene Coexpression Network in Maize Using RNA-Seq Data. *Plant physiology*, **175**, 568–583.

Huang, Y.Y., Cho, S.T., Haryono, M., & Kuo, C.H. (2017b) Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLoS ONE*, **12**, 1–16.

Huber, K.T., Oxelman, B., Lott, M., & Moulton, V. (2006) Reconstructing the Evolutionary History of Polyploids from Multilabeled Trees. *Molecular Biology and Evolution*, **23**, 1784–1791.

Huelsenbeck, J.P. & Rannala, B. (2004) Frequentist Properties of Bayesian Posterior Probabilities of Phylogenetic Trees Under Simple and Complex Substitution Models. *Systematic Biology*, **53**, 904–913.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., & Bollback, J.P. (2001) Bayesian Inference of Phylogeney and Its Impact on Evolutionary Biology. *Science*, **294**, 2310–2314.

Huson, D.H. & Scornavacca, C. (2012) Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology*, **61**, 1061–1067.

IBI (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

Idziak, D., Hazuka, I., Poliwczak, B., Wiszynska, A., Wolny, E., & Hasterok, R. (2014) Insight into the Karyotype Evolution of Brachypodium Species Using Comparative Chromosome Barcoding. *PLoS ONE*, **9**, e93503.

Illumina Inc (2017) An Introduction to Next-Generation Sequencing Technology. 1–6. https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.

Inda, L.A., Sanmart, I., Buerki, S., & Catalán, 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, S.W.L. & Everett, J. (2000) *Grasses: Systematics and Evolution.* Csiro Publishing, 416 pp.

Janiak, A., Kwa, M., & Szarejko, I. (2015) Gene expression regulation in roots under drought. *Journal of Experimental Botany*, **67**, 1003–1014.

Jansen, R.K. & Ruhlman, T.A. (2012) Plastid genomes of seed plants. *Genomics of chloroplast and mitochondria* (ed. by R. Bock and V. Knoop), pp. 103–126. Springer,

Jenkins, D.G., Carey, M., Czerniewska, J., et al. (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, **33**, 315–320.

Jiao, W.B. & Schneeberger, K. (2017) The impact of third generation genomic technologies on plant genome assembly. *Current Opinion in Plant Biology*, **36**, 64–70.

Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., et al. (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biology*, **13**, 1–14.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, 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-Mack, J., & DePamphilis, C.W. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.

Johnston, J.S., Bennett, M.D., Rayburn, A.L., Galbraith, D.W., & Price, H.J. (1999) Reference standards for determination of DNA content of plant nuclei. *American Journal of Botany*, **86**, 609–613.

Jones, G., Sagitov, S., & Oxelman, B. (2013) Statistical Inference of Allopolyploid Species Networks in the Presence of Incomplete Lineage Sorting. *Systematic Biology*, **62**, 467–478.

Joshi, R., Wani, S.H., Singh, B., Bohra, A., Dar, Z.A., Lone, A.A., Pareek, A., & Singla-Pareek, S.L. (2016) Transcription Factors and Plants Response to Drought Stress: Current

Understanding and Future Directions. *Frontiers in plant science*, **7**, 1–15.

Judd, W.S., Campbell, C.S., Kellogg, E.A., Stevens, P.F., & Donoghue, M.J. (2008) *Plant Systematics. A Phylogenetic Approach.* Sinauer Associates, Sunderland, Massachusetts, USA.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., & Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**, 587–589.

Kamata, N., Okada, H., Komeda, Y., & Takahashi, T. (2013) Mutations in epidermis-specific HD-ZIP IV genes affect floral organ identity in *Arabidopsis thaliana*. *The Plant Journal*, **75**, 430–440.

Kamneva, O.K., Syring, J., Liston, A., & Rosenberg, N.A. (2017) Evaluating allopolyploid origins in strawberries (Fragaria) using haplotypes generated from target capture sequencing. *BMC Evolutionary Biology*, **17**, 1–19.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, **45**, D353–D361.

Kanehisa, M. & Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, **28**, 27–30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, **44**, D457–D462.

Kannan, L. & Wheeler, W.C. (2012) Maximum Parsimony on Phylogenetic networks. *Algorithms for Molecular Biology*, **7**, 1–10.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data.

Bioinformatics, **28**, 1647–1649.

Kellogg, E.A. (2001) Update on Evolution Evolutionary History of the Grasses. *Plant Physiology*, **125**, 1198–1205.

Kellogg, E.A. (2015a) *The Families and Genera of Vascular Plants. Vol. XIII. Flowering Plants. Monocots. Poaceae.* Springer, New York.

Kellogg, E.A. (2015b) *Brachypodium distachyon* as a Genetic Model System. *Annual reviews of genetics*, **49**, 1–20.

Khan, M.A. & Stace, C.A. (1999) Breeding relationships in the genus *Brachypodium* (Poaceae: Pooideae). *Nordic Journal of Botany*, **19**, 257–269.

Kidd, K.K. & Sgaramella-Zonta, L.A. (1971) Phylogenetic Analysis: Concepts and Methods. *American Journal of Human Genetics*, **23**, 235–252.

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L., & Paterson, A.H. (2016) Plant Science Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Science*, **242**, 14–22.

Kim, D., Langmead, B., & Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.

Krijgsman, W. (2002) The Mediterranean: Mare Nostrum of Earth sciences. *Earth and Planetary Science Letters*, **205**, 1–12.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., & Marra, M.A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Kumar, S., Stecher, G., & Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870-1874.

Langfelder, P. & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 1–13.

Langfelder, P. & Horvath, S. (2010) Available at: https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/ModulePreservation/Tutorials/glossaryTable.pdf.

Leary, N.A.O., Wright, M.W., Brister, J.R., et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, **44**, 733–745.

Leigh, J.W. & Bryant, D. (2015) POPART: full-feature software for haplotype network construction. *Molecular in Ecology and Evolution*, **6**, 1110–1116.

Leitch, I.J. & Bennett, M.D. (2004) Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*, **82**, 521–536.

Leonelli, S. & Ankeny, R.A. (2013) What makes a model organism? *Endeavour*, **37**, 209–212.

Levin, D.A. (2013) The timetable for allopolyploidy in flowering plants. *Annals of Botany*, **112**, 1201–1208.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li, M., Copeland, A., & Han, J. (2011a) DUK – A Fast and Efficient Kmer Matching Tool. *Lawrence Berkeley National Laboratory. LBNL Paper LBNL-4516E-Poster p*, https://www.osti.gov/servlets/purl/1016000.

Li, M., Copeland, A., & Han, J. (2011b) DUK - A Fast and Efficient Kmer Based Sequence Matching Tool. http://duk.sourceforge.net/.

Li, S., Pearl, D.K., & Doss, H. (2000) Phylogenetic Tree Construction Using Markov Chain Monte Carlo. *Journal of the American Statistical Association*, **95**, 493–508.

Li, W. & Cui, X. (2014) A Special Issue on Plant Stress Biology: From Model Species to Crops. *Molecular Plant*, **7**, 755–757.

Li, Y., Zuo, S., Zhang, Z., Li, Z., Han, J., Chu, Z., Hasterok, R., & Wang, K. (2018) Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *The Plant Journal*, **93**, 1088–1101.

Liang, H. & Hilu, K.W. (1996) Application of the matK gene sequences to grass systematics. *Canadian Journal of Botany*, **74**, 125–134.

Librado, P. & Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Linder, H.P. & Barker, N.P. (2014) Does polyploidy facilitate long-distance dispersal? *Annals of Botany*, **113**, 1175–1183.

Lipscomb, D. (1998) Basics of Cladistic Analysis. http://taxonomy.zoology.gla.ac.uk/teaching/Cladistics.pdf.

Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., & Guan, X. (2012a) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC genomics*, **13**, 715.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012b) Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, **2012**, 1–11.

Liu, Y., Schröder, J., & Schmidt, B. (2013) Musket: A multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics*, **29**, 308–315.

Lohman, B.K., Weber, J.N., & Bolnick, D.I. (2016) Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources*, **16**, 1315–1321.

Lohse, M., Drechsel, O., Kahlau, S., & Bock, R. (2013) OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic acids research*, **41**, 575–581.

López-Alvarez, D., López-Herranz, M.L., Betekhtin, A., & Catalán, P. (2012) A DNA Barcoding Method to Discriminate between the Model Plant *Brachypodium distachyon* and Its Close Relatives *B. stacei* and *B. hybridum* (Poaceae). *PLoS ONE*, **7**, e51058.

López-Alvarez, D., Manzaneda, A.J., Rey, P.J., Giraldo, P., Benavente, E., Allainguillaume, J., Mur, L., Caicedo, A.L., Hazen, S.P., Breiman, A., Ezrati, S., & Catalan, P. (2015) Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. *American Journal of Botany*, **102**, 1073–1088.

López-Álvarez, D., Zubair, H., Beckmann, M., Draper, J., & Catalán, P. (2017) Diversity and association of phenotypic and metabolomic traits in the close model grasses *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Annals of Botany*, **119**, 545–561.

Lowry, D.B., Modliszewski, J.L., Wright, K.M., Wu, C.A., & Willis, J.H. (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical transactions of the Royal Society*, **363**, 3009–3021.

Luo, M.C., Deal, K.R., Akhunov, E.D., et al. (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proceedings of the National Academy of Sciences*, **106**, 15780–15785.

Lyons, C.W.P. & Scholthof, K.-B.G. (2016) *Brachypodium* as an *Arabidopsis* for the grasses: Are we there yet? *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models* (ed. by J.P. Vogel), pp. 327–342. Springer, Switzerland.

Ma, P.F., Zhang, Y.X., Zeng, C.X., Guo, Z.H., & Li, D.Z. (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Systematic Biology*, **63**, 933–950.

Ma, X.F. & Gustafson, J.P. (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenetic and Genome Research*, **109**, 236–249.

Madlung, A. (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity*, **110**, 99–104.

Mairal, M., Pokorny, L., Aldasoro, J.J., Alarcón, M.L., & Sanmartín, I. (2015) Ancient vicariance and climate-driven extinction explain continental-wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). *Molecular ecology*, **24**, 1335–1354.

Manzaneda, A.J., Rey, P.J., Anderson, J.T., Raskin, E., Weiss-Lehman, C., & Mitchell-Olds, T. (2015) Natural variation, differentiation, and genetic trade-offs of ecophysiological traits in response to water limitation in *Brachypodium distachyon* and its descendent allotetraploid *B. hybridum* (Poaceae). *Evolution*, **69**, 2689–2704.

Manzaneda, A.J., Rey, P.J., Bastida, J.M., Weiss-Lehman, C., Raskin, E., & Mitchell-Olds, T. (2012) Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytologist*, **193**, 797–805.

Mao, H., Wang, H., Liu, S., Li, Z., Yang, X., Yan, J., Li, J., Tran, L.P., & Qin, F. (2015) A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nature Communications*, **6**, 1–13.

Mao, L., Hemert, J.L. Van, Dash, S., & Dickerson, J.A. (2009) Arabidopsis gene co-expression network and its functional modules. *BMC Bioinformatics*, **10**, 1–24.

Marcussen, T., Heier, L., Brysting, A.K., Oxelman, B., & Jakobsen, K.S. (2015) From Gene Trees to a Dated Allopolyploid Network: Insights from the Angiosperm Genus *Viola* (Violaceae). *Systematic Biology*, **64**, 84–101.

Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X., & Olsen, O.A. (2014) Ancient hybridizations among the ancestral genomes of bread wheat. *Science*, **345**, .

Mardis, E.R. (2013) Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, **6**, 287–303.

Marques, I., Shiposha, V., López-Alvarez, D., Manzaneda, A.J., Hernandez, P., Olonova, M., & Catalán, P. (2017) Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*. *BMC Evolutionary Biology*, **17**, 1–14.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, **1**, 1–5.

Martínez, L.M., Fernández-ocaña, A., Rey, P.J., Salido, T., Amil-ruiz, F., & Manzaneda, A.J. (2018) Variation in functional responses to water stress and differentiation between natural allopolyploid populations in the *Brachypodium distachyon* species complex. *Annals of Botany*, **00**, 1–14.

Masalia, R.R., Bewick, A.J., & Burke, J.M. (2017) Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. *PLoS ONE*, **12**, e0182289.

Mascher, M., Gundlach, H., Himmelbach, A., et al. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature*, **544**, 427–433.

Mathews, S. & Sharrock, R.A. (1996) The phytochrome gene family in grasses (Poaceae): A phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. *Molecular Biology and Evolution*, **13**, 1141–1150.

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.

Matsuoka, Y., Takumi, S., & Nasuda, S. (2014) Genetic Mechanisms of Allopolyploid Speciation Through Hybrid Genome Doubling. Novel Insights from Wheat (*Triticum* and *Aegilops*) Studies. *International Review of Cell and Molecular Biology* (ed. by K.W. Jeon), pp. 199–258. Academic Press.

Matsuoka, Y., Yamazaki, Y., Ogihara, Y., & Tsunewaki, K. (2002) Whole Chloroplast Genome Comparison of Rice, Maize, and Wheat: Implications for Chloroplast Gene Diversification and Phylogeny of Cereals. *Molecular Biology and Evolution*, **19**, 2084–2091.

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.

Mattioli, R., Costantino, P., & Trovato, M. (2009) Proline accumulation in plants. Not only stress. *Plant Signaling and Behavior*, **4**, 1016–1018.

Mau, B., Newton, M.A., & Larget, B. (1999) Bayesian Phylogenetic Inference via Markov Chain Monte Carlo Methods. *Biometrics*, **55**, 1–12.

Mayrose, I., Zhan, S.H., Rothfels, 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.

Medgyesy, P., Fejes, E., & Maliga, P. (1985) Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. *Proceedings of the National Academy of Sciences of the United States of America*, **82**, 6960–6964.

Meijer, P.T. & Krijgsman, W. (2005) A quantitative analysis of the desiccation and refilling of the Mediterranean during the Messinian Salinity Crisis. *Earth and Planetary Science Letters*, **240**, 510–520.

Meimberg, H., Rice, K.J., Milan, N.F., Njoku, C.C., & McKay, J.K. (2009) Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany*, **96**, 1262–1273.

Metzker, M.L. (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31–46.

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 African-Eurasian convergent plate boundary zone. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **196**, 209–228.

Meyer, E., Aglyamova, G. V., & Matz, M. V. (2011) Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology*, **20**, 3599–3616.

Meyer, S. & Haeseler, A. Von (2003) Identifying Site-Specific Substitution Rates. *Molecular Biology and Evolution*, **20**, 182–189.

Mi, H., Muruganujan, A., & Thomas, P.D. (2013) PANTHER in 2013: modeling the evolution of gene function, and other gene attribute , in the context of phylogenetic trees. *Nucleic Acids Research*, **41**, D377–D386.

Miao, Z., Han, Z., Zhang, T., Chen, S., & Ma, C. (2017) A systems approach to a spatio-temporal understanding of the drought stress response in maize. *Scientific Reports*, **7**, 1–14.

Middleton, C.P., Senerchia, N., Stein, N., Akhunov, E.D., Keller, B., Wicker, T., & Kilian, B. (2014) Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the triticeae tribe. *PLoS ONE*, **9**, e85761.

Minaya, M., Díaz-Pérez, A., Mason-Gamer, R., Pimentel, M., & Catalán, P. (2015) Evolution of the beta-amylase gene in the temperate grasses: Non-purifying selection, recombination, semiparalogy, homeology and phylogenetic signal.

Molecular Phylogenetics and Evolution, **91**, 68–85.

Minaya, M., Hackel, J., Namaganda, M., Brochmann, C., Vorontsova, M.S., Besnard, G., & Catalán, P. (2017) Contrasting dispersal histories of broad- and fine-leaved temperate Loliinae grasses: range expansion, founder events, and the roles of distance and barriers. *Journal of Biogeography*, 1–14.

Minh, B.Q., Nguyen, M.A.T., & von Haeseler, A. (2013) Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, **30**, 1188–1195.

Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T., & Shinozaki, K. (2011) Global Landscape of a Co-Expressed Gene Network in Barley and its Application to Gene Discovery in Triticeae Crops. *Plants and Cell Physiology*, **52**, 785–803.

Morris, L.M. & Duvall, M.R. (2010) The chloroplast genome of *Anomochloa marantoidea* (Anomochlooideae; Poaceae) comprises a mixture of grass-like and unique features. *American Journal of Botany*, **97**, 620–627.

Mun, B.-G., Lee, S.-U., Park, E.-J., Kim, H.-H., Hussain, A., Muhammad Imran, Q., Lee, I.-J., & Yun, B.-W. (2017) Analysis of transcription factors among differentially expressed genes induced by drought stress in Populus davidiana. *3 Biotech*, **7**, 1–12.

Mur, L.A.J., Allainguillaume, J., Catalán, P., Hasterok, R., Jenkins, G., Lesniewska, K., Thomas, I., & Vogel, J. (2011) Exploiting the *Brachypodium* tool box in cereal and grass research. *New Phytologist*, **191**, 334–347.

Murat, F., Xu, J., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., & Messing, J. (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Research*, **20**, 1545–1557.

Nadalin, F., Vezzi, F., & Policriti, A. (2012) GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics*, **13**, S8.

Nadot, S., Bajon, R., & Lejeune, B. (1994) The Chloroplast Gene Rps4 as a Tool for the Study of Poaceae Phylogeny. *Plant Systematics and Evolution*, **191**, 27–38.

Nakashima, K., Ito, Y., & Yamaguchi-Shinozaki, K. (2009) Transcriptional Regulatory Networks in Response to Abiotic Stresses in *Arabidopsis* and Grasses. *Plant Physiology*, **149**, 88–95.

Nakashima, K., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Frontiers in plant science*, **5**, 1–7.

Nakato, R. & Gotoh, O. (2010) Cgaln: fast and space-efficient whole-genome alignment. *BMC Bioinformatics*, **11**, 1-14.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., & Minh, B.Q. (2014) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, **32**, 268–274.

Nguyen, N.T.T., Contreras-Moreira, B., Castro-Mondragon, J.A., Santana-Garcia, W., Ossio, R., Robles-Espinoza, C.D., Bahin, M., Collombet, S., Vincens, P., Thieffry, D., Helden, J. Van, Medina-Rivera, A., & Thomas-Chollier, M. (2018) RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Research*, **1**, 1–6.

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.

Van Nieuwerburgh, F., Thompson, R.C., Ledesma, J., Deforce, D., Gaasterland, T., Ordoukhanian, P., & Head, S.R. (2012) Illumina mate-paired DNA sequencing-library preparation using Cre-Lox recombination. *Nucleic Acids Research*, **40**, e24.

Niu, X., Guan, Y., Chen, S., & Li, H. (2017) Genome-wide analysis of basic helix-loop- helix (bHLH) transcription factors in *Brachypodium distachyon*. *BMC Genomics*, **18**, 1–20.

Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M., & Henry, R.J. (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnology Journal*, **9**, 328–333.

Noirot, M., Charrier, A., Stoffelen, P., & Anthony, F. (2015) Reproductive isolation , gene flow and speciation in the former Coffea subgenus: a review. *Trees. Structure and function*, 1–12.

Novikova, P.Y., Hohmann, N., Nizhynska, V., et al. (2016) Sequencing of the genus Arabidopsis identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*, **48**, 1077–1082.

O'Mara, F.P. (2012) The role of grasslands in food security and climate change. *Annals of Botany*, **110**, 1263–1270.

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: Biological Sciences*, **361**, 173–194.

Otto, F. (1992) Preparation and Staining of Cells for High-Resolution DNA Analysis. *Flow Cytometry and Cell Sorting* (ed. by A. Radbruch), pp. 65–68. Springer Laboratory, Berlin, Heidelberg.

Otto, S.P. (2007) The Evolutionary Consequences of Polyploidy. *Cell*, **131**, 452–462.

Otto, S.P. & Whitton, J. (2000) Polyploid incidence and evolution. *Annual Review of Genetics*, 401–37.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., & Buell, R.C. (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research*, **35**, 883–887.

Palisot de Beauvois, A.-M.-F.-J. (1812) *Essai d'une nouvelle agrostographie.* Paris.

Panchy, N., Lehti-shiu, M., & Shiu, S. (2016) Evolution of Gene Duplication in Plants. *Plant physiology*, **171**, 2294–2316.

Paradis, E., Claude, J., & Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Patel, D.A., Zander, M., Dalton-Morgan, J., & Batley, J. (2015) Advances in Plant Genotyping: Where the Future Will Take Us. *Plant Genotyping. Methods and Protocols* (ed. by J. Batley), pp. 1–11. Springer Science+Business Media, New York.

Paterson, A.H., Bowers, J.E., & Chapman, B.A. (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences*, **101**, 9903–9908.

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

Peng, X., Zhao, Y., Cao, J., Zhang, W., Jiang, H., Li, X., Ma, Q., & Zhu, S. (2012) CCCH-Type Zinc Finger Family in Maize: Genome-Wide Identification , Classification and Expression Profiling under Abscisic Acid and Drought Treatments. *PLoS ONE*, **7**, e40120.

Perea, C., Fernando, J., Hoz, D. La, Cruz, D.F., Lobaton, J.D., Izquierdo, P., Quintero, J.C., Raatz, B., & Duitama, J. (2016) Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC genomics*, **17**, suppl. 5.

Pereira, A. (2016) Plant Abiotic Stress Challenges from the Changing Environment. *Frontiers in plant science*, **7**, 1–3.

Peterson, D.G., Tomkins, J.P., Frisch, D.A., Wing, R.A., & Paterson, A.H. (2000) Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide. *Journal of Agricultural Genomics*, **5**, 1–100.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D.T.J., Manuel, M., Wörheide, G., & Baurain, D. (2011) Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biology*, **9**, e1000602.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., & Pachter, L. (2017a) Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods*, **14**, 687–690.

Pimentel, M., Escudero, M., Sahuquillo, E., Minaya, M.Á., & Catalán, P. (2017b) Are diversification rates and chromosome evolution in the temperate grasses (Pooideae) associated with major environmental changes in the Oligocene-Miocene? *PeerJ*, **5**, e3815.

Pokorny, L., Oliván, G., & Shaw, A.J. (2011) Phylogeographic Patterns in Two Southern Hemisphere Species of *Calyptrochaeta* (Daltoniaceae, Bryophyta). *Systematic Botany*, **36**, 542–553.

Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J.R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S., & Tangphatsornruang, S. (2015) Genomics Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*, **105**, 288–295.

Porras-Huratdo, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., & Lareu, M. V (2013)

An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, **4**, 1–13.

Posada, D. & Crandall, K.A. (1998) MODELTEST: Testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Prasad, V., Strömberg, C.A.E., Alimohammadian, H., & Sahni, A. (2005) Dinosaur Coproliites and the Early Evolution of Grasses and Grazers. *Science*, **310**, 1177–1180.

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. (2014) Tracer. .

Ramsey, J. & Schemske, D.W. (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, **29**, 467–501.

Rannala, B. & Yang, Z. (1996) Probability distribution of molecular evolutionary trees: A new method of phylogenetic inference. *Journal of Molecular Evolution*, **43**, 304–311.

Ranocha, P., Denance, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Köhler, L., Pouzet, C., Renou, J.-P., Sundberg, B., Boernjan, W., & Goffner, D. (2010) Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *The Plant Journal*, **63**, 469–483.

Ream, T.S., Woods, D.P., Schwartz, C.J., Sanabria, C.P., Mahoy, J.A., Walters, E.M., Kaeppler, H.F., & Amasino, R.M. (2014) Interaction of photoperiod and vernalization determines flowering time of *Brachypodium distachyon*. *Plant physiology*, **164**, 694–709.

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) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**, 4–14.

Reuter, J.A., Spacek, D. V., & Snyder, M.P. (2015) High-Throughput Sequencing Technologies. *Molecular Cell*, **58**, 586–597.

Rey, P.J., Manzaneda, A.J., & Alcántara, J.M. (2017) The interplay between aridity and competition determines colonization ability, exclusion and ecological segregation in the heteroploid *Brachypodium distachyon* species complex. *New Phytologist*, **215**, 85–96.

Rieux, A. & Balloux, F. (2016) Inferences from tip-calibrated phylogenies: a review and a practical guide. *Molecular Ecology*, **25**, 1911–1924.

Ritchie, S.C., Watts, S., Fearnley, L.G., Holt, K.E., Abraham, G., & Inouye, M. (2016) A Scalable Permutation Approach Reveals Replication and Preservation Patterns of Network Modules in Large Datasets. *Cell Systems*, **3**, 71–82.

Robertson, I.H. (1981) Chromosome numbers in *Brachypodium* Beauv. (Gramineae). *Genetica*, **56**, 55–60.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-garcia, S.A., Mcmullen, M.D., Holland, J.B., Buckler, E.S., & Gardner, C.A. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, **14**, R55.

Ronaghi, M., Uhlén, M., & Nyrén, P. (1998) A sequencing method based on real-time pyrophosphate. *Science*, **281**, 363–365.

Ronen, R., Boucher, C., Chitsaz, H., & Pevzner, P. (2012) SEQuel: Improving the accuracy of genome assemblies. *Bioinformatics*, **28**, 188–196.

Ronquist, F., Huelsenbeck, J., & Teslenko, M. (2011) MrBayes Version 3.2 Manual: Tutorials and Model Summaries. 1–103.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rubio, V., Bustos, R., Irigoyen, M.L., Cardona-López, X., Rojas-Triana, M., & Paz-Ares, J. (2009) Plant hormones and nutrient signaling. *Plant Molecular Biology*, **69**, 361–373.

Rutschmann, F. (2006) Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions*, **12**, 35–48.

Rzhetsky, A. & Nei, M. (1992) A Simple Method for Estimating and Testing Minimum-Evolution Trees. *Molecular biology and evolution*, **9**, 945–967.

Saarela, J.M., Bull, R.D., Paradis, M.J., Ebata, S.N., Peterson, P.M., Soreng, R.J., & Paszko, B. (2017) Molecular phylogenetics of cool-season grasses in the subtribes Agrostidinae, Anthoxanthinae, Aveninae, Brizinae, Calothecinae, Koeleriinae and Phalaridinae (Poaceae, Pooideae, Poeae, Poeae chloroplast group 1). *PhytoKeys*, **87**, 1–139.

Saarela, J.M., Burke, S. V., Wysocki, W.P., Barrett, M.D., Clark, L.G., Craine, J.M., Peterson, P.M., Soreng, R.J., Vorontsova, M.S., & Duvall, M.R. (2018) A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ*, **6**, e4299.

Saarela, J.M., Wysocki, W.P., Barrett, C.F., Soreng, R.J., Davis, J.I., Clark, L.G., Kelchner, S.A., Pires, J.C., Edger, P.P., Mayfield, D.R., & Duvall, M.R. (2015) Plastid phylogenomics of the cool-season grass subfamily: Clarification of relationships among early-diverging tribes. *AoB PLANTS*, **7**, 1–27.

Saitou, N. & Nei, M. (1987) The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular biology and evolution*, **4**, 406–425.

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.

Sánchez-Ken, J.G. & Clark, 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.

Sancho, R., Cantalapiedra, C.P., López-Alvarez, D., Gordon, S.P., Vogel, J.P., Catalán, P., & Contreras-Moreira, B. (2018) Comparative plastome genomics and phylogenomics of *Brachypodium*: Flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist*, **218**, 1631–1644.

Sanger, F., Nicklen, S., & Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, **74**, 5463–5467.

Sanmartín, I. (2012) Historical Biogeography: Evolution in Time and Space. *Evolution: Education and Outreach*, **5**, 555–568.

Sanmartin, 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.

Sasaki, T. & Antonio, B.A. (2004) Rice genome as a model system for cereals. *Cereal Genomics* (ed. by P.K. Gupta and R.K. Varshney), pp. 535–557. Kluwer Academic Publishers, Netherlands.

Saski, C., Lee, S.-B., Siri, F., Chittibabu, G., Jansen, R.K., Luo, H., Tomkins, J., Rognli, O.A., Daniell, H., & Clarke, J.L. (2007) Complete chloroplast genome sequences of *Hordeum vulgare, Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theoretical and Applied Genetics*, **115**, 571–590.

Scheben, A., Batley, J., & Edwards, D. (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*, **15**, 149–161.

Schippmann, U. (1990) *Brachypodium boissieri*. An endemic grass species of southern Spain. *Lagascalia*, **15**, 179–188.

Schippmann, U. (1991) Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera*, **45**, 249 pp.

Schneider, J., Winterfeld, G., Hoffmann, M.H., & Röser, M. (2011) Duthieeae, a new tribe of grasses (Poaceae) identified among the early diverging lineages of subfamily Pooideae: molecular phylogenetics, morphological delineation, cytogenetics, and biogeography. *Systematics and Biodiversity*, **9**, 27–44.

Scholthof, K.-B.G., Irigoyen, S., Catalán, P., & Mandadi, K.K. (2018) *Brachypodium*: A monocot grass model system for plant biology. Plant Cell. [In press].

Schuh, R.T. (2000) *Biological Systematics. Principles and applications.* Cornell University Press, United States of America, 328 pp.

Schuster, S.C. (2008) Next-generation sequencing transforms today's biology. *Nature Methods*, **5**, 16–18.

Schwartz, C.J., Doyle, M.R., Manzaneda, A.J., Rey, P.J., Mitchell-Olds, T., & Amasino, R.M. (2010) Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. *Bioenergy Research*, **3**, 38–46.

Sebastian, A. & Contreras-Moreira, B. (2014) footprintDB: a database of transcription factors with annotated cis elements and binding interfaces. *Bioinformatics*, **30**, 258–265.

Serin, E.A.R., Nijveen, H., Hilhorst, H.W.M., & Ligterink, W. (2016) Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Genetics*, **7**, 1–18.

Shendure, J. & Ji, H. (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.

Shi, Y. (1991) Molecular studies of the evolutionary relationships of *Brachypodium* (Poaceae). Ph.D thesis, University of Leicester.

Shi, Y., Draper, J., & Stace, C. (1993) Ribosomal DNA variation and its phylogenetic implication in the genus *Brachypodium* (Poaceae). *Plant Systematics and Evolution*, **188**, 125–138.

Shinozaki, K. & Yamaguchi-Shinozaki, K. (2007) Gene networks involved in drought stress response and tolerance. *Journal of Experimental Botany*, **58**, 221–227.

Shiposha, V., Catalán, P., Olonova, M., & Marques, I. (2016) Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands. *PeerJ*, **4**, e2407.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., Mcwilliam, H., Remmert, M., Söding, J., Thompson, J.D., & Higgins, D.G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systematics Biology*, **7**, 1–6.

Silvertown, J., Servaes, C., Biss, P., & Macleod, D. (2005) Reinforcement of reproductive isolation between adjacent populations in the Park Grass Experiment. *Heredity*, **95**, 198–205.

Sneath, P.H.A. & Sokal, R.R. (1973) *Numerical Taxonomy. The principles and practice of* *numerical classification.* W. H. Freeman and Company, San Francisco, 588 pp.

Sokal, R. & Michener, C. (1958) A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin*, **38**, 1409–1438.

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. (2014a) Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. *New Phytologist*, **202**, 1105–1117.

Soltis, D.E., Visger, C.J., Blaine Marchant, D., & Soltis, P.S. (2016) Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany*, **103**, 1146–1166.

Soltis, D.E., Visger, C.J., & Soltis, P.S. (2014b) The polyploidy revolution then...and now: Stebbins revisited. *American Journal of Botany*, **101**, 1057–1078.

Soltis, P.S. & Soltis, D.E. (2016) Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology*, **30**, 159–165.

Soreng, R., Davidse, G., Peterson, P., Zuloaga, F., Judziewicz, E., Filgueiras, T., Morrone, O., & Romaschenko, K. (2014) Available at: http://www.tropicos.org/docs/meso/WWP A WORLDWIDE CLASSIFICATION OF POACEAE TROPICOS version JUN 30 2017.htm.

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. *The Botanical Review*, **64**, 1–85.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Teisher, J.K., Clark, L.G., Barberá, P., Gillespie, L.J., & Zuloaga, F.O. (2017) A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution*, **55**, 259–290.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.I., & Morrone, O. (2015) A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, **53**, 117–137.

Stamatakis, A. (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–

2690.

Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stebbins, G.L. (1949) The evolutionary significance of natural and artificial polyploids in the family Gramineae. *Hereditas*, **35**, 461–458.

Stebbins, G.L. (1985) Polyploidy, hybridization and the invasion of new habitats. *Annals of the Missouri Botanical Garden*, **72**, 824–832.

Steinwand, M.A., Young, H.A., Bragg, J.N., Tobias, C.M., & Vogel, J.P. (2013) *Brachypodium sylvaticum*, a Model for Perennial Grasses: Transformation and Inbred Line Development. *PLoS ONE*, **8**, e75180.

Stetter, M.G. & Schmid, K.J. (2017) Molecular Phylogenetics and Evolution Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution*, **109**, 80–92.

Strien, M.J. Van, Holderegger, R., & Heck, H.J. Van (2014) Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, **114**, 27–37.

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*, **102**, 11980–11984.

Strömberg, C.A.E. (2011) Evolution of Grasses and Grassland Ecosystems. *Annual Review of Earth and Planetary Sciences*, **39**, 517–544.

Stuart, J.M., Segal, E., Koller, D., & Kim, S.K. (2003) A Gene Coexpression Network for Global Discovery of Conserved Genetic Modules. *Science*, **302**, 249–255.

Suda, J., Kyncl, T., & Jarolímová, V. (2005) Genome size variation in Macaronesian angiosperms: forty percent of the Canarian endemic flora completed. *Plant Systematics and Evolution*, **252**, 215–238.

Swofford, D.L. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4b10.* Sinauer Associates, Sunderland, Massachusetts, USA.

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., Fukuda, H., Yamaguchi-Shinozaki, K., & Shinozaki, K. (2018) A small peptide modulates stomatal control via abscisic acid in long-distance signalling. *Nature*, **556**, 235–238.

Talavera, S. (1978) Aportacion al estudio cariologico de las gramineas españolas. *Lagascalia*, **7**, 133–142.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tandonnet, S. & Torres, T.T. (2017) Traditional versus 3′ RNA-seq in a non-model species. *Genomics Data*, **11**, 9–16.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., & Paterson, A.H. (2008) Synteny and Collinearity in Plant Genomes. *Science*, **320**, 486–489.

Teixeira Torres, T., Metta, M., Ottenwälder, B., & Schlötterer, C. (2008) Gene expression profiling by massively parallel sequencing. *Genome Research*, **18**, 172–177.

The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Communications*, **25**, 25–29.

The Gene Ontology Consortium (2017) Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, **45**, D331–D338.

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, **45**, D158–D169.

Thomas, G.W.C., Ather, S.H., & Hahn, M.W. (2017) Gene-Tree Reconciliation with MUL-Trees to Resolve Polyploidy Events. *Systematic Biology*, **0**, 1–12.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, **13**, 2129–2141.

Thorvaldsdóttir, H., Robinson, J.T., & Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**, 178–192.

Thudi, M., Li, Y., Jackson, S.A., May, G.D., & Varshney, R.K. (2012) Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional*

Genomics, **11**, 3–11.

Tyler, L., Lee, S.J., Young, N.D., Deiulio, G.A., Benavente, E., Reagon, M., Sysopha, J., Baldini, R.M., Troìa, A., Hazen, S.P., & Caicedo, A.L. (2016) Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. *Plant Genome*, **9**, 1–55.

Uozu, S., Ikehashi, H., Ohmido, N., Ohtsubo, H., Ohtsubo, E., & Fukui, K. (1997) Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Molecular Biology*, **35**, 791–799.

Vanderlaan, T.A., Ebach, Malte, C., Williams, D.M., & Wilkins, J.S. (2013) Defining and redefining monophyly: Haeckel, Hennig, Ashlock, Nelson and the proliferation of definitions. *Austraian Systematics Botany*, **26**, 347–355.

Vaughan, D.A. (1994) *The wild relatives of rice. A genetic resources handbook.* IRRI, Manila, 137 pp.

Vicentini, A., Barber, J.C., 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.

Vinuesa, P., Ochoa-Sánchez, L.E., & Contreras-Moreira, B. (2018) GET _ PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies , used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Frontiers in Microbiology*, **9**, 1-22.

Vogel, J. & Bragg, J. (2009) *Brachypodium distachyon*, a New Model for the Triticeae. *Genetics and Genomics of the Triticeae. Plant Genetics and Genomics: Crops and Models 7* (ed. by C. Feuillet and G.J. Muehlbauer), pp. 427–449. Springer.

Vogel, J. & Hill, T. (2008) High-efficiency Agrobacterium-mediated transformation of *Brachypodium distachyon* inbred line Bd21-3. *Plant Cell Reports*, **27**, 471–478.

Vogel, J.P. (2016) The Rise of *Brachypodium* as a Model System. *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models* (ed. by J.P. Vogel), pp. 1–8. Springer.

Vogel, J.P., Garvin, D.F., Leong, O.M., & Hayden, D.M. (2006) Agrobacterium-mediated transformation and inbred line development in the model grass *Brachypodium distachyon. Plant Cell, Tissue and Organ Culture*, **84**, 199–211.

Vogel, J.P., Garvin, D.F., Mockler, T.C., et al. (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

Vogel, J.P., Tuna, M., Budak, H., Huo, N., Gu, Y.Q., & Steinwand, M.A. (2009) Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. *BMC plant biology*, **9**, 88.

von Bothmer, R., Jacobsen, N., Baden, C., Bagger Jørgensen, R., & Linde-Laursen, I. (1995) *An ecogeographical study of the genus Hordeum.* IPGRI, Rome.

Wang, R.-J., Cheng, C.-L., Chang, C.-C., Wu, C.-L., Su, T.-M., & Chaw, S.-M. (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC evolutionary biology*, **8**, 36.

Wang, Y. & Ma, H. (2015) Step-wise and lineage-specific diversification of plant RNA polymerase genes and origin of the largest plant-specific subunits. *New Phytologist*, **207**, 1198–1212.

Wang, Z., Gerstein, M., & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.

Washburn, J.D., Schnable, J.C., Conant, G.C., Brutnell, T.P., Shao, Y., Zhang, Y., Ludwig, M., Davidse, G., & Pires, J.C. (2017) Genome-Guided Phylo-Transcriptomic Methods and the Nuclear Phylogentic Tree of the Paniceae Grasses. *Scientific Reports*, **7**, 13528.

Waters, D.L.E., Nock, C.J., Ishikawa, R., Rice, N., & Henry, R.J. (2012) Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecology and Evolution*, **2**, 211–217.

Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag, New York.

Wisecaver, J.H., Borowsky, A.T., Tzin, V., Jander, G., Kliebenstein, D.J., & Rokas, A. (2017) A Global Co-expression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *Plant Cell*, **25**, 944–959.

Wolfe, C.J., Kohane, I.S., & Butte, A.J. (2005) Systematic survery reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics*, **6**, 1–10.

Wolny, E. & Hasterok, R. (2009) Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Annals of Botany*, **104**, 873–881.

Wolny, E., Lesniewska, K., Hasterok, R., & Langdon, T. (2011) Compact genomes and complex evolution in the genus *Brachypodium*. *Chromosoma*, **120**, 199–212.

Woods, D.P., Mckeown, M.A., Dong, Y., Preston, J.C., & Amasino, R.M. (2016) Evolution of VRN2/GhD7- Like Genes in Vernalization-Mediated Repression of Grass. *Plant Physiology*, **170**, 1–12.

Woods, D.P., Ream, T.S., & Amasino, R.M. (2014) Memory of the vernalized state in plants including the model grass *Brachypodium distachyon*. *Frontiers in plant science*, **5**, 99.

Wright, S. (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wu, Z.Q. & Ge, S. (2012) The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Molecular Phylogenetics and Evolution*, **62**, 573–578.

Wysocki, W.P., Clark, L.G., Attigala, L., Ruiz-Sanchez, E., & Duvall, M.R. (2015) Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC evolutionary biology*, **15**, 50.

Xiang, Q.P., Wei, R., Shao, Y.Z., Yang, Z.Y., Wang, X.Q., & Zhang, X.C. (2015) Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **82**, 1–14.

Yang, Y. & Smith, S.A. (2014) Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics. *Molecular Biology and Evolution*, **31**, 3081–3092.

Yu, H., Jiao, B., & Liang, C. (2017) High-quality rice RNA-seq-based co-expression network for predicting gene function and regulation. *bioRxiv*, https://www.biorxiv.org/content/early/2017/05/15/138040.

Yuan, H. & Liu, D. (2008) Signaling components involved in plant responses to phosphate starvation. *Journal of Integrative Plant Biology*, **50**, 849–859.

Zerbino, D.R. (2010) Using the Columbus extension to Velvet. http://gensoft.pasteur.fr/docs/velvet/1.1.02/Columbus_manual.pdf.

Zerbino, D.R. & Birney, E. (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.

Zhang, B. & Horvath, S. (2005) A General Framework for Weighted Gene Co-expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, **4**, Article 17.

Zhang, W. (2000) Phylogeny of the grass family (Poaceae) from rpl16 intron sequence data. *Molecular Phylogenetics and Evolution*, **15**, 135–146.

Zhao, Q., Feng, Q., Lu, H., et al. (2018) Pan-genome analysis highlihgts the extent of genomic variation in cultivated and wild rice. *Nature Genetics*, **50**, 278–284.

Zhou, L., Wang, S.-B., Jian, J., Geng, Q.-C., Wen, J., Song, Q., Wu, Z., Li, G.-J., Liu, Y.-Q., Dunwell, J.M., Zhang, J., Feng, J.-Y., Niu, Y., Zhang, L., Ren, W.-L., & Zhang, Y.-M. (2015) Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Scientific Reports*, **5**, 1–8.

Zhou, S., Yan, B., Li, F., Zhang, J., Zhang, J., Ma, H., Liu, W., Lu, Y., Yang, X., Li, X., Liu, X., & Li, L. (2017) RNA-Seq Analysis Provides the First Insights into the Phylogenetic Relationship and Interspecific Variation between *Agropyron cristatum* and Wheat. *Frontiers in Plant Science*, **8**, 1–13.

Appendix I: Supporting Information of Chapter 1

Methods/Results S1: Expanded materials and methods and results

Taxon sampling

The three annual species are largely distributed in the circumMediterranean region (*B. distachyon*, *B. stacei*, *B. hybridum*), whereas the 17 perennial taxa show either large [Eurasian (*B. pinnatum* 2x, 4x*, B. sylvaticum*), Mediterranean (*B. retusum*), American (*B. mexicanum*)] or restricted disjunct [W Mediterranean (*B. phoenicoides*), C Mediterranean (*B. genuense*), E Mediterranean (*B. glaucovirens*), S Spain (*B. boissieri*), Canarian (*B. arbuscula*), W European (*B. rupestre* 2x, 4x), S African (*B. bolusii*), tropical and S African (*B. flexum*), Madagascar (*B. madagascariense*), Taiwan (*B. kawakamii*), and S-SE Asia - Malesia (*B. sylvaticum* var. *pseudodistachyon*)] geographic distributions (Fig. 1).

Six poorly known taxa (35.3% of the total taxonomic diversity) were studied phylogenetically (*B. bolusii*, *B. flexum*, *B. genuense*, *B. kawakamii*, *B. madagascariense*, *B. sylvaticum* var. *pseudodistachyon*). Our study also included representatives of both diploid and allotetraploid cytotypes of the perennial *B. pinnatum* and *B. rupestre* species that were only used in the Bayesian (MrBayes) phylogeny and haplotype network analyses. Chromosomal and ploidy data was collected for most samples (Table S1), though some poorly known species have not been karyologically studied yet. We sampled from three to ten geographically distinct populations of each taxon, except for a few extremely isolated species that were represented by one or two accessions (Table S1).

DNA extraction, amplification, cloning and sequencing

The plastid data included the 3-end coding region of the NAD dehydrogenase subunit F (*ndh*F) gene and the *trn*L(UAA) intron – *trn*L(UAA) exon – *trn*L(UAA)/*trn*F(GAA) spacer (*trn*LF) region, which were amplified and sequenced in all samples following the procedures indicated in Catalán et al. (2012). The nuclear multicopy data included the sequences of the external transcribed spacer (ETS) and the internal transcribed spacer (ITS) of the ribosomal DNA repeat unit, and the nuclear single copy gene data consisted of coding and intron regions of the GIGANTEA (GI) gene. DNA isolation, amplification, cloning and sequencing was done following the procedures indicated in Catalán *et al.* (2012) and López-Alvarez *et al.* (2012). Five clones per sample were sequenced for each locus in both diploid and polyploid taxa, aiming to detect all potential ribotypes and homeologous copies.

A total of 973 new *Brachypodium* sequences [411 ETS (Genbank accession codes KP709080-KP709491, 269 ITS (KP709492-KP709761), 160 GI (KP709897-KP710057), 67 ndhF (KP709762-KP9829), 66 trnLF (KP709830-KP709896)] generated in the present study were aligned with sequences obtained in our previous studies and others retrieved from Genbank and were used in the phylogenetic analysis (Table S1). A total of 1154 DNA sequences from the three nuclear (ETS, ITS, GI) and the two plastid (*ndh*F, *trn*LF) loci were used to reconstruct the phylogeny of *Brachypodium*. Multiple sequence alignments were performed separately for each data set using the Clustal algorithm option of Geneious v. R.8.0.2 and adjusted manually. The final data sets consisted of 431 sequences/682 aligned positions for ETS, 368/645 for ITS, 280/831 for GI, 95/564 for ndhF, and 100/941 for trnLF. The non-recombinant *ndh*F + *trn*LF plastid (cpDNA) sequences were concatenated into a combined 105/1505 data set. Data matrices used in exploratory phylogenetic analyses consisted of reduced haplotypic aligned data sets of 199 haplotypes for ETS, 159 for ITS, 114 for GI, and 44 for the concatenated cpDNA, where identical redundant sequences were previously removed.

Phylogenetic and haplotypic network analyses

Exploratory phylogenetic analyses were first performed with the reduced haplotypic ETS, ITS, GI and cpDNA data sets, aiming to recover the evolutionary history of the *Brachypodium* lineages supported by each separate gene, to detect nuclear homeologous copies in the polyploids, and to estimate the levels of interspecific haplotype sharing in different groups and genes. Phylogenetic trees were computed through Bayesian Inference (BI) methods. All the conducted searches excluded gaps from the analysis and used other pooid representatives and *Oryza* (Oryzoideae) as outgroups. BI was computed in MrBAYES v. 3.1.2 (Ronquist & Huelsenbeck, 2003), imposing the GTR + Γ (nst = 6 and rates = invgamma) model, selected as the optimal model for the four data sets based on the Aikake criterion implemented in jMODELTEST (Guindon & Gascuel, 2003; Darriba et al., 2012). Two runs were performed, each with 5 000 000 generations, sampling trees every 1000 generations, and imposing a burn-in option of 1250 trees per run once stability in the likelihood values was attained. Convergence of parameters was analysed with TRACER v. 1.6 (Rambaut et al., 2014), being consistent with ESS values >200. The 3750 saved trees of each search were used to compute the respective Bayesian all-compatible consensus trees where the posterior probability values of branches were interpreted as a measure of nodal support. Haplotypic networks were constructed to infer the genealogical relationships of the *Brachypodium* haplotypes (species and samples) obtained from each separate data set using statistical parsimony approaches (Clement et al., 2002) computed with POPART (Leigh & Bryant, 2015).

Phylogenetic and biogeographic results

The statistical parsimony haplotypic networks (Figs. 3A-3D) were highly congruent with those constructed through Bayesian methods (Catalán et al., 2016b) and did not include the single-clone allelic copies. The cpDNA haplotypic network consisted of 44 haplotypes (Fig. 3A) and was relatively well resolved for the early divergences of the monophyletic *B. boissieri*, *B. stacei*, *B. mexicanum* and *B. distachyon* clusters, each separated by a number of mutational steps. These divergences were highly supported in the corresponding BI phylogenetic tree (data not shown). The *B. hybridum* haplotypes were shared with its *B. stacei* parent. However, the cluster of the recently evolved core perennial species showed a lack of genealogical and taxonomic structure, denoted by the high number of interspecific shared haplotypes (with some haplotypes shared by up to six species, and an ambiguous resolution, manifested in two internal loops and few internal mutational steps (Fig. 3A).

The ITS and ETS haplotypic networks (Figs. 3B, 3C) and BI phylogenies (data not shown), constructed, respectively, with 159 and 199 haplotypes, were congruent in the separate early divergences of the *B. boissieri*, *B. stacei*, *B. mexicanum* and *B. distachyon* lineages and in the complex reticulate structure of the core perennials group. They further detected the early divergences of the *B. bolusii* / *B. flexum*, *B. arbuscula* and *B. retusum* lineages within the core perennials clade (Figs. 3B, 3C) and the clustering of endemic East Asia – Madagascar (*B. sylvaticum* [China] / *B. kawakamii*, *B. madagascariense*) (Fig. 3C) and East Asia–New Guinea (*B. kawakamii* / *B. sylvaticum* var. *pseudodistachyon*) haplotypes in their respective ETS and ITS regional subnetworks. The introgression and homoplasy levels detected by these loci were much higher than those detected by the plastid data within the core perennial cluster, and mostly affected the Eurasian and Mediterranean species. Thus, the most common ITS haplotype was shared by 10 species (Fig. 3B) and the most common ETS haplotype by 6 species (Fig. 3C). Both loci detected co-inherited *B. stacei*-type and *B. distachyon*-type parental copies in *B. hybridum*, those from the latter parent being more frequent (Figs. 3B, 3C).

The GI haplotypic network (Fig. 3D) and the BI phylogeny (data not shown), constructed with 114 haplotypes, also supported the early divergence of the *B. boissieri*, *B. mexicanum*, *B. stacei* and *B. distachyon* lineages and the reticulation of the recent core perennials clade, though relationships varied with respect to those observed in the ITS and ETS networks and trees and were overall less resolved. The level of potential introgression detected by the GI network was apparently very high, showing a most common haplotype shared by samples from 8 perennial species (Fig. 3D). In contrast, the GI clones detected the highest number of co-inherited ancestral- vs. recently evolved-type homeologous copies among the perennial *Brachypodium* allopolyploid species. Interestingly, highly divergent GI sequences of *B. boissieri*, *B. retusum*, and *B. phoenicoides* were nested within both the early split '*B. boissieri*' cluster and the recently split core perennial cluster. The *B. hybridum* individuals showed homeologous copies from each *B. stacei* and *B. distachyon* parent. The analyses also recovered two close but separate homeologous lineages within the early divergent *B. mexicanum* (Figs. 3D).

The derived allotetraploid (heteroploid) origin of the annual *B. hybridum* from diploid *B. stacei* and *B. distachyon* ancestors (Figs. 3B-D, C2-C4) is firmly supported by all nuclear loci, as it is the only allopolyploid species showing 100% support values of all its nuclear alleles to the two respective out-core parental terminal branches of the species tree (Fig. 4). Our dating analysis also confirms the recent origin of *B. hybridum* in the Quaternary (0.03 Ma; Table 1), supporting its neopolyploid status (Catalán et al., 2012). Dinh Thi *et al.* (2016) recreated a synthetic *B. hybridum* allotetraploid from specific crosses and artificial chromosome doubling, confirming the likely occurrence of past bidirectional crosses between parental genome donors resulting in the allotetraploid individuals (López-Alvarez et al., 2012). Because all the examined individuals correspond to the same taxonomic species *B. hybridum*, and show similar morphology and stability in their chromosome number 2n = 30 (López-Alvarez et al., 2012; Catalán et al., 2016a), we conclude that multiple bidirectional hybridizations and genome doublings have resulted in the same speciation process for this neo-allotetraploid species, paralleling similar cases hypothesized for other temperate Mediterranean annual grasses (e. g., *Aegylops triuncialis, A. cylindrica*; (Meimberg et al., 2009)). No evidence of backcrossing of *B. hybridum* with any of its parents has been found to date, suggesting that allopolyploidization effectively contributed to the reproductive isolation of the allotetraploid from its progenitors and to the genomic, phenotypic and ecological stabilization of the new species (Catalán et al., 2012, 2016a; Lopez-Alvarez et al., 2015; López-Álvarez et al., 2017).

In contrast to *B. hybridum*, the identification of the genome donors of the perennial allopolyploid species is more complex. Chromosome counts of 2n = 40 and duplicated single copy gene allelic dosages indicated that *B. mexicanum* could be a tetraploid (Wolny et al., 2011). Our minimum evolution and species network analyses strongly support the presence of a *B. stacei*-type allele (*STACEI*) in *B. mexicanum* and the likely presence of one ancestral genome allele (*ANCESTRAL*) (Figs. 4), supporting its purported allotetraploidy and homoploid chromosome base numbers of x=10. The existence of a *B. stacei*-like homeologous genome in *B. mexicanum* could explain the shared biological, morphological and genomic features of this short-rhizomatose species and its closely related annual relative (e. g., self-compatibility, non-rhizomatous habit, protein and DNA families; (Catalán et al., 2016a)). Phylogenetic analysis of single-copy nuclear genes (CAL, GI) detected different but close homeologous copies in *B. mexicanum* (Wolny et al., 2011; Catalán et al., 2012, 2016a); our current study confirms this and has also found shared plastid genes in *B. mexicanum* and *B. stacei* (Fig. 3A). These results suggest that *B. mexicanum* could be a homoploid allotetraploid species originated from the cross of two closely related ancestral *Brachypodium* diploid lineages of x=10, of which the *B. stacei*-type lineage probably acted as maternal parent.

Grafting allelic copies of the remaining polyploid or unknown ploidy *Brachypodium* species was restricted to the recent stem branch and internal branches of the core perennial clade or just to core branches. *SYLVATICUM* and *PINNATUM* were potentially involved in the origins of at least four allopolyploid core perennial species (Figs. 4 and 5). *SYLVATICUM* (*B. sylvaticum/genuense*-like) could be one of the potential genome donors of all allopolyploids, except *B. hybridum* and *B. mexicanum*, and *PINNATUM* (*B.*

*pinnatum/rupestre*2x-like) could be a donor to *B. flexum*, *B. kawakamii*, and allopolyploids *B. retusum* and *B. phoenicoides* (Figs. 4 and 5).

Independent and combined analysis of cloned GI, ITS and ETS sequences have identified out-core and core copies in the core perennial *B. boissieri* (Figs. 3B-D), indicating that this species imherited both ancestral and recent homeologous genomes. Our dated chronogram indicates the early divergence of the x=10 *B. stacei* and *B. mexicanum* lineages (N_{ST}, 6.8 Ma), followed by that of x=5 *B. distachyon* lineage (N_{DS}, 5.1Ma), which predated that of the x=8, 9 core perennial clade lineages (N_{AR}, 2.4 Ma) (Fig. 5), supporting the descendant (x=10 → x=5) and ascendant disploidy (x=5 → x=9) scenario of karyotypic evolution in *Brachypodium* (Betekhtin et al., 2014). The identity of the ancestral and recent genome donors of the perennial allopolyploids still remains unknown and would require deeper genomic analyses. The allotetraploid *B. rupestre* 4x and *B. pinnatum* 4x cytotypes might constitute separate species, paralleling the case of the segregated annual species of the diploid-allopolyploid *B. distachyon* complex (Catalán et al., 2012); however, their current sampling shortage prevents their use in biogeographical analysis.

Supporting Tables

Table S1. List of *Brachypodium* taxa used in the phylogenetic and biogeographic study. Information on accession code, locality, chromosome number (2n), chromosome base number (x), ploidy and Genbank accession numbers of the studied ndhF, trnLF, ITS, ETS and GIGANTEA (GI) genes are indicated for each sample.

Taxon	Code	Locality	2n	x	ploidy	ndhF	trnLF	ITS	ETS	GI
Annual *Brachypodium* taxa										
B. distachyon	Bdis17	Iraq: Salah ad Din, 4 km from Salahuddin, USDA Bd21, type.	2n=10	5	2x	JN187631	JN187656	JN187608; KP709528-KP709532	KP709119-KP709123	-
B. distachyon	Bdis18	Turkey: Kiresehir, Kaman. ABR1	2n=10	5	2x	KP709768	KP709839	JX665548-JX665550	KP709124-KP709128	-
B. distachyon	Bdis19	France: Herault, Octon. ABR2 (Bdis306)	2n=10	5	2x	JN187636	JN187661	JN187613		-
B. distachyon	Bdis20	Slovenia: Lubjana. ABR9 (Bdis384)	2n=10	5	2x	JN187638	JN187663	–	KP709129	-
B. distachyon	Bdis21	Spain: Huesca, Ibieca, Foces. PC&LM Bdis400	2n=10	5	2x	JN187640	JN187665	JN187615; JX665553-JX665557	KP709130-KP709134	-
B. distachyon	Bdis22	Spain: Huesca, Jaca, Guasillo. PC&LM Bdis401	2n=10	5	2x	JN187641	JN187666	JN187593; JX665559-JX665563	KP709135-KP709139	-
B. distachyon	Bdis23	Spain: Albacete, Alcaraz. CS Bd115F	2n=10	5	2x	KP709769	JX665854	JX665565-JX665569	KP709140-KP709144	JX666047
B. hybridum	Bhyb8	Spain: Teruel, Calaceite. Bdis41	2n=30	15	4x	KP709773	JX665898	JX665608-JX665612	KP709164-KP709168	JX666089
B. hybridum	Bhyb9	France: Corsica, Bonifacio. ABR112 (Bdis383)	2n=30	15	4x	JN187637	JN187662	JX665613-JX665618	KP709169-KP709173	-
B. hybridum	Bhyb10	Portugal: Lisboa 40. ABR Bdis385	2n=30	15	4x	JN187639	JN187664	JN187614; JX665620-JX665623	KP709174-KP709178	JX666091-JX666095
B. hybridum	Bhyb11	Portugal: Lisboa, ABR113. Type	2n=30	15	4x	JN187632	JN187658	JN187610; JX665625-JX665627	KP709179-KP709184	-
B. hybridum	Bhyb12	France: Aude. ABR110	2n=30	15	4x	JN187633	JN187657	JN187609 ; KP709533-KP709537	KP709185-KP709190	-
B. hybridum	Bhyb13	Afghanistan: USDA PI219965, ABR117	2n=30	15	4x	JN187635	JN187660	JN187612; JX665630-JX665632	KP709191-KP709195	JX666098-JX666104
B. hybridum	Bhyb14	Spain: Zaragoza, La Alfranca. PC&LM:Bdis402	2n=30	15	4x	JN187642	JN187667	JX665634-JX665638	KP709196-KP709201	-
B. hybridum	Bhyb15	Spain: Girona, Cadaques. PC&LM:Bdis403	2n=30	15	4x	JN187643	JN187668	JN187617; JX665640-JX665644	KP709202-KP709207	-
B. hybridum	Bhyb22	Spain: Almeria, Cabo de Gata. ST SEV268823	2n=30	15	4x	KP709774	JX665912	JX665661-JX665665	KP709208-KP709212	-
B. hybridum	Bhyb26	Spain: Jaen,La Cimbarra. AM	2n=30	15	4x	-	-	-	-	JX666115
B. hybridum	Bhyb41	Turkey: HB & JV BdTR6B	2n=30	15	4x	-	-	-	-	JX666144-JX666145
B. hybridum	Bhyb811	Spain: Almeria, Abrucena. ST SEV268811	2n=30	15	4x	-	-	-	KP709145-KP709149	-
B. stacei	Bsta1	Spain: Formentera, Torrent, ABR114, type	2n=20	10	2x	JN187634	JN187659	JN187611; JX665763-JX665766	KP709150-KP709153	HQ890969
B. stacei	Bsta3	Spain: Jaen, Baeza. CS Bd114F	2n=20	10	2x	KP709770	JX666001	JX665769-JX665773	KP709154-KP709158	JX666228
B. stacei	Bsta4	Spain: Granada, Moclín. CS Bd129F	2n=20	10	2x	KP709771	JX666002	JX665775-JX665779	KP709159-KP709163	JX666229,JX666231
B. stacei	Bsta5	Spain: Alicante, Cabo de La Nao. CS Bd483F	2n=20	10	2x	KP709772	JX666003	JX665781-JX665785	-	-

Perennial *Brachypodium* taxa										
B. arbuscula	Barb96	Spain: Canary isles, Gomera. CAS.	2n=18	9	2x	KP709762	KP709830	KP709492-KP709496	KP709080-KP709085	-
B. arbuscula	Barb405	Spain: Canary isles, Tenerife, Teno. PC.	2n=18	9	2x	KP709763	-	KP709497	-	-
B. arbuscula	Barb500	Spain: Canary isles, Gomera, Barranco las Rosas. CA.	2n=18	9	2x	JN187629	JN187654	JN187606;KP709498-KP709502	KP709086-KP709090	KP709897-KP709901
B. arbuscula	Barb501	Spain: Canary isles (cultivated in JBVC Botanical Garden).	2n=18	9	2x	KP709764	KP709831	KP709503-KP709507	KP709091-KP709095	JN589948/JN589949B;KP709902-KP709906
B. boissieri	Bboi1	Spain: Granada, Sierra Nevada, Monachil, Dornajo. PC.	cf 2n=42, 46	-	6x-8x?	JN187630	JN187655	JN187607; KP709508-KP709512	KP709096-KP709100	KP709907-KP709912
B. boissieri	Bboi2	Spain: Granada: Sierra Nevada, Huescar-Sierra, Dornajo. PC.	cf 2n=42, 46	-	6x-8x?	-	KP709832	-	KP709101	KP709913-KP709914
B. boissieri	Bboi3	Spain: Granada: Sierra de Huetor, Puerto de La Mora. JM.	cf 2n=42, 46	-	6x-8x?	KP709765	KP709833	KP709513-KP709517	KP709102-KP709106	KP709915-KP709920
B. boissieri	Bboi4	Spain: Granada: Sierra de Tejeda-Almijara,Tajos de la Chapa. JM.	cf 2n=42, 46	-	6x-8x?	KP709766	KP709834	KP709518-KP709522	KP709107-KP709111	-
B. bolusii	Bbol1	South Africa. Kew_8706.	-	-	-	-	KP709835	-	KP709112	-
B. bolusii	Bbol46	South Africa: Natal, Drakernsbergs, Organ Pipes Pass. PC SA046	-	-	-	KP709767	KP709836	KP709523-KP709527	KP709113-KP709118	KP709921-KP709922
B. flexum	Bflex38	South Africa: KwaZulu Natal, Weza Forest. PC SA038	-	-	-	KP709775	KP709838	KP709538-KP709543	KP709213-KP709217	KP709923-KP709924
B. flexum	Bflex43	South Africa: KwaZulu Natal, Didima. PC SA043	-	-	-	KP709776	KP709839	KP709544-KP709547	KP709218-KP709223	KP709925
B. genuense	Bgen1	Italy: Parco Nazionale della Majella. Giardino Botanico D. Brescia	2n=18	9	2x	KP709777	KP709840	KP709548-KP709550	KP709224-KP709228	KP709926
B. glaucovirens	Bgla224	Turkey: Between Soma and Akhisar.	2n=16	8	2x	KP709778	KP709841	KP709551-KP709553	KP709229-KP709234	KP709927
B. kawakamii	Bkaw1	Taiwan: Taitung county, Kuan-Shan-Ling-Shan. BLS.	-	-	-	KP709779	KP709842	KP709554-KP709559	KP709235-KP709245	KP709928-KP709932
B. madagascariense	Bmad2	Madagascar: Ankaratra Mts, Tsiafajavjona Mt. MP.	-	-	-	KP709780	KP709843	KP709560-KP709563	KP709246-KP709251	-
B. mexicanum	Bmex295	Mexico: Mexico, Texcoco. CAS.	2n=40	-	4x?	KP709781	KP709844	KP709564-KP709568	KP709252-KP709256	-
B. mexicanum	Bmex347	Mexico: Hidalgo, Sierra de Pachuca. CAS.	2n=40	-	4x?	JN187644	JN187669	JN187619; KP709569-KP709573	JN187596; KP709257-KP709261	HQ890968/HQ890971
B. mexicanum	Bmex502	Venezuela: Merida, Laguna de Coromoto. PC	2n=40	-	4x?	KP709782	KP709845	KP709574-KP709579	KP709262-KP709265	KP709933-KP709937
B. mexicanum	Bmex647	Mexico. CAS.	2n=40	-	4x?	KP709783	KP709846	KP709580-KP709584	KP709266-KP709270	KP709938-KP709939
B. phoenicoides	Bpho2	France: Var, Escalet. CAS.	2n=28	14	4x	KP709784	KP709847	KP709585	KP709271	KP709940-KP709944
B. phoenicoides	Bpho6	Spain: Huesca, Panzano, Sierra Guara. PC.	2n=28	14	4x	-	-	-	-	KP709945-KP709949
B. phoenicoides	Bpho39	France: Var, Montferrat. CAS.	2n=28	14	4x	JN187645	JN187670	JN187620	JN187597	-
B. phoenicoides	Bpho88	Portugal. CAS.	2n=28	14	4x	KP709785	KP709848	KP709586	KP709272-KP709277	HQ890975/HQ890981; KP709950-KP709954
B. phoenicoides	Bpho414	Portugal: Algarve. CAS.	2n=28	14	4x	KP709786	KP709849	KP709588-KP709589	KP709278-KP709283	KP709955-KP709959
B. phoenicoides	Bpho503	Spain: Almeria, Sorbas. PC.	2n=28	14	4x	KP709787	KP709850	KP709590	KP709284-KP709289	KP709960-KP709969
B. phoenicoides	Bpho504	Spain: Zaragoza, Alagón, Grisén. PC.	2n=28	14	4x	KP709788	-	-	KP709290	-
B. phoenicoides	Bpho507	Spain: Huesca, EPS. PC PC93	2n=28	14	4x	KP709789	KP709851	KP709587	KP709301-KP709306	KP709980-KP709984
B. phoenicoides	Bpho505	Spain: Huesca. USDA:PI 89817	2n=28	14	4x	KP709790	KP709852	KP709591-KP709595	KP709291-KP709295	KP709970-KP709974
B. phoenicoides	Bpho506	Spain: Toledo, Los Yeyenes.USDA:PI 318961	2n=28	14	4x	KP709791	KP709853	KP709596-KP709600	KP709296-KP709300	KP709975-KP709979

B. pinnatum	Bpin8	England: Sussex, Fairlight. CAS.	2n=28	14	4x	JN187646	JN187671	JN187621	-	-
B. pinnatum	Bpin11	England: Sussex, Fairlight. CAS & PC.	2n=28	14	4x	KP709793	KP709855	KP709611	-	-
B. pinnatum	Bpin229	Iran. CAS.	2n=18	9	2x	KP709794	KP709856	KP709612	KP709313-KP709318	-
B. pinnatum	Bpin243	Russia. CAS.	2n=18	9	2x	KP709795	KP709857	KP709613	KP709319-KP709324	-
B. pinnatum	Bpin281	Russia. CAS.	2n=18	9	2x	KP709796	KP709858	KP709614-KP709619	KP709325-KP709330	-
B. pinnatum	Bpin412	England: Wiltshire, Somerford Common. CAS & PC.	2n=28	14	4x	-	-	KP709620	KP709331	-
B. pinnatum	Bpin413	England: Gloucestershire, Halfmoon. CAS & PC.	2n=28	14	4x	KP709797	KP709859	KP709621-KP709626	KP709332-KP709336	-
B. pinnatum	Bpin418	England: Wiltshire, Swindon, Elborough Bridge. CAS & PC.	-	-	-	KP709798	KP709860	KP709627	KP709337-KP709342	-
B. pinnatum	Bpin501	Slovenia: Kamnik Alps, Cerklje, Ravne. Locotype. NJ.	-	-	-	KP709799	KP709861	KP709628-KP709633	KP709343-KP709348	-
B. pinnatum	Bpin502	Iraq. USDA:PI 185135	2n=16	8	2x	KP709800	KP709862	KP709634-KP709638	KP709349-KP709352	KP709985-KP709989
B. pinnatum	Bpin503	Turkey: Istanbul, 19 miles west of Istanbul. USDA: PI 251445	2n=28	147	4x	KP709801	KP709863	KP709639-KP709643	KP709353-KP709357	HQ890966/HQ890978;KP709990-KP709992
B. pinnatum	Bpin505	Norway. USDA: PI 345982	2n=18	9	2x	KP709802	KP709864	KP709644-KP709648	KP709358-KP709361	HQ890980; KP709993-KP709994
B. pinnatum	Bpin506	Ireland. USDA: PI 430277	2n=28	147	4x	KP709803	KP709865	KP709649-KP709653	KP709362-KP709366	KP709995-KP709996
B. retusum	Bret1	Spain: Huesca, Fraga, San Simon. PC.	2n=36	-	6x?	JN187647	JN187672	JN187622;KP709654-KP709655	KP709367-KP709371	-
B. retusum	Bret116	France: Aude, Pyrenees. CAS.	2n=36	-	6x?	KP709804	KP709866	-	KP709372	-
B. retusum	Bret171	Spain: Granada. PC.	2n=36	-	6x?	KP709805	KP709867	-	KP709373	-
B. retusum	Bret363	Spain: Alicante. PC.	2n=36	-	6x?	-	KP709868	KP709656	-	-
B. retusum	Bret403	Spain: Zaragoza. PC.	2n=36	-	6x?	KP709806	-	KP709672	-	-
B. retusum	Bret400	Spain: Huesca, Angüés, Bas. PC.	2n=36	-	6x?	KP709807	KP709869	KP709657-KP709661	KP709374-KP709378	KP709997-KP710006
B. retusum	Bret401	Spain: Balearic isles, Mallorca, Formentor. PC & DL.	2n=36	-	6x?	KP709808	KP709870	KP709662-KP709666	KP709379-KP709383	KP710007-KP710017
B. retusum	Bret402	Italy: Campania, Peninsula Sorrentina, Punta Campanella. PC.	2n=36	-	6x?	KP709809	KP709871	KP709667-KP709671	KP709384-KP709388	KP710018-KP710022
B. retusum	Bret3	Greece:4195. Wolny *et al* 2010	2n=38	-	-	-	-	-	-	HQ890965/HQ890967/HQ890979
B. rupestre	Brup4	France: Ambleteuse. CAS.	2n=28	14	4x	KP709792	KP709854	KP709601-KP709610	KP709307-KP709312	-
B. rupestre	Brup5	Spain: Huesca, Jaca, Castiello de Jaca, Aratorés. PC.	2n=28	14	4x	-	-	-	-	KP710023-KP710027
B. rupestre	Brup144	Germany. JM.	2n=28	14	4x	KP709810	KP709872	KP709673-KP709678	KP709389-KP709393	-
B. rupestre	Brup182	Croatia: Istria. CAS.	2n=28	14	4x	KP709811	KP709873	KP709679	KP709394	-
B. rupestre	Brup417	England: Wiltshire. CAS & PC.	-	-	-	-	KP709874	KP709680	-	-
B. rupestre	Brup420	England: Wiltshire, Morgan's hill. CAS & PC.	-	-	-	-	-	-	KP709395	-
B. rupestre	Brup421	England: Wiltshire. CAS & PC.	-	-	-	KP709812	KP709875	KP709681-KP709686	KP709396-KP709401	-
B. rupestre	Brup430	Slovenia: Notranjsko, Cerknica. NJ.	-	-	-	KP709813	KP709876	KP709687-KP709692	KP709402-KP709407	KP710028
B. rupestre	Brup431	Spain: Huesca, Canfranc, Somport. PC.	-	-	-	KP709814	KP709877	-	KP709408	-
B. rupestre	Brup437	Spain: Navarra, Betelu. PC PC2393.	-	-	-	JN187648	JN187673	JN187623;KP709693-KP709697	KP709425-KP709429	-
B. rupestre	Brup433	Italy.	2n=36?	9?	4x?	KP709815	KP709878	KP709698-KP709702	KP709409-KP709413	-
B. rupestre	Brup434	Ukraine:Krym. USDA: PI 639821	2n=18	9	2x	KP709816	KP709879	KP709703-KP709704	KP709414-KP709419	KP710029-KP710031
B. rupestre	Brup435	Russia. USDA: PI 440172	2n=18	9	2x	KP709817	KP709880	KP709705-KP709709	KP709420-KP709424	HQ890964; KP710032-KP710036

B. sylvaticum	Bsyl	England: Leicester (cultivated)	-	-	-	-	-	KP709710	-	-
B. sylvaticum	Bsyl2	England: Wiltshire, Sommerford, Someford Common. CAS & PC.	2n=18	9	2x	KP709818	-	-	KP709430	-
B. sylvaticum	Bsyl4	England: Leicester (cultivated)	2n=18	9	2x	-	KP709881	KP709711	-	-
B. sylvaticum	Bsyl28	Belgium: Liége (cultivated at the Botanical Garden). CAS.	2n=18	9	2x	KP709819	-	-	-	-
B. sylvaticum	Bsyl131	Hungary. CAS.	2n=18	9	2x	JN187649	JN187674	JN187624	KP709431-KP709435	-
B. sylvaticum	Bsyl372	Spain: Balearic isles, Mallorca, near Lluc. PC & DL.	2n=18	9	2x	-	KP709882	KP709712	KP709436	-
B. sylvaticum	Bsyl416	England:Leicestershire, Leicester. CAS & PC.	2n=18	9	2x	KP709820	KP709883	-	KP709437-KP709442	-
B. sylvaticum	Bsyl419	England:Wiltshire, Somerford Common. CAS & PC.	2n=18	9	2x	KP709821	KP709884	KP709713-KP709714	KP709443-KP709448	-
B. sylvaticum	Bsyl422	Slovakia: Ruzomberok. CAS.	2n=18	9	2x	-	-	-	-	-
B. sylvaticum	Bsyl449	China:Sichuan, RS B4570BH.	-	-	-	KP709829	KP709892	KP709753-KP709757	KP709486-KP709491	KP710057
B. sylvaticum	Bsyl440	Spain:Canary Isles, La Palma. PC.	2n=18	9	2x	KP709822	KP709885	KP709715-KP709720	KP709449-KP709454	-
B. sylvaticum	Bsyl441	Spain: Canary Isles, Gomera. MA 682635	2n=18	9	2x	-	-	KP709721	-	-
B. sylvaticum	Bsyl442	Spain: Huelva, Aracena. SEV299/09	2n=18	9	2x	KP709823	KP709886	KP709722-KP709727	KP709455-KP709460	KP710037-KP710041
B. sylvaticum	Bsyl443	Spain. USDA: PI 237792	2n=18	9	2x	KP709824	KP709887	KP709728-KP709732	KP709461-KP709465	KP710042-KP710045
B. sylvaticum	Bsyl444	Australia. USDA: PI 297868	2n=18	9	2x	KP709825	KP709888	KP709733-KP709737	KP709466-KP709470	HQ890976/HQ890977; KP710046-KP710051
B. sylvaticum	Bsyl445	Spain: Avila, Gredos, Candeleda. USDA: PI 318962	2n=18	9	2x	KP709826	KP709889	KP709738-KP709742	KP709471-KP709475	KP710052-KP710056
B. sylvaticum	Bsyl446	Iran: Ardebil, on east side of grade to Astara. USDA: PI 380758	2n=18	9	2x	KP709827	KP709890	KP709743-KP709747	KP709476-KP709480	-
B. sylvaticum	Bsyl447	Iran: east of Gorgan. USDA: PI 268222	2n=18	9	2x	KP709828	KP709891	KP709748-KP709752	KP709481-KP709485	-
B. sylvaticum cf. var. *pseudodistachyon*	Bsyl450	New Guinea. Kew MWC8191	-	-	-	-	KP709893	KP709758	-	-
B. sylvaticum cf. var. *pseudodistachyon*	Bsyl451	New Guinea. Kew MWC8192	-	-	-	-	KP709894	KP709759	-	-
B. sylvaticum cf. var. *pseudodistachyon*	Bsyl452	New Guinea.Kew MWC8193	-	-	-	-	KP709895	KP709760	-	-
B. sylvaticum cf. var. *pseudodistachyon*	Bsyl453	New Guinea.Kew MWC8194	-	-	-	-	KP709896	KP709761	-	-
outgroups										
Oryza sativa		Genbank	2n=24		2x	AY522330	AY522330	AF169230	EF661831	FJ235799
Melica ciliata		Spain: Huesca: Fraga. PC PC1793	2n=18		2x	JN187625	JN187650	JN187602	JN187580	-
Glyceria declinata		Spain: Cantabria: Picos de Europa. PC PC2893	2n=20		2x	JN187626	JN187651	JN187603	JN187581	-
Secale cereale		GenBank	2n=14		2x	EU012710.1	EU013658.1	AF303400	AJ315034.1	-
Festuca pratensis		England: Wilshire, Calne, CAS & PC, and GenBank	2n=14		2x	JN187627	JN187652	JN187604	JN187582	FN376854
Lolium perenne		England: Leicester, CAS & PC, and GenBank	2n=14		2x	JN187628	JN187653	JN187605	JN18758	DQ534010
Hordeum vulgare		GenBank				-	-	-	-	BJ481891

Abbreviations: Collectors: AB & SE, Adina Breiman and Smadar Ezrati; AC, Ana Caicedo; AM, Antonio Manzaneda; BLS, Bing-Ling Shih; CA, Carlos Aedo; CAS, Clive A Stace; CS, Consuelo Soler; DL, Diana López-Alvarez; HB, Hikmet Budak; JM, Jöchen Müller; JV, John Vogel; LM, Luis Mur; MP, Manuel Pimentel; MRR, Mohammad Reza Rahiminejad; NJ, Neg Jogan; PC, Pilar Catalán; RS, Robert Soreng, SH, Samuel Hazen; ST, Salvador Talavera. Herbaria and Germplasm Banks: ABR, Aberystwyth (UK); INRA, Institut National de la Recherche Agronomique (France); JBVC, Jardin Botánico Viera y Clavijo (Spain); MA, Real Jardin Botanico de Madrid (Spain); RH, Reading Herbarium (UK); USDA, United States Department of Agriculture (US).

Table S2. Dispersal rate matrices reflecting the palaeogeographic connectivity among the study areas in each historical scenario (time slices TSI, TSII, TSIII). Areas: A) western Mediterranean; B) eastern Mediterranean + SW Asia; C) western Eurasia (from Atlantic to Urals); D) eastern Eurasia (from Urals to Pacific and eastern Asia); E) Canary Isles; F) America (from Mexico to Peru-Bolivia); G) Africa (Tropical Africa and South Africa); H) Madagascar; I) Taiwan; J) Malesia (including Papua-New Guinea).

TSI: Middle Miocene (Langhian) to Late Miocene (Tortonian); 16.2 - 7.2 Ma)

	A	B	C	D	E	F	G	H	I	J
A	-	0.75	1	0.5	0.5	0.01	0.25	0.01	0.01	0.01
B		-	0.5	1	0.25	0.01	0.5	0.25	0.01	0.01
C			-	1	0.25	0.01	0.25	0.01	0.01	0.01
D				-	0.01	0.75	0.25	0.01	1	0.75
E					-	0.01	0.01	0.01	0.01	0.01
F						-	0.01	0.01	0.01	0.01
G							-	0.5	0.01	0.01
H								-	0.01	0.01
I									-	0.75
J										-

TSII: Late Miocene (Messinian) – Pliocene (7.2 – 2.6 Ma)

	A	B	C	D	E	F	G	H	I	J
A	-	1	1	0.5	0.5	0.01	0.01	0.01	0.01	0.01
B		-	0.75	1	0.25	0.01	0.75	0.01	0.01	0.01
C			-	1	0.25	0.01	0.5	0.25	0.01	0.01
D				-	0.01	0.01	0.5	0.25	1	0.5
E					-	0.01	0.01	0.01	0.01	0.01
F						-	0.01	0.01	0.01	0.01
G							-	0.75	0.01	0.01
H								-	0.01	0.01
I									-	0.5
J										-

TSIII: Pleistocene – Present (2.6 – 0 Ma)

	A	B	C	D	E	F	G	H	I	J
A	-	1	1	0.5	0.5	0.01	0.01	0.01	0.01	0.01
B		-	1	1	0.25	0.01	0.75	0.01	0.01	0.01
C			-	1	0.25	0.01	0.5	0.25	0.01	0.01
D				-	0.01	0.01	0.5	0.25	1	1
E					-	0.01	0.01	0.01	0.01	0.01
F						-	0.01	0.01	0.01	0.01
G							-	0.75	0.01	0.01
H								-	0.01	0.01
I									-	0.5
J										-

Appendix II: Supporting Information of Chapter 2

Methods and Results S1

Two approaches were developed to extract and analyse the information of homeologous subgenomes of polyploid species. The first approach (A) was conducted with the aim of extracting SNPs by mapping genomic/transcriptomic sequence reads against three diploid reference genomes. The second approach (A) was developed to recover core genes/transcripts expressed in *Brachypodium* species and to label homeologous sequences according to their placement within the consensus phylogram of diploid species. They are described in detail below.

(A) Detailed description of reference-genome syntenic mapping pipeline (figs. 1a, b, c)

We have developed a set of bioinformatic tools (*mapcoords*, *vcf2alignment* and *vcf2alignment_synteny*) for extracting, filtering and aligning single nucleotide polymorphisms (SNPs) from sequence reads mapped on several reference genomes. Our pipeline accepts both genomic and transcriptomic sequences, being able to combine data from different sources mapped on the same reference genome. The complete protocol is available at https://github.com/eead-csic-compbio/vcf2alignment.

Perl script *vcf2alignment* filters SNPs according to parameters such as coverage, missing data per site, bi or/and multi-allelic loci, homozygous/heterozygous, constant or/and polymorphic sites. This tool produces datasets of aligned SNPs in FASTA format.

Perl script *mapcoords* produces a table of syntenic positions extracted from whole-genome alignments of two species computed with CGaln (http://www.iam.u-tokyo.ac.jp/chromosomeinformatics/rnakato/cgaln).

Perl script *vcf2alignment_synteny* combines SNPs data set from *vcf2alignment* with syntenic positions produced by *mapcoords*, producing a dataset of aligned SNPs which distinguishes SNPs mapped on each reference genome. Thus, the script extracts and aligns syntenic SNPs. This tool produces also aligned SNPs in FASTA format.

In the current study, we have mapped all paired-end reads on three concatenated reference genomes of *Brachypodium* (*Brachypodium distachyon v3.1*, line Bd21 from

Irak (IBI, 2010), *Brachypodium stacei*, line ABR114 from Formentera, Balearic Isles, Spain (*Brachypodium stacei* v1.1 *DOE-JGI,* http://phytozome.jgi.doe.gov/) and *Brachypodium sylvaticum*, line Ain-1 from Tunisia (*Brachypodium sylvaticum* v1.1 DOE-JGI, http://phytozome.jgi.doe.gov). These genomes show different divergence times, being *B. stacei* the most ancestral, *B. distachyon* intermediate and *B. sylvaticum* the most recently evolved. Using several reference genomes with different divergence times we aim to recover SNPs from different ancestors of allopolyploids.

The general workflow is illustrated in fig. 1a, b, including the pre-processing of paired-end reads, mapping, variant calling and filtering of SNPs (*vcf2alignment*). The steps of extracting syntenic positions (*mapcoords*) and SNPs matching those syntenic positions (*vcf2aligmnet_synteny*) are summarized on the right side. This figure shows two generic reference genomes, named “Master Genome” and “Secondary genome”, but in our study we actually used two secondary genomes.

A.1. Mapping paired-end reads on reference genome/s

We mapped all paired-end reads on three concatenated reference genomes in order to recover SNPs from each species. Alignment tools used were bwa 0.7.12-r1039 (Li & Durbin, 2009) and hisat2-2.0.4 (Kim et al., 2015) for GBS and RNA-seq paired-end reads, respectively. Alignment files (SAM format) were converted to BAM format and sorted by samtools-1.3.1 software (Li & Durbin, 2009; Li et al., 2009).

A.2. Converting BAM sorted files to VCF files

We generated variant calling format files (VCF) for each sorted BAM and carried out SNP calling using bcftools-1.3.1 (Danecek et al., 2011). We kept biallelic, multiallelic and all alternative alleles present in the alignments at this stage. Finally, we merged all VCF files from each sample into one file. The merged VCF file is the starting point in order to filter, extract and combine the SNPs from all samples. Duplicate VCF lines with different variants for one position were removed with script *rm_double_lines*.

A.3. Filtering SNPs and converting to FASTA file using *vcf2alignment*

Perl script *vcf2alignment* was used to filter SNPs according to the following criteria: 10xcoverage, including bi and multi-allelic loci, only homozygous sites. A FASTA file of aligned SNPs and a LOG file with the statistics of SNPs per chromosome and sample are produced (fig. S1). The species phylogeny was obtained from this FASTA file.

A.4. Whole-genome alignment of reference genomes and extraction of equivalent/syntenic positions using *mapcoords*

We used “soft-masked” versions of all complete genomes to conduct whole-genome alignment, and removed unassembled contigs or scaffolds.

Brachypodium distachyon v3.1 was used as “Master Genome”. All three genome assemblies were aligned with Cgaln software (Nakato & Gotoh, 2010) using parameters -r (both strand), –fc (filter colony to extract consistent set), -cons (filter inconsistent HSPs at the HSP-chaining), block size (-BS) of 10,000 and X-drop-off at block-level (-X) of 12,000 nucleotides. *Brachypodium stacei* and *B. sylvaticum*, defined as “Secondary Genomes”, were aligned against *B. distachyon*.

Parameters k-mer size, block size (BS) and drop-off at block-level (X) were customized to improve the alignments. Recovered syntenic regions of *B. stacei* and *B. sylvaticum* aligned to *B. distachyon* reference genome are shown in the graphic of fig. 1b, c.

Script *mapcoords* was used to produce a table of 0-based equivalent coordinates in TSV format. Perl one-liners were used to extract the equivalent (syntenic) coordinates whose position matched a “valid locus” recovered by *vcf2alignment* from data set including constant sites.

A.5. Filtering and extracting syntenic SNPs using *vcf2alignment_synteny*

SNPs matching syntenic sites were filtered according to the same criteria of step 3. We recovered all “valid loci” for each sample respect to each reference genome, noted as “_Bdis” (*B. distachyon*), _Bsta (*B. stacei*) and _Bsyl (*B. sylvaticum*) using *vcf2alignment_synteny* (fig. 1b). A FASTA file of aligned syntenic SNPs was produced. Subgenomes of diploid species were collapsed to one line with script *collapse_aln*. Residual SNPs of _Bdis subgenome in *B. stacei* sample, _Bsta subgenome of *B. distachyon* sample and _Bsyl subgenome of *B. hybridum* samples were removed. The phylogeny of subgenomes was obtained from this FASTA file.

Benchmark and considerations on mapping sequences on several concatenated reference genomes

In order to validate our approach and test the biases produced when SNPs are extracted from sequence reads mapped on alternative genomes, we checked different combinations of reference genomes. RNA-seq reads were mapped on three different pairs of concatenated reference genomes: *B. stacei* + *B. sylvaticum*, *B. distachyon* + *B. sylvaticum* and *B. distachyon* + *B. stacei*. The results were compared to those obtained when all three *B. distachyon* + *B. stacei* + *B. sylvaticum* genomes were concatenated.

Most SNPs from core perennial species were recovered from reads mapped to *B. sylvaticum* (fig. S9a, b) and, less frequently, *B. distachyon* chromosomes (fig. S9c). *Brachypodium stacei* showed the fewest number of matches except in *B. hybridum*. These observations are congruent with the evolutionary position of the reference genomes with respect to the samples analyzed here. Indeed, *Brachypodium sylvaticum* is the most recently evolved, while *B. distachyon* is intermediate and *B. stace*i is the most ancestral species of the three reference genomes. Thus, core perennial species, the most recently evolved clade, displayed most SNPs in the *B. sylvaticum* reference. Moreover, *Brachypodium hybridum* had roughly the same number of SNPs mapped on its ancestral progenitor *B. stacei* and *B. distachyon*.

In the course of these benchmarks, subgenome trees showed some differences depending on the references used (figs. S10). However, phylograms of subgenomes using *B. stacei* + *B. sylvaticum* (fig. S10a) and *B. distachyon + B. stacei* (fig. S10c) as concatenated reference genomes showed very similar topologies, with the exception of recently ancestral copy of *B. retusum*, whose placement was more ancestral mapping on *B. distachyon + B. stacei* than *B. stacei* + *B. sylvaticum.* This incongruence is likely a consequence of the used reference, as *Brachypodium distachyon* is more ancestral than *B. sylvaticum*, it resulted in B. retusum_Bdis being placed in a more ancestral position than B. retusum_Bsyl.

When reads were mapped on all three concatenated references, we recovered three putative copies of *B. retusum* (fig. 4). However, the putative intermediate copy could not be detected when using only two concatenated genomes. In general, the topology arising with three concatenated reference genomes was congruent but more informative than with combinations of two references. In addition, the subgenome tree produced by concatenating *B. distachyon* + *B. sylvaticum* (fig. S10b) exhibited a bias, as it placed *B. distachyon* more ancestral than *B. stacei*. This is presumably a consequence of mapping ancestral samples to more recently evolved genomes such as *B. distachyon* and *B. sylvaticum*.

RNA-seq data used in synteny analyses included polymorphic (informative) and constant sites because we noted that constant sites also contributed to recover more putative subgenomes (fig. 4) than only informative sites.

We now discuss ways to reduce the bias due to mapping on concatenated reference genomes, maximizing the information about the putative copies (subgenomes) and progenitor of allopolyploids using our approach *vcf2alignment/vcf2alignment_synteny*.

- Previous knowledge of the phylogeny is required to choose the most suitable reference genomes, including quantity and species. Thus, we have checked our approach using two or three reference genomes with different combinations of species. The best option was mapping on three reference genomes, recovering more putative copies (such as the allohexaploid *B. retusum*), with congruent positions of positive control samples (*B. distachyon*, *B. stacei* and *B. hybridum*) and high support values (fig. 4). Our tests using two concatenated references were suitable using the most ancestral (*B. stacei*) and the most recently evolved (*B. sylvaticum*) available species genomes as reference (fig. S10a). Comparable results were recovered using the most ancestral (*B. stacei*) and intermediate evolved (*B. distachyon*) available species genomes (fig. S10c). However, incongruent results were obtained using the intermediate (*B. distachyon*) and recently evolved (*B. sylvaticum*) species genomes as reference (fig. S10b).
- Positive controls as different lines of the same species (*B. phoenicoides*-Bpho6 and *B. phoenicoides*-B422) or/and allopolyploid samples with known progenitors and evolutionary history (such as the *Brachypodium distachyon* complex) should be included in the analyses if they are available.
- Diploid samples with more than 90% of reads, and SNPs, matching a single reference genomes can be simplified removing the other residual mappings. In our tests, *B. distachyon* and *B. stacei* reads mapped back mostly to their respective reference genomes (98-99%, see table S5; S6). Consequently, as a result of this, artificial putative subgenomes _Bsta and _Bsyl in *B. distachyon*, and _Bdis and Bsyl in *B. stacei*, were removed. The same protocol can be conducted with allopolyploids whose evolutionary history is known if both progenitors are included in the analyses.

- Putative subgenomes of diploid samples of unknown evolution and progenitors are collapsed in one line according to syntenic position using script *collapse_aln*.
- Data set for synteny analyses has been evaluated using both informative (data not shown) and informative plus constant sites. More putative subgenomes were recovered using informative plus constant sites data set.

(B) Detailed description of the pipeline to label core transcripts of homeologous subgenomes from allopolyploid species (figs. 2a, b; 3a, b)

We have developed a set of bioinformatic tools (*trim_MSA_block, reroot_tree, check_diploids, check_lineages_polyploids and make_lineage_stats*) for filtering multiple sequence alignment, rooting and sorting phylograms, checking diploid skeleton, and labeling homeologous subgenomes of allopolyploid species respect to nearest diploid species.

B.1. Filtering multiple sequence alignments (MSA) and defining the compact block of sequences from diploid species

Core transcript clusters recovered by GET_HOMOLOGES_EST were first depurated using script annotate_cluster.pl –collapse 20 (included in that software suite). . Then, script *trim_MSA_block* was ran to obtain a compact block of unique sequences for each diploid species, removing short and fragmented sequences (<100 bp), and isoforms. Alignments which did not include diploid species (*B. stacei, B. distachyon, B. arbuscula* and *B. pinnatum* or *B. sylvaticum*) and both outgroups (*Hordeum vulgare* and *Oryza sativa*) were removed. We recovered 1,786 MSAs from a total of 3,324 core clusters. Finally, polyploid sequences which not overlap at least 50% of the diploid block were removed; surviving MSAs were further processed by trimAl v1.4.rev15 using -zautomated1 option to remove spurious sequences or poorly aligned regions. Those selected and filtered alignments were used for the phylogenomic analysis of homeologous subgenomes.

B.2. Building, rooting and sorting of core transcripts trees

Phylogenetic trees for each core transcript cluster were conducted by IQ-TREE using ModelFinder (Kalyaanamoorthy et al., 2017) to model selection. Trees were rooted and sorted in decreasing order by Perl script *reroot_tree*.

B.3. Checking congruent placement of diploid species for each phylogenetic tree

Phylogenetic trees were checked attending to the placement of diploid species. Previous analyses have recovered the evolutionary emplacement of diploid species included in this study (figs. S2; S3; S4), showing two topologies as the most congruent evolved scenarios, from the most ancestral to recent, *B. stacei, B. distachyon, B. arbuscula, B. pinnatum/B. sylvaticum* or *B. sylvaticum/B. pinnatum*. The diploid skeleton was checked for each phylogenetic tree with script *check_diploids*, recovering 397 core transcripts cluster MSAs, and the corresponding trees, with congruent diploid topology.

B.4. Checking and labeling homeologous subgenomes of allopolyploids species

Polyploid tips of trees with congruent diploid topology [consensus topology from GBS, RNA-seq (figs. S3, S4, S5) species tree, and statistics of positions of each diploid species in the gene trees (fig. 5a)] were analysed to label them using as reference the nearest diploid species (ancestor, descendant and sister diploid species). Thus, we defined the following rules to label each polyploid tips (fig. 3a, b):

- A → *Brachypodium* stem branch.
- B → *B. stacei* sister lineage or *B. stacei* sister branch sister lineage.
- C → *B. distachyon* sister lineage or *B. distachyon* sister branch sister lineage.
- D → *B. arbuscula* sister lineage or any other core-perenial clade nested lineage.

Those rules can be adapted, for example, detailing more specific diploid nodes (see table S8):

- A →*Brachypodium* stem branch.
- B → *B. stacei* sister sister lineage.
- C → *B. stacei* sister branch sister lineage lineage.
- D → *B. distachyon* sister lineage.
- E → *B. distachyon* sister branch sister lineage.
- F → *B. arbuscula* sister lineage.
- G → *B. arbuscula* sister branch sister lineage.
- H → *B. pinnatum* or *B. sylvaticum* sister lineage.
- I → *B. pinnatum* or *B. sylvaticum* sister branch sister lineage.

This script generates three output files:

- MSA labeled FASTA reduced → complete gapped sequences, with subgenomes which were not recovered not included.
- MSA labeled FASTA → complete gapped sequences, with subgenomes not recovered included as gap-only MSA lines. Those files were eventually concatenated to build a large multi-gene MSA.
- Gene tree labeled → phylograms of gene clusters with polyploid tips labeled in Newick format.

B.5. Labeled concatenated MSA, FASTA and statistical report

Statistical reports were generated with script *make_lineage_stats* to check the number of subgenomes recovered for each allopolyploid species. Subgenomes with values (number of subgenomes per species) under the threshold 10% of genes studied were removed from large multi-gene alignment. Thus, we only conserved the most representative subgenomes with the objective of recovering the most plausible ancestors of allopolyploid species.

All labeled MSA FASTA files were concatenated with script *concat_alignment* from the GET_PHYLOMARKERS suite (Vinuesa et al., 2018) and the less representative subgenomes removed.

This large multi-gene FASTA file was used for subsequent dating and phylogenomics analyses.

Considerations about labeling core transcripts from allopolyploid homeologous subgenomes.

Our approach is supported by previous knowledge about diploid species because all downstream analyses are based on a congruent and robust phylogenetic tree of diploid species which were uses to label allopolyploid homeologous subgenomes.

In order to test our approach, different levels of trimming/filtering were conducted. We concluded that strict filtering has to be conducted if one wants to build a compact block of diploid sequences to cover the complete alignment. Redundant and incomplete sequences have to be removed to avoid incongruent tips (e.g. polytomies). This seems important particularly for de-novo assembled transcripts.

We used positive controls to validate our pipeline. In particular, *B. hybridum* (4x) is an allopolyploid whose genitors are not extinct and are included in diploid skeleton: *B.*

stacei (2x) and *B. distachyon* (2x). As expected, we recovered sister emplacement of homeologous subgenomes of *B. hybridum*, with B. stacei-type sequences sister to *B. stacei* and B. distachyon-type sequences sister to *B. distachyon*. As second positive control we added two ecotypes of the same species (*B. phoenicoides*-Bpho6 and *B. phoenicoides* B422) to confirm the same phylogenetic emplacement of both homeologous subgenomes in all species/subgenomes trees.

Finally, it is useful to know the ploidy of the species studied to fix the rules of labelling polyploid tips and to remove the poorly represented subgenomes by adjusting the number of subgenomes to ploidy level. Exceptions could be found respect to correlation between ploidy and homeologous subgenomes recovered, as *B. retusum* (tetra/hexa-allopolyploid) and *B. boissieri* (hexa/octopolyploid), with four and three homeologous subgenomes recovered, respectively. The ploidy of those species are not confirmed yet and the expression level of certain transcripts could be over/under-expressed with respect to homeologous subgenomes.

Supplemental phylogenomics and dating results

Phylogenomics based on reference-genome synteny mapping: the *Brachypodium* species and subgenome trees

The GBS-based tree recovered an incongruent ancestral position of *B. distachyon* with respect to *B. stacei*, though this tree reconstructed *B. pinnatum* 4x (only sampled for GBS) as the sister lineage to the other core perennial allotetraploids (*B. rupestre* 4x/*B. phoenicoides*) (fig. S2b). Independent re-analyses with NGSEP and GIbPSs tools validated the topology but also failed to produce consistent splits between *B. stacei* and *B. distachyon* and within the core perennial clade (fig. S4a, b).

Dating the *Brachypodium* plastome tree

Dating analysis was conducted with the *Brachypodium* 31 core plastome gene data using BEAST 2.5.0, the same priors and calibration points imposed in the 397 nuclear core transcripts analysis, and running 20,000,000 MCMC, aiming to compare the nodal divergence estimations with those of the nuclear core gene tree. A maximum clade credibility (MCC) tree was computed after discarding 1% of the respective saved trees as burn-in. The dated MCC plastome cladogram (data not shown) inferred a split age for the *Brachypodium* stem node in the Mid-Oligocene (32.0 Ma) slightly older than that of the nuclear dated tree but both showing overlapping HDP intervals. however, the estimated ages for the splits of *Brachypodium* crown node and all subsequent divergences were considerably younger than the estimations inferred in the nuclear tree. Due to the smaller size and lower rate of mutation of the plastome sequences, only the nuclear estimations will be further considered.

Supporting Tables

Table S1. Filtered paired-end (PE) and single-end (SE) reads used to build the respective RNA-seq and GBS data sets of the *Brachypodium* species, cytotypes and outgroup taxa under study. Newly generated data are indicated in bold. Crosses and asterisks indicate transcriptome and genomic data obtained in other studies, respectively]. Sources of accessions are indicated in table 1.

	Filtered Paired-end reads	
Samples	**RNA-seq**	**GBS**
B.arbuscula	**25,461,838**	**1,099,183**
B.boissieri	**23,602,582**	**718,919**
***B.distachyon* Bd21*†**	**24,523,648**	**1,386,369**
***B.hybridum* ABR113**	**-**	**813,379**
***B.hybridum* BdTR6g**	**13,876,186**	**-**
B.mexicanum	**19,464,557**	**447,502**
***B.phoenicoides* Bpho6**	**26,550,837**	**840,138**
***B.phoenicoides* B422**	**22,318,373**	**1,024,389**
***B.pinnatum* 2x**	**23,012,037**	**1,365,497**
***B.pinnatum* 4x**	**-**	**1,483,157**
B.retusum	**25,579,987**	**1,100,569**
B.rupestre	**20,203,954**	**907,919**
***B.stacei* ABR114**	**-**	**707,377**
***B.stacei* TE4.3**	**10,439,505**	**-**
B.sylvaticum Sin1*	-	186,404,851
B.sylvaticum Cor†	55,725,304	-
B.sylvaticum Esp†	50,687,151	-
B.sylvaticum Gre†	47,313,815	-
*Oryza sativa**†	33,377,660	178,359,840
*Hordeum vulgare**†	23,922,873 (SE)	391,220,241

Table S2. Statistics of the assembled transcripts obtained from the *Brachypodium* species and ecotypes under study using Trinity assembler. Genes correspond to Trinity components, while transcripts include all assembled isoforms. Contig N50 indicates that at least half of all assembled nucleotides are in transcript contigs of at least the detected N50 length value. Sources of accessions are indicated in table 1.

				Stats based on ALL transcript contigs				Stats based on ONLY LONGEST ISOFORM per 'GENE'			
Samples	**Total Trinity genes**	**Total Trinity transcripts**	**% GC**	**Contig N50**	**Median contig length**	**Average contig length**	**Total assembled nucleotides**	**Contig N50**	**Median contig length**	**Average contig length**	**Total assembled nucleotides**
B.arbuscula	81,409	124,549	48.41	1,507	555	909.59	113,288,406	1,306	396	741.57	60,370,486
B.boissieri	85,265	132,525	48.85	1,258	483	791.07	104,835,946	1,123	371	676.95	57,720,048
***B.hybridum* BdTR6g**	72,025	102,627	49.73	1,112	430	709.31	72,794,463	980	348	616.78	44,423,843
B.mexicanum	92,817	137,735	49.34	1,121	423	709.82	97,767,065	905	338	594.9	55,217,218
***B.phoenicoides* Bpho6**	108,114	160,417	48.03	1,359	511	842.17	135,098,959	1,190	393	712.14	76,992,275
***B.phoenicoides* B422**	94,758	139,769	48.48	1,310	471	802.07	112,104,432	1,098	356	667.83	63,282,223
***B.pinnatum* 2x**	72,371	108,186	47.94	1,538	552	918.81	99,402,628	1,337	394	747.45	54,093,562
B.retusum	101,941	154,466	48.6	1,217	464	765.63	118,264,125	999	354	631.91	64,418,005
B.rupestre	86,845	119,677	48.81	1,185	414	726.02	86,888,429	989	342	620.08	53,850,943
***B.stacei* TE4.3**	55,366	72,611	49.31	1,352	469	810.36	58,840,986	1,211	374	702.31	38,883,912

Table S3. Plastome data set of the *Brachypodium* species and cytotypes and outgroup taxa (*Hordeum vulgare*, *Oryza sativa*) under study. **(A)** Paired-end (PE) (plus singled-end (SE) in *Hordeum vulgare*) reads "fished" by DUK and number of SNPs extracted with the vcf2alignment. **(B)** Statistic of the assembled transcripts obtained using Trinity assembler. The Total Trinity 'genes' correspond to components. Contig N50 indicates that at least half of the assembled nucleotides are in transcript contigs of at least N=50 length value. Accession codes correspond to those indicated in table 1.

(A)

Accession	Barb	Bboi	Bdis Bd21	Bhyb BdTR6g	Bmex	Bpho6	Bpho B422	Bpin 2x	Bret	Brup	Bsta TE4.3	Bsyl Cor	Bsyl Esp	Bsyl Gre	Osat	Hvul
PE reads	104,241	917,555	353,520	89,166	73,881	234,188	104,126	84,482	107,968	64,543	146,354	332,462	452,975	231,309	11,284,095	12,726
SNPs	1,743	2,139	1,845	912	740	1,970	1,410	1,116	1,644	937	1,078	1,832	2,031	1,571	1600	66

(B)

				Stats based on ALL transcript contigs				Stats based on ONLY LONGEST ISOFORM per 'GENE'			
Samples	Total Trinity 'genes'	Total Trinity transcripts	% GC	Contig N50	Median contig length	Average contig length	Total assembled nucleotides	Contig N50	Median contig length	Average contig length	Total assembled nucleotides
B.boissieri	129	162	37.18	3,646	428.5	1,315.6	213,056	4,062	375	144.40	147,628
B.mexicanum	127	157	39.08	1,848	392	832.49	130,701	1,240	339	688.89	87,489
B.phoenicoides Bpho6	137	176	37.58	3,808	395	1,257.19	221,265	4,049	355	1064.61	145,851
***B.phoenicoides* B422**	121	149	37.49	3,315	413	1,075.58	160,261	2,376	330	881.39	106,648
***B.pinnatum* 2x**	138	164	37.71	2,797	405.5	954.68	156,568	1,656	391.5	848.99	117,160
B.retusum	148	179	37.46	2,905	431	1,087.41	194,646	2,442	373,5	951.16	140,772
B.rupestre	128	159	38.74	1,792	410	908.50	144,452	1,266	381	772.36	98,862

Table S4. Global statistics of paired-end (PE) and single-end (SE) reads filtered from RNA-seq **(A)** and GBS **(B)** data obtained from the *Brachypodium* species and cytotypes and for the outgroup species under study mapped on three concatenated reference genomes *Brachypodium distachyon* Bd21 – *B. stacei* ABR114 – *B. sylvaticum* Ain1 (quality map threshold ≥ 30). Figures about the total number of mapped reads, supplementary reads (reads showing chimeric, fused or non-linear alignments), final mapped reads (total mapped reads with supplementary reads removed), read1 and read2 (final mapped reads split into paired-end (PE) read1 and read2), properly paired reads (reads with correct insert size and orientation; percentage in parenthesis), mate PE mapped (final mapped pair-end reads with singletons removed), singletons (only one read mapped from each pair; percentage in parenthesis), mates mapped to different chr (paired-end reads mapped onto different chromosomes) are indicated for the studied accessions. None of the mapped reads aligned to more than one site. None of them were duplicated. Sources and abbreviations of accessions correspond to those indicated in table 1.

(A)

Reads	Barb	Bboi	Bdis	Bhyb BdTR6g	Bmex	Bpho6	Bpho B422	Bpin 2x	Bret	Brup	Bsta TE4.3	Bsyl Cor	Bsyl Esp	Bsyl Gre	Osat	Hvul
total mapped	36,092,762	21,155,034	45,410,629	23,982,182	15,820,114	35,346,553	31,528,441	34,428,054	30,122,778	27,598,056	18,534,866	84,166,875	77,734,464	69,468,199	18,397,692	1,395,176 (SE)
final mapped reads	36,092,762	21,155,034	45,410,629	23,982,182	15,820,114	35,346,553	31,528,441	34,428,054	30,122,778	27,598,056	18,534,866	84,166,875	77,734,464	69,468,199	18,397,692	-
read1	18,018,982	10,546,428	22,724,464	11,995,902	7,916,236	17,645,078	15,756,275	17,175,609	15,000,872	13,808,112	9,269,335	42,105,476	38,877,870	34,740,532	9,550,201	-
read2	18,073,780	10,608,606	22,686,165	11,986,280	7,903,878	17,701,475	15,772,166	17,252,445	15,121,906	13,789,944	9,265,531	42,061,399	38,856,594	34,727,667	8,847,491	-
properly paired	27,076,308 (75.02%)	13,914,376 (65,77%)	44,540,920 (98.08%)	14,114,050 (58.85%)	7,883,524 (49.83%)	27,036,690 (76.49%)	24,011,444 (76.16%)	26,400,160 (76.68%)	21,590,692 (71.68%)	15,948,782 (57.79%)	12,153,698 (65.57%)	74,655,906 (88.70%)	68,606,894 (88.26%)	59,823,454 (86.12%)	13,066,116 (71.02%)	-
mate PE mapped	33,820,915	18,396,278	45,076,951	23,760,685	13,838,802	33,207,657	29,733,766	32,694,170	27,272,729	26,448,990	18,380,713	79,482,855	73,805,075	65,370,764	15,262,548	-
singletons	2,271,847 (6.29%)	2,758,756 (13.04%)	333,678 (0.73%)	221,497 (0.92%)	1,981,312 (12.52%)	2,138,896 (6.05%)	1,794,675 (5.69%)	1,733,884 (5.04%)	2,850,049 (9.46%)	1,149,066 (4.16%)	154,153 (0.83%)	4,684,020 (5.57%)	3,929,389 (5.05%)	4,097,435 (5.90%)	3,135,144 (17.04%)	-
mates mapped to different chr	2,142,648	1,806,882	267,752	337,075	1,393,131	1,821,362	1,740,492	1,800,848	2,117,600	1,144,184	161,413	3,446,406	3,171,888	3,262,392	1,120,314	-

(B)

Reads	Barb	Bboi	Bdis	Bhyb ABR113	Bmex	Bpho6	Bpho B422	Bpin-2x	Bpin-4x	Bret	Brup	Bsta ABR114	Bsyl Sin1	Osat	Hvul
total mapped	1,706,066	894,379	2,592,717	1,499,365	491,789	1,274,155	1,552,738	2,102,169	2,246,079	1,527,626	1,392,447	1,352,317	234,188,489	6,432,465	26,882,873
supplementary	13,056	5,835	10,324	8,869	2,077	11,514	14,759	16,866	30,312	11,084	13,167	6,616	2,773,085	4,081	30,539
final mapped reads	1,693,010	888,544	2,582,393	1,490,496	489,712	1,262,641	1,537,979	2,085,303	2,215,767	1,516,542	1,379,280	1,345,701	231,415,404	6,428,384	26,852,334
read1	847,531	444,541	1,292,876	745,969	245,007	631,917	769,620	1,044,033	1,108,997	758,602	690,362	673,680	115,924,193	3,208,000	13,733,172
read2	845,479	444,003	1,289,517	744,527	244,705	630,724	768,359	1,041,270	1,106,770	757,940	688,918	672,021	115,491,211	3,220,384	13,119,162
properly paired	1,596,420 (94.29%)	853,960 (96.11%)	2,445,511 (94.70%)	1414906 (94.93%)	473,578 (96.71%)	1,197,578 (94.85%)	1,447,282 (94.10%)	1,980,441 (94.97%)	1,997,573 (90.15%)	1,435,812 (94.68%)	1,297,618 (94.08%)	1,287,163 (95.65%)	218,152,218 (94,27%)	5,003,476 (77.38%)	22,908,834 (85.31%)
mate PE mapped	1,685,515	882,592	2,577,270	1,486,796	486,420	1,256,091	1,529,280	2,076,015	2,195,798	1,505,494	1,371,325	1,343,251	230,117,338	5,141,505	23,420,657
singletons	7,495 (0.44%)	5,952 (0.67%)	5,123 (0.20%)	3700 (0.25%)	3,292 (0.67%)	6,550 (0.52%)	8,699 (0.57%)	9,288 (0.45%)	19,969 (0.90%)	11,048 (0.73%)	7,955 (0.58%)	2,450 (0.18%)	1,298,066 (0,56%)	1,286,879 (20.02%)	3,431,677 (12.78%)
mates mapped to different chr	47,392	20,714	53,830	37,258	9,657	35,244	48,350	52,352	133,017	44,086	43,485	29,789	9,191,107	121,725	479,703

Table S5. Percentages of paired-end (PE) and single-end (SE) reads filtered from RNA-seq **(A)** and GBS **(B)** data mapped on each of the three reference genomes of *Brachypodium* (*B. distachyon* Bd21, *B. stacei* ABR114, *B. sylvaticum* Ain1) (quality map threshold ≥ 30). Sources and abbreviations of accessions correspond to those indicated in table 1.

(A)

Reference/accession	Barb	Bboi	Bdis Bd21	Bhyb BdTR6g	Bmex	Bpho6	Bpho B422	Bpin 2x	Bret	Brup	Bsta TE4.3	Bsyl Cor	Bsyl Esp	Bsyl Gre	Osat	Hvul
B. distachyon Bd21	14.6	32.4	99.4	48.0	30.0	13.3	13.3	12.6	19.2	12.0	1.0	8.5	8.8	10.8	50.8	36.7
B. stacei ABR114	6.5	28.0	0.3	50.7	33.9	6.3	6.3	5.9	14.8	5.9	98.1	4.1	4.1	5.4	24.8	32.0
B. sylvaticum Ain1	78.9	39.6	0.3	1.3	36.1	80.4	80.4	81.5	66.0	82.1	0.9	87.4	87.1	83.8	24.4	31.3

(B)

Reference/accession	Barb	Bboi	Bdis Bd21	Bhyb ABR113	Bmex	Bpho6	Bpho B422	Bpin 2x	Bpin 4x	Bret	Brup	Bsta ABR114	Bsyl Sin1	Osat	Hvul
***B. distachyon* Bd21**	13.9	29.0	99.8	49.4	28.2	13.4	13.3	11.9	11.9	19.7	12.2	0.0	7.0	28.9	34.7
***B. stacei* ABR114**	6.1	27.2	0.15	49.4	33.2	6.3	6.2	5.4	5.3	15.6	5.8	99.9	2.8	34.3	35.4
***B. sylvaticum* Ain1**	80.0	43.8	0.05	1.2	38.6	80.3	80.5	82.7	82.8	64.7	82.0	0.1	90.2	36.8	29.9

Table S6. Number of SNPs (percentage in parenthesis) extracted from RNA-seq **(A)** and GBS **(B)** data of the *Brachypodium* species and cytotypes and outgroup taxa under study mapped onto each of the three *Brachypodium* reference genomes (*B. distachyon* Bd21, *B. stacei* ABR114 and *B. sylvaticum* Ain1) using the *vcf2alignment* tool. Sources and abbreviations of accessions correspond to those indicated in table 1.

(A)

Reference/ accession	Barb	Bboi	Bdis Bd21	Bhyb BdTR6g	Bmex	Bpho6	Bpho B422	Bpin 2x	Bret	Brup	Bsta TE4.3	Bsyl Cor	Bsyl Esp	Bsyl Gre	Osat	Hvul
B. distachyon	92,355	112,238	211,388	156,072	77,862	84,027	77,638	73,269	95,053	49,997	2,607	67,784	67,107	78,218	892	9,082
Bd21	(23.3)	(35.1)	(99.8)	(55.0)	(34.1)	(23.6)	(23.2)	(20.8)	(26.8)	(20.2)	(1.8)	(18.3)	(17.7)	(21.6)	(39.1)	(32.2)
B. stacei	39,822	85,300	208	123,179	70,842	36,613	33,792	31,596	69,595	21,723	141,304	30,786	29,904	36,238	685	9,824
ABR114	(10.1)	(26.7)	(0.1)	(43.4)	(31.0)	10.3)	(10.1)	(9.0)	(19.6)	(8.8)	(96.7)	(8.3)	(7.9)	(10.0)	(30.1)	(34.9)
B. sylvaticum	263,767	122,295	317	4,580	79,779	235,754	222,972	247,409	190,569	175,475	2,245	271,292	282,380	247,507	701	9,249
Ain1	(66.6)	(38.2)	(0.4)	(1.6)	(34.9)	(66.1)	(66.7)	(70.2)	(53.6)	(71.0)	(1.5)	(73.4)	(74.4)	(68.4)	(30.8)	(32.9)
TOTAL	**395,944**	**319,833**	**211,913**	**283,831**	**228,483**	**356,394**	**334,402**	**352,274**	**355,217**	**247,195**	**146,156**	**369,862**	**379,391**	**361,963**	**2,278**	**28,155**

(B)

Reference/ accession	Barb	Bboi	Bdis Bd21	Bhyb ABR113	Bmex	Bpho6	Bpho B422	Bpin 2x	Bpin 4x	Bret	Brup	Bsta ABR114	Bsyl Sin1	Osat	Hvul
B. distachyon	5,537	8,514	15,262	13,671	6,662	6,421	6,795	5,323	6,699	9,539	6,598	7	5,019	543	229
Bd21	(16.8)	(29.1)	(99.9)	(52.7)	(28.4)	(18.0)	(17.9)	(15.8)	(17.3)	(23.4)	(18.2)	(0.1)	(11.6)	(14.8)	(14.0)
B. stacei	2,389	7,575	5	11,433	7,167	2,907	3,083	2,236	3,114	7,410	3,018	11,790	2,343	1,012	521
ABR114	(7.2)	(25.9)	(0.0)	(44.0)	(30.6)	(8.1)	(8.2)	(6.6)	(8.1)	(18.1)	(8.3)	(99.8)	(5.4)	(27.5)	(31.8)
B. sylvaticum	25,121	13,138	17	857	9,607	26,426	27,994	26,181	28,878	23,898	26,667	11	36,033	2,123	888
Ain1	(76.0)	45.0)	(0.1)	(3.3)	(41.0)	(73.9)	(73.9)	(77.6)	74.6)	(58.5)	(73.5)	(0.1)	(83.0)	(57.7)	(54.2)
TOTAL	**33,047**	**29,227**	**15,284**	**25,961**	**23,436**	**35,754**	**37,872**	**33,740**	**38,691**	**40,847**	**36,283**	**11,808**	**43,395**	**3,678**	**1,638**

Table S7. Aligned data sets showing number of aligned SNPs (total aligned sites, informative sites, and constant sites) obtained from filtered RNAseq and GBS data of the *Brachypodium* species and cytotypes and outgroup taxa under study processed with different bioinformatic pipelines. Best-fit models and species used to root the trees, selected by IQ-TREE software and imposed in the respective phylogenomic analyses of each of the six aligned data sets (RNA-seq: *vcf2alignment,* NGSEP, *vcf2alignment_synteny;* GBS: *vcf2alignment,* GIbPSs, NGSEP), are indicated. The different data sets were used in alternative phylogenomic reconstructions.

Data set	**Software**	**Total aligned sites**	**Informative sites**	**Constant sites**	**Best-fit model**	**Root**
RNA-seq	*vcf2alignment*	708,356	190,003	0	GTR+ASC	*Oryza sativa*
	NGSEP	2,319,362	637,001	978,367	GTR	*Oryza sativa*
	vcf2alignment_ synteny (including constant sites)	28,563,327	505,512	27,681,446	TVM+R4	*Oryza sativa*
GBS	*vcf2alignment*	71,831	17,439	0	TVM+ASC	*Oryza sativa*
	GIbPSs	51,427	31,365	0	TVM+ASC+R3	*Brachypodium stacei* ABR114
	NGSEP	326,084	37,290	82,769	TVM	*Oryza sativa*

Table S8. Number (#) and percentage (%) of genes representing putative homeologous subgenomes detected in the studied allopolyploid *Brachypodium* species and cytotypes by our *multigene-based phylogenomic* pipeline using aligned core transcripts. Subgenomes were classified into four types (A to D) according to the ancestral-sister-descendant branches of diploid backbone tree lineages where allopolyploid allelic copies representing potential subgenomes where grafted to. Subgenomes represented by less than 15% of the total number of transcripts were discarded. Asterisks indicate subgenomic transcripts removed from downstream phylogenomic and dating analyses. The most representative subgenomes of each allopolyploid taxon are highlighted in bold.

Code	*Subgenome type*		Bboi		Bhyb		Bmex		Bpho6		B422		Bret		Brup	
			#	%	#	%	#	%	#	%	#	%	#	%	#	%
A	*Ancestral (Brachypodium* stem branch*))*		**159**	***49***	3*	*1*	**171**	***59***	4*	*1*	3*	*1*	**77**	***22***	2*	*1*
B	*B.stacei* type	*B.stacei* sister branch	**83**	*2*	**182**	***53***	**79**	*6*	4*	*0*	4*	*0*	**57**	*1*	5*	*0*
		B.stacei - B. distachyon branch		***25***		*1*		***21***		*1*		*1*		***16***		*1*
C	*B.distachyon* type	*B.distachyon* sister branch	**58**	*3*	147	***44***	34*	*2*	**75**	*0*	**62**	*0*	**98**	*1*	**50**	*0*
		B.distachyon - B.arbuscula branch		***16***		*1*		*10*		***20***		***17***		***27***		***13***
D	Core perennial type	*B.arbuscula* sister branch	15*	*1*	2*	*0*	6*	*0*	**284**	*3*	**289**	*3*	**115**	*2*	**310**	*3*
		B.arbuscula -B. pinnatum/B. sylvaticum branch		*3*		*0*		*1*		*14*		*16*		*10*		*17*
		B.pinnatum/B.sylvaticum sister branch		*0*		*0*		*0*		*14*		*15*		*7*		*21*
		B.pinnatum or *B.sylvaticum* sister branch		*1*		*0*		*1*		***47***		***47***		***14***		***44***

Table S9. Annotations of exclusive accessory gene clusters retrieved from transcriptome data of the *Brachypodium* species and cytotypes under study using the GET_HOMOLOGUES-EST pipeline and the Pfam, RefSeq and SwissProt reference databases. Only those with annotations in at least two databases (Pfam, RefSeq, SwissProt) are shown. NA (data non available)

(A). Annotated gene clusters present in group ANI1 (*B. distachyon*, *B. hybridum, B. stacei, B. boissieri, B. mexicanum)* and absent in the remaining species of *Brachypodium* (ANI2). The total number of transcript clusters entering the annotation pipeline was 14.

cluster ID	Pfam	RefSeq	SwissProt
225819	NA	serine carboxypeptidase-like	Precursor of serine carboxypeptidase-like
202662	IBR domain, a half RING-finger domain	E3 ubiquitin-protein ligase RNF144A	E3 ubiquitin-protein ligase RNF144A
193821	SHQ1 protein; HIT zinc finger	PREDICTED: zinc finger HIT domain-containing protein 2 isoform X2	Zinc finger HIT domain-containing protein 2
190960	DnaJ domain; Protein of unknown function (DUF3752)	NA	DnaJ homolog subfamily C member 5B
186157	Cation transporter/ATPase, N-terminus; E1-E2 ATPase; haloacid dehalogenase-like hydrolase	PREDICTED: plasma membrane ATPase	ATPase 1, plasma membrane-type
187796	Nucleolar complex-associated protein; CBF/Mak21 family	PREDICTED: LOW QUALITY PROTEIN: nucleolar complex protein 3 homolog	Nucleolar complex protein 3 homolog
203669	Cofilin/tropomyosin-type actin-binding protein	PREDICTED: actin-depolymerizing factor 10	Actin-depolymerizing factor 6
185873	NLI interacting factor-like phosphatase; Double-stranded RNA binding motif; Double-stranded RNA binding motif	PREDICTED: RNA polymerase II C-terminal domain phosphatase-like 1 isoform X3	RNA polymerase II C-terminal domain phosphatase-like 2
215897	Myb-like DNA-binding domain; Myb-like DNA-binding domain	PREDICTED: transcription factor MYB44-like	Transcription factor MYB105

(B). Annotated gene clusters absent in group ANI1 (*B. distachyon*, *B. hybridum, B. stacei, B. boissieri, B. mexicanum)* and expressed in the remaining species of *Brachypodium* (ANI2). The total number of transcript clusters entering the annotation pipeline was 52.

cluster ID	Pfam	RefSeq	SwissProt
24990	Replication factor-A protein 1, N-terminal domain; Replication factor-A C terminal domain	PREDICTED: replication protein A 70 kDa DNA-binding subunit A isoform	Replication protein A 70 kDa DNA-binding subunit C
30112	DNA photolyase; Alpha/beta hydrolase family	NA	Precursor of pheophytinase, chloroplastic
28652	NA	PREDICTED: E3 ubiquitin-protein ligase XB3	Putative E3 ubiquitin-protein ligase XBAT31
20703	Protein kinase domain; Flavin-binding monooxygenase-like	PREDICTED: probable LRR receptor-like serine/threonine-protein kinase At4g29180	Receptor-like serine/threonine-protein kinase SD1-6
11653	Leucine rich repeat N-terminal domain; Leucine Rich Repeat; Leucine rich repeat; Protein kinase domain	PREDICTED: putative receptor-like protein kinase At3g47110	LRR receptor-like serine/threonine-protein kinase GSO2
17843	NB-ARC domain	PREDICTED: putative disease resistance RPP13-like protein 3	Putative disease resistance protein RGA4
9931	NA	PREDICTED: disease resistance protein RGA2-like isoform X2	Putative disease resistance protein RGA4
34523	3-oxo-5-alpha-steroid 4-dehydrogenase	3-oxo-5-alpha-steroid 4-dehydrogenase 1-like	3-oxo-5-alpha-steroid 4-dehydrogenase 2
52155	NA	PREDICTED: 26S protease regulatory subunit 6B homolog	26S protease regulatory subunit 6B homolog
34292	Peroxidase	PREDICTED: peroxidase 5-like	Precursor of peroxidase 3; Rare cold-inducible protein
15106	EamA-like transporter family	PREDICTED: WAT1-related protein At4g30420-like	Protein WALLS ARE THIN 1

(C). Annotated gene clusters present in perennial (*B. arbuscula, B. boissieri, B. mexicanum, B. phoenicoides, B. pinnatum, B. retusum, B. rupestre,* and *B. sylvaticum*) and absent in annual (*B. distachyon*, *B. hybridum* and *B. stacei*) species of *Brachypodium*. The total number of transcript clusters entering the annotation pipeline was 30.

cluster ID	Pfam	RefSeq	SwissProt
22314	GAG-pre-integrase domain	Retrovirus-related Pol polyprotein from transposon TNT 1-94	Retrovirus-related Pol polyprotein from transposon TNT 1-94
16095	NA	Sorcin-like	Sorcin
22513	NA	GDSL esterase/lipase At1g71691-like	GDSL esterase/lipase At5g55050
11155	Protein tyrosine kinase; Leucine rich repeat	Cysteine-rich receptor-like protein kinase	G-type lectin S-receptor-like serine/threonine-protein kinase RKS1
2215	Homeobox domain	Homeobox-leucine zipper protein ROC6-like	Homeobox-leucine zipper protein HDG1 (GLABRA-like)
34561	Leucine rich repeat N-terminal domain; Protein kinase domain	Probable LRR receptor-like serine/threonine-protein kinase	Probable inactive leucine-rich repeat receptor kinase XIAO
40411	Phosphatidylinositol 3- and 4-kinase; Ubiquitin family	Phosphatidylinositol 4-kinase gamma 4-like	Phosphatidylinositol 4-kinase gamma
44401	FAR1 DNA-binding domain; MULE transposase domain; SWIM zinc finger	Protein FAR1-RELATED SEQUENCE 5-like	Protein FAR1-RELATED SEQUENCE 6
44440	S-locus glycoprotein domain; Protein kinase domain	G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	G-type lectin S-receptor-like serine/threonine-protein kinase LECRK1
582	NB-ARC domain	Putative disease resistance protein RGA4	Putative disease resistance protein RGA4
9829	alpha/beta hydrolase fold	Probable carboxylesterase 15	Probable carboxylesterase 15

(D). Annotated gene clusters present in annual (*B. distachyon*, *B. hybridum* and *B. stacei*) and absent in perennial (*B. arbuscula, B. boissieri, B. mexicanum, B. phoenicoides, B. pinnatum, B. retusum, B. rupestre* and *B. sylvaticum*) species of *Brachypodium*. The total number of transcript clusters entering the annotation pipeline was 49.

cluster ID	Pfam	RefSeq	SwissProt
266830	Tubulin/FtsZ family, GTPase domain	Cell division protein FtsZ homolog 1, chloroplastic	Cell division protein FtsZ
269661	JAB1/Mov34/MPN/PAD-1 ubiquitin protease	26S proteasome non-ATPase regulatory subunit 7 homolog A	26S proteasome non-ATPase regulatory subunit 7 homolog A
270801	PAP_fibrillin	Probable plastid-lipid-associated protein 13, chloroplastic	Probable plastid-lipid-associated protein 13, chloroplastic
279010	NA	Ripening-related protein 3-like	Ripening-related protein 3
265956	NA	Zinc finger CCCH domain-containing protein 30	Zinc finger CCCH domain-containing protein 30
279496	Carbohydrate esterase, sialic acid-specific acetylesterase	Probable carbohydrate esterase At4g34215	Probable carbohydrate esterase At4g34215
283472	Dirigent-like protein	Dirigent protein 21-like	Dirigent protein 21
287158	NA	Non-specific lipid-transfer protein 2-like	Non-specific lipid-transfer protein 2
291032	Ubiquitin carboxyl-terminal hydrolase, family 1	Ubiquitin carboxyl-terminal hydrolase isozyme L3-like	Ubiquitin carboxyl-terminal hydrolase 3
292401	RNA polymerase beta subunit	DNA-directed RNA polymerases IV and V subunit 2-like	DNA-directed RNA polymerase D subunit 2b
300951	Lytic transglycolase; Pollen allergen	Expansin-A1-like	Expansin-A1
305847	Terpene synthase, N-terminal domain	LOW QUALITY PROTEIN: alpha-humulene synthase-like	Alpha-humulene synthase
301110	Legume lectin domain; Protein kinase domain	L-type lectin-domain containing receptor kinase IV.1-like	L-type lectin-domain containing receptor kinase
302923	Hsp70 protein	Chaperone protein DnaK-like	NA
317052	U-box domain	U-box domain-containing protein 34-like	NA
317065	Pollen proteins Ole e I like	Pistil-specific extensin-like protein	NA
319872	SAM dependent carboxyl methyltransferase	Anthranilate O-methyltransferase 2-like	Anthranilate O-methyltransferase 2
336721	NA	BTB/POZ domain-containing protein At2g46260-like	NA
332400	XS zinc finger domain	Factor of DNA methylation 1-like	Factor of DNA methylation 1

(E). Annotated gene clusters present in polyploid (*B. boissieri, B. hybridum*, *B. mexicanum, B. phoenicoides, B. retusum* and *B. rupestre* 4x) and absent in diploid (*B. arbuscula, B. distachyon*, *B. pinnatum* 2x, *B. stacei* and *B. sylvaticum*) species of *Brachypodium*. The total number of transcript clusters entering the annotation pipeline was 14.

cluster ID	Pfam	RefSeq	SwissProt
54789	NA	Amino acid permease 3-like	Amino acid permease
35021	Leucine Rich Repeat	Receptor kinase-like protein Xa21	Receptor kinase-like protein Xa21
2161	No apical meristem (NAM) protein; Amino acid kinase family; Calmodulin binding protein-like	Aspartokinase 1, chloroplastic-like	Aspartokinase 1, chloroplastic
54848	NA	tRNA modification GTPase MnmE	NA
69099	Phosphofructokinase	ATP-dependent 6-phosphofructokinase 6-like	ATP-dependent 6-phosphofructokinase 6

(F). Annotated gene clusters present in *B. mexicanum*, *B. boissieri* and *B. retusum* and absent in remaining studied species and cytotypes of *Brachypodium.* The total number of transcript clusters entering the annotation pipeline was 143.

Cluster ID	Pfam	RefSeq	SwissProt
196786	Tryptophan synthase alpha chain	Tryptophan synthase alpha chain	Indole-3-glycerol phosphate lyase, chloroplastic
194344	Protein SCAI	Putative cell division cycle ATPase	LEC14B homolog; Cytochrome c-type biogenesis CcmH-like mitochondrial protein
197044	-	Flap endonuclease GEN-like 2	Flap endonuclease GEN-like 2
205139	Seven in absentia protein family	E3 ubiquitin-protein ligase SINAT5-like	E3 ubiquitin-protein ligase SINAT3
207344	-	E3 ubiquitin-protein ligase MIEL1-like	E3 ubiquitin-protein ligase MIEL1
208843	-	Protein ECERIFERUM 1-like	Protein ECERIFERUM 1
215074	-	Eukaryotic translation initiation factor 3 subunit B-like	Eukaryotic translation initiation factor 3 subunit B
223465	-	ABC transporter E family member 2-like	ABC transporter E family member 1
224357	alpha/beta hydrolase fold	Probable carboxylesterase 8	Probable carboxylesterase 8

229751	-	Ubiquitin carboxyl-terminal hydrolase 2-like	-
231111	BURP domain	Uncharacterized protein LOC100844634	BURP domain-containing protein 11
233160	Peroxidase	Cationic peroxidase SPC4-like	Cationic peroxidase SPC4; Peroxidase 12
233483	-	DNA (cytosine-5)-methyltransferase 1-like	DNA (cytosine-5)-methyltransferase CMT3
254636	D-mannose binding lectin	G-type lectin S-receptor-like serine/threonine-protein kinase At5g35370	G-type lectin S-receptor-like serine/threonine-protein kinase At5g35370
257319	-	sarcoplasmic reticulum histidine-rich calcium-binding protein-like	-

(G). Annotated gene clusters present all studied species and cytotypes of *Brachypodium* except *B. boissieri, B. mexicanum* and *B. retusum.* The total number of transcript clusters entering the annotation pipeline was 8.

Cluster ID	**Pfam**	**RefSeq**	**SwissProt**
16828	HAD superfamily, subfamily IIIB (Acid phosphatase)	Stem 28 kDa glycoprotein	Stem 28 kDa glycoprotein
24872	Glycolipid transfer protein (GLTP)	Accelerated cell death 11	Accelerated cell death 11
15104	Peptide methionine sulfoxide reductase	Peptide methionine sulfoxide reductase A5	Peptide methionine sulfoxide reductase A5
16007	JAB1/Mov34/MPN/PAD-1 ubiquitin protease; Maintenance of mitochondrial structure and function	Eukaryotic translation initiation factor 3 subunit F	Eukaryotic translation initiation factor 3 subunit F
5718	Apoptosis inhibitory protein 5 (API5)	Apoptosis inhibitor 5-A-like	Apoptosis inhibitor 5
28356	Universal stress protein family	Universal stress protein YxiE	Universal stress protein A-like protein

Supporting Figures

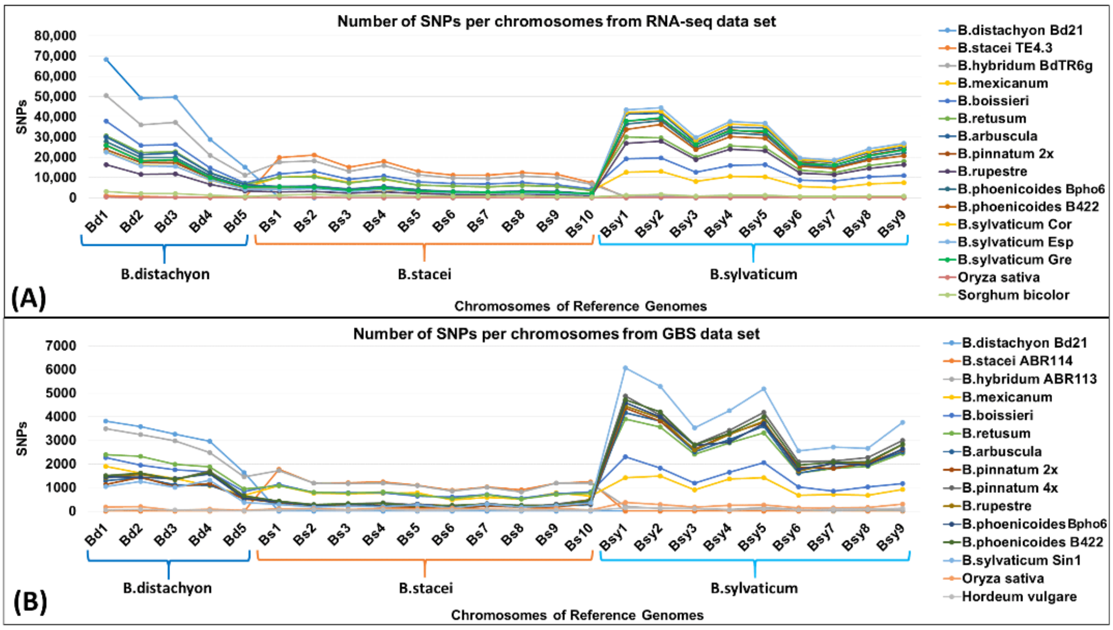


Figure S1. Plots showing number of RNA-seq and GBS SNPs extracted from the *Brachypodium* species, cytotypes and ecotypes under study mapped onto the chromosomes of the three concatenated reference genomes [*B. distachyon* (chromosomes Bd1-Bd5) + *B. stacei* (chromosomes Bs1-Bs10) + *B. sylvaticum* (chromosomes Bsy1-Bsy9)] with the *vcf2alignment* tool. **(A)** RNA-seq data. **(B)** GBS data.

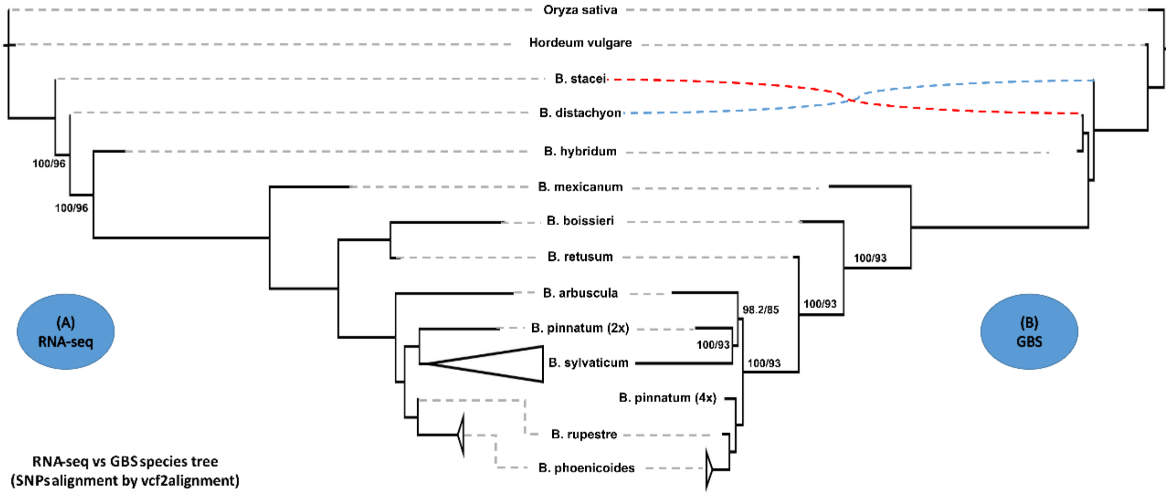


Figure S2. The *Brachypodium* maximum likelihood (ML) species tree constructed with IQTREE based on RNA-seq **(A)** and of GBS **(B)** SNP data extracted from the samples under study using the *vcf2alignment* tool. *Oryza sativa* was used to root the tree. SH-aLRT/UltraFast Bootstrap support values <99 are shown on branches. Incongruences between the RNA-seq and the GBS trees in the topological positions of *B. stacei* and *B. distachyon* are indicated with color lines.

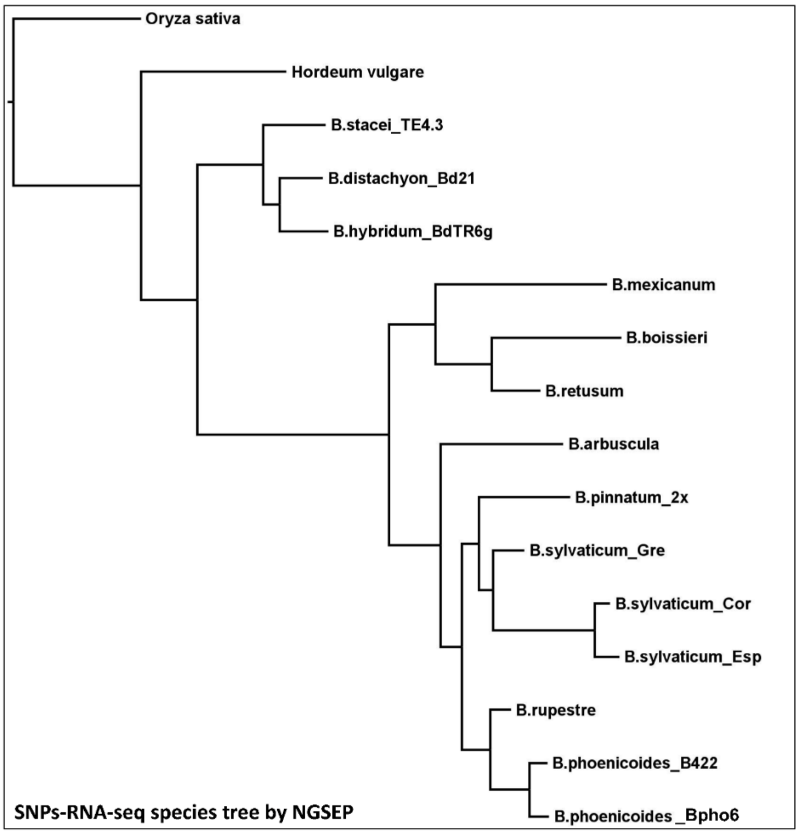


Figure S3. Validation of *vcft2alignment* ML species tree based on RNA-seq SNP data through phylogenomic reconstruction using the NGSEP software. SH-aLRT/UltraFast Bootstrap support values <99 are shown on branches. *Oryza sativa* was used to root the tree. The topology of the NGSEP tree was highly congruent with that obtained from the *vcf2alignment* approach (fig. S2a).

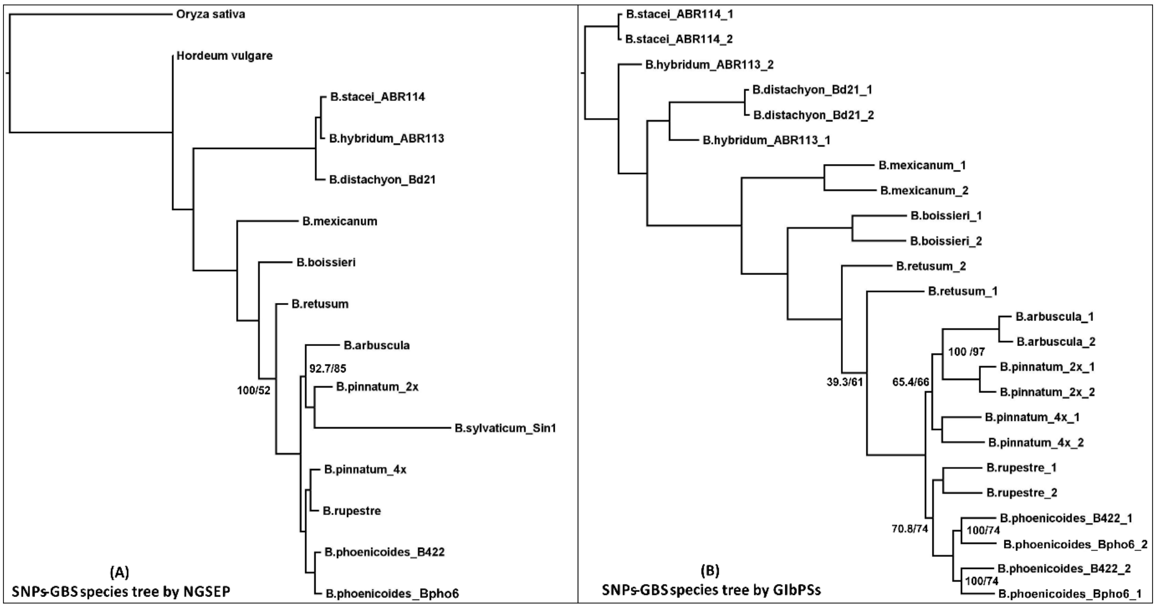


Figure S4. Validation of vcf2alignment ML species tree based on GBS data through phylogenomic reconstruction using the NGSEP **(A)** and GIbPSs **(B)** software. *Oryza sativa* and *B. stacei* ABR114 were used to root the NGSEP and GIbPSs trees, respectively. SH-aLRT/UltraFast Bootstrap support values <99 are shown on branches. The topology of the NGSEP tree was highly congruent with that obtained from the *vcf2alignment* approach (fig. S2b). The GIbPSs tree also recovered an overall congruent topology though it did not totally resolve the two parental subgenomes of the control allopolyploid species *Brachypodium hybridum*.

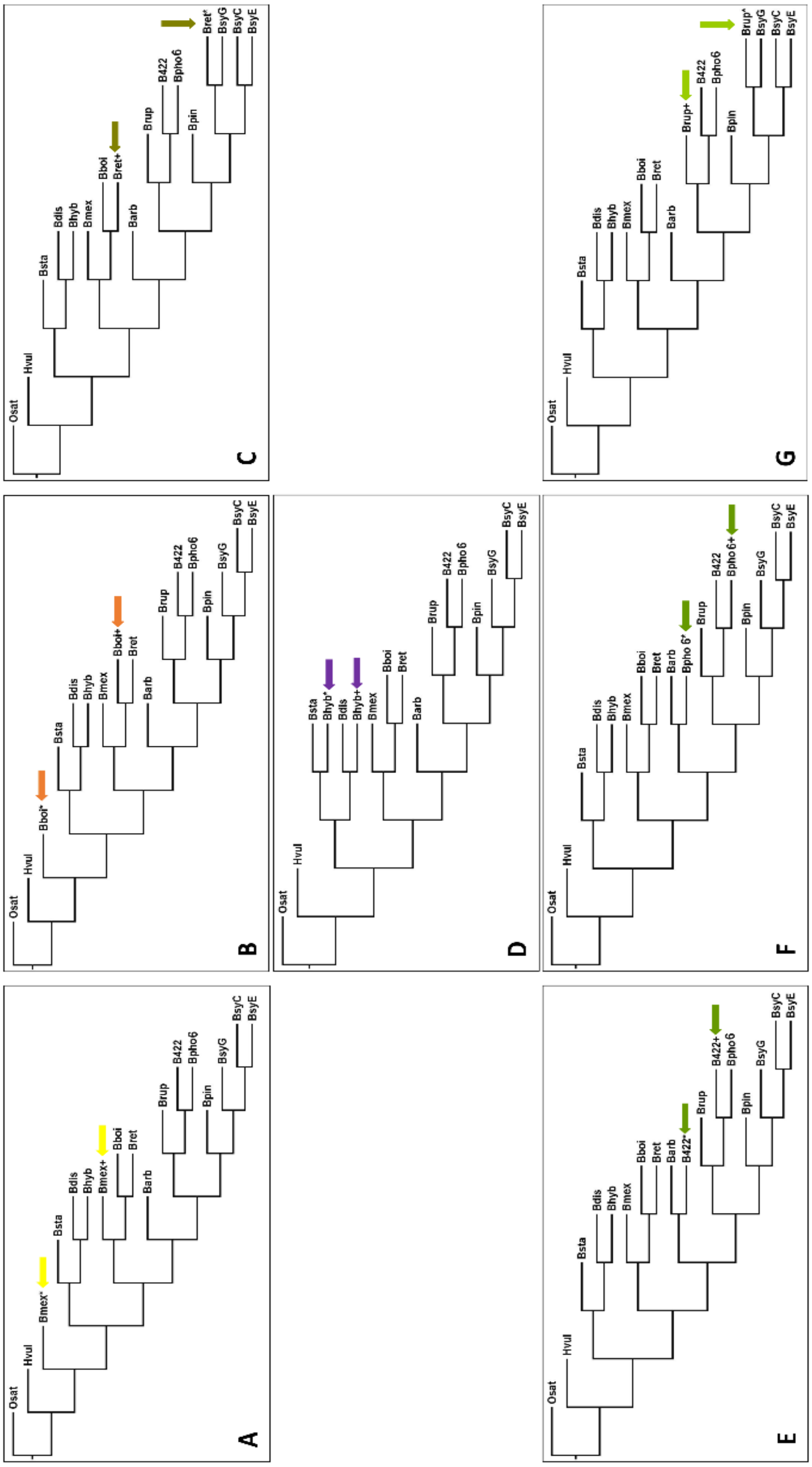


Figure S5. Partial validation of the *Brachypodium* subgenome tree based on core cluster transcripts using the minimum parsimony score option implemented in the GRAMPA software checking for potential polyploidization scenarios (a maximum number of two events could be detected by the program) of allopolyploid *Brachypodium* species. A total of 3,173 core gene clusters were used in the GRAMPA analysis. Best parsimony Multi-labeled trees obtained for *B. mexicanum* (Bmex) **(A)**, (*B. boissieri* (Bboi) **(B)**, *B. retusum* (Bret) **(C)**, *B. hybridum* (Bhyb) **(D)**, *B. phoenicoides* B422 **(E)**, *B. phoenicoides* Bpho6 **(F)**, *B. rupestre* **(G)**. *Oryza sativa* was used to root the trees. Arrows indicate the two putative homeologous (subgenomic) lineages inferred to have contributed to each allopolyploid accession.

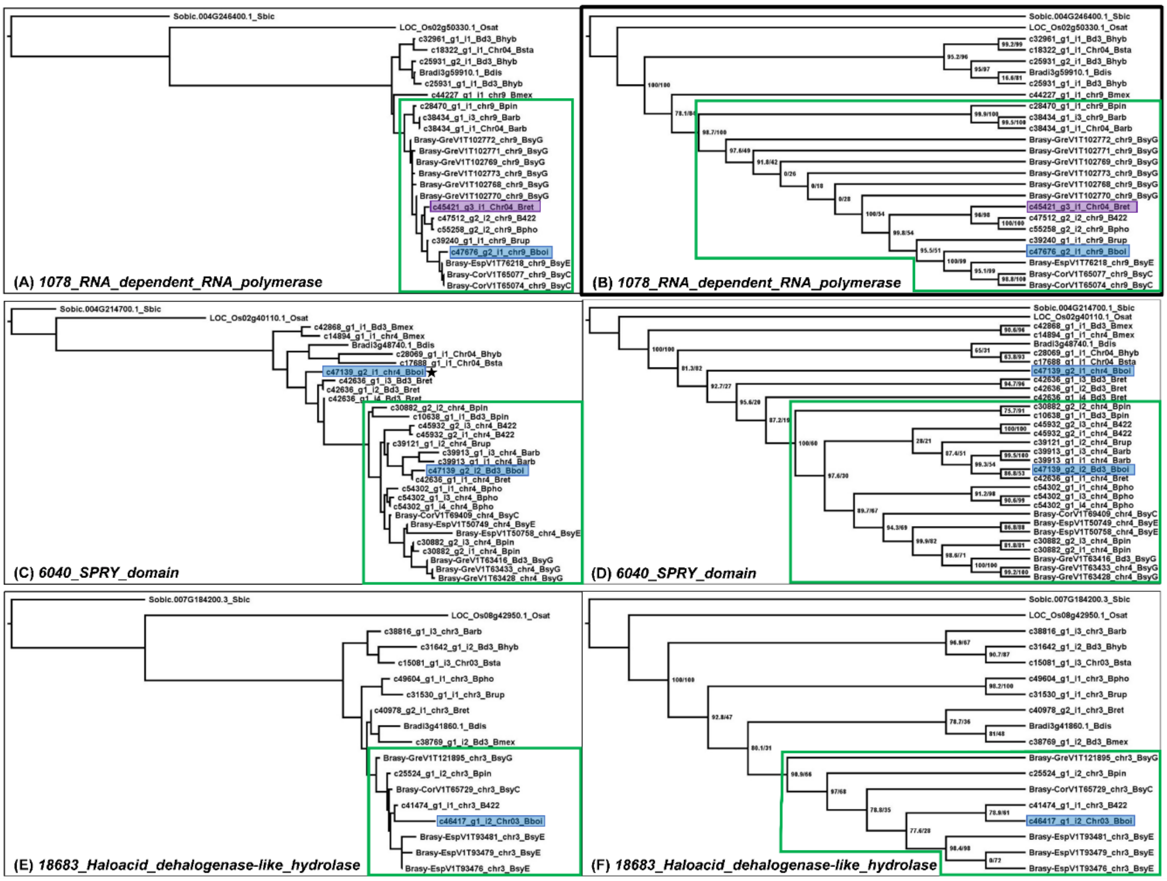

(A) 1078_RNA_dependent_RNA_polymerase
(B) 1078_RNA_dependent_RNA_polymerase
(C) 6040_SPRY_domain
(D) 6040_SPRY_domain
(E) 18683_Haloacid_dehalogenase-like_hydrolase
(F) 18683_Haloacid_dehalogenase-like_hydrolase

Figure S6. Selection of three *Brachypodium* maximum likelihood IQTREE gene trees showing *B. boissieri* allelic copies nested within a strongly supported core perennial *Brachypodium* clade: 1078_RNA_dependent_RNA_polymerase **(A, B)**, 6040_SPRY_domain **(C, D)**, 18683_Haloacid_dehalogenase-like_hydrolase **(E, F)**. *Sorghum bicolor* and *Oryza sativa* were used as outgroups.

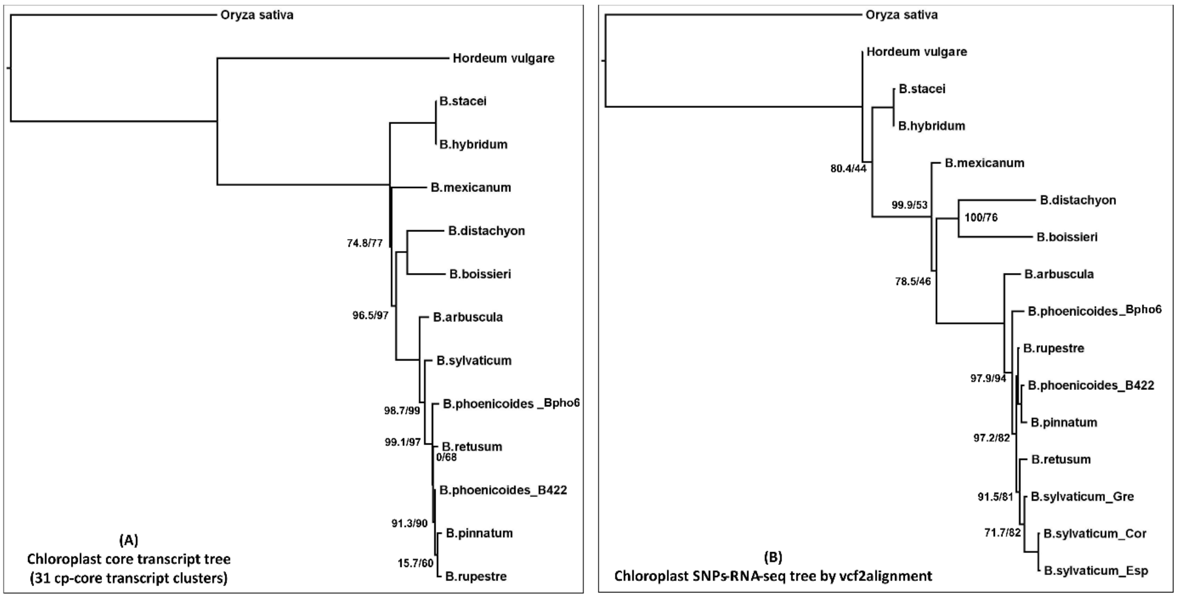


Figure S7. *Brachypodium* maximum likelihood IQTREE plastome trees based on analysis of RNA-seq data. **(A)** Species tree constructed from 31 plastome core transcripts. **(B)** Species tree constructed from SNPs extracted from reads mapped onto the *B. stacei* (ABR114; NC_036837) plastome. *Oryza sativa* was used to root the trees. SH-aLRT/UltraFast Bootstrap supports <99 score are showed on branches.

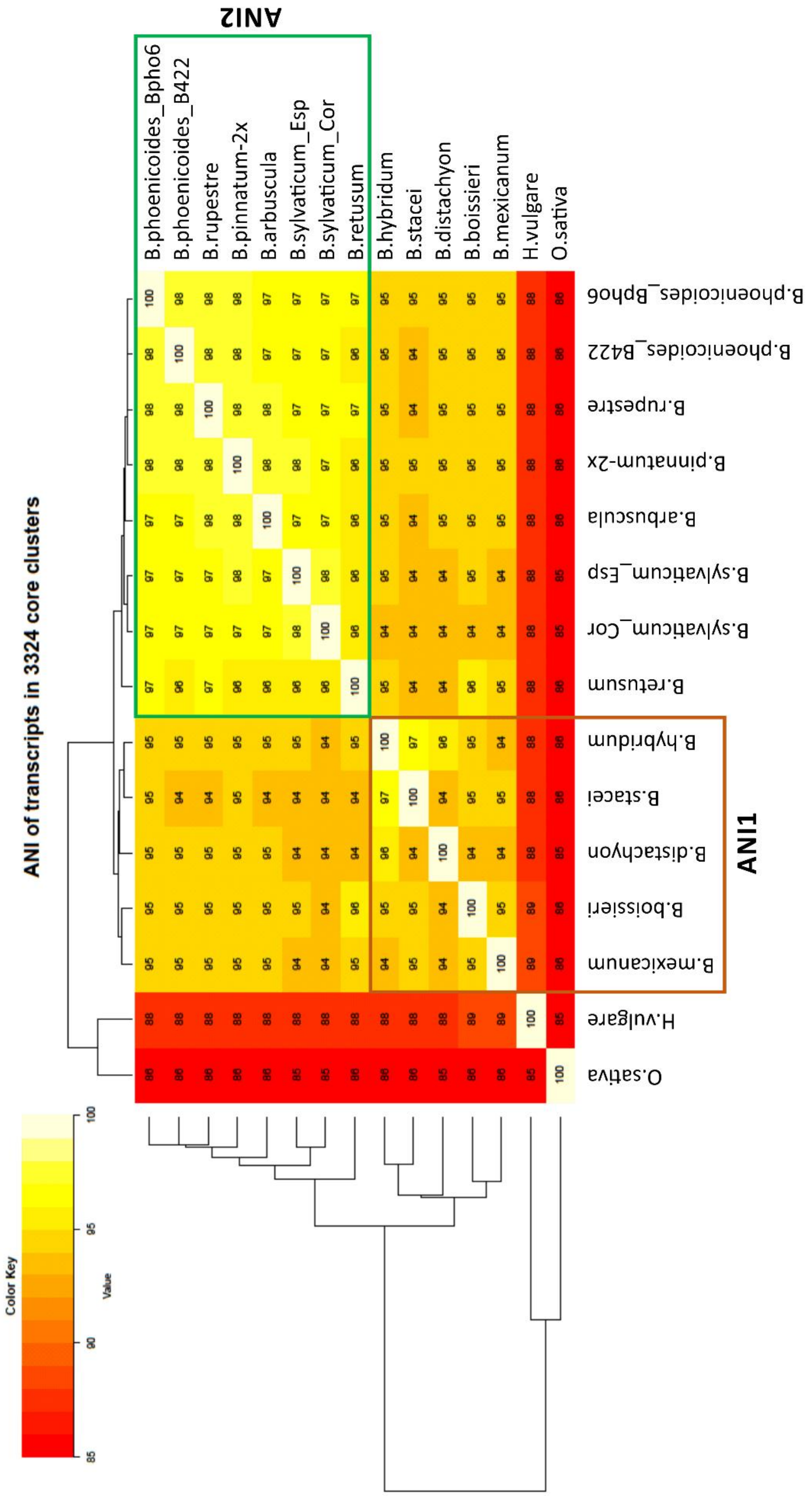


Figure S8. Heat-map and hierarchical clustering of 3,324 *Brachypodium*, *Oryza* and *Hordeum* core transcripts clusters using Average nucleotide identity (AIN) matrix computed with the GET_HOMOLOGUES_EST pipeline. Two main *Brachypodium* groups were detected *(B. distachyon + B. stacei + B. hybridum + B. mexicanum + B. boissieri*: brown square; *B. retusum + B. sylvaticum + B. arbuscula + B. pinnatum + B. rupestre + B. phoenicoides*: green square).

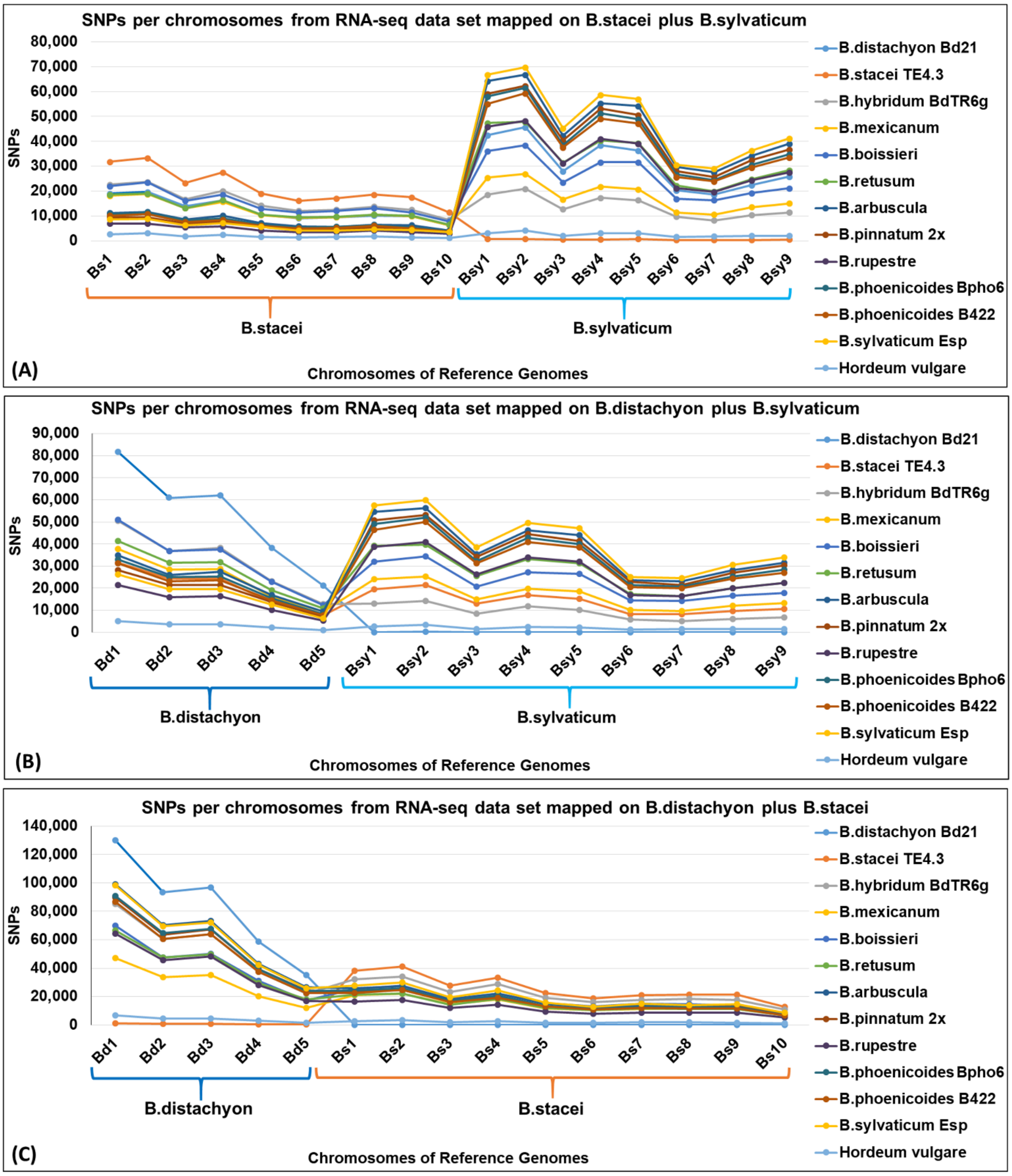


Figure S9. Graphics of SNPs from RNA-seq sequences mapped on different combination of parse-concatenated reference genomes extracted and filtered by *vcf2alignment* tool. **(A)** *B. stacei* plus *B. sylvaticum*, **(B)** *B. distachyon* plus *B. sylvaticum* and **(C)** *B. distachyon* plus *B. stacei*.

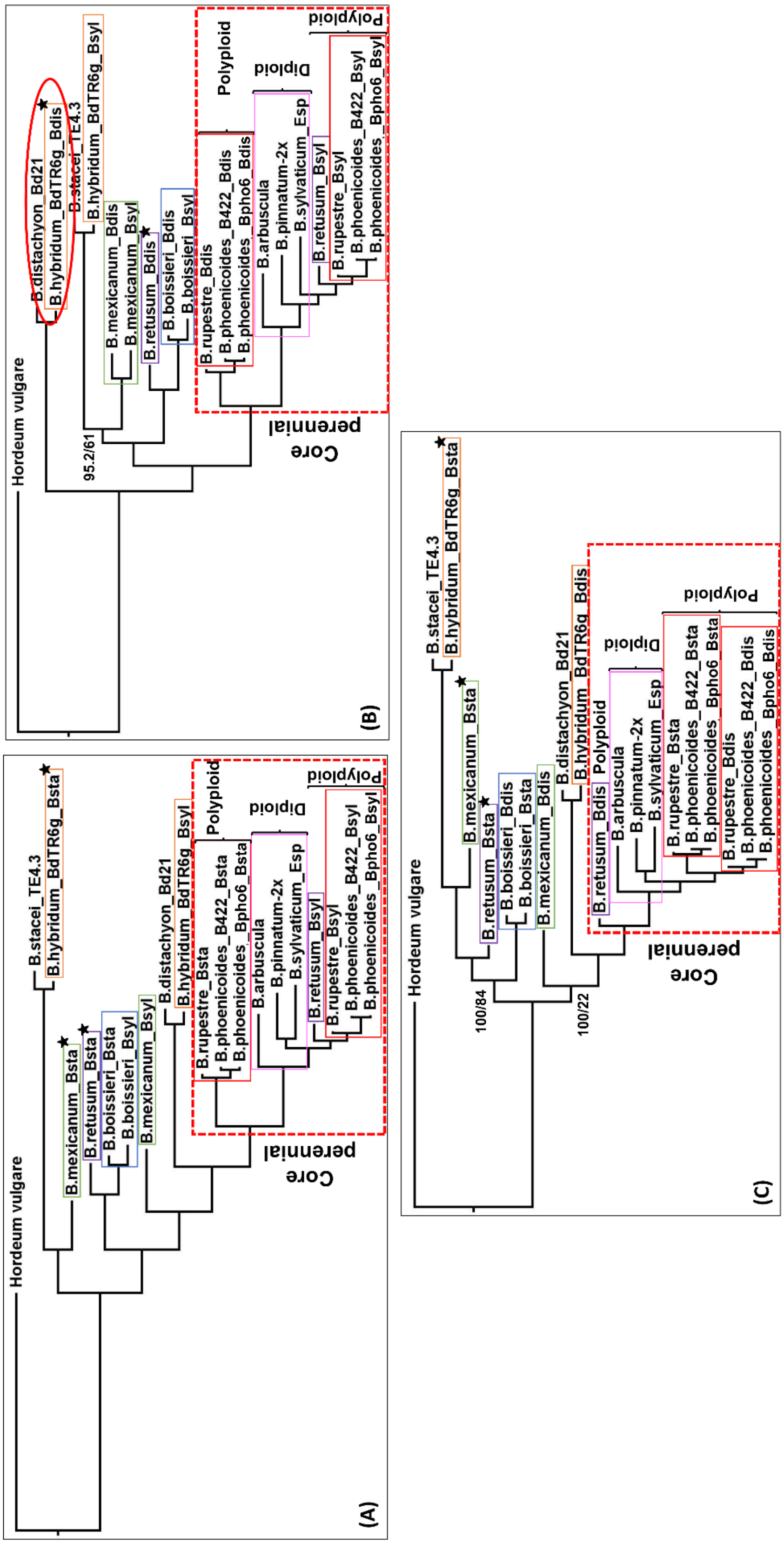


Figure S10. Subgenomes tree of SNPs from RNA-seq mapped on different combinations of parse reference genomes (*B. distachyon*-Bdis; *B. stacei-Bsta*; *B. sylvaticum*-Bsyl) extracted, filtered and aligned by *vcf2alignment_synteny*. **(A)** *B. stacei* plus *B. sylvaticum*, **(B)** *B. distachyon* plus *B. sylvaticum* and **(C)** *B. distachyon* plus *B. stacei*. Stars indicate the most putative ancestral copies of each species. Incongruent positions in the phylogram and low support in the cladogram are marked by a red circle. SH-aLRT/UltraFast Bootstrap supports are showed in the cladograms.

Appendix III: Supporting Information of Chapter 3

Methods S1: Detailed description of the plastome automated assembly pipeline

A pipeline for the automated assembly and annotation of plastomes was developed (Fig. S1). This workflow employs a large set of bioinformatics software packages (Table S1). First, DUK (http://duk.sourceforge.net) is used to extract putative plastid reads from WGS reads. The Next steps involve quality control and filtering of raw sequencing reads using FastQC v.0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and Trimmomatic v.0.32 (Bolger *et al*., 2014) respectively. Substitution errors can be optionally corrected with Musket v.1.0.6 (Liu et al., 2013). These trimming and filtering steps result in paired and single reads which can be managed using split_pairs v.0.5 (https://github.com/eead-csic-compbio/split_pairs). Further quality control can be performed by assessing orientation and insert size of paired reads, after mapping them to a reference genome with BWA v.0.7.8 (Li & Durbin, 2009). Contig assembly can be performed with Velvet v.1.2.07 *de novo* assembler (Zerbino & Birney, 2008) or with Columbus module of Velvet (Zerbino, 2010) for reference-guided assembly, attempting to resolve inverted repeats (IRs). Scaffolds are constructed using SSPACE Basic v.2.0 (Boetzer et al., 2011), and Gapfiller v.1.11 (Boetzer and Pirovano 2012; Nadalin *et al*., 2012) is used to gap-fill them using all available paired-end and mate-pair (reverse complement) reads. Potential overlaps among scaffold ends are confirmed with custom Perl scripts and BLAST v.2.2.28+ (Camacho, 2013).

Assembly and annotation of Brachypodium plastomes

Several Pooideae plastomes were used to infer background plastid k-mer distributions with DUK (Table S2). Plastomes of *B. stacei* (ABR114) and *B. hybridum* (ABR113) were *de novo* assembled with k-mer length 47 and 0.5M paired-end reads (insert size=250bp). Paired-end libraries of insert size=500bp were then used for scaffolding and gap filling. As the resulting scaffolds contained only one inverted repeat segment (IRa) with approximately double depth of coverage, IRb was crafted by duplicating IRa. Final plastomes were obtained by merging the scaffolds, which had overlapping ends of length 40-46. These curated plastomes were first aligned to the Bd21 plastome, to validate the general structure of the assembly, and then verified experimentally: the main scaffold junctions, i. e., 1,161 bp insertion and *rps*19 deletion, were validated by

PCR and/or Sanger sequencing (see below, Figure S2d and Table S5). Note that the *rps*19 deletion was detected as it sits right in the LSC – IRb junction and was thus correctly assembled with short reads. Finally, the original read libraries were mapped back to the assembled plastomes in order to be visually inspected using IGV v.2.3.8 software (Thorvaldsdóttir *et al*, 2013).

The remaining plastomes were assembled both *de novo* and reference-guided (with Columbus module) and the best strategy was chosen for each accession. For *B. distachyon* accessions, the plastome of ecotype Bd21 (NC_011032.1) was used as a reference to pre-map reads. However, for *B.hybridum* accessions Pob1 and BdTR6G the chosen reference was the *B. stacei* ABR114 plastome. For each accession, optimal assembly parameters, i.e. number of input reads (from 0.5M to 2M reads in steps of 0.5M) and k-mer length (from 47 to 87), were estimated with VelvetOptimiser v.2.2.5 (http://www.vicbioinformatics.com/software.velvetoptimiser.shtml). After scaffolding and gap filling, some scaffolds were further merged by checking overlapping ends. Finally, any remaining errors were detected and corrected with help from SEQuel v.1.0.2 (Ronen et al., 2012), and by visual inspection of the original sequence reads mapped onto the assembled scaffolds using IGV. As to the inverted repeats, they were assembled separatedly when guided by the reference plastome. In those cases where *de novo* assemblies were superior, IRb was manually duplicated as explained earlier.

Protein-coding genes and transfer RNAs in *B. distachyon* (ABR6 ecotype), *B. stacei* (ABR114 ecotype) and *B. hybridum* (ABR113 ecotype) plastomes were identified and annotated using cpGAVAS web version (Liu et al., 2012a) and BLAST v.2.2.28+ (Camacho, 2013) tools, with extensive manual curation. These annotations were then exported and adapted to the remaining genome assemblies with Perl script _annot_fasta_from_gbk.pl, documented at https://github.com/eead-csic-compbio/chloroplast_assembly_protocol.

All protein-coding and tRNA genes were further aligned and validated by comparison with *B. distachyon* Bd21 (NC_011032.1) reference plastome. A circular gene map of the plastid genome was generated using OrganellarGenomeDRAW web version (Lohse et al., 2013) and the similarities and differences among all assembled genomes were analyzed using script _check_matrix.pl (https://github.com/eead-csic-compbio/chloroplast_assembly_protocol) and illustrated with Circos software v.0.69 (Krzywinski et al., 2009) using *B. distachyon* ABR6 line as reference and window size = 100 (Fig. 2).

Validation of plastid assemblies by PCR and Sanger sequencing

Junctions between IR-LSC, LSC-IR, IR-SSC and SSC-IR regions of *B. stacei* ABR114 and *B. hybridum* ABR113 plastomes were amplified and sequenced. Besides, the deletion of one *rps*19 copy (180 bp) in the junction between LSC and IR in *B. stacei* and *hybridum* lines (see Results) was confirmed by amplifying, gel electrophoresis and Sanger sequencing in all *B. hybridum* and *B. stacei* lines and *B. distachyon* Bd21 (Fig. S2d). Primers (Table S6) were designed using Geneious v.8.1.4 software (Kearse et al., 2012).

Each 25 µL PCR reaction contained the following: 2.5 µL of KAPA Taq buffer A with MgCl$_2$, 2.5 µL, MgCl$_2$, 0.63 µL of dNTPs, 0.25 µL KAPA Taq DNA Polymerase, 0.5 µL of each primer (10 µM), 17.12 µL of Milli-Q water and 1 µL template DNA. PCR conditions were 3 min at 94°C, followed by 30 cycles of 94°C for 30 secs, 65ºC for 45 secs, and 72°C for 1 min and finally 72ºC for 7 min.

The obtained amplicon sequences are shown below, with amplicon numbers 1, 2, 3 and 4 corresponding to primers 1 & 2, 3 & 4, 5 & 6 and 7 & 8, respectively. Primer sequences are in Table S6:

>ABR113-1

tgtggaccacccccatgggggcggtgaagggaaagcccccattggtagaaaaaaacccacaaccccttggggttatcctgc
gcttggaagaagaactaggaaaaggaaaaaatatagcgatagttttattcttcgtcgccgtaagtaaatacgtaactagga
atatggaaaattgcatttttgaatttgcaataatgcgatgggcgaacgacgggaattgaacccgcgcatggtggattcaca
atccactgccttgatccacttggctacatccgccccttatccagctacaggatttttctctttttccattcatcattattctattta
ttctgacctccatacttcgatcgagatattggacatagattgccgctctttaaaaaggaaagaaatacccaatatcttgctag
aacaagatattgggtatttctcgctttcctttcttcaaaaattcttatatgttagcggaaaaaccttatccattaatagcgggaa
cttcaagagcagctagatctagagggaagttgtgagcattacgttcgtgcattacttccataccaagattagcacggttgatg
atatcagcccaagtattaataacgcgaccttgactatcaactacagattggttgaaattgaatccatttaggttgaacgccat
agtactaatacctaaagcagtgaaccagattcctactacaggccaagcagccaagaagaagtgtaaagaacgagagttgt
tgaaactagcatattggaagattaatc

>ABR113-2

gacctaccataggatttgttatgtaaataggtatatgttcctttccattatgaattgcgattgtatggccaaccattgttggtag
aatgctagatgcccgggaccacgttactattgtttctttctcctccttcatattgaccttttctatttttgccaataaatgatgagc
tacaaaaggattcgtttttttcgtgtcacagctgattactcctttttccttttaaagagcggcaatctatgtccaatatctcga
tcgaagtatggaggtcagaataaatagaataatgatgaatggaaaaaagagaaaaatcctgtagctggataaggggcgg
atgtagccaagtggatcaaggcagtggattgtgaatccaccatgcgcgggttcaattcccgtcgttcgcccatcgcattattg

caaattcaaaaaatgcaattttccatattcctagttacgtatttacttacggcgacgaagaataaaactatcgctatattttttc
cttttcctagttcttcttccaagcgcaggataaccccaaggggttgtgggtttttttctaccaatgggggctttccctt

>ABR113-3

tattcgggagcagtaatttaatcgttcgaattttttttcttattttatttagtagccttatagtagtcttagattttgcattttgatga
gcctcgttttgaggaattcatggaataatgaattaaggaagaaaggatatgagtctaccgcttacaagaaaagatctcatga
tagtcaatatgggccctcaacacccatcaatgcatggtgttcttcgactgatcgttactctcgatggtgaagatgttattgattg
tgaacctatattagggtatttacacagaggaatggaaaaaatcgcggaaaaccgaacgattatacaatacttaccttatgta
acacggtagaaaagagacctggaaattccttcagttaagaaagaaaaaagaataaaaaaacagatacataacatagaaa
aaagaataaataagacgagattcgccctcccctacatatttaatttcttctcctatacaaaaactagcaagacctactccatt
ggtaatcccatcaatgacacccttatcgaaaaactccgtgagttcggttaatcctcttatacccagggtaaagaccccactat
agaaaatatctatataaccacgattatatgaccaactgtatatctttttttttacttgatcccaaaagttctttttgggacttcctt
tgtaaaaggaattttgtaaatc

>ABR113-4

aataatcacagaaatctaaacatttctcgatccatccataaggtagatcggcggctactcctccgatgcgaaagtaattgtgc
atcattcgcatacctgtagcagcttcaaatagatcatatattaactctctctctaaaaatgtagaaaaaaggagtctgtgcg
ccgagatccgccataaaaggcccaagccataacaagtgagaagctatacggctcaactctaacataattaccctaatatag
ctggctctttggggtatttgaatattttccaagaattctggtgcatttaccgttattgcttctgtaaacatagtagctaaataatc
ccaccgtgttacataaggtaagtattgtataatcgttcggttttccgcgatttttttccattcctctgtgtaaataccctaatatag
gttcacaatcaataacatcttcaccatcgagagtaacgatcagtcgaagaacaccatgcattgatgggtgttgagggcccat
attgactatcatgagatcttttcttgtaagcggtagactcatatcctttcttccttaattcattattccatgaattcctcaaaacga
ggctcatcaaaatgcaaaatctaagactactataaggctactaaataaaataagaaaaaaaattcgaacgattaaattactg
ctcccgaatattcaactgactgattaatttcttataacgtactctatttttctttgccaaataagccagcaaacgtcgacgttttc
ccaaaagtcttcgtagacctctttccgatgaaaaatctttttttgtgtaattccaatg

>ABR114-1

gtggctaggtaaacgccccatagtaagaggggtagttatgaaccctgtggaccacccccccatgggggcggtgaagggaaa
gcccccattggtagaaaaaaacccacaaccccttggggttatcctgcgcttggaagaagaactaggaaaaggaaaaaata
tagcgatagttttattcttcgtcgccgtaagtaaatacgtaactaggaatatggaaaattgcatttttttgaatttgcaataatgc
gatgggcgaacgacgggaattgaacccgcgcatggtggattcacaatccactgccttgatccacttggctacatccgcccct
tatccagctacaggatttttctctttttttccattcatcattattctatttattctgacctccatacttcgatcgagatattggacata
gattgccgctctttaaaaaggaaagaaatacccaatatcttgctagaacaagatattgggtatttctcgctttcctttcttcaaa
aattcttatatgttagcggaaaaaccttatccattaatagcgggaacttcaagagcagctagatctagagggaagttgtgag
cattacgttcgtgcattacttccataccaagattagcacggttgatgatatcagcccaagtattaataacgcgaccttgactat
caactacagattggttgaaattgaatccatttaggttgaacgccatagtactaatacctaaagcagtgaaccagattcctact
acaggccaagcagccaagaagaagtgtaaagaacgagagttgttgaaactagcatattgg

>ABR114-2

ctaccataggatttgttatgtaaataggtatatgttcctttccattatgaattgcgattgtatggccaaccattgttggtagaatg
ctagatgcccgggaccacgttactattgtttctttctcctccttcatattgaccttttctatttttgccaataaatgatgagctaca
aaaggattcgttttttttcgtgtcacagctgattactcctttttttcctttttaaagagcggcaatctatgtccaatatctcgatcga
agtatggaggtcagaataaatagaataatgatgaatggaaaaaagagaaaaatcctgtagctggataaggggcggatgta
gccaagtggatcaaggcagtggattgtgaatccaccatgcgcgggttcaattcccgtcgttcgcccatcgcattattgcaaat
tcaaaaaatgcaattttccatattcctagttacgtatttacttacggcgacgaagaataaaactatcgctatatttttttcctttttc
ctagttcttcttccaagcgcaggataaccccaaggggttgtgggtttttttctaccaatgggggctttcccttcaccgcccccat

>ABR114-3

tttagtagccttatagtagtcttagattttgcattttgatgagcctcgttttgaggaattcatggaataatgaattaaggaagaa
aggatatgagtctaccgcttacaagaaaagatctcatgatagtcaatatgggccctcaacacccatcaatgcatggtgttctt
cgactgatcgttactctcgatggtgaagatgttattgattgtgaacctatattagggtatttacacagaggaatggaaaaaat
cgcggaaaaccgaacgattatacaatacttaccttatgtaacacggtagaaaagagacctggaaattccttcagttaagaa
agaaaaaagaataaaaaaacagatacataacatagaaaaagaataaataagacgagattcgccctcccccctacatattt
aatttcttctcctatacaaaaactagcaagacctactccattggtaatcccatcaatgacacccttatcgaaaaactccgtga
gttcggttaatcctcttatacccagggtaaagaccccactatagaaaatatctatataaccacgattatatgaccaactgtata
tcttttttttacttgatcccaaaagttctttttgggacttcctttgtaaaaggaattttg

>ABR114-4

gttgatattcacaactccccgtaaaaaataatcacagaaatctaaacatttctcgatccatccataaggtagatcggcggcta
ctcctccgatgcgaaagtaattgtgcatcattcgcatacctgtagcagcttcaaatagatcatatattaactctctctctctaaa
aatgtagaaaaaaggagtctgtgcgccgagatccgccataaaaggcccaagccataacaagtgagaagctatacggctca
actctaacataattaccctaatatagctggctctttggggtatttgaatattttccaagaattctggtgcatttaccgttattgctt
ctgtaaacatagtagctaaataatcccaccgtgttacataaggtaagtattgtataatcgttcggttttccgcgatttttccatt
cctctgtgtaaataccctaatataggttcacaatcaataacatcttcaccatcgagagtaacgatcagtcgaagaacaccatg
cattgatgggtgttgagggcccatattgactatcatgagatcttttcttgtaagcggtagactcatatcctttcttccttaattca
ttattccatgaattcctcaaaacgaggctcatcaaaatgcaaaatctaagactactataaggctactaaataaaataagaaa
aaaattcgaacgattaaattactgctcccgaatattcaactgactgattaatttcttataacgtactctatttttctttgccaaat
aagccagcaaacgtcgacgttttcccaaaagtcttcgtagacctctttccgatgaaaaatctttttgtgtaattccaatgtga
a

Validation of rps19 deletion by Sanger sequencing (see Fig. S2d)

Primers 1 & 2 in Table S6 were also used to confirm the deletion of one *rps*19 copy in IRb, obtaining the following amplicon sequences:

>Bd21 (B. distachyon)

cccccatgggggcggtgaagggaaagcccccattggtagaaaaaaacccacaaccccttggggttatcctgcgcttggaa
gaagaactaggaaaaggaaaaaatatagcgatagttttattcttcgtcgccgtaagtaaatacgtaactaggaatatggaa
aattgcatttttttgcatttgcaataatgcgatgggcgaacgacgggaattgaacccgcgcatggtggattcacaatccactgc
cttgatccacttggctacatccgccccttatccagctacaggatttttctctttttttccattcatcattattctatttattctgacctc
catacttcgatcgagatattggacatagattgccgctctttaaaaaggaaaaaaaaggagtaatcagctgtgacacgaaaa
aaaacgaatccttttgtagctcatcatttattggcaaaaatagaaaaggtcaatatgaaggaggagaaagaaacaatagta
acgtggtcccgggcatctagcattctacccacaatggttggccatacaatcgcgattcataatggaaaggaacatatacta
tttacataacaaatcctatggtaggtcgcaaattgggggaattcgtacctactcggcatttcacgagttatgaaaatgcaaga
aaggatactaaatctcgtcgttaactgaattcagaatagaaagattcagaataaaaaaaagatgcaaagtaaagaaatacc
caatatcttggtagaacaagatattgggtatttctcgctttcttttcttcaaaaattcttatatgttagcggaaaaaccttatcca
ttaatagcgggaacttcaagagcagctagatctagagggaagttgtgagcattacgttcgtgcattacttccataccaagatt
agcacggttgatgatatcagcccaagtattaataacgcgaccttgactatcaactacagattggttgaaattgaatccattta
ggttgaacgccatagtactaatacctaaagcagtgaaccagattcctactacaggccaagcagcc

> ABR114 (B. stacei)

gtggctaggtaaacgccccatagtaagaggggtagttatgaaccctgtggaccacccccccatgggggcggtgaagggaaa
gcccccattggtagaaaaaaacccacaaccccttggggttatcctgcgcttggaagaagaactaggaaaaggaaaaaata
tagcgatagttttattcttcgtcgccgtaagtaaatacgtaactaggaatatggaaaattgcatttttttgaatttgcaataatgc
gatgggcgaacgacgggaattgaacccgcgcatggtggattcacaatccactgccttgatccacttggctacatccgcccct
tatccagctacaggatttttctctttttttccattcatcattattctatttattctgacctccatacttcgatcgagatattggacata

gattgccgctctttaaaaaggaaagaaatacccaatatcttgctagaacaagatattgggtatttctcgctttcctttcttcaaa
aattcttatatgttagcggaaaaaccttatccattaatagcgggaacttcaagagcagctagatctagagggaagttgtgag
cattacgttcgtgcattacttccataccaagattagcacggttgatgatatcagcccaagtattaataacgcgaccttgactat
caactacagattggttgaaattgaatccatttaggttgaacgccatagtactaatacctaaagcagtgaaccagattcctact
acaggccaagcagccaagaagaagtgtaaagaacgagagttgttgaaactagcatattgg

>ABR113 (B. hybridum)

tgtggaccacccccatgggggcggtgaagggaaagcccccattggtagaaaaaaacccacaaccccttggggttatcctgc
gcttggaagaagaactaggaaaaggaaaaaatatagcgatagttttattcttcgtcgccgtaagtaaatacgtaactagga
atatggaaaattgcatttttttgaatttgcaataatgcgatgggcgaacgacgggaattgaacccgcgcatggtggattcaca
atccactgccttgatccacttggctacatccgccccttatccagctacaggatttttctctttttccattcatcattattctattta
ttctgacctccatacttcgatcgagatattggacatagattgccgctctttaaaaaggaaagaaatacccaatatcttgctag
aacaagatattgggtatttctcgctttcctttcttcaaaaattcttatatgttagcggaaaaaccttatccattaatagcgggaa
cttcaagagcagctagatctagagggaagttgtgagcattacgttcgtgcattacttccataccaagattagcacggttgatg
atatcagcccaagtattaataacgcgaccttgactatcaactacagattggttgaaattgaatccatttaggttgaacgccat
agtactaatacctaaagcagtgaaccagattcctactacaggccaagcagccaagaagaagtgtaaagaacgagagttgt
tgaaactagcatattggaagattaatc

> BdTR6G (B. hybridum)

gggtagttatgaaccctgtggaccacccccatgggggcggtgaagggaaagcccccattggtagaaaaaaacccacaacc
ccttggggttatcctgcgcttggaagaagaactaggaaaaggaaaaaatatagcgatagttttattcttcgtcgccgtaagta
aatacgtaactaggaatatggaaaattgcatttttttgaatttgcaataatgcgatgggcgaacgacgggaattgaacccgcg
catggtggattcacaatccactgccttgatccacttggctacatccgccccttatccagctacaggatttttctctttttccattc
atcattattctatttattctgacctccatacttcgatcgagatattggacatagattgccgctctttaaaaaggaaagaaatacc
caatatcttgctagaacaagatattgggtatttctcgctttcctttcttcaaaaattcttatatgttagcggaaaaaccttatcca
ttaatagcgggaacttcaagagcagctagatctagagggaagttgtgagcattacgttcgtgcattacttccataccaagatt
agcacggttgatgatatcagcccaagtattaataacgcgaccttgactatcaactacagattggttgaaattgaatccattta
ggttgaacgccatagtactaatacctaaagcagtgaaccagattcctactacaggccaagcagccaagaagaagtgtaaa
gaacgagagt

>Pob1 (B. hybridum)

ccctgtggaccacccccatgggggcggtgaagggaaagcccccattggtagaaaaaaacccacaaccccttggggttatc
ctgcgcttggaagaagaactaggaaaaggaaaaaatatagcgatagttttattcttcgtcgccgtaagtaaatacgtaacta
ggaatatggaaaattgcatttttttgaatttgcaataatgcgatgggcgaacgacgggaattgaacccgcgcatggtggattc
acaatccactgccttgatccacttggctacatccgccccttatccagctacaggatttttctctttttccattcatcattattctat
ttattctgacctccatacttcgatcgagatattggacatagattgccgctctttaaaaaggaaagaaatacccaatatcttgct
agaacaagatattgggtatttctcgctttcctttcttcaaaaattcttatatgttagcggaaaaaccttatccattaatagcggg
aacttcaagagcagctagatctagagggaagttgtgagcattacgttcgtgcattacttccataccaagattagcacggttga
tgatatcagcccaagtattaataacgcgaccttgactatcaactacagattggttgaaattgaatccatttaggttgaacgcc
atagtactaatacctaaagcagtgaaccagattcctactacaggccaagcagccaagaagaagtgtaaagaacgagagtt
gt

Supporting Tables

Table S1. List of *B. distachyon*, *B. stacei* and *B. hybridum* accessions studied.
Origin of samples: ABR1 - ABR7 (Brachyomics collections (C. Stace & P. Catalán), Aberystwyth, UK); BdTR_ accessions (Filiz et al., 2009); Bd1-1, Bd2-3, Bd3-1, Bd18-1, Bd21 (Vogel et al., 2006); Bd21-3 (Vogel & Hill, 2008); Adi_, Gaz_, Tek_, Bis_, Kah_, Koz_ (Vogel et al., 2009); Foz1, Mig3, Mon3, Mur1, Uni2 (Mur et al., 2011). * BdTR6G cited as ‘B. distachyon’ in GRIN-Global; Filiz et al., (2009) described it as a “polyploid line”. **Bd30-1, developed by D. Garvin from material collected by A. Manzaneda. IL (inbred line). Web sites:

https://gold.jgi.doe.gov/biosamples?Study.GOLD%20Study%20ID=Gs0033763

https://gold.jgi.doe.gov/projects?page=6&Biosample.Biosample+Name=Brachypodium+distachyon&count=100

https://www.google.com/fusiontables/DataSource?docid=1EQ1jPj9PJdBj4_or3zQGtIThS2YSUjli6Jcd1Q#rows:id=1

Accession	Collection location	GOLD Biosample ID (JGI)	SRX accession	SRA sample	Elevation (masl)	Latitude	Longitude
ABR2	Hérault, France	Gb0017122	SRX182920	SRS360668	371	43° 36' 15.343" N	3° 15' 46.580" E
ABR3	Aísa, Huesca, Spain	Gb0017178	SRX2021046	SRS1615737	1928	42° 10' 49.8" N	0° 4’ 23.2" W
ABR4	Arén, Huesca, Spain	Gb0017179	SRX2021047	SRS1615738	480	42° 15' 45.54" N	0° 43' 0.48" E
ABR5	Jaca, Huesca, Spain	Gb0017180	SRX182894	SRS360645	828	42° 34' 23.45" N	0° 33' 49.39" W
ABR6	Los Arcos, Navarra, Spain	Gb0017181	SRX298413	SRS438486	484	42° 34' 27.48" N	2° 11' 5.39" W
ABR7	Otero, Valladolid, Spain	Gb0017232	SRX2021065	SRS1615752	725	41° 35' 23.86" N	4° 45' 24.26" W
ABR8	Siena, Italy	Gb0017233	SRX874557 SRX874558	SRS844147	272	43° 18' 52.423" N	11° 19' 10.902" E
Adi10	Adiyaman, Turkey	Gb0009975	SRX185151	SRS361658	510	37° 46' 14.5" N	38° 21' 8.2" E
Adi12	Adiyaman, Turkey	Gb0009864	SRX2020035	SRS1615304	510	37° 46' 14.5" N	38° 21' 8.2" E
Adi2	Adiyaman, Turkey	Gb0017235	SRX2020496	SRS1615344	510	37° 46' 14.5" N	38° 21' 8.2" E
Arn1	Arén, Huesca, Spain	Gb0009976	SRX298355	SRS438433	681	42° 15' 23.44" N	0° 43' 47.46" E
Bd1-1	Soma, Manisa, Turkey	Gp0001144	SRX060134 SRX060135 SRX060136 SRX116611	SRS190935	141	39° 11' 27.44" N	27° 36' 28.59" E
Bd18-1	Kaman, Kırşehir Province, Turkey	Gb0009918	SRX2020043 SRX2020044	SRS1615310	1101	39° 22' 4.25" N	33° 43' 48.91" E

Bd2-3	Iraq	Gb0009943	SRX2020036	SRS1615305	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd21	near Salakudin, Iraq	Gb0012676	SRX2020505 SRX2020506	SRS1615350	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd21-3	near Salakudin, Iraq	Gp0039861	SRX119501 SRX146215	SRS291714 SRS312328	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd3-1	Iraq	Gp0001284	SRX117923	SRP001538	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd30-1**	Dilar, Granada, Spain	Gp0001821	SRX116649 SRX059915	SRS190910	1220	36° 59' 25.76" N	3° 33' 31.44" W
BdTR10C	Turkey	Gb0009946	SRX185149	SRS361656	1288	37° 46' 41.64" N	31° 53' 5.68" E
BdTR11A	Turkey	Gb0017236	SRX2020493	SRS1615342	986	38° 25' 0.42" N	28° 1' 52.75" E
BdTR11G	Kirklareli, Turkey	Gb0017237	SRX2020507	SRS1615351	124	41° 25' 17.86" N	27° 28' 36.81" E
BdTR11I	Turkey	Gb0009945	SRX2020031 SRX2020032	SRS1615301	363	39° 44' 17.39" N	28° 2' 24.71" E
BdTR12C	Turkey	Gp0009928	SRX059779 SRX059780	SRS190847	1035	39° 44' 53.45" N	34° 39' 1.15" E
BdTR13A	Ankara, Turkey	Gb0017238	SRX183318	SRS360828	787	39° 45' 23.35" N	32° 25' 56.46" E
BdTR13C	Ankara, Turkey	Gb0009863	SRS1615294	SRS1615294	1192	39° 24' 46.28" N	32° 59' 17.24" E
BdTR1I	Aydin, Turkey	Gb0017239	SRX183383	SRS360859	841	38° 5' 35.03" N	28° 34' 59.02" E
BdTR2B	Turkey	Gb0012677	SRX2020498 SRX2020499	SRS1615346	667	40° 4' 55.55" N	31° 19' 52.01" E
BdTR2G	Ankara, Turkey	Gb0009917	SRX185148	SRS361655	1596	40° 23' 37.13" N	32° 59' 7.32" E
BdTR3C	Turkey	Gb0009942	SRX2020033	SRS1615302	1957	36° 46' 58.92" N	32° 57' 46.71" E
BdTR5I	Turkey	Gb0009974	SRX2020021	SRS1615296	1596	40° 23' 37.13" N	32° 59' 7.32" E
BdTR7A	Yozgat, Turkey	Gb0017240	SRX183377	SRS360854	1035	39° 44' 53.45" N	34° 39' 1.15" E
BdTR8I	Turkey	Gb0017241	SRX181206 SRX181207 SRX181208 SRX181209	SRS359840	2385	37° 6' 31.87" N	34° 4' 17.06" E
BdTR9K	Eskişehir, Turkey	Gb0009919	SRX2020011	SRS1615290	932	39° 45' 10.62" N	30° 47' 19.07" E
Bis1	Bismil, Turkey	Gp0017242	SRX2020040 SRX2020041 SRX2020042	SRS1615309	529	37° 52' 35.6" N	41° 0' 54.3" E
Foz1	Foz de Lumbier, Navarra, Spain	Gp0009893 (Mig1) Project ID 404167	SRX2020038	SRS1615307	434	42º 38' 11.44" N	1º 18' 17.42" W
Gaz8	Gaziantep, Turkey	Gb0009947	SRX185147	SRS361654	891	37° 7' 39.8" N	37° 23' 26.9" E
Jer1	Ermita de San Jerónimo, Huesca, Spain	Gp0009916 (Mon1) Project ID 404166	SRX2020045	SRS1615311	418	42° 3' 16.56" N	0° 0' 44.57" W
Kah1	Kahta, Turkey	Gp0017374	SRX2020494 SRX2020495	SRS1615343	665	37° 44' 2.3" N	38° 32' 0.2" E

Kah5	Kahta, Turkey	Gb0017182	SRX2020497	SRS1615345	665	37° 44' 2.3" N	38° 32' 0.2" E
Koz1	Kozluk Turkey	Gb0012678	SRX183517	SRS360986	853	38° 9' 8.2.6" N	41° 36' 34.8" E
Koz3	Kozluk, Turkey	Gp0009991	SRX059781 SRX059782	SRS190848	853	38° 9' 8.2.6" N	41° 36' 34.8" E
Luc1 (G31i1)	Ermita de Santa Lucía, Berdún, Huesca, Spain	Gp0017244	SRX1869528	SRS1520207	597	42° 36' 36.18" N	0° 53' 35.48" W
Mig3	San Miguel de Foces, Ibieca, Huesca, Spain	Gb0017183	SRX182705	SRS360564	572	42° 8' 52.76" N	0° 11' 41.89" W
Mon3	Puerto de Pallaruelo, Castejón de Monegros, Zaragoza, Spain	Gb0017184	SRX182916	SRS360665	515	41° 39' 4.75" N	0° 12' 37.51" W
Mur1	Castillo de Mur, Lleida, Spain	Gb0009944	SRX181229	SRS359860	487	42° 06' 18" N	0° 51' 23" E
Per1 (G30i1)	Puerto del Perdón, Navarra, Spain	Gp0017243	SRX1869283	SRS1520047	742	42° 44' 13.34" N	1° 44' 58.6" W
S8iiC	Zaidín, Huesca, Spain	Gb0017185	SRX2021637	SRS1616272	144	41° 36' 19.3" N	0° 08' 38.4" E
Sig2	Sigüés, Zaragoza, Spain	Gp0009900 (Sig1) Project ID 404169	SRX2020046	SRS1615312	524	42° 36' 46.55" N	1° 0' 52.38" W
Tek2	Tekirdag, Turkey	Gb0012679	SRX183516	SRS360985	20	41° 0' 40.1" N	27° 31' 8.8" E
Tek4	Tekirdag, Turkey	Gb0017188	SRX2020048 SRX2020047	SRS1615313	20	41° 0' 40.1" N	27° 31' 8.8" E
Uni2	Escuela Politécnica Superior, Huesca, Spain	Gb0017189	SRX2021048	SRS1615739	480	42° 7' 3.98" N	0° 26' 42.81" W
RON4 (RON2)	Roncal, Navarra, Spain	Gp0039823	SRX711596	SRS710321	594	42° 46' 50" N	0° 57' 48'' W
Bd29-1	Krym, Ukraine	-	-	-	260	44° 30' 55" N	33° 33' 23" E
Pob1 (G32i2) *B. hybridum*	Poblado de San Antonio, Calaceite, Teruel, Spain	Gp0017245	SRX1869527	SRS1520206	573	41° 0' 16.99" N	0° 11' 6.72" E

BdTR6G* *B. hybridum*	Turkey	Gb0022615	SRX716913	SRS712683	872	39° 45' 15'' N	33° 31' 16'' E
ABR113 *B. hybridum*	Lisbon, Portugal	Gp0016929	SRX299256 SRX874525 SRX874526 SRX874527 SRX874528 SRX874529 SRX874530 SRX1971039 SRX1971040 SRX1971041 SRX1971042 SRX1971043 SRX1971044 SRX1971045 SRX1971046 SRX1971047 SRX1971048 SRX1971049	SRS439049 SRS844137	187	38° 46' 58.775" N	9° 15' 1.757" W
ABR114 *B. stacei*	Torrent, Formentera , Spain	Gp0016930	SRX299239 SRX874533 SRX874534 SRX874535 SRX874536 SRX874537 SRX874538 SRX874539 SRX874540 SRX1970692 SRX1970693 SRX1970694 SRX1970695 SRX1970696 SRX1970697 SRX1970698	SRS439047 SRS844132	122	39° 28' 35.350" N	2° 49' 55.448" E

Table S2. Grass plastomes employed in evolutionary and genomic analyses.
(1). Genomes used in ML (RAxML) and BI (MrBayes) phylogenomic analyses. (2). Genomes used in Bayesian nested dating analysis (BEAST). (3). Genomes used to infer background k-mer distributions (DUK).

Species	Accession	GI	RAxML[1]	BEAST[2]	DUK[3]
Acidosasa purpurea	NC_015820.1	340034177	x	x	x
Aegilops bicornis cultivar Clae57	NC_024831.1	685508511	x	x	
Aegilops cylindrica	NC_023096.1	568246973	x	x	
Aegilops geniculata	NC_023097.1	568244975	x	x	
Aegilops kotschyi cultivar TA1980	NC_024832.1	699008472	x	x	
Aegilops longissima cultivar TA1924	NC_024830.1	685508428	x		
Aegilops searsii cultivar TA1926	KJ614413.1	667754557	x	x	
Aegilops sharonensis cultivar TA1995	NC_024816.1	697964657	x	x	
Aegilops speltoides var. ligustica cultivar AE918	KJ614404.1	667753810	x	x	
Aegilops tauschii cultivar AL8/78	KJ614412.1	667754474	x	x	
Agrostis stolonifera	NC_008591.1	118430280	x	x	x
Ammophila breviligulata voucher CAN:Peterson 20867	NC_027465.1	884998160	x	x	
Ampelocalamus calcareus	NC_024731.1	675155489	x	x	
Ampelodesmos mauritanicus voucher B:Royl & Schiers s.n.	NC_027466.1	884998245	x	x	
Anomochloa marantoidea	NC_014062.1	295065706	x	x	x
Anthoxanthum odoratum voucher CAN:Saarela 500	NC_027467.1	884998329	x	x	
Arundinaria appalachiana	NC_023934.1	608787536	x	x	
Arundinaria gigantea	NC_020341.1	452849461	x	x	
Arundinaria tecta	NC_023935.1	608787620	x	x	
Avena sativa voucher CAN:Saarela 775	NC_027468.1	884998413	x	x	
Bambusa emeiensis	NC_015830.1	340034430	x	x	x
Bambusa oldhamii	NC_012927.1	253729536	x	x	x
Brachyelytrum aristosum voucher BH:J.I. Davis 777	NC_027470.1	884998582	x	x	
Brachypodium distachyon Bd21	NC_011032.1	194033128			x
Briza maxima voucher CAN:Saarela 284	NC_027471.1	884998669	x	x	

Bromus vulgaris voucher CAN:Saarela 822	NC_027472.1	884998754	x	x	
Chimonocalamus longiusculus	NC_024714.1	675154211	x	x	
Coix lacryma-jobi	NC_013273.1	260677373	x	x	x
Dactylis glomerata voucher CAN:Saarela 496	NC_027473.1	884998837	x	x	
Dendrocalamus latiflorus	NC_013088.1	255961360	x	x	x
Deschampsia antarctica	NC_023533.1	589229800	x	x	
Diarrhena obovata voucher BH:J.I. Davis 756	NC_027474.1	884998922	x	x	
Fargesia nitida	NC_024715.1	675154294	x	x	
Fargesia spathacea	NC_024716.1	675154378	x	x	
Fargesia yunnanensis	NC_024717.1	675154462	x	x	
Ferrocalamus rimosivaginus	NC_015831.1	340034515	x	x	x
Festuca altissima	NC_019648.1	427436954	x	x	
Festuca arundinacea	NC_011713.2	255961284			x
Festuca arundinacea voucher CAN:Saarela 331	KM974751.1	768805826	x	x	
Festuca ovina	NC_019649.1	426406618	x	x	
Festuca pratensis	NC_019650.1	427437051	x	x	
Gaoligongshania megalothyrsa	NC_024718.1	675154546	x	x	
Gelidocalamus tessellatus	NC_024719.1	675154630	x	x	
Helictochloa hookeri voucher CAN:Saarela 18359	NC_027469.1	884998498	x	x	
Hierochloe odorata voucher A:E.A. Kellogg s.n.	NC_027475.1	884999006	x	x	
Hordeum jubatum voucher CAN:Saarela 18478	NC_027476.1	884999091	x	x	
Hordeum vulgare subsp. vulgare cultivar Barke	KC912687.1	521300931	x	x	
Hordeum vulgare subsp. vulgare cultivar Morex	EF115541.1	118201020			x
Indocalamus longiauritus	NC_015803.1	339906432	x	x	x
Indocalamus wilsonii	NC_024720.1	675154714	x	x	
Indosasa sinica	NC_024721.1	675154798	x	x	
Lecomtella madagascariensis	NC_024106.1	662020661	x	x	
Leersia tisserantii	JN415112.1	346228283	x	x	
Lolium multiflorum	NC_019651.1	427437197	x	x	
Lolium perenne	NC_009950.1	159106843	x	x	x
Melica mutica voucher US:W.J. Kress & M. Butts 04-7461	NC_027477.1	884999174	x	x	

Melica subulata voucher CAN:Saarela 836	NC_027478.1	884999258	x	x	
Oligostachyum shiuyingianum	NC_024722.1	675154881	x	x	
Olyra latifolia	KF515509.1	628098861	x	x	
Oryza nivara	NC_005973.1	50233947	x	x	
Oryza rufipogon	KF428978.1	552954453	x	x	
Oryza sativa (indica cultivar-group) isolate 93-11	AY522329.1	42795473	x	x	x
Oryza sativa (japonica cultivar-group) isolate PA64S	AY 522331.1	42795601			x
Oryzopsis asperifolia voucher CAN:Saarela 430	NC_027479.1	884999342	x	x	
Panicum virgatum chloroplast	NC_015990.1	345895196	x	x	x
Phaenosperma globosum voucher BH:J.I. Davis 779	NC_027480.1	884999427	x	x	
Phalaris arundinacea voucher CAN:Saarela 973	NC_027481.1	884999512	x	x	
Pharus lappulaceus	NC_023245.1	570700293	x	x	
Pharus latifolius	NC_021372.1	511347561	x	x	
Phleum alpinum voucher CAN:Saarela 1234	NC_027482.1	884999596	x	x	
Phyllostachys edulis	NC_015817.1	340034006	x	x	x
Phyllostachys nigra var. henonis	NC_015826.1	340034345	x	x	x
Phyllostachys propinqua	NC_016699.1	374249330	x	x	
Phyllostachys sulphurea	NC_024669.1	671743764	x	x	
Piptochaetium avenaceum voucher CAN:R.J. Soreng & K. Romaschenko 430	NC_027483.1	884999681	x	x	
Pleioblastus maculatus chloroplas	NC_024723.1	675155300	x	x	
Poa palustris voucher CAN:Saarela 1080	NC_027484.1	884999765	x	x	
Puccinellia nuttalliana	NC_027485.1	884999850	x	x	
Puelia olyriformis	NC_023449.1	586929210	x	x	
Rhynchoryza subulata	NC_016718.1	374249599	x	x	
Saccharum hybrid cultivar NCo 310	NC_006084.1	50812505			x
Saccharum hybrid cultivar SP-80-3280	NC_005878.2	50198865	x		
Sarocalamus faberi	NC_024713.1	675154126	x	x	
Secale cereale	NC_021761.1	525782195	x	x	
Sorghum bicolor	NC_008602.1	118614470	x	x	x
Stipa hymenoides voucher CAN:Saarela 725	NC_027464.1	884998075	x	x	

Thamnocalamus spathiflorus	NC_024724.1	675155405	x	x	
Torreyochloa pallida voucher CAN:Saarela 1110	NC_027486.1	884999935	x	x	
Trisetum cernuum voucher CAN:Saarela 876	NC_027487.1	885000020	x	x	
Triticum aestivum	NC_002762.1	14017551			x
Triticum aestivum cultivar Chinese Spring TA3008	KJ614396.1	667753146	x	x	
Triticum monococcum subsp. aegilopoides	KC912692.1	521301327	x	x	
Triticum timopheevii cultivar TA0941	KJ614407.1	667754059	x	x	
Triticum turgidum cultivar TA2801	KJ614399.1	667753395	x	x	
Triticum urartu cultivar PI428335	KJ614411.1	667754391	x	x	
Yushania levigata	NC_024725.1	675154964	x	x	
Zea mays	X86563.2	11990232	x	x	x

Table S3. Flowering time classes classified according to Ream *et al*. (2014).

Classes	**Flowering time (days)**	**Photoperiod requirements (hours)**	**Weeks vernalization (wV)**
Extremely Rapid Flowering (ERF)	< 30	20	NV
Rapid Flowering (RF)	30-35	20	NV
Intermediate Rapid Flowering (IRF)	50-60	20	NV
Intermediate Delayed Flowering (IDF)	50	20	2-4
Delayed Flowering (DF)	20-30	20	6-8
Extremely Delayed Flowering (EDF)	60	20	10

NV, No vernalization

Table S4. Bioinformatic tools used in the assembly and annotation of *Brachypodium* plastomes and in their evolutionary and genomic analyses.

Bioinformatics tools	**Brief description**	**References**
	Plastid assembly	
DUK	DUK - A fast and efficient kmer based sequence matching too.	(Li et al., 2011b)
FastQC v.0.10.1	FastQC is a quality control application for high throughput sequence data. It reads in sequence data in a variety of formats and can either provide an interactive application to review the results of several different QC checks, or create an HTML based report which can be integrated into a pipeline.	(Andrews, 2010)
Trimmomatic v.0.32	Flexible trimmer for Illumina sequence data	(Bolger et al., 2014)
Musket v.1.0.6	Multistage k-mer spectrum-based error corrector for Illumina sequence data.	(Liu et al., 2013)
BWA v.0.7.8	Fast and accurate short read alignment with Burrows–Wheeler transform.	(Li & Durbin, 2009)
VelvetOptimiser v.2.2.5	Multi-threaded Perl script for automatically optimising the three primary parameter options (K, -exp_cov, -cov_cutoff) for the Velvet de novo sequence assembler.	(Gladman & Seemann, 2012)
Velvet v.1.2.07 Columbus module	Algorithms for de novo short read assembly using de Bruijn graphs.	(Zerbino & Birney, 2008; Zerbino, 2010)
SSPACE Basic v.2.0	Stand-alone scaffolder of pre-assembled contigs using paired-read data.	(Boetzer et al., 2011)
GapFiller v.1.11	*De novo* assembly approach to fill the gap within paired reads.	(Boetzer & Pirovano, 2012; Nadalin *et al.*, 2012)
BLAST v.2.2.28+	The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.	(Camacho, 2013)
SEQuel v.1.0.2	Tool for correcting errors (i.e., insertions, deletions, and substitutions) in contigs output from assembly. The algorithm behind SEQuel makes use of a graph structure called the	(Ronen et al., 2012)

	positional “de Bruijn” graph, which models k-mers within reads while incorporating their approximate positions into the model.	
SAMtools v.0.1.18	SAMtools implements various utilities for post-processing alignments in the SAM format, such as indexing, variant caller and alignment viewer, and thus provides universal tools for processing read alignments.	(Li et al., 2009)
IGV v.2.3.8	High-performance viewer that efficiently handles large heterogeneous data sets, while providing a smooth and intuitive user experience at all levels of genome resolution.	(Thorvaldsdóttir *et al.*, 2013)
Alignment and viewer		
MAFFT v.7.031b	Multiple sequence alignment program.	(Katoh & Standley, 2013)
MEGA v.7.0.14	The Molecular Evolutionary Genetics Analysis (MEGA) software is developed for comparative analyses of DNA and protein sequences	(Kumar et al., 2016)
SeaView v.4	Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building.	(Gouy et al., 2010)
Geneious v.8.1.4	Integrated and extendable desktop software platform for the organization and analysis of sequence data.	(Kearse et al., 2012)
trimAl v.1.2rev59	Tool for automated alignment trimming in large-scale phylogenetic analyses	(Capella-Gutiérrez *et al.*, 2009)
Annotation and drawing		
CpGAVAS (web)	Integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences.	(Liu et al., 2012a)
Organellar GenomeDRAW (web)	Tool for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets	(Lohse et al., 2013)
Circos v.0.69	Visualization tool to the identification and analysis of similarities and differences arising from comparison of genomes	(Krzywinski et al., 2009)
Phylogenetic, haplotypic and genomic diversity analyses		
JModelTest v.2.1.7	Tool to carry out statistical selection of best-fit models of nucleotide substitution	(Guindon & Gascuel, 2003;

		Darriba *et al.*, 2012)
RAxML v.8.1.17	Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies.	(Stamatakis, 2014)
MrBayes v.3.2.2	MrBayes 3 performs Bayesian phylogenetic analysis combining information from different data partitions or subsets evolving under different stochastic evolutionary models.	(Ronquist & Huelsenbeck, 2003; Ronquist et al., 2011)
BEAST v.1.8.2	Bayesian Evolutionary Analysis Sampling Trees	(Drummond et al., 2012)
BEAUti v.1.8.2	A simple user interface for creating input files to run BEAST	(Drummond et al., 2012)
Tracer v.1.6	Tracer is graphical tool for visualization and diagnostics of MCMC output.	(Rambaut et al., 2014)
TCS v.1.21	Phylogenetic network estimation using statistical parsimony.	(Clement et al., 2000)
Structure v.2.3.2	Free software package for using multi-locus genotype data to investigate population structure	(Pritchard et al., 2000)
RDP v.4.56	Recombination detection program that implements an extensive array of methods for detecting and visualizing recombination events.	(Martin et al., 2015)
OrgConv v.1.1	Computer package developed for detection of gene conversion between mitochondrial and chloroplast homologous genes.	(Hao, 2010)
In-house Scripts		
split pairs v.0.5	Efficient kseq-based program to sort and find paired reads within FASTQ/FASTA files, with the ability to edit headers with the power of Perl-style regular expressions.	https://github.com/eead-csic-compbio/chloroplast_assembly_protocol
annot_fasta_rom_gbk.pl	Tool for transferring features annotated on a reference GenBank file to another sequence (in FASTA forma)	
_check_matrix.pl	Script to analyze DNA polymorphisms along pre-aligned cp genomes. Produces data files to be used as tracks with Circos software.	
Chloroplast_Assembly_Protocol	A set of scripts for the assembly of chloroplast genomes out of whole-genome sequencing reads.	

Table S5. Comparative ptDNA data of the assembled *B. distachyon*, *B. hybridum* and *B. stacei* plastomes and Embl/ENA accession numbers.

(1) k-mer → length of k-mers in the best assembly. (2) C → number of contigs assembled (Velvet output). (3) L_c → length of the longest contig. (4) S → number of scaffolds assembled (SSPACE output). (5) L_s → length of the longest scaffold. (6) De novo assemblies as opposed to reference-guided assemblies. (7) L_{Total} → Total length of the assembled genome at the end of process (including missing data, Ns). (8) N → missing data in percent. * Original de novo assemblies combining automated scaffolding, visual inspection, and validated by Sanger sequencing (see Methods S1).

Accession ID (Embl/ENA)	**Ecotypes**	**k-mer (1)**	**C (2)**	**L_c (bp) (3)**	**Median coverage depth**	**S (4)**	**L_s (bp) (5)**	**De novo assemblies (6)**	**L_{Total} (bp) (7)**	**Ambiguous Base – Ns (%) (8)**
LT558583	ABR2	59	7	98,942	108	1	134,840		135,170	0.1
LT558584	ABR3	47	3	79,476	299	2	101,121	X	135,147	-
LT558585	ABR4	59	8	98,896	241	1	134,958		135,138	<0.1
LT558586	ABR5	81	8	98,932	81	1	134,954		135,187	0.1
LT222229	ABR6	47	3	79,487	239	2	101,032	X	135,159	-
LT558587	ABR7	59	9	98,883	250	3	134,772		135,125	0.1
LT558588	ABR8	47	8	68,322	350	3	134,878		135,214	<0.1
LT558590	Adi2	71	8	98,841	76	1	135,065		135,186	0.1
LT558633	Adi10	47	9	68,281	266	4	70,688	X	135,155	0.1
LT558632	Adi12	47	6	79,494	259	3	105,959	X	135,186	0.3
LT558591	Arn1	47	7	84,060	165	1	134,976		135,116	<0.1
LT558592	Bd1-1	47	10	87,720	211	3	135,053		135,039	0.1
LT558595	Bd18-1	59	7	98,859	96	3	99,686		135,191	0.4
LT558593	Bd2-3	81	10	92,493	84	8	93,342		135,186	3.4
LT558596	Bd21-3	47	1	135,186	302	1	135,232		135,186	<0.1
LT558597	Bd21C (control)	47	12	82,896	145	2	134,579		135,202	0.3
LT558598	Bd29-1	47	8	34,312	96	3	79,485	X	135,049	<0.1
LT558594	Bd3-1	47	8	98,932	101	2	136,368		135,186	<0.1
LT558599	Bd30-1	47	12	88,757	75	2	120,904		135,133	-
LT558606	BdTR10C	59	11	90,871	258	6	102,044		135,186	4.8
LT558607	BdTR11A	59	9	68,319	92	2	134,785		135,156	0.2
LT558608	BdTR11G	59	20	72,915	159	13	134,379		135,157	0.3
LT558609	BdTR11I	47	13	80,716	166	9	81,807		135,157	0.4

LT558610	BdTR12C	47	7	98,923	75	1	135,174		135,186	-
LT558631	BdTR13C	81	15	81,100	83	11	92,314		135,048	3.4
LT558611	BdTR13A	47	10	68,288	168	3	134,851		135,044	<0.1
LT558600	BdTR1I	81	10	50,454	83	1	135,164		135,186	0.1
LT558601	BdTR2B	73	11	68,321	192	3	135,029		135,185	0.1
LT558602	BdTR2G	81	8	68,314	78	2	118,743		135,186	4.4
LT558634	BdTR3C	47	22	70,632	271	15	75,137		134,991	0.1
LT558635	BdTR5I	59	3	79,506	187	2	101,051	X	135,186	<0.1
LT558604	BdTR7A	59	13	44,684	254	5	134,443		135,141	0.4
LT558605	BdTR8I	59	10	98,911	243	3	134,620		135,159	0.5
LT558636	BdTR9K	59	10	92,492	97	6	93,497		135,186	5.7
LT558612	Bis1	47	7	87,717	175	1	134,788		135,044	0.1
LT558613	Foz1	47	7	98,906	164	1	135,020		135,149	0.1
LT558582	Gaz8	47	23	24,678	508	14	38,929		135,187	4.3
LT558614	Jer1	47	3	79,492	133	2	101,037	X	135,161	-
LT558615	Kah1	47	8	98,853	105	1	134,976		135,186	<0.1
LT558616	Kah5	81	8	68,314	82	1	134,821		135,186	0.3
LT558617	Koz1	47	8	98,924	161	2	135,043		135,186	0.1
LT558618	Koz3	47	8	52,928	72	1	135,155		135,186	-
LT558619	Luc1	59	8	98,890	109	2	135,015		135,132	0.1
LT558620	Mig3	47	7	98,863	164	1	134,819		135,116	0.1
LT558621	Mon3	47	9	98,910	167	3	134,977		135,140	<0.1
LT558622	Mur1	47	14	82,907	163	8	102,460		135,174	1.0
LT558623	Per1	59	9	80,513	241	2	134,552		135,175	0.4
LT558625	RON4	59	6	98,901	316	1	135,080		135,144	<0.1
LT558626	S8iiC	47	12	98,846	278	7	111,777		135,145	1.8
LT558627	Sig2	59	11	98,906	243	4	134,745		135,149	0.1
LT558629	Tek2	47	8	98,911	163	2	134,887		135,159	0.2
LT558628	Tek4	47	4	79,476	271	4	79,522	X	135,159	<0.1
LT558630	Uni2	47	7	68,229	251	5	79,475	X	135,106	<0.1
LT222230	ABR113	47	7	49,506	230	*	*	X*	136,327	-
LT558624	Pob1	59	4	47,139	107	1	136,402		136,327	-
LT558603	BdTR6G	87	2	136,298	313	2	136,384		136,326	-
LT558589	ABR114	47	6	42,837	272	*	*	X*	136,330	-

Table S6. Primers used for amplification and sequencing of IRa and IRb junction regions and of the IR *rps*19 copy.

Primer name	Sequencing
1_IRb_LSC_Forward	AGCCGGATCTAAGTGTTGGC
2_IRb_LSC_Reverse	GCTCATGGTTATTTTGGCCGAT
3_LSC_IRa_Forward	ATTCCCCCAATTTGCGACCT
4_LSC_IRa_Reverse	TGTTGGCTAGGTAAACGCCC
5_IRa_SSC_Forward	ACGTTTGCTGGCTTATTTGGC
6_IRa_SSC_Reverse	TCTGTAAGTCTAGYTATCCTCGGT
7_SSC_IRb_Forward	ACTCGCTCAACTCGTTCCAA
8_SSC_IRb_Reverse	AAGATACGGAGACTTGCTTCACA

Table S7. Polymorphisms found in inter and intra-specific comparisons of the *B. distachyon*, *B. stacei* and *B. hybridum* plastomes. **(a).** Polymorphisms found between the plastomes of *B. distachyon* inbred line Bd21 (NC_011032.1) uploaded by Bortiri *et al.* (2008) and *B. distachyon* inbred line Bd21 assembled in present study. Note that our newly assembled Bd21 plastome has better supporting evidence than the NC_011032.1 plastome, as most mutations detected in our assembly have great read depth-coverage and were also found in a large number of plastomes of the other studied *B. distachyon* accessions. *Annotated insertions. **Poly-A region highly variable. **(b).** Characteristics of the 133 genes found in the *B. distachyon*, *B. stacei* and *B. hybridum* assembled plastomes, and in the B. distachyon Bd21 (NC_011032.1) reference plastome, annotated according to the best assembled *B. distachyon* ABR6 plastome (excluding the IRb region). **(c).** Indels reported in rpl23 and rps19 gene copies in several plastomes of grasses. **(d).** rpl23 pseudogene output obtained from Blastx searches of the *B. stacei/B. hybridum* 1,161 kbp insert into annotated plastomes of several grasses. **(e).** Polymorphisms detected among the assembled plastomes of the *B. stacei* and *B. hybridum* accessions.

(a)

Regions	Mutation (SNPs, indels*)	Brachypodium distachyon Bd21				Region and mutation (synonymous/non-synonymous)
		NC_011032.1 (Bortiri *et al.*, 2008)	Current study Bd21-control (Bd21C)	Consensus between our assembled Bd21-control and the other assembled lines of B. distachyon	Depth of coverage (x) of Bd21-control polymorphisms	
LARGE SINGLE COPY (LSC)	Indel AGG	1- 3	-	52/52	-	Intergenic
	Indel C	10	-	52/52	-	Intergenic
	Indel G	24	-	52/52	-	Intergenic
	Indel G	27	-	52/52	-	Intergenic
	Indel T	53	-	52/52	-	Intergenic
	SNP	G (593)	T (586)	52/52	16,302	psbA - synonymous
	Indel G	3,208	-	52/52	-	Intergenic
	Indel T	3,578	-	52/52	-	Intergenic
	SNP	A (3,921)	C (3,913)	37/52	11,536	Intergenic
	Indel T	-	6,753	52/52	14,463	psbK – non-synonymous
	Indel A	-	6,757	52/52	14,351	psbK – non-synonymous
	SNP	C (6,783)	T (6,777)	52/52	14,490	psbK – non-synonymous

	Indel A	-	6,784	52/52	14,325	psbK – non-synonymous
	Indel AAAAAA	11,141 – 11,146	-	**	-	Intergenic
	Indel C	12,333	-	52/52	-	Intergenic
	Indel T	-	17,447	52/52	16,325	Intergenic
	SNP	T (29,074)	C (29,063)	52/52	16,750	rpoC2 – non-synonymous
	SNP	A (29,276)	G (29,265)	52/52	13,439	rpoC2 – synonymous
	SNP	T (29,396)	G (29,385)	52/52	8,711	rpoC2 – synonymous
	Indel C	-	29,487	52/52	13,702	rpoC2 – non-synonymous
	Indel A	29,504	-	52/52	-	rpoC2 – non-synonymous
	Indel G	-	36,487	52/52	14,617	Intergenic
	Indel C	-	36,490	52/52	14,950	Intergenic
	SNP	G (40,576)	A (40,567)	52/52	11,318	psaA – synonymous
	Indel A	-	54,273	52/52	16,419	Intergenic
	SNP	K (70,065)	T (70,058)	52/52	14,966	Intergenic
	Missing data	N (70,595)	A (70,588)	52/52	15,477	Intergenic
	Indel T	-	70,628	52/52	16,033	Intergenic
	Indel T	-	77,483	52/52	13,873	Intergenic
	Indel A	-	77,492	52/52	14,002	Intergenic
	Indel A	-	77,494	52/52	13,879	Intergenic
	Indel C	78,925	-	52/52	-	Intergenic
	Indel T	-	78,929	46/52	12284	Intergenic
	SNP	G (78,944)	T (78,941)	52/52	12,543	Intergenic
	Indel A	-	78,950	52/52	12,884	Intergenic
INVERTED REPEAT (IR)	Indel T	-	98,920	46/52	14,041	Intergenic
	Indel T	-	98,924	46/52	14,316	Intergenic
	Indel A	-	98,928	46/52	14,421	Intergenic
	Indel G	98,948	-	52/52	-	Intergenic
	SNP	A (100,189)	C (100,189)	31/52	219	Intergenic
SHORT SINGLE COPY (SSC)	Indel A	-	103,494	**	-	Intergenic
INVERTED REPEAT (IR)	Indel T	-	115,719	49/52	12,988	Intergenic
	Missing data	-	117,709 – 117,948	-	-	Intergenic
	Indel T	123,317	-	-	-	16S ribosomal RNA
	Missing data	-	126,895 – 126,979	-	-	Intergenic
	Missing data	-	134,020 – 134,113	-	-	Intergenic

(b)

Gene	Coordinate	accessions compared	Length (bp)	Missing data (Ns)	Indels	SNPs	Parsimony informative	Non-synonymous	Missing in 122 grass ptDNA genomes
psbA	88	58	1062	0	0	6	5	6	1
matK	**1639**	**58**	**1536**	**0**	**0**	**30**	**28**	**30**	**2**
rps16	4401	58	258	0	0	4	4	4	3
psbK	6687	58	186	0	0	12	4	12	0
psbI	7283	58	111	0	0	0	0	0	0
psbD	8588	58	1062	0	0	10	9	10	0
psbC	9597	58	1422	0	0	6	6	6	1
psbZ	11627	58	189	0	0	0	0	1	2
psbM	16357	58	105	0	0	0	0	0	0
petN	17258	58	90	0	0	0	0	0	2
rpoB	**19549**	**58**	**3231**	**0**	**0**	**31**	**31**	**31**	**0**
rpoC1	22812	58	2049	0	0	22	22	22	0
rpoC2	**25070**	**58**	**4443**	**0**	**0**	**70**	**63**	**70**	**1**
rps2	29819	58	711	0	0	10	9	10	0
atpI	30785	58	744	0	0	3	3	3	0
atpH	31915	58	246	0	0	1	1	1	0
atpF	32576	58	552	0	0	2	2	2	2
atpA	34044	58	1524	0	0	14	13	14	0
rps14	36110	58	312	0	0	2	2	2	2
psaB	36567	58	2205	0	0	14	13	14	0
psaA	38797	58	2253	0	0	15	13	15	0
ycf3	41685	58	513	0	0	3	3	3	3
rps4	44631	58	606	0	0	2	2	2	0
ndhJ	48071	58	480	0	0	1	1	1	0
ndhK	48656	58	738	0	3	7	6	7	1
ndhC	49384	58	363	0	0	4	4	4	0
atpE	51608	58	414	0	0	3	3	3	0
atpB	52018	58	1497	0	0	9	9	9	2
rbcL	54300	57	1431	0	0	15	15	15	0
psaI	56192	58	111	0	0	1	1	1	3
ycf4	56622	58	558	0	0	7	7	7	1
cemA	57605	58	693	0	0	5	5	5	2
petA	58522	58	963	0	0	4	4	4	0
psbJ	60319	58	123	0	0	0	0	0	3
psbL	60572	58	117	0	0	1	1	1	4
psbF	60711	58	120	0	0	0	0	0	2
psbE	60841	58	252	0	0	1	1	1	2
petL	62370	58	96	0	0	0	0	0	5
petG	62639	58	114	0	0	0	0	0	4
psaJ	63500	58	129	0	0	1	1	1	5
rpl33	64070	58	201	0	0	1	1	1	3
rps18	64579	58	492	0	3	6	6	6	5

rpl20	65227	58	360	0	0	7	7	7	4
rps12	66278	58	363	0	0	0	0	0	0
rps12-2	66278	58	375	0	0	0	0	0	0
clpP	66533	58	651	0	0	3	3	3	0
psbB	67696	58	1527	0	0	12	12	12	0
psbT	69409	58	108	0	0	2	1	2	3
psbN	69565	58	132	0	0	0	0	0	1
psbH	69800	58	222	0	0	3	3	3	1
petB	70915	58	699	0	0	4	4	4	2
petD	72516	58	525	0	0	3	3	3	2
rpoA	73250	58	1020	0	0	15	15	15	0
rps11	74334	58	432	0	0	4	4	4	1
rpl36	74953	58	114	0	0	1	1	1	1
infA	75173	58	324	0	0	5	5	5	0
rps8	75577	58	411	0	0	4	4	4	0
rpl14	76129	58	372	0	0	3	3	3	1
rpl16	76587	58	411	0	0	4	4	4	0
rps3	78193	58	720	0	0	12	11	12	0
rpl22	78970	58	450	0	0	6	5	6	0
rps19	79493	58	282	1	0	2	2	3	3
rpl2	80038	58	792	0	431	1	1	1	1
rpl23	81540	58	282	0	0	1	1	1	0
ndhB	85219	58	1533	0	0	3	3	3	3
rps7	87762	56	471	0	0	2	2	2	1
rps15	100385	58	273	0	0	1	1	1	0
ndhF	**101014**	**57**	**2225**	**0**	**15**	**59**	**57**	**59**	**1**
rpl32	104086	58	181	0	12	9	9	9	0
ccsA	105139	58	969	0	0	19	18	19	1
ndhD	106272	58	1503	0	0	17	16	17	0
psaC	107894	58	246	0	0	5	5	5	0
ndhE	108586	57	306	27	0	4	4	31	0
ndhG	109100	58	531	0	0	6	6	6	0
ndhI	109881	58	543	0	0	9	8	9	0
ndhA	110517	58	1089	0	0	15	14	15	1
ndhH	112643	58	1182	0	0	16	16	16	0

(c)

			***rpl*23**		***rps*19**	
Specie	**Accession**	**Length (bp)**	**Functional gene copies annotated in plastid**	**Annotation of rpl23 (pseudogen or functional gene) in "the insert"**	**Plastid copies annotated**	**Annotation of rps19 copies**
Acidosa purpurea	NC_015820	139,697	2	*pseudogene not annotated*	2	
Agrostis stolonifera	NC_008591	136,584	3 rpl23 copies	56,532-56,816 rbcL-rpl23-psaI YP_874745	2	
Anomochloa marantoidea	NC_014062	138,412	2	pseugene rpl23 absent	*not annotated*	*not annotated*
Bambusa emeiensis	NC_015830	139,493	2	*pseudogene not annotated*	2	
Bambusa oldhamii	NC_012927	139,350	2	*pseudogene not annotated*	2	
Brachypodium distachyon Bd21	NC_011032	135,199	2	pseugene rpl23 absent	2	
Brachypodium stacei ABR114	**current study**	**136,330**	**2+1 pseudo**	**56,338-56,565 rbcL - rpl23 pseudogene - psaI** ***(Insert 56,336-57,496)***	**1**	**80,961-81,242**
Brachypodium hybridum ABR113	**current study**	**136,327**	**2+1 pseudo**	**56,337-56,564 rbcL - rpl23 pseudogene - psaI (Insert 56,335-57,495)**	**1**	**80,958-81,239**
Coix lacryma-jobi	NC_013273	140,745	2+1 pseudo	58,900-59,163 rbcL - rpl23 pseudogen - accD pseudogene - psaI	2	
Dendrocalamus latiflorus	NC_013088	139,394	2	*pseudogene not annotated*	2	
Ferrocalamus rimosivaginus	NC_015831	139,467	2	*pseudogene not annotated*	2	
Festuca arundinacea	NC_011713	136,048	2	*pseudogene not annotated*	2	
Hordeum vulgare subsp. vulgare cultivar Morex	EF115541	136,462	2+1	56,648-56,925 similar to rpl23 rbcL-rpl23 misc_feature-psaI	2	
Indocalamus longiauritus	NC_015803	139,668	2	*pseudogene not annotated*	2	
Lolium perenne	NC_009950	135,282	2	*pseudogene not annotated*	2	
Oryza sativa (indica cultivar-	AY522329	134,496	2	*pseudogene not annotated*	2	134,175-134,456

group) isolate 93-11						“similar to ribosomal protein S19"
Oryza sativa (japonica cultivar-group) isolate PA64S	AY 522331	134,551	2	*pseudogene not annotated*	2	80,649-80,930 134,226-134,507 "similar to ribosomal protein S19"
Panicum virgatum	NC_015990	139,619	2	*pseudogene not annotated*	2	
Phyllostachys edulis	NC_015817	139,679	2	*pseudogene not annotated*	2	
Phyllostachys nigra var. henonis	NC_015826	139,839	2	*pseudogene not annotated*	2	
Saccharum hybrid cultivar NCo 310	NC_006084	141,182	2+1	59,179-59,421 rbcL-rpl23 pseudogene-psaI	2	
Sorghum bicolor	NC_008602	140,754	3 rpl23 copies	59,411-59,701 rbcL-rpl23-psaI YP_899416	2	
Triticum aestivum	NC_002762	134,545	2+1	55,636-56,919 rbcL-rpl23 pseudogene-psaI	2	
Zea mays	X86563	140,384	2	*pseudogene not annotated*	2	

(d)

Description	**Max. score**	**Total score**	**Query cover**	**E value**	**Ident.**	**Accession**
ribosomal protein L23 [Bambusa oldhamii]	146	146	19%	3.00E-39	93%	YP_003029781.1
ribosomal protein L23 [Aristida purpurea]	144	144	19%	1.00E-38	92%	YP_009072631.1
ribosomal protein L23 [Greslania sp. McPherson 19217]	144	144	19%	2.00E-38	93%	YP_009135152.1
ribosomal protein L23 [Agrostis stolonifera]	144	144	19%	2.00E-38	93%	YP_874779.1
ribosomal protein L23 [Hordeum vulgare subsp. vulgare]	144	144	19%	3.00E-38	93%	AGP50796.1
ribosomal protein L23 [Zea mays]	143	143	19%	4.00E-38	92%	NP_043068.1
ribosomal protein L23 [Oryza sativa Japonica Group]	143	143	19%	4.00E-38	92%	NP_039429.1
ribosomal protein L23 [Olyra latifolia]	143	143	19%	4.00E-38	92%	YP_009033485.1
Putative ribosomal protein L23 from chromosome 10 chloroplast insertion [Oryza sativa Japonica Group]	143	143	19%	5.00E-38	92%	AAM08579.1
ribosomal protein L23 [Oryza sativa Indica Group]	144	144	19%	7.00E-38	92%	AER12861.1

(e)

B. stacei ABR114		**B. hybridum ABR113**		**B. hybridum BdTR6G**		**B. hybridum Pop1**	
Position	**Mutation**	**Position**	**Mutation**	**Position**	**Mutation**	**Position**	**Mutation**
1,552	indel T			1,552	indel T	-	-
-	-	-	-	7,697	substitution G (T)	-	-
-	-	-	-	70,902	**substitution (psbT gene - synonymous) T (G)**	-	-
-	-	106,202	substitution A (T)	-	-	-	-
112,971	substitution T (G)	-	-	-	-	-	-
-	-	134,462	**substitution (rpl23 gene – non – synonymous) C (G)**	-	-	-	-

Table S8. List of *B. distachyon* ptDNA haplotypes found across the 53 analyzed ecotypes' plastomes.

	Haplotypes (SNPs only, indels excluded)	**Haplotypes (SNPs and indels)**
Total # haplotypes	32	36
Unique haplotypes	26	30
H1	13 (Adi10; Adi12; Bd21-3; Bd2-3; Bd3-1; BdTR12C; BdTR1I; BdTR2B; BdTR2G; BdTR5I; BdTR9K; Kah1; Koz1)	11 (Adi12; Bd21-3; Bd2-3; Bd3-1; BdTR12C; BdTR1I; BdTR2B; BdTR2G; BdTR9K; Kah1; Koz1)
H2	2 (BdTR10C; Kah5)	2 (BdTR10C; Kah5)
H3	3 (BdTR11A; BdTR11G; BdTR11I)	3 (BdTR11A; BdTR11G; BdTR11I)
H4	4 (BdTR13A; BdTR13C; BdTR3C; Bis1)	2 (BdTR13A; Bis1)
H5	3 (BdTR8I; Tek2; Tek4)	3 (BdTR8I; Tek2; Tek4)
H6	2 (Foz1; Sig2)	2 (Foz1; Sig2)

Table S9. Percentages of membership of 53 *B. distachyon* ecotypes' plastome profiles to optimal K= 2 and K= 4 Bayesian genomic groups.

		Bayesian genomic groups					
Phylogenetic groups	**Ecotypes**	**K2**		**K4**			
		Group 1	**Group 2**	**Group 1**	**Group 2**	**Group 3**	**Group 4**
EDF+	Bd29-1	1	0	0	0.999	0	0
	Bd1-1	1	0	0	0.999	0	0
	BdTR13A	1	0	0	0.999	0	0
	BdTR13C	1	0	0	0.999	0	0
	BdTR3C	1	0	0	0.999	0	0
	Bis1	1	0	0	0.999	0	0
	BdTR7A	0.999	0.001	0.001	0.999	0	0
	BdTR11A	1	0	0	0.999	0	0
	BdTR11G	1	0	0	0.999	0	0
	BdTR11I	1	0	0	0.999	0	0
	BdTR8I	1	0	0	0.999	0	0
	Tek2	1	0	0	0.999	0	0
	Tek4	1	0	0	0.999	0	0
S+	Arn1	0.561	0.439	0.001	0.001	0.997	0.001
	Mon3	0.569	0.431	0	0	0.999	0
	ABR8	0.001	0.999	0.996	0	0	0.003
	Jer1	0.001	0.999	0.997	0	0	0.002
	ABR2	0.001	0.999	0.996	0	0.001	0.003
	S8iiC	0.001	0.999	0.996	0	0	0.003
	ABR6	0	1	0.998	0	0	0.002
	RON4	0	1	0.998	0	0	0.001
	ABR5	0	1	0.998	0	0	0.001
	Mur1	0	1	0.998	0	0	0.001
	Per1	0	1	0.998	0	0	0.001
	Foz1	0	1	0.998	0	0	0.001
	Sig2	0	1	0.998	0	0	0.001
	Mig3	0	1	0.994	0	0	0.005
	ABR4	0	1	0.996	0	0	0.003
	ABR7	0	1	0.996	0	0	0.003
	Bd30-1	0	1	0.996	0	0	0.003
	Luc1	0	1	0.996	0	0	0.004
T+	Bd21C	0	1	0.002	0	0	0.997
	ABR3	0	1	0.002	0	0	0.997
	Uni2	0	1	0.002	0	0	0.997
	Bd18-1	0	1	0.002	0	0	0.997
	Gaz8	0	1	0.001	0	0	0.998
	Koz3	0	1	0.001	0	0	0.998
	Adi10	0	1	0.001	0	0	0.998
	Adi12	0	1	0.001	0	0	0.998
	Bd21-3	0	1	0.001	0	0	0.998

Bd2-3	0	1	0.001	0	0	0.998
Bd3-1	0	1	0.001	0	0	0.999
BdTR12C	0	1	0.001	0	0	0.998
BdTR1I	0	1	0.001	0	0	0.998
BdTR2B	0	1	0.001	0	0	0.998
BdTR2G	0	1	0.001	0	0	0.998
BdTR5I	0	1	0.001	0	0	0.998
BdTR9K	0	1	0.001	0	0	0.998
Kah1	0	1	0.001	0	0	0.998
Koz1	0	1	0.001	0	0	0.998
Adi2	0	1	0.001	0	0	0.998
BdTR10C	0	1	0.001	0	0	0.999
Kah5	0	1	0.001	0	0	0.998

Table S10. Pairwise Tamura-Nei raw and phylogenetically-based patristic genetic distances between 3 *Brachypodium* and 91 grass plastomes. Patristic distances were calculated in the best ML tree (Fig. S5a, b).

			B. stacei (ABR114)	B. hybridum (ABR113)	B. distachyon (ABR6)	B. stacei (ABR114)	B. hybridum (ABR113)	B. distachyon (ABR6)
			Pastristic Tamura-Nei			Raw Tamura-Nei		
	Anomochlooideae	Anomochloa marantoidea	0.211	0.211	0.211	0.079	0.079	0.079
	Pharoideae	Pharus lappulaceus	0.156	0.156	0.156	0.059	0.059	0.060
		Pharus latifolius	0.157	0.157	0.158	0.060	0.060	0.061
	Puelioideae	Puelia olyriformis	0.102	0.102	0.102	0.040	0.040	0.041
PACMAD	Panicoideae	Lecomtella madagascariensis	0.133	0.133	0.133	0.050	0.050	0.051
		Panicum virgatum	0.143	0.143	0.144	0.054	0.054	0.055
		Zea mays	0.148	0.148	0.148	0.056	0.056	0.057
		Coix lacryma-jobi	0.148	0.148	0.148	0.056	0.056	0.056
		Sorghum bicolor	0.145	0.145	0.145	0.055	0.055	0.055
		Saccharum hybrid	0.142	0.142	0.142	0.054	0.054	0.054
BOP	Oryzoideae	Rhynchoryza subulata	0.119	0.119	0.120	0.047	0.047	0.047
		Leersia tisserantii	0.139	0.139	0.139	0.054	0.054	0.054

	Oryza rufipogon	0.133	0.133	0.133	0.051	0.051	0.051
	Oryza sativa	0.133	0.133	0.134	0.051	0.051	0.051
	Oryza nivara	0.134	0.134	0.134	0.051	0.051	0.052
Bambusoideae	Olyra latifolia	0.118	0.118	0.118	0.047	0.047	0.048
	Dendrocalamus latiflorus	0.095	0.095	0.096	0.038	0.038	0.039
	Bambusa emeiensis	0.094	0.094	0.095	0.038	0.038	0.038
	Bambusa oldhamii	0.097	0.097	0.097	0.039	0.039	0.039
	Ampelocalamus calcareus	0.099	0.099	0.099	0.039	0.039	0.040
	Gaoligongshania megalothyrsa	0.098	0.098	0.098	0.039	0.039	0.039
	Ferrocalamus rimosivaginus	0.096	0.096	0.096	0.038	0.038	0.038
	Gelidocalamus tessellatus	0.096	0.096	0.096	0.038	0.038	0.038
	Arundinaria gigantea	0.097	0.097	0.097	0.038	0.038	0.038
	Arundinaria appalachiana	0.096	0.096	0.096	0.038	0.038	0.038
	Arundinaria tecta	0.097	0.097	0.097	0.038	0.038	0.039
	Acidosasa purpurea	0.095	0.095	0.096	0.038	0.038	0.038
	Pleioblastus maculatus	0.095	0.095	0.096	0.038	0.038	0.038
	Indosasa sinica	0.095	0.095	0.096	0.038	0.038	0.038
	Oligostachyum shiuyingianum	0.095	0.095	0.096	0.038	0.038	0.038
	Indocalamus wilsonii	0.096	0.096	0.097	0.038	0.038	0.039
	Chimonocalamus longiusculus	0.096	0.096	0.097	0.038	0.038	0.039
	Thamnocalamus spathiflorus	0.096	0.096	0.096	0.038	0.038	0.039
	Sarocalamus faberi	0.095	0.095	0.095	0.037	0.037	0.038
	Fargesia yunnanensis	0.095	0.095	0.095	0.037	0.037	0.038
	Indocalamus longiauritus	0.095	0.095	0.095	0.038	0.038	0.038
	Yushania levigata	0.095	0.095	0.095	0.038	0.038	0.038
	Fargesia nitida	0.095	0.095	0.095	0.038	0.038	0.038
	Fargesia spathacea	0.095	0.095	0.095	0.038	0.038	0.038
	Phyllostachys propinqua	0.095	0.095	0.096	0.038	0.038	0.038
	Phyllostachys edulis	0.095	0.095	0.095	0.038	0.038	0.038
	Phyllostachys nigra	0.095	0.095	0.095	0.038	0.038	0.038

		Phyllostachys sulphurea	0.095	0.095	0.095	0.038	0.038	0.038
Pooideae	Brachyelytreae	Brachyelytrum aristosum	0.104	0.104	0.105	0.041	0.041	0.041
	Phaenospermateae	Phaenosperma globosum	0.077	0.077	0.077	0.031	0.031	0.032
	Stipeae	Stipa hymenoides	0.078	0.078	0.078	0.031	0.031	0.032
		Piptochaetium avenaceum	0.074	0.074	0.074	0.030	0.030	0.030
	Ampelodesmeae	Ampelodesmos mauritanicus	0.070	0.070	0.071	0.028	0.028	0.029
	Stipeae	Oryzopsis asperifolia	0.070	0.070	0.071	0.028	0.028	0.029
	Meliceae	Melica mutica	0.095	0.095	0.095	0.037	0.037	0.037
		Melica subulata	0.097	0.097	0.097	0.037	0.037	0.038
	Diarheneae	Diarrhena obovata	0.055	0.055	0.055	0.023	0.023	0.023
	Poeae + Aveneae	Avena sativa	0.099	0.099	0.099	0.039	0.039	0.039
		Trisetum cernuum	0.095	0.095	0.095	0.037	0.037	0.037
		Phalaris arundinacea	0.089	0.089	0.089	0.035	0.035	0.035
		Torreyochloa pallida	0.087	0.087	0.087	0.035	0.035	0.035
		Anthoxanthum odoratum	0.105	0.105	0.105	0.041	0.041	0.041
		Hierochloe odorata	0.088	0.088	0.088	0.035	0.035	0.035
		Briza maxima	0.098	0.098	0.099	0.039	0.039	0.039
		Agrostis stolonifera	0.092	0.092	0.092	0.036	0.036	0.036
		Ammophila breviligulata	0.085	0.085	0.085	0.034	0.034	0.034
		Puccinellia nuttalliana	0.095	0.095	0.095	0.037	0.037	0.038
		Phleum alpinum	0.090	0.090	0.091	0.036	0.036	0.036
		Poa palustris	0.095	0.095	0.095	0.037	0.037	0.037
		Helictochloa hookeri	0.099	0.099	0.099	0.039	0.039	0.039
		Dactylis glomerata	0.098	0.098	0.099	0.039	0.039	0.039
		Deschampsia antarctica	0.092	0.092	0.093	0.037	0.037	0.037
		Festuca ovina	0.100	0.100	0.100	0.039	0.039	0.039
		Festuca altissima	0.092	0.092	0.092	0.036	0.036	0.036
		Festuca arundinacea	0.101	0.101	0.102	0.039	0.039	0.040
		Festuca pratensis	0.103	0.103	0.103	0.040	0.040	0.041

	Lolium multiflorum	0.104	0.104	0.104	0.040	0.040	0.041
	Lolium perenne	0.104	0.104	0.104	0.040	0.040	0.041
Triticeae + Bromeae	Bromus vulgaris	0.092	0.092	0.092	0.036	0.036	0.036
	Hordeum jubatum	0.092	0.092	0.092	0.035	0.035	0.036
	Hordeum vulgare	0.094	0.094	0.094	0.036	0.036	0.037
	Secale cereale	0.090	0.090	0.090	0.035	0.035	0.036
	Triticum monococcum	0.090	0.089	0.090	0.035	0.035	0.035
	Triticum urartu	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops tauschii	0.089	0.089	0.090	0.035	0.035	0.035
	Aegilops cylindrica	0.089	0.089	0.090	0.035	0.035	0.035
	Aegilops geniculata	0.089	0.089	0.090	0.035	0.035	0.036
	Aegilops bicornis	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops kotschyi	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops sharonensis	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops longissima	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops searsii	0.089	0.089	0.089	0.035	0.035	0.035
	Aegilops speltoides	0.089	0.089	0.089	0.035	0.035	0.035
	Triticum timopheevii	0.089	0.088	0.089	0.035	0.035	0.035
	Triticum turgidum	0.089	0.089	0.089	0.035	0.035	0.035
	Triticum aestivum	0.089	0.089	0.089	0.035	0.035	0.035

Supporting Figures

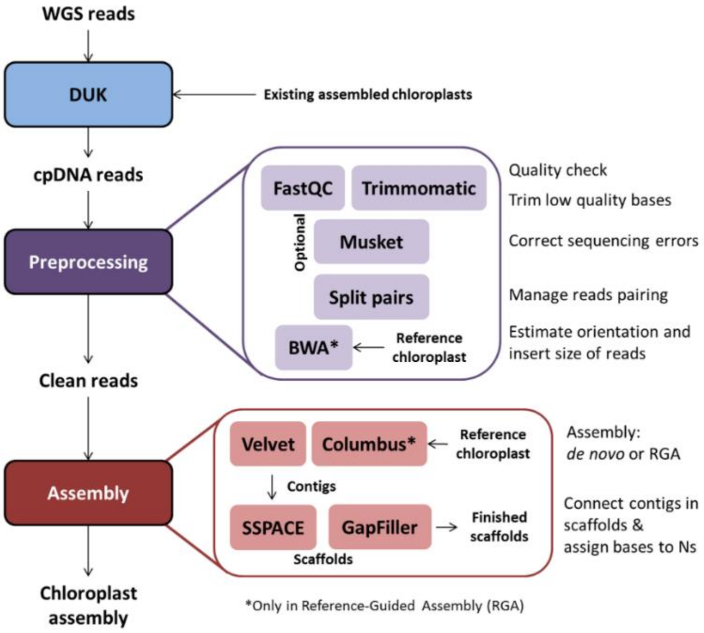


Figure S1. Pipeline used for the assembly of the *Brachypodium* plastomes.

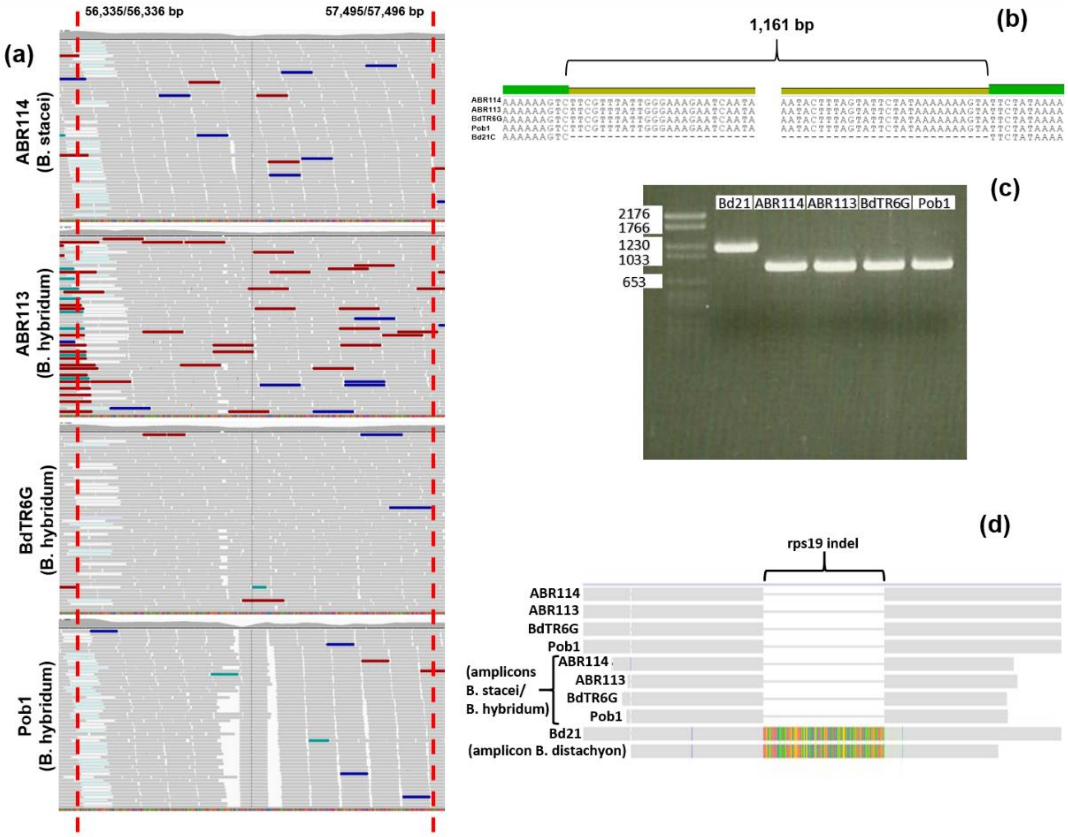


Figure S2. Evidence of major indels found among the *B. distachyon*, *B. stacei* and *B. hybridum* plastomes. **(a).** IGV image of the psaI - rbcL insert region (1,161 bp) found in the assembled *B. stacei* and *B. hybridum* plastomes. **(b).** Alignment of the insert region in *B. stacei*, *B. hybridum* and *B. distachyon* (Bd21C) ecotypes. **(c).** Electrophoresis gel showing the amplified LSC-IRb junction region (including deletion of one rps19 copy). **(d).** Evidence of rps19 indel found among B. distachyon, B. stacei and B. hybridum plastomes.

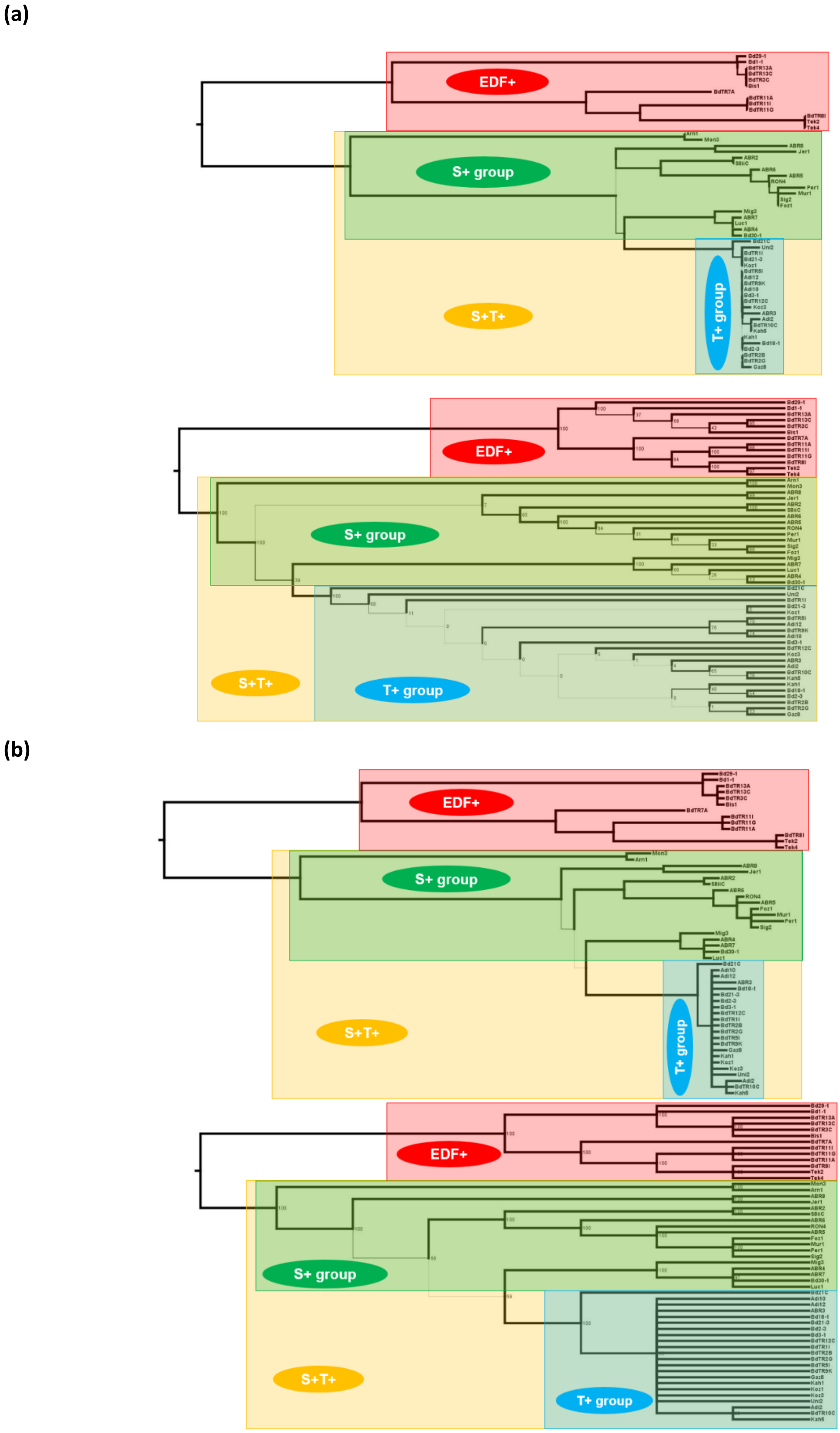


Figure S3. Phylogenomic analysis of *B. distachyon* plastomes. **(a).** Maximum likelihood ptDNA phylogenomic tree and cladogram of 53 *Brachypodium* distachyon ecotypes computed with RAxML. Thickness of branches indicates bootstrap support (thick, 90-100%; intermediate, 70-89%; thin, <70%). **(b).** Bayesian ptDNA 50% majority rule consensus phylogenomic tree and cladogram of 53 *Brachypodium* distachyon ecotypes computed with MrBayes. Thickness of branches indicates posterior probability support (thick, 0.95-1; intermediate, 0.90-0.94.; thin, <0.90).

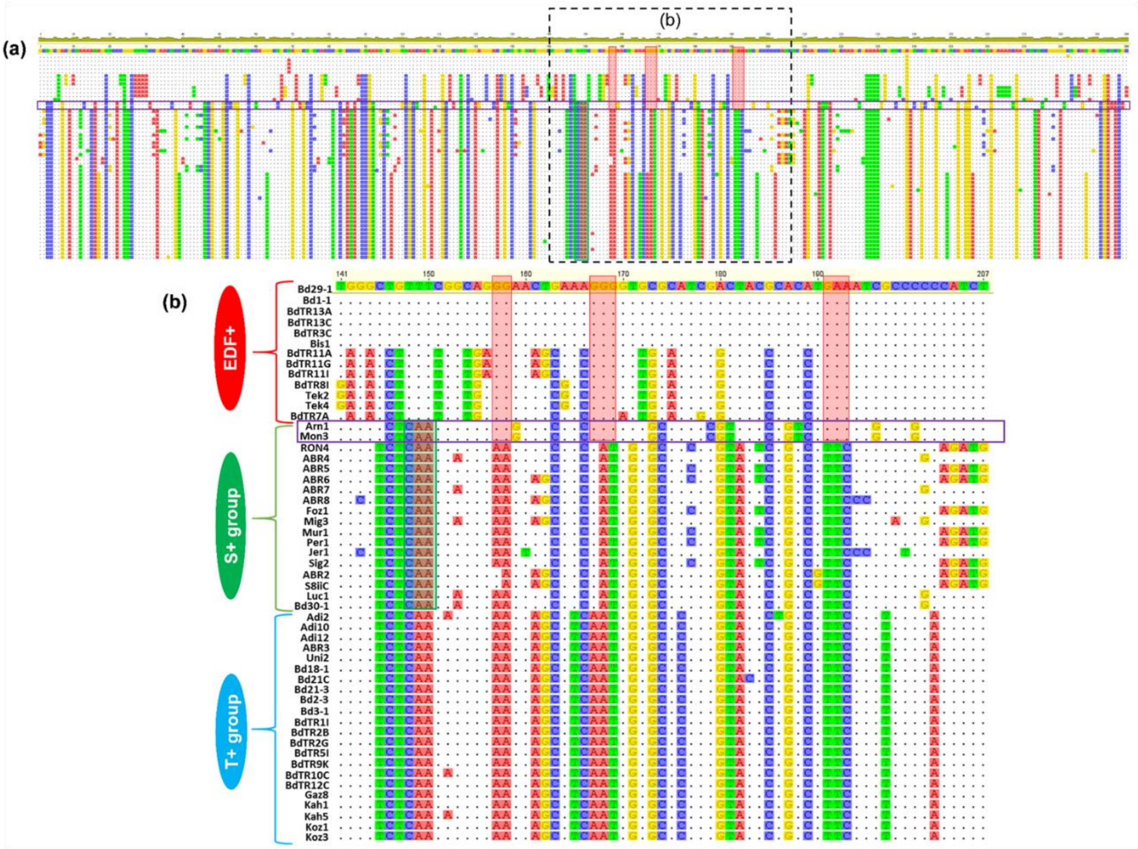
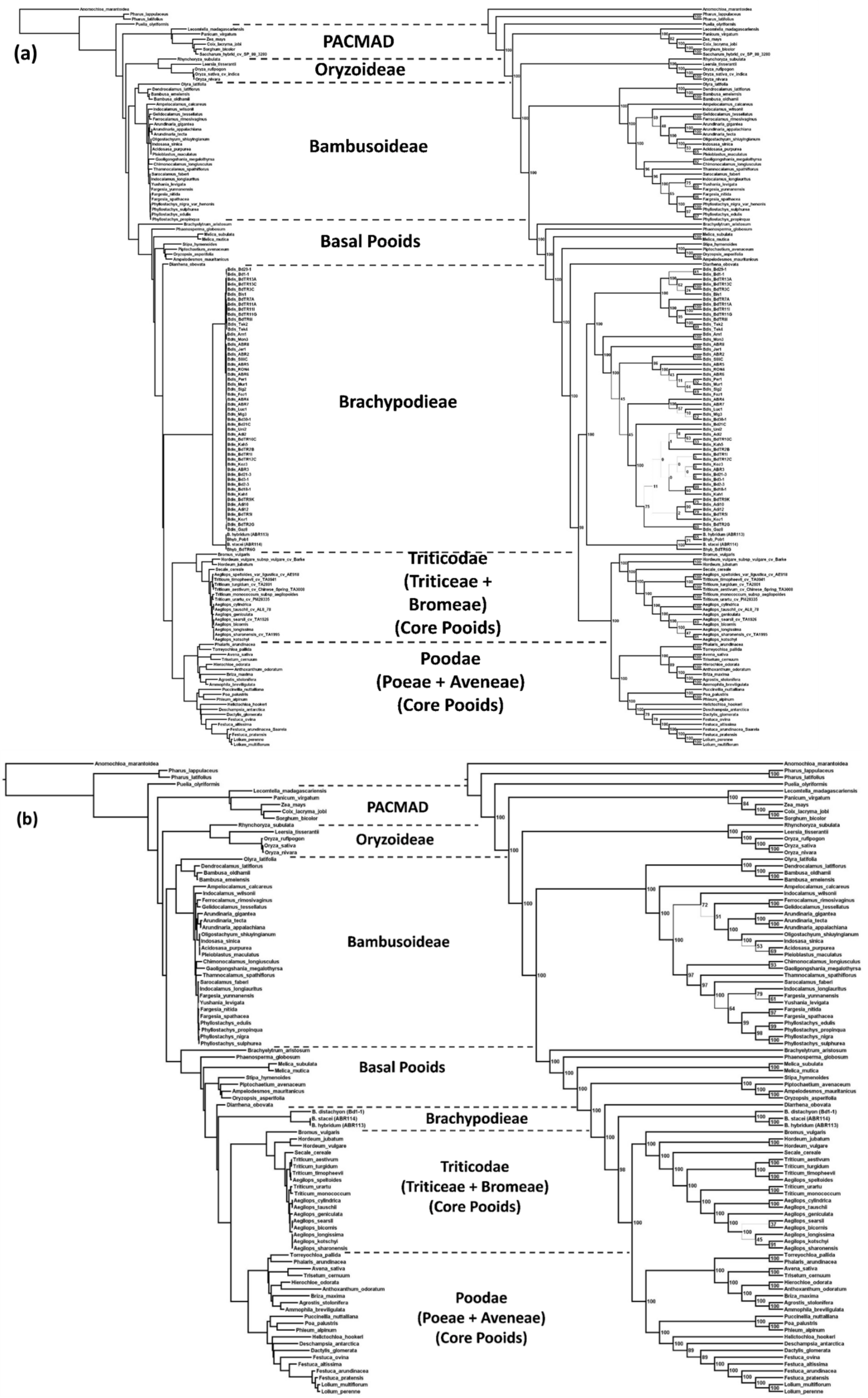
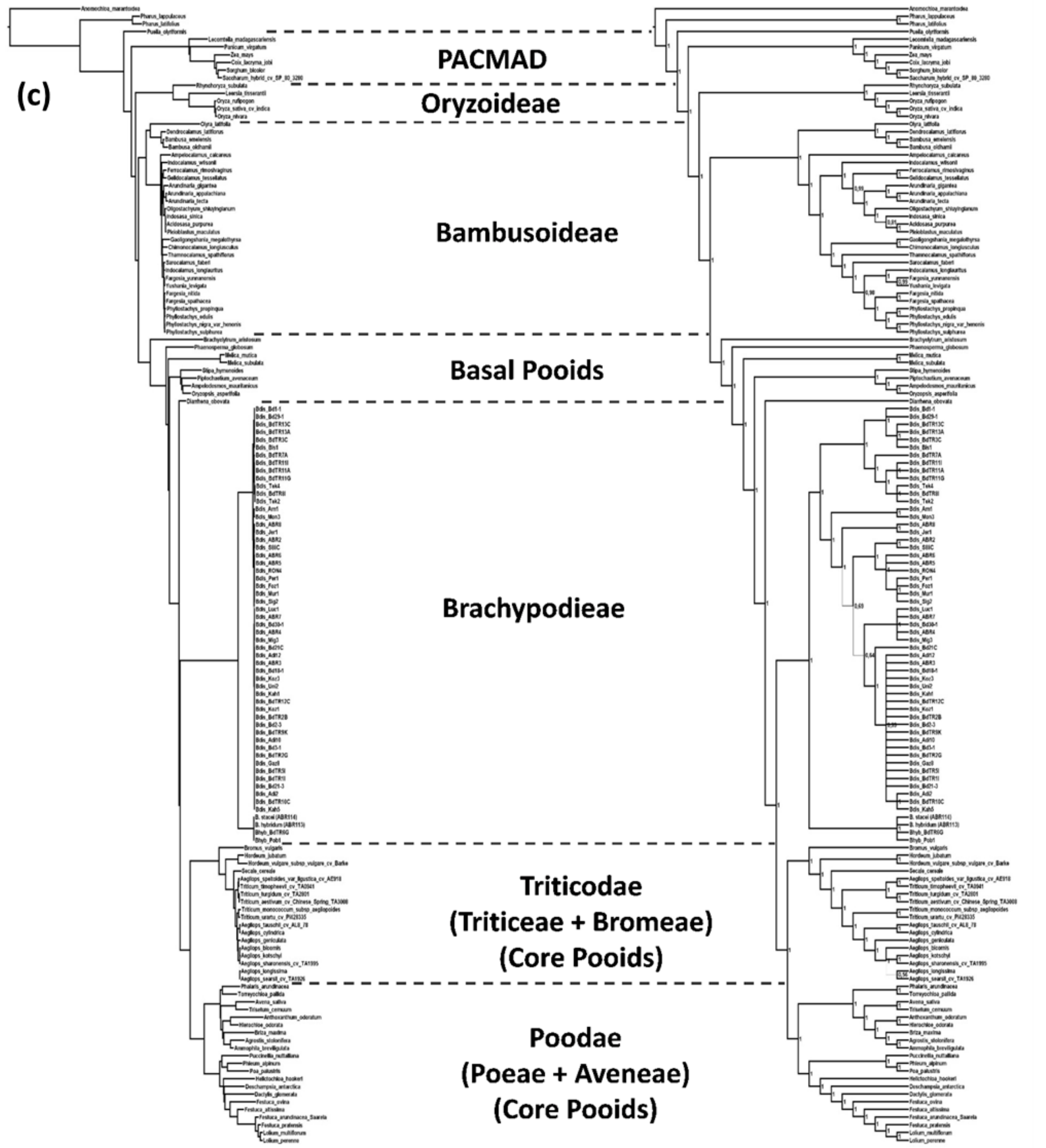


Figure S4. Potential recombination events detected in the plastomes of the introgressed *B. distachyon* Arn1 and Mon3 ecotypes.

(a). Aligned data matrix of 298 polymorphic positions found across the 53 studied B. distachyon plastomes. **(b).** Detail of the recombinant region (polymorphic positions 141 – 207) indicating potential micro-recombination events (Red rectangle: positions shared between Arn1 and Mon3 and EDF+ clade. Green rectangle: positions shared between Arn1 and Mon3 and S+ group).


(a)
PACMAD
Oryzoideae
Bambusoideae
Basal Pooids
Brachypodieae
Triticodae
(Triticeae +
Bromeae)
(Core Pooids)
Poodae
(Poeae + Aveneae)
(Core Pooids)
(b)
PACMAD
Oryzoideae
Bambusoideae
Basal Pooids
Brachypodieae
Triticodae
(Triticeae + Bromeae)
(Core Pooids)
Poodae
(Poeae + Aveneae)
(Core Pooids)


(c)
PACMAD
Oryzoideae
Bambusoideae
Basal Pooids
Brachypodieae
Triticodae
(Triticeae + Bromeae)
(Core Pooids)
Poodae
(Poeae + Aveneae)
(Core Pooids)
Anomochloa_marantoidea
Pharus_lappulaceus
Pharus_latifolius
Puelia_olyriformis
Lecomtella_madagascariensis
Panicum_virgatum
Zea_mays
Coix_lacryma_jobi
Sorghum_bicolor
Saccharum_hybrid_cv_SP_80_3280
Rhynchoryza_subulata
Leersia_tisserantii
Oryza_rufipogon
Oryza_sativa_cv_indica
Oryza_nivara
Olyra_latifolia
Dendrocalamus_latiflorus
Bambusa_emeiensis
Bambusa_oldhamii
Ampelocalamus_calcareus
Indocalamus_wilsonii
Ferrocalamus_rimosivaginus
Gelidocalamus_tessellatus
Arundinaria_gigantea
Arundinaria_appalachiana
Arundinaria_tecta
Oligostachyum_shiuyingianum
Indosasa_sinica
Acidosasa_purpurea
Pleioblastus_maculatus
Gaoligongshania_megalothyrsa
Chimonocalamus_longiusculus
Thamnocalamus_spathiflorus
Sarocalamus_faberi
Indocalamus_longiauritus
Fargesia_yunnanensis
Yushania_levigata
Fargesia_nitida
Fargesia_spathacea
Phyllostachys_propinqua
Phyllostachys_edulis
Phyllostachys_nigra_var_henonis
Phyllostachys_sulphurea
Brachyelytrum_aristosum
Phaenosperma_globosum
Melica_mutica
Melica_subulata
Stipa_hymenoides
Piptochaetium_avenaceum
Ampelodesmos_mauritanicus
Oryzopsis_asperifolia
Diarrhena_obovata
Bdis_Bd1-1
Bdis_Bd29-1
Bdis_BdTR13C
Bdis_BdTR13A
Bdis_BdTR3C
Bdis_Bis1
Bdis_BdTR7A
Bdis_BdTR11I
Bdis_BdTR11A
Bdis_BdTR11G
Bdis_Tek4
Bdis_BdTR8I
Bdis_Tek2
Bdis_Arn1
Bdis_Mon3
Bdis_ABR8
Bdis_Jer1
Bdis_ABR2
Bdis_S8IIC
Bdis_ABR6
Bdis_ABR5
Bdis_RON4
Bdis_Per1
Bdis_Foz1
Bdis_Mur1
Bdis_Sig2
Bdis_Luc1
Bdis_ABR7
Bdis_Bd30-1
Bdis_ABR4
Bdis_Mig3
Bdis_Bd21C
Bdis_Adi12
Bdis_ABR3
Bdis_Bd18-1
Bdis_Koc3
Bdis_Uni2
Bdis_Kah1
Bdis_BdTR12C
Bdis_Koz1
Bdis_BdTR2B
Bdis_Bd2-3
Bdis_BdTR9K
Bdis_Adi10
Bdis_Bd3-1
Bdis_BdTR2G
Bdis_Gaz8
Bdis_BdTR5I
Bdis_BdTR1I
Bdis_Bd21-3
Bdis_Adi2
Bdis_BdTR10C
Bdis_Kah5
B. stacei (ABR114)
B. hybridum (ABR113)
Bhyb_BdTR6G
Bhyb_Pob1
Bromus_vulgaris
Hordeum_jubatum
Hordeum_vulgare_subsp_vulgare_cv_Barke
Secale_cereale
Aegilops_speltoides_var_ligustica_cv_AE910
Triticum_timopheevii_cv_TA0041
Triticum_turgidum_cv_TA2001
Triticum_aestivum_cv_Chinese_Spring_TA3008
Triticum_monococcum_subsp_aegilopoides
Triticum_urartu_cv_PI428335
Aegilops_tauschii_cv_AL8_78
Aegilops_cylindrica
Aegilops_geniculata
Aegilops_bicornis
Aegilops_kotschyi
Aegilops_sharonensis_cv_TA1995
Aegilops_longissima
Aegilops_searsii_cv_TA1926
Phalaris_arundinacea
Torreyochloa_pallida
Avena_sativa
Trisetum_cernuum
Anthoxanthum_odoratum
Hierochloe_odorata
Briza_maxima
Agrostis_stolonifera
Ammophila_breviligulata
Puccinellia_nuttalliana
Phleum_alpinum
Poa_palustris
Helictochloa_hookeri
Deschampsia_antarctica
Dactylis_glomerata
Festuca_ovina
Festuca_altissima
Festuca_arundinacea_Saarela
Festuca_pratensis
Lolium_multiflorum
Lolium_perenne
1
0,95
0,91
0,99
0,98
0,63
0,64
0,56

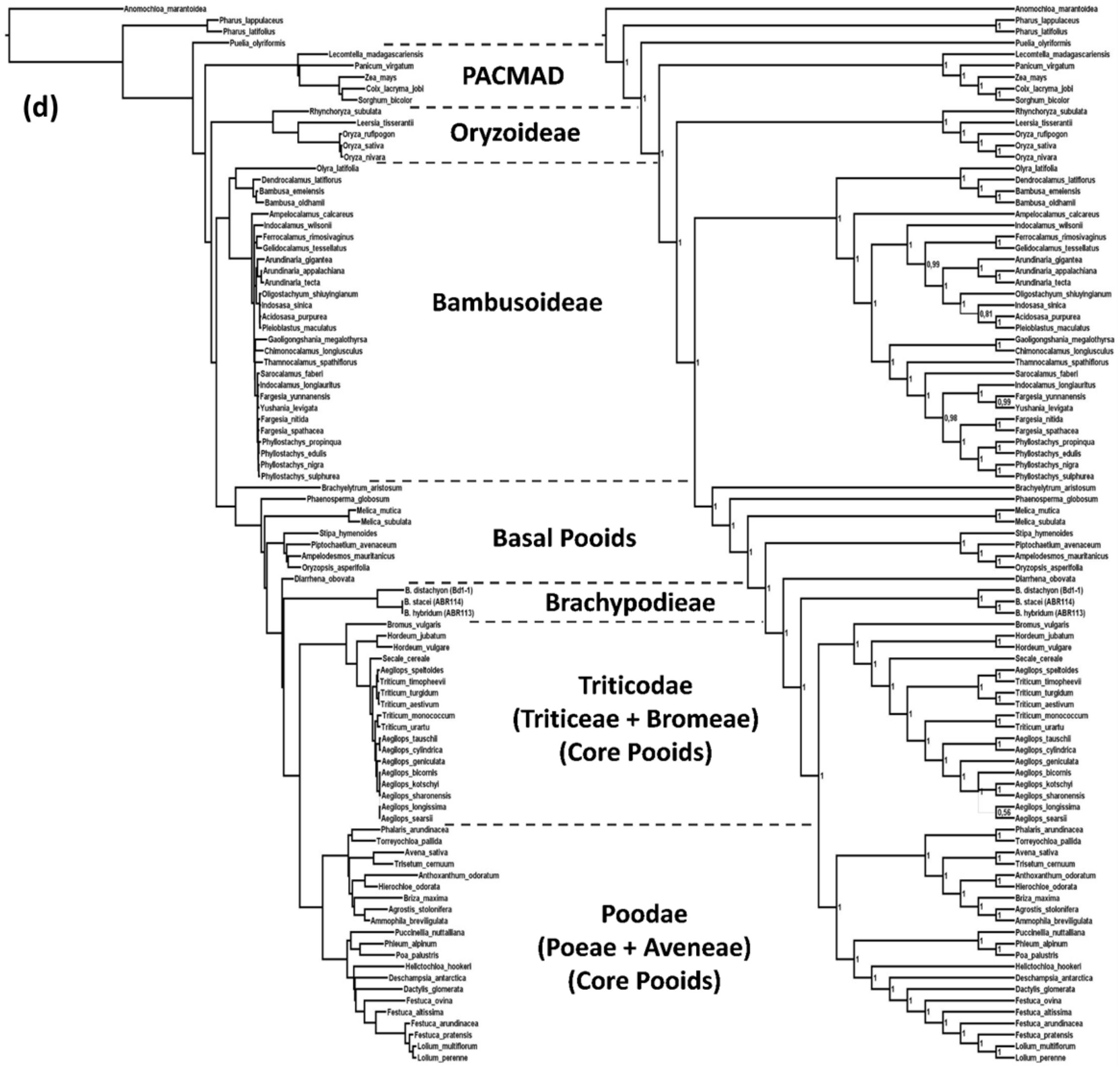


Figure S5. Plastome phylogenomic analysis of Poaceae.

(a). Maximum likelihood ptDNA phylogenomic tree and cladogram of 95 Poaceae lineages, including one *B. stacei* and three *B. hybridum* accessions, and 53 *Brachypodium distachyon* lineages computed with RAxML. Thickness of branches indicates bootstrap support (thick, 90-100%; intermediate, 70-89%; thin, <70%). **(b).** Maximum likelihood ptDNA phylogenomic tree and cladogram of 93 grass lineages plus one accession each of *B. distachyon*, *B. stacei* and *B. hybridum* computed with RAxML. Thickness of branches indicates bootstrap support (thick, 90-100%; intermediate, 70-89%; thin <70%). **(c).** Bayesian ptDNA 50% majority rule consensus phylogenomic tree and cladogram of 95 Poaceae lineages, including one *B. stacei* and three *B. hybridum* accessions, and 53 *Brachypodium* distachyon lineages computed with MrBayes. Thickness of branches indicates bootstrap support (thick, 0.95-1; intermediate, 0.90-0.94.; thin grey, <0.90). **(d).** Bayesian ptDNA 50% majority rule consensus phylogenomic tree and cladogram of 93 grass lineages plus one accession each of *B. distachyon, B. stacei* and *B. hybridum* computed with RAxML. Thickness of branches indicates bootstrap support (thick, 0.95-1; intermediate, 0.90-0.94.; thin grey, <0.90)

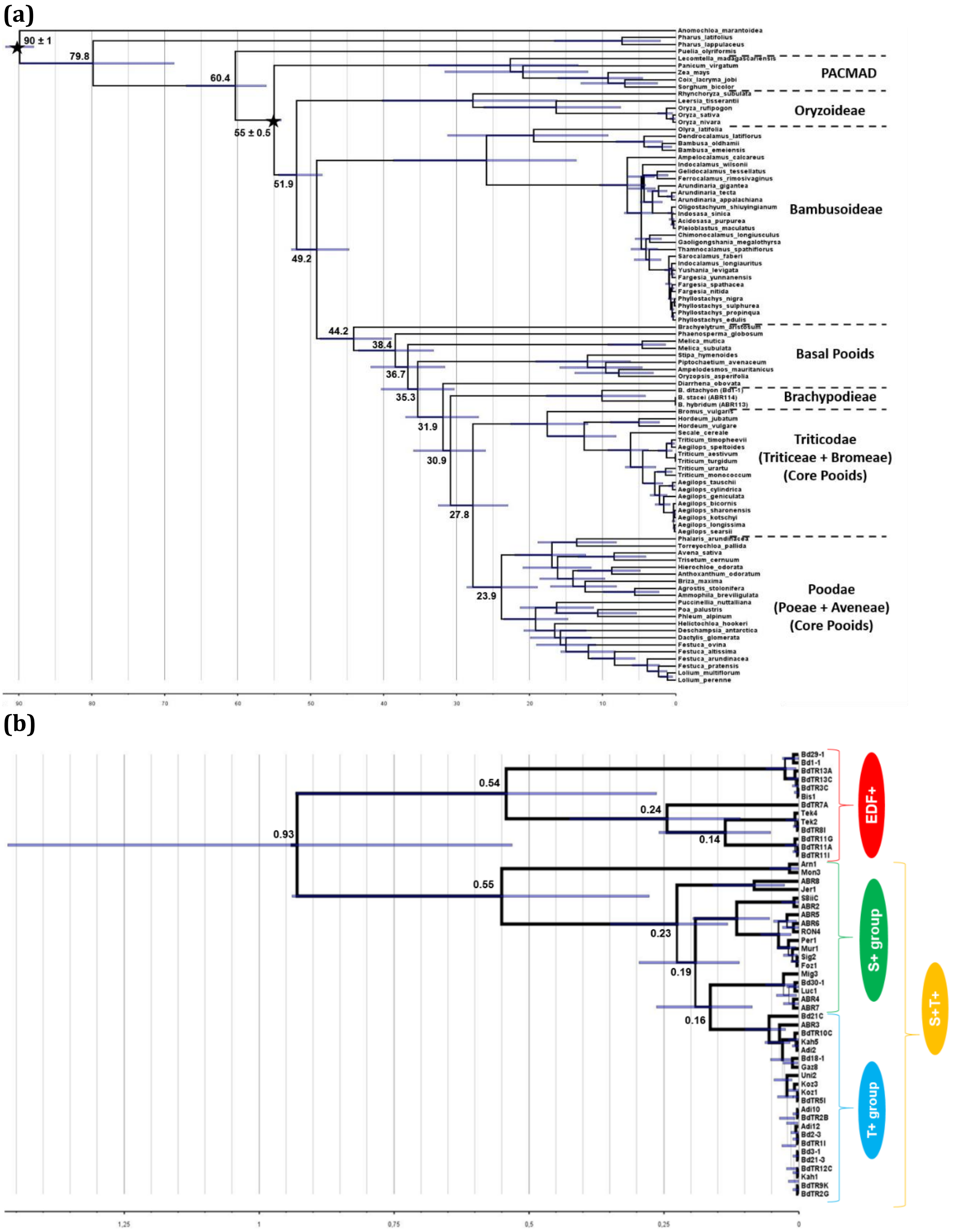


Figure S6. BEAST nested dating analysis of Poaceae (above-species) and *B. distachyon* (below-species) plastome sequences.

(a). BEAST nested dated chronogram of 93 above-species grass plastomes showing the estimated divergence times, HPD ranges (bars) of each node. Stars indicate nodal calibration priors (ages) for the Poaceae and BOP+PACMAD clades. **(b).** BEAST nested dated chronogram of 53 below-species B. distachyon plastomes showing divergence times, HPD ranges (bars) and posterior probability support (thick, 0.95-1; intermediate, 0.90-0.94.; thin, <0.90) of each node.

Appendix IV: Supporting Information of Chapter 4

Supporting Tables

Table S1. Natural accessions of *Brachypodium distachyon* used in the study. Information on elevation (meters above sea level, masl), latitude and longitude of collection sites.

Accession	Collection location	Elevation (masl)	Latitude	Longitude
ABR2	Hérault, France	371	43° 36' 15.343" N	3° 15' 46.580" E
ABR3	Aísa, Huesca, Spain	1928	42° 10' 49.8" N	0° 4’ 23.2" W
ABR4	Arén, Huesca, Spain	480	42° 15' 45.54" N	0° 43' 0.48" E
ABR5	Jaca, Huesca, Spain	828	42° 34' 23.45" N	0° 33' 49.39" W
ABR6	Los Arcos, Navarra, Spain	484	42° 34' 27.48" N	2° 11' 5.39" W
ABR8	Siena, Italy	272	43° 18' 52.423" N	11° 19' 10.902" E
Adi10	Adiyaman, Turkey	510	37° 46' 14.5" N	38° 21' 8.2" E
Adi12	Adiyaman, Turkey	510	37° 46' 14.5" N	38° 21' 8.2" E
Adi2	Adiyaman, Turkey	510	37° 46' 14.5" N	38° 21' 8.2" E
Bd1-1	Soma, Manisa, Turkey	141	39° 11' 27.44" N	27° 36' 28.59" E
Bd18-1	Kaman, Kırşehir Province, Turkey	1101	39° 22' 4.25" N	33° 43' 48.91" E
Bd21	near Salakudin, Iraq	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd21-3	near Salakudin, Iraq	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd2-3	Iraq	42	33° 45' 39.18" N	44° 24' 11.07" E
Bd30-1	Dilar, Granada, Spain	1220	36° 59' 25.76" N	3° 33' 31.44" W
Bd3-1	Iraq	42	33° 45' 39.18" N	44° 24' 11.07" E
BdTR10c	Turkey	1288	37° 46' 41.64" N	31° 53' 5.68" E
BdTR11g	Kirklareli, Turkey	124	41° 25' 17.86" N	27° 28' 36.81" E
BdTR11i	Turkey	363	39° 44' 17.39" N	28° 2' 24.71" E
BdTR12c	Turkey	1035	39° 44' 53.45" N	34° 39' 1.15" E
BdTR13A	Ankara, Turkey	787	39° 45' 23.35" N	32° 25' 56.46" E
BdTR1i	Aydin, Turkey	841	38° 5' 35.03" N	28° 34' 59.02" E
BdTR2b	Turkey	667	40° 4' 55.55" N	31° 19' 52.01" E
BdTR2g	Ankara, Turkey	1596	40° 23' 37.13" N	32° 59' 7.32" E
BdTR3c	Turkey	1957	36° 46' 58.92" N	32° 57' 46.71" E
BdTR5i	Turkey	1596	40° 23' 37.13" N	32° 59' 7.32" E
BdTR9k	Eskişehir, Turkey	932	39° 45' 10.62" N	30° 47' 19.07" E
Bis-1	Bismil, Turkey	529	37° 52' 35.6" N	41° 0' 54.3" E
Kah-1	Kahta, Turkey	665	37° 44' 2.3" N	38° 32' 0.2" E
Kah-5	Kahta, Turkey	665	37° 44' 2.3" N	38° 32' 0.2" E
Koz-1	Kozluk, Turkey	853	38° 9' 8.2.6" N	41° 36' 34.8" E
Koz-3	Kozluk, Turkey	853	38° 9' 8.2.6" N	41° 36' 34.8" E
Ron-2	Roncal, Navarra, Spain	594	42° 46' 50" N	0° 57' 48'' W

Table S2. RNA sequencing data and drought/water experimental design information. Raw SE (raw single-end reads). Filtered SE (single-end reads filtered by Trimmomatic). date (variables sorted by sequencing dates: “a”, “b”, “c”, “d”, “e”, “f”, “g”, “h”, “i”). Treatment (drought: D; water: W).

ecotypes (accessions)	Raw SE	Filtered SE	date	Treatment
ABR2	6,601,127	1,667,289	e	D
ABR2	5,746,463	2,222,313	f	D
ABR2	4,804,333	3,147,028	a	D
ABR2	4,224,149	2,595,889	d	D
ABR2	6,489,984	2,092,468	f	W
ABR2	1,784,829	1,063,952	g	W
ABR2	9,874,691	3,725,483	b	W
ABR2	3,846,175	3,164,341	c	W
ABR3	4,625,149	2,900,061	f	D
ABR3	3,151,382	2,788,129	g	D
ABR3	5,725,171	4,069,591	a	D
ABR3	5,398,049	4,056,956	d	D
ABR3	3,653,050	937,687	f	W
ABR3	11,283,377	5,216,959	f	W
ABR3	3,858,632	2,705,509	d	W
ABR3	2,157,282	1,444,249	e	W
ABR4	7,774,753	2,116,971	e	D
ABR4	1,978,027	1,439,774	h	D
ABR4	8,113,578	5,410,834	b	D
ABR4	8,290,472	6,252,911	e	D
ABR4	4,772,086	4,265,809	g	W
ABR4	3,402,337	2,799,591	h	W
ABR4	7,152,112	3,173,759	b	W
ABR4	2,723,962	1,649,000	b	W
ABR5	1,518,507	1,059,416	g	D
ABR5	6,700,459	3,350,008	h	D
ABR5	8,415,687	3,613,083	a	D
ABR5	6,372,579	5,203,682	d	D
ABR5	4,243,346	3,411,799	g	W
ABR5	3,074,709	2,519,731	h	W
ABR5	2,960,971	1,520,797	e	W
ABR5	6,207,230	5,076,436	e	W
ABR6	38,184,219	5,211,877	f	D
ABR6	7,337,310	4,028,305	h	D
ABR6	8,407,721	4,506,524	b	D
ABR6	6,849,023	4,180,661	d	D
ABR6	5,861,023	2,762,096	f	W
ABR6	4,014,346	2,894,224	g	W
ABR6	1,572,086	726,106	d	W
ABR6	3,770,757	3,115,437	e	W
ABR8	2,180,204	1,433,019	f	D
ABR8	5,980,806	4,808,335	h	D

ABR8	4,735,655	3,098,240	d	D
ABR8	1,859,648	1,440,451	e	D
ABR8	3,581,107	2,791,425	e	W
ABR8	5,048,561	4,447,005	h	W
ABR8	2,869,311	2,334,201	a	W
ABR8	3,629,754	2,176,382	d	W
Adi-10	9,078,234	7,870,859	g	D
Adi-10	11,252,107	8,816,557	h	D
Adi-10	5,609,457	2,720,611	b	D
Adi-10	6,469,852	3,074,409	b	D
Adi-10	5,166,031	2,863,430	f	W
Adi-10	4,485,655	3,153,124	h	W
Adi-10	7,357,845	5,076,624	c	W
Adi-10	4,229,599	2,998,656	d	W
Adi-12	3,315,119	2,365,916	g	D
Adi-12	5,397,127	3,730,142	h	D
Adi-12	9,718,366	5,938,112	b	D
Adi-12	6,297,978	4,527,496	d	D
Adi-12	5,337,436	1,167,870	f	W
Adi-12	7,483,059	4,806,165	f	W
Adi-12	5,214,017	2,600,489	b	W
Adi-12	3,053,578	1,436,316	b	W
Adi-2	7,134,806	5,658,048	h	D
Adi-2	1,954,259	1,614,746	h	D
Adi-2	6,758,134	3,379,245	b	D
Adi-2	3,395,135	2,943,975	d	D
Adi-2	2,906,937	2,464,347	e	W
Adi-2	4,695,911	3,948,557	g	W
Adi-2	7,454,548	6,916,945	e	W
Bd1-1	4,922,864	2,007,767	f	D
Bd1-1	6,601,540	5,082,930	g	D
Bd1-1	6,601,084	2,998,518	a	D
Bd1-1	1,983,400	1,149,142	d	D
Bd1-1	4,801,056	4,145,854	e	W
Bd1-1	4,995,276	3,971,153	g	W
Bd1-1	4,890,577	2,222,735	b	W
Bd18-1	1,209,291	852,834	g	D
Bd18-1	4,300,181	3,294,256	d	D
Bd18-1	7,005,425	1,677,013	e	W
Bd18-1	2,234,251	1,486,291	g	W
Bd18-1	6,450,898	3,558,449	c	W
Bd18-1	11,453,626	6,881,259	c	W
Bd21	9,012,664	6,908,706	g	D
Bd21	8,393,000	6,685,084	g	D
Bd21	4,388,610	2,454,344	c	D
Bd21	3,209,402	2,440,369	e	D
Bd21	3,890,732	3,443,385	e	W
Bd21	2,731,571	1,598,308	g	W

Bd21	5,577,765	3,037,099	c	W
Bd21	7,866,176	6,578,328	d	W
Bd21-3	4,373,022	1,664,189	f	D
Bd21-3	11,184,245	8,830,207	h	D
Bd21-3	3,725,637	2,509,645	d	D
Bd21-3	7,060,460	4,794,417	d	D
Bd21-3	2,704,777	1,424,564	g	W
Bd21-3	9,189,178	3,010,421	b	W
Bd21-3	5,597,966	3,521,383	e	W
Bd2-3	6,426,336	3,985,982	e	D
Bd2-3	7,233,290	5,270,217	g	D
Bd2-3	5,158,114	2,190,269	a	D
Bd2-3	4,562,719	3,905,164	d	D
Bd2-3	4,698,918	4,091,484	g	W
Bd2-3	6,104,643	5,397,405	g	W
Bd2-3	4,932,312	4,315,897	d	W
Bd2-3	672,413	377,700	d	W
Bd30-1	15,450,940	8,648,687	f	D
Bd30-1	7,739,481	6,118,678	g	D
Bd30-1	1,926,448	1,521,180	c	D
Bd30-1	5,678,800	4,691,441	e	D
Bd30-1	17,729,697	13,269,740	f	W
Bd30-1	5,724,855	5,259,550	g	W
Bd30-1	4,774,069	3,231,139	a	W
Bd30-1	6,729,545	4,714,110	a	W
Bd3-1	7,153,771	3,363,646	f	D
Bd3-1	7,847,984	5,946,898	f	D
Bd3-1	6,279,816	5,731,603	e	D
Bd3-1	1,516,791	1,088,093	f	W
Bd3-1	4,861,752	1,968,962	f	W
Bd3-1	2,790,729	2,020,517	a	W
Bd3-1	7,500,512	5,520,814	a	W
BdTR10c	4,833,695	3,959,199	f	D
BdTR10c	8,262,740	3,977,196	f	D
BdTR10c	1,993,453	1,323,181	d	D
BdTR10c	6,191,972	4,901,492	e	D
BdTR10c	2,634,263	1,940,567	e	W
BdTR10c	5,139,972	1,885,709	f	W
BdTR10c	2,498,383	1,355,283	d	W
BdTR10c	1,790,136	1,192,950	d	W
BdTR11g	6,265,037	4,947,291	f	D
BdTR11g	8,131,691	6,832,907	h	D
BdTR11g	5,063,729	2,509,484	c	D
BdTR11g	7,408,854	6,200,830	d	D
BdTR11g	4,149,389	756,407	f	W
BdTR11g	8,603,073	6,133,854	g	W
BdTR11g	3,415,642	2,633,429	a	W
BdTR11g	6,786,389	4,707,163	b	W

BdTR11i	5,819,425	842,218	f	D
BdTR11i	2,378,944	1,592,929	g	D
BdTR11i	4,136,864	2,389,466	d	D
BdTR11i	2,559,294	1,126,434	d	D
BdTR11i	11,864,916	6,251,158	e	W
BdTR11i	4,231,383	3,079,181	i	W
BdTR11i	5,684,403	2,986,219	b	W
BdTR11i	2,671,800	1,639,304	e	W
BdtR12c	6,176,408	1,353,488	e	D
BdTR13A	3,360,398	2,395,414	e	D
BdTR13A	1,894,752	1,223,359	g	D
BdTR13A	8,824,689	5,023,187	c	D
BdTR13A	1,860,137	1,349,117	e	D
BdTR13A	9,513,938	3,604,562	e	W
BdTR13A	5,531,061	1,859,726	f	W
BdTR13A	4,519,318	3,247,705	a	W
BdTR13A	3,950,096	3,098,152	e	W
BdTR1i	4,696,550	2,983,773	g	D
BdTR1i	5,377,545	4,540,394	g	D
BdTR1i	6,541,054	4,913,208	d	D
BdTR1i	2,626,574	1,432,016	d	D
BdTR1i	5,647,332	5,068,792	g	W
BdTR1i	5,969,862	5,321,731	g	W
BdTR1i	4,937,157	3,015,216	c	W
BdTR1i	4,294,152	3,640,765	e	W
BdTR2b	5,387,175	4,634,162	e	D
BdTR2b	4,369,920	4,017,001	g	D
BdTR2b	5,347,090	3,816,668	a	D
BdTR2b	7,664,367	6,172,380	d	D
BdTR2b	5,527,842	4,695,446	e	W
BdTR2b	4,076,062	3,451,682	h	W
BdTR2b	851,547	518,275	d	W
BdTR2b	3,216,653	2,214,486	d	W
BdTR2g	8,019,598	1,877,329	e	D
BdTR2g	7,613,459	2,207,658	e	D
BdTR2g	1,660,817	847,456	d	D
BdTR2g	5,672,244	5,066,239	e	D
BdTR2g	8,509,693	4,559,083	g	W
BdTR2g	6,190,328	4,640,646	i	W
BdTR2g	3,343,180	1,915,583	d	W
BdTR2g	5,106,301	3,857,120	e	W
BdTR3c	8,983,763	4,631,760	e	D
BdTR3c	5,783,183	3,418,183	f	D
BdTR3c	11,749,346	5,279,440	c	D
BdTR3c	3,845,280	2,589,362	e	D
BdTR3c	6,719,760	1,732,909	e	W
BdTR3c	10,425,380	5,456,807	f	W
BdTR3c	4,636,494	2,930,343	c	W

BdTR3c	1,864,046	955,753	d	W
BdTR5i	6,147,955	5,646,130	g	D
BdTR5i	6,615,513	6,141,256	g	D
BdTR5i	9,911,861	7,003,424	c	D
BdTR5i	5,101,192	4,085,464	c	D
BdTR5i	10,175,612	3,244,779	e	W
BdTR5i	8,170,908	7,550,963	g	W
BdTR5i	21,415,184	16,945,841	e	W
BdTR9k	7,259,406	1,829,780	e	D
BdTR9k	4,411,548	930,976	f	D
BdTR9k	6,494,900	4,214,763	a	D
BdTR9k	4,568,566	3,358,484	b	D
BdTR9k	5,057,129	3,277,704	e	W
BdTR9k	438,159	290,059	h	W
BdTR9k	1,547,011	848,560	d	W
BdTR9k	1,551,245	818,788	d	W
Bis-1	10,011,946	3,363,014	e	D
Bis-1	6,751,170	5,619,119	h	D
Bis-1	12,089,934	8,812,063	d	D
Bis-1	1,718,896	1,275,179	e	D
Bis-1	2,480,717	1,920,222	e	W
Bis-1	3,737,209	2,567,122	e	W
Bis-1	1,997,874	1,089,328	d	W
Bis-1	3,721,297	2,853,085	e	W
Kah-1	3,626,318	2,577,194	e	D
Kah-1	7,985,749	6,827,694	g	D
Kah-1	7,086,549	3,085,245	b	D
Kah-1	3,852,565	2,846,075	e	D
Kah-1	4,006,900	3,496,340	e	W
Kah-1	5,137,740	3,398,223	f	W
Kah-1	2,338,856	1,504,107	e	W
Kah-1	8,404,640	6,208,590	e	W
Kah-5	10,194,420	4,100,570	f	D
Kah-5	2,280,340	1,927,109	g	D
Kah-5	6,221,252	3,284,009	b	D
Kah-5	3,528,651	3,016,609	d	D
Kah-5	5,593,350	1,861,012	f	W
Kah-5	5,239,891	4,620,182	g	W
Kah-5	2,577,573	1,856,909	b	W
Kah-5	4,445,867	3,605,558	e	W
Koz-1	3,505,791	2,654,375	f	D
Koz-1	3,753,891	2,934,115	h	D
Koz-1	2,186,858	989,478	d	D
Koz-1	3,372,454	2,456,387	d	D
Koz-1	6,847,592	5,801,961	h	W
Koz-1	1,347,983	440,861	h	W
Koz-1	4,448,408	3,044,017	d	W
Koz-1	2,776,127	2,116,334	e	W

Koz-3	4,185,182	3,591,463	e	D
Koz-3	3,676,421	2,971,293	g	D
Koz-3	2,038,664	1,698,405	c	D
Koz-3	8,083,293	6,657,015	d	D
Koz-3	9,774,753	4,793,443	f	W
Koz-3	3,739,971	3,045,871	g	W
Koz-3	6,580,350	2,869,464	c	W
Koz-3	9,788,997	6,103,038	c	W
Ron-2	4,058,417	3,049,059	e	D
Ron-2	4,734,388	1,587,136	f	D
Ron-2	8,603,600	5,384,036	b	D
Ron-2	4,578,734	573,134	f	W
Ron-2	4,806,195	2,029,483	f	W
Ron-2	5,141,488	2,478,911	a	W
Ron-2	3,283,641	2,784,591	e	W

Table S3. Percentage of transcripts of drought and water modules with negative k_{diff} values (difference between intra and inter-modular connectivity) ID: numerical identifier of modules. Colors of modules correspond to those indicated in fig. 2

ID	module's color	Drought	Water
1	turquoise	78.1	94.1
2	blue	45.8	91.4
3	brown	93.4	97.9
4	yellow	31.0	40.1
5	green	68.3	91.4
6	red	100	83.5
7	black	92.1	100
8	pink	100.0	96.7
9	magenta	99.1	87.4
10	purple	83.2	73.7
11	greenyellow	100	100
12	tan	100	100
13	salmon	100	63.7
14	cyan	70.2	100
15	midnightblue	100	100
16	lightcyan	100	100
17	grey60	100	100
18	lightgreen	100	100
19	lightyellow	95.3	100
20	royalblue	100	100
21	darkred	96.8	100
22	darkgreen	100	100
23	darkturquoise	95.0	100
24	darkgrey	100	100
25	orange	90.6	100
26	darkorange	100	100

27	white	91.1	100
28	skyblue	100	100
29	saddlebrown	100	100
30	steelblue	100	100
31	paleturquoise	100	-
32	violet	100	-
33	darkolivegreen	100	-
34	darkmagenta	100	-
35	sienna3	100	-
36	yellowgreen	100	-
37	skyblue3	100	-
38	plum1	100	-

Table S4. Statistics of topological features (Connectivity, Scaled connectivity, cluster coefficient, maximum adjacency ratio (MAR), Density, Centralization and Heterogeneity) of drought and water networks.

Mininum (Min.) and Maximun (Max.) range, first (1st Qu.) and third (3rd Qu.) quartiles, median and mean values are detailed for connectivity, scaled connectivity, cluster coefficient and MAR variables in all cases. Density (Dens), Centralization (Cent) and Heterogeneity (Het) averaged values are indicated for the Drought and Water networks

Network	*Statistics*	Connectivity	Scaled Connectivity	ClusterCoef	MAR	Dens	Cent	Het
DROUGHT	***Min.***	0.68	0.00275	0.00215	0.00117	0.00198	0.01317	0.97250
	1st Qu.	11.08	0.04463	0.01277	0.01623			
	Median	21.94	0.08838	0.02087	0.04671			
	Mean	32.49	0.13087	0.02851	0.07043			
	3rd Qu.	42.69	0.17196	0.03263	0.09441			
	Max.	248.27	1.00000	0.32533	0.63040			
WATER	***Min.***	1.09	0.00337	0.00323	0.00161	0.00292	0.01677	0.93200
	1st Qu.	17.63	0.05464	0.01525	0.02168			
	Median	33.42	0.10360	0.02407	0.05357			
	Mean	47.91	0.14852	0.03974	0.07198			
	3rd Qu.	62.75	0.19451	0.03515	0.09278			
	Max.	322.59	1.00000	0.28749	0.56581			

Table S5. Hub transcripts and genes of modules detected in the drought and water networks. ID: numerical identifier of modules. Colors of modules correspond to those indicated in fig. 2.

		drought		**water**	
ID	**color**	**transcripts**	**genes**	**transcripts**	**genes**
1	turquoise	54	40	55	42
2	blue	111	96	22	20
3	brown	71	66	20	14
4	yellow	110	80	322	251
5	green	61	55	54	41
6	red	30	23	100	86
7	black	13	12	30	22
8	pink	3	2	36	28
9	magenta	11	8	54	46
10	purple	32	23	83	68
11	greenyellow	13	11	1	1
12	tan	15	15	37	34
13	salmon	13	11	68	59
14	cyan	53	42	36	24
15	midnightblue	0	0	1	1
16	lightcyan	12	5	9	8
17	grey60	21	12	4	2
18	lightgreen	15	14	9	8
19	lightyellow	22	18	1	1
20	royalblue	6	1	35	28
21	darkred	6	4	11	6
22	darkgreen	8	1	10	4
23	darkturquoise	11	6	14	13
24	darkgrey	17	15	9	9
25	orange	9	7	11	10
26	darkorange	6	5	0	0
27	white	23	22	12	10
28	skyblue	8	7	11	1
29	saddlebrown	10	8	9	1
30	steelblue	9	1	8	1
31	paleturquoise	1	1	-	-
32	violet	3	2	-	-
33	darkolivegreen	3	3	-	-
34	darkmagenta	10	8	-	-
35	sienna3	9	1	-	-
36	yellowgreen	7	1	-	-
37	skyblue3	6	1	-	-
38	plum1	8	1	-	-

Table S6. Occupancy of clustered and non-clustered genes **(A)** and hub genes **(B)** from the drought and water networks in the core, soft-core and shell genes compartments. Non-redundant (non-redun) indicates number of genes found in the modules that matched the pan-genome matrix excluding duplicated genes. Genes number (#) and percentage (%) are indicated for core (present in all 33 studied accessions), soft-core (present in 32 or 31 accessions) and shell genes (present in 30 or less accessions. The percentage were computed with respect to non-redundant pan-genes. ID: numerical identifier of modules

(A)

	drought genes occupancy (H)							**Water genes occupancy (H)**						
ID	**non-redun**	**Core**		**Soft-core**		**Shell**		**non-redun**	**Core**		**Soft-core**		**Shell**	
		#	**%**	**#**	**%**	**#**	**%**		**#**	**%**	**#**	**%**	**#**	**%**
0	3545	2437	68.7	498	14.0	610	17.2	2087	1358	65.1	298	14.3	431	20.7
1	1188	901	75.8	151	12.7	136	11.4	1676	1255	74.9	238	14.2	183	10.9
2	775	548	70.7	141	18.2	86	11.1	1114	792	71.1	137	12.3	185	16.6
3	748	547	73.1	112	15.0	89	11.9	986	695	70.5	129	13.1	162	16.4
4	622	443	71.2	110	17.7	69	11.1	942	679	72.1	135	14.3	128	13.6
5	595	373	62.7	87	14.6	135	22.7	872	648	74.3	146	16.7	78	8.9
6	508	373	73.4	69	13.6	66	13.0	603	452	75.0	110	18.2	41	6.8
7	470	357	76.0	56	11.9	57	12.1	497	338	68.0	87	17.5	72	14.5
8	384	290	75.5	66	17.2	28	7.3	569	402	70.7	101	17.8	66	11.6
9	359	245	68.2	75	20.9	39	10.9	490	358	73.1	88	18.0	44	9.0
10	287	212	73.9	42	14.6	33	11.5	332	232	69.9	58	17.5	42	12.7
11	308	230	74.7	47	15.3	31	10.1	243	167	68.7	35	14.4	41	16.9
12	302	204	67.5	49	16.2	49	16.2	262	215	82.1	33	12.6	14	5.3
13	235	157	66.8	41	17.4	37	15.7	267	145	54.3	66	24.7	56	21.0
14	235	160	68.1	45	19.1	30	12.8	205	147	71.7	32	15.6	26	12.7
15	187	133	71.1	33	17.6	21	11.2	161	108	67.1	27	16.8	26	16.1
16	149	96	64.4	19	12.8	34	22.8	182	143	78.6	22	12.1	17	9.3
17	139	113	81.3	15	10.8	11	7.9	120	93	77.5	13	10.8	14	11.7
18	168	141	83.9	23	13.7	4	2.4	150	96	64.0	26	17.3	28	18.7
19	142	96	67.6	27	19.0	19	13.4	93	67	72.0	11	11.8	15	16.1
20	113	73	64.6	23	20.4	17	15.0	100	78	78.0	12	12.0	10	10.0
21	121	84	69.4	24	19.8	13	10.7	86	54	62.8	24	27.9	8	9.3
22	103	72	69.9	19	18.4	12	11.7	80	52	65.0	21	26.3	7	8.8
23	109	80	73.4	15	13.8	14	12.8	85	60	70.6	16	18.8	9	10.6
24	101	60	59.4	31	30.7	10	9.9	82	60	73.2	14	17.1	8	9.8
25	67	46	68.7	12	17.9	9	13.4	53	38	71.7	7	13.2	8	15.1
26	70	45	64.3	11	15.7	14	20.0	43	28	65.1	5	11.6	10	23.3
27	74	35	47.3	18	24.3	21	28.4	53	32	60.4	14	26.4	7	13.2
28	55	47	85.5	6	10.9	2	3.6	33	22	66.7	8	24.2	3	9.1
29	55	36	65.5	13	23.6	6	10.9	26	15	57.7	1	3.8	10	38.5
30	32	24	75.0	4	12.5	4	12.5	22	13	59.1	6	27.3	3	13.6
31	35	18	51.4	9	25.7	8	22.9	-	-	-	-	-	-	-
32	29	16	55.2	10	34.5	3	10.3	-	-	-	-	-	-	-
33	33	25	75.8	5	15.2	3	9.1	-	-	-	-	-	-	-
34	37	26	70.3	7	18.9	4	10.8	-	-	-	-	-	-	-
35	29	15	51.7	1	3.4	13	44.8	-	-	-	-	-	-	-
36	28	23	82.1	2	7.1	3	10.7	-	-	-	-	-	-	-
37	25	15	60.0	4	16.0	6	24.0	-	-	-	-	-	-	-
38	16	10	62.5	4	25.0	2	12.5	-	-	-	-	-	-	-

(B)

	drought hub genes occupancy (H)							Water hub genes occupancy (H)						
ID	non-redun	Core		Soft-core		Shell		non-redun	Core		Soft-core		Shell	
		#	%	#	%	#	%		#	%	#	%	#	%
1	40	27	67.5	7	17.5	6	15.0	41	33	80.5	5	12.2	3	7.3
2	96	55	57.3	30	31.3	11	11.5	15	6	40.0	1	6.7	8	53.3
3	66	46	69.7	14	21.2	6	9.1	14	13	92.9	1	7.1	0	0.0
4	80	54	67.5	18	22.5	8	10.0	245	182	74.3	31	12.7	32	13.1
5	34	9	26.5	2	5.9	23	67.6	41	30	73.2	7	17.1	4	9.8
6	23	19	82.6	1	4.3	3	13.0	86	54	62.8	24	27.9	8	9.3
7	12	10	83.3	1	8.3	1	8.3	22	17	77.3	3	13.6	2	9.1
8	2	2	100.0	0	0.0	0	0.0	28	22	78.6	6	21.4	0	0.0
9	8	6	75.0	2	25.0	0	0.0	46	37	80.4	7	15.2	2	4.3
10	23	19	82.6	4	17.4	0	0.0	67	46	68.7	17	25.4	4	6.0
11	11	8	72.7	1	9.1	2	18.2	1	0	0.0	1	100.0	0	0.0
12	15	10	66.7	1	6.7	4	26.7	34	29	85.3	4	11.8	1	2.9
13	8	5	62.5	1	12.5	2	25.0	59	30	50.8	14	23.7	15	25.4
14	42	28	66.7	7	16.7	7	16.7	24	15	62.5	6	25.0	3	12.5
15	0	0	0.0	0	0.0	0	0.0	1	1	100.0	0	0.0	0	0.0
16	5	4	80.0	0	0.0	1	20.0	8	7	87.5	1	12.5	0	0.0
17	12	9	75.0	3	25.0	0	0.0	2	2	100.0	0	0.0	0	0.0
18	14	11	78.6	3	21.4	0	0.0	8	5	62.5	1	12.5	2	25.0
19	17	10	58.8	6	35.3	1	5.9	1	0	0.0	1	100.0	0	0.0
20	1	1	100.0	0	0.0	0	0.0	28	19	67.9	3	10.7	6	21.4
21	4	2	50.0	2	50.0	0	0.0	6	2	33.3	4	66.7	0	0.0
22	1	1	100.0	0	0.0	0	0.0	4	2	50.0	2	50.0	0	0.0
23	6	6	100.0	0	0.0	0	0.0	13	9	69.2	2	15.4	2	15.4
24	15	5	33.3	9	60.0	1	6.7	9	5	55.6	4	44.4	0	0.0
25	4	2	50.0	1	25.0	1	25.0	6	4	66.7	0	0.0	2	33.3
26	5	3	60.0	2	40.0	0	0.0	0	0	0.0	0	0.0	0	0.0
27	22	10	45.5	5	22.7	7	31.8	9	6	66.7	3	33.3	0	0.0
28	7	7	100.0	0	0.0	0	0.0	1	1	100.0	0	0.0	0	0.0
29	8	5	62.5	1	12.5	2	25.0	1	1	100.0	0	0.0	0	0.0
30	1	1	100.0	0	0.0	0	0.0	1	1	100.0	0	0.0	0	0.0
31	1	1	100.0	0	0.0	0	0.0	-	-	-	-	-	-	-
32	2	1	50.0	1	50.0	0	0.0	-	-	-	-	-	-	-
33	3	2	66.7	1	33.3	0	0.0	-	-	-	-	-	-	-
34	8	6	75.0	1	12.5	1	12.5	-	-	-	-	-	-	-
35	1	1	100.0	0	0.0	0	0.0	-	-	-	-	-	-	-
36	1	1	100.0	0	0.0	0	0.0	-	-	-	-	-	-	-
37	1	1	100.0	0	0.0	0	0.0	-	-	-	-	-	-	-
38	1	1	100.0	0	0.0	0	0.0	-	-	-	-	-	-	-

Table S7. Statistically significant biological processes (Fisher's Exact test with False Discovery Rate (FDR) threshold > 0.05) of drought and water network modules based on the identity of transcripts assigned to modules. ID: numerical identifier of modules.

ID	drought network	water network
1	mRNA splicing, via spliceosome; response to endoplasmic reticulum stress; negative regulation of cellular macromolecule biosynthetic process; negative regulation of gene expression; establishment of protein localization to organelle; protein folding; intracellular protein transport; vesicle-mediated transport; cellular macromolecule catabolic process; RNA modification; secondary metabolic process	maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); formation of translation preinitiation complex; ribosomal large subunit assembly; fatty acid beta-oxidation; endonucleolytic cleavage of tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); chromatin remodeling; maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); regulation of translational initiation; maturation of 5.8S rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); mitochondrial translation; ribosomal small subunit assembly; mRNA transport; translational elongation; histone acetylation; RNA export from nucleus; rRNA modification; mRNA splicing, via spliceosome; ribonucleoprotein complex export from nucleus; regulation of cellular component organization; protein folding; establishment of protein localization to organelle; vesicle-mediated transport; protein complex subunit organization; protein phosphorylation; signal transduction; secondary metabolic process; drug catabolic process; plant-type cell wall organization
2	adenylate cyclase-modulating G-protein coupled receptor signaling pathway; tryptophan biosynthetic process; Golgi vesicle budding; regulation of membrane lipid distribution; protein N-linked glycosylation; calcium ion transport; organophosphate ester transport; tricarboxylic acid cycle; dicarboxylic acid metabolic process; glucose transmembrane transport; glucose import; nicotinamide nucleotide metabolic process; nucleotide phosphorylation; transmembrane receptor protein serine/threonine kinase signaling pathway; protein folding; anion transport; intracellular protein transport; protein phosphorylation; regulation of transcription, DNA-templated; RNA modification; nucleic acid phosphodiester bond hydrolysis	methylglyoxal catabolic process to lactate; spliceosomal complex assembly; mitochondrial ATP synthesis coupled electron transport; transcription initiation from RNA polymerase II promoter; nucleosome assembly; tricarboxylic acid cycle; endosomal transport; protein transmembrane transport; protein targeting; establishment of protein localization to organelle; vesicle-mediated transport; translation
3	box C/D snoRNP assembly; rRNA export from nucleus; histone exchange; assembly of large subunit precursor of preribosome; maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); endonucleolytic cleavage to generate mature 3'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA); formation of translation preinitiation complex; ribosomal large subunit	mRNA splicing, via spliceosome; intracellular signal transduction

	export from nucleus; ribosomal large subunit assembly; endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); regulation of translational initiation; ribosomal small subunit assembly; spliceosomal snRNP assembly; translational elongation; rRNA base methylation; de novo' protein folding; mitochondrial gene expression; RNA secondary structure unwinding; chaperone-mediated protein folding; tRNA aminoacylation for protein translation; protein localization to organelle; signal transduction; protein phosphorylation; cell wall organization or biogenesis; cellular response to chemical stimulus; response to organic substance	
4	photosystem II stabilization; photosynthesis, light harvesting in photosystem I; photorespiration; protein-chromophore linkage; glycine metabolic process; gluconeogenesis; pentose-phosphate shunt; translational termination; response to reactive oxygen species; cellular response to oxidative stress; response to light stimulus; coenzyme biosynthetic process; cell redox homeostasis; carboxylic acid biosynthetic process; lipid biosynthetic process; organophosphate biosynthetic process; oxidation-reduction process; regulation of macromolecule metabolic process; RNA metabolic process; nucleic acid phosphodiester bond hydrolysis	protein transport; ribonucleoprotein complex biogenesis; cellular macromolecule localization; translation
5	photosynthesis	response to zinc ion; mitochondrial electron transport, cytochrome c to oxygen; regulation of sequestering of zinc ion; signal peptide processing; intra-Golgi vesicle-mediated transport; regulation of actin cytoskeleton organization; ATP hydrolysis coupled proton transport; ER to Golgi vesicle-mediated transport; retrograde vesicle-mediated transport, Golgi to ER; regulation of cellular component size; protein targeting to ER; protein N-linked glycosylation; response to endoplasmic reticulum stress; small GTPase mediated signal transduction; ATP synthesis coupled proton transport; protein folding; mitochondrion organization; nicotinamide nucleotide metabolic process; monocarboxylic acid biosynthetic process; cell redox homeostasis; fatty acid metabolic process; cellular amino acid metabolic process; regulation of transcription, DNA-templated; RNA modification; nucleic acid phosphodiester bond hydrolysis
6	glycolipid biosynthetic process; intracellular protein transmembrane transport; protein targeting; establishment of protein localization to organelle	chorismate metabolic process; protein N-linked glycosylation; glucose transmembrane transport; glucose import; monovalent inorganic cation transport; cell surface receptor signaling

		pathway; cation transmembrane transport; inorganic ion transmembrane transport; protein phosphorylation; nitrogen compound transport; RNA modification
7	cellular respiration; protein folding; purine ribonucleoside monophosphate metabolic process; purine ribonucleotide metabolic process; intracellular protein transport; organic acid metabolic process	protein repair; photosynthesis; carboxylic acid metabolic process
8	carboxylic acid catabolic process; ER to Golgi vesicle-mediated transport; carboxylic acid biosynthetic process; organic substance transport	L-proline biosynthetic process; glycogen biosynthetic process; carbohydrate catabolic process; nucleoside monophosphate metabolic process; localization
9	L-proline biosynthetic process; cold acclimation; response to water deprivation; macromolecule modification	plant-type primary cell wall biogenesis; tetrahydrofolate interconversion; cinnamic acid biosynthetic process; cellulose biosynthetic process; L-phenylalanine catabolic process; xylan biosynthetic process; regulation of jasmonic acid mediated signaling pathway; starch metabolic process; lignin catabolic process; response to wounding; cytoskeleton organization; microtubule-based process; cell wall organization; cellular protein metabolic process; regulation of gene expression; transcription, DNA-templated; nucleic acid phosphodiester bond hydrolysis; RNA processing; RNA modification
10	response to heat; protein folding	L-serine biosynthetic process; tryptophan biosynthetic process; toxin catabolic process; glutathione metabolic process; drug transport; cellular component organization or biogenesis
11	peptidyl-serine dephosphorylation; peptidyl-diphthamide biosynthetic process from peptidyl-histidine; tRNA N2-guanine methylation; mitochondrial respiratory chain complex IV assembly; snoRNA 3'-end processing; cytoplasmic translation; ribosomal large subunit assembly; snRNA metabolic process; rRNA modification; maturation of LSU-rRNA; protein localization to membrane; maturation of 5.8S rRNA; protein targeting to mitochondrion; maturation of SSU-rRNA; translational initiation; nuclear-transcribed mRNA catabolic process; protein import; protein phosphorylation; signal transduction	No statistically significant results
12	No statistically significant results	obsolete chloroplast ribulose bisphosphate carboxylase complex biogenesis; positive regulation of superoxide dismutase activity; menaquinone biosynthetic process; chaperone cofactor-dependent protein refolding; glutaminyl-tRNAGln biosynthesis via transamidation; DNA-templated transcription, termination; response to unfolded protein; heme biosynthetic process; tRNA aminoacylation for protein translation; RNA phosphodiester bond hydrolysis, endonucleolytic; photosynthesis; methylation;

		regulation of macromolecule metabolic process; catabolic process; protein phosphorylation; signal transduction
13	No statistically significant results	stress-activated protein kinase signaling cascade; activation of protein kinase activity; signal transduction by protein phosphorylation; protein polyubiquitination; transcription, DNA-templated; regulation of transcription, DNA-templated
14	transmembrane receptor protein serine/threonine kinase signaling pathway; intracellular signal transduction; protein phosphorylation	cellular response to oxidative stress; glucose metabolic process; serine family amino acid metabolic process; regulation of RNA metabolic process; regulation of cellular macromolecule biosynthetic process; photosystem II stabilization; photosynthetic electron transport in photosystem I; photorespiration; chlorophyll biosynthetic process; carbon fixation
15	No statistically significant results	No statistically significant results
16	No statistically significant results	box H/ACA snoRNA 3'-end processing; box C/D snoRNA 3'-end processing; histone glutamine methylation; mRNA pseudouridine synthesis; snRNA pseudouridine synthesis; histone arginine methylation; peptidyl-arginine methylation, to asymmetrical-dimethyl arginine; S-adenosylmethionine metabolic process; rRNA pseudouridine synthesis; formation of translation preinitiation complex; regulation of translational initiation; ribosomal large subunit assembly; maturation of LSU-rRNA; ribosomal small subunit assembly; maturation of SSU-rRNA; tRNA modification; RNA methylation; protein import; mitochondrial transport; establishment of protein localization to organelle; protein phosphorylation; signal transduction
17	No statistically significant results	No statistically significant results
18	obsolete chloroplast ribulose bisphosphate carboxylase complex biogenesis; menaquinone biosynthetic process; positive regulation of superoxide dismutase activity; chaperone cofactor-dependent protein refolding; plastid translation; PSII associated light-harvesting complex II catabolic process; protein repair; response to unfolded protein; tRNA aminoacylation for protein translation; isoprenoid biosynthetic process; RNA modification; nucleic acid phosphodiester bond hydrolysis; cellular protein modification process; regulation of transcription, DNA-templated; cellular response to stimulus	No statistically significant results
19	No statistically significant results	No statistically significant results
20	No statistically significant results	No statistically significant results
21	plant-type primary cell wall biogenesis; cellulose biosynthetic process; cell wall organization	photosynthesis, light harvesting in photosystem I; protein-chromophore linkage; response to light stimulus; lipid biosynthetic process; cellular

		lipid metabolic process; small molecule biosynthetic process
22	No statistically significant results	No statistically significant results
23	response to karrikin; flavonoid biosynthetic process; plant-type primary cell wall biogenesis; cellulose biosynthetic process; cell wall organization	No statistically significant results
24	No statistically significant results	protein folding; cellular metabolic process
25	No statistically significant results	No statistically significant results
26	No statistically significant results	No statistically significant results
27	No statistically significant results	No statistically significant results
28	No statistically significant results	No statistically significant results
29	No statistically significant results	No statistically significant results
30	No statistically significant results	No statistically significant results
31	No statistically significant results	-
32	No statistically significant results	-
33	cellular response to phosphate starvation	-
34	No statistically significant results	-
35	No statistically significant results	-
36	No statistically significant results	-
37	No statistically significant results	-
38	No statistically significant results	-

Table S8. P-values of statistics computed for drought and water modules based on all isoforms data set by the permutation test. ID: numerical identifier of modules. Avg. weight (the average magnitude of edge weights in the water (test) dataset: or how connected nodes in the module are to each other on average); coherence (the proportion of variance in the module data explained by the module's summary profile vector in the water (test) dataset); cor.cor (concordance of the correlation structure); cor.degree (concordance of the weighted degree of nodes between the two datasets, drought (discovery) and water (test) dataset); cor.contrib (concordance of the node contribution between the two dataset); avg.cor (average magnitude of the correlation coefficients of the module in the water (test) dataset); avg.contrib (average magnitude of the node contribution in the water (test) dataset). P-value > 0.01 are indicated in bold.

ID	avg.weight	coherence	cor.cor	cor.degree	cor.contrib	avg.cor	avg.contrib
1	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
2	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
3	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
4	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
5	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
6	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
7	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
8	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
9	1E-04	**0.0179**	1E-04	1E-04	1E-04	1E-04	1E-04
10	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
11	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
12	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
13	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
14	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
15	1E-04	1E-04	1E-04	1E-04	**0.0252**	1E-04	1E-04
16	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
17	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
18	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
19	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
20	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
21	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
22	1E-04	1E-04	1E-04	1E-04	**0.9951**	1E-04	1E-04
23	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
24	1E-04	0.0002	1E-04	0.0002	0.0038	1E-04	1E-04
25	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
26	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
27	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
28	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
29	1E-04	1E-04	1E-04	0.0011	0.0081	1E-04	1E-04
30	1E-04	1E-04	1E-04	1E-04	**1**	1E-04	0.5236
31	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
32	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
33	1E-04	**0.0153985**	1E-04	1E-04	**0.87761**	0.0003	**0.0726**
34	1E-04	1E-04	1E-04	1E-04	0.0003	1E-04	1E-04
35	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
36	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
37	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
38	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04

Table S9. WGCN analyses from primary transcript data set (number and percentages), percentage of transcripts with negative k_{diff} (difference between intra and inter-modular connectivity), and number hub nodes in drought **(A)** and water **(B)** conditions. Numerical (ID) and color identifier modules Zero or grey module correspond to non-assigned primary transcripts.

(A)

		primary transcripts		negative k_{diff}	hub nodes
ID	**color**	**#**	**%**	**%**	**#**
0	grey	2554	25.9	NA	-
1	turquoise	1941	19.7	14.9	57
2	blue	714	7.2	42.4	88
3	brown	591	6.0	64.3	39
4	yellow	516	5.2	32.9	64
5	green	404	4.1	93.3	7
6	red	334	3.4	100.0	9
7	black	294	3.0	99.3	10
8	pink	291	2.9	100.0	5
9	magenta	289	2.9	100.0	45
10	purple	256	2.6	100.0	12
11	greenyellow	243	2.5	76.5	22
12	tan	215	2.2	67.4	37
13	salmon	165	1.7	90.9	15
14	cyan	157	1.6	94.9	14
15	midnightblue	147	1.5	100.0	4
16	lightcyan	117	1.2	100.0	1
17	grey60	116	1.2	100.0	5
18	lightgreen	93	0.9	100.0	4
19	lightyellow	84	0.9	100.0	13
20	royalblue	84	0.9	96.4	7
21	darkred	75	0.8	100.0	5
22	darkgreen	66	0.7	86.4	19
23	darkturquoise	59	0.6	91.5	7
24	darkgrey	39	0.4	100.0	6
25	orange	31	0.3	100.0	10

(B)

		primary transcripts		negative k_{diff}	hub nodes
ID	**color**	**#**	**%**	**%**	**#**
0	grey	1812	18.3	NA	-
1	turquoise	913	9.2	39.2	29
2	blue	908	9.2	77.2	29
3	brown	801	8.1	96.1	34
4	green	733	7.4	4.4	181
5	yellow	733	7.4	97.8	38
6	red	536	5.4	95.9	5
7	black	515	5.2	73.6	71
8	pink	401	4.1	64.1	44
9	magenta	278	2.8	67.6	59
10	purple	277	2.8	91.3	33
11	greenyellow	269	2.7	96.3	25
12	tan	232	2.3	48.7	54
13	salmon	206	2.1	100.0	9
14	cyan	203	2.1	98.5	15
15	midnightblue	188	1.9	100.0	16
16	lightcyan	181	1.8	98.9	6
17	grey60	169	1.7	100.0	2
18	lightgreen	114	1.2	100.0	11
19	lightyellow	97	1.0	100.0	10
20	royalblue	79	0.8	100.0	16
21	darkred	74	0.7	100.0	11
22	darkgreen	72	0.7	100.0	13
23	darkturquoise	51	0.5	98.0	7
24	darkgrey	33	0.3	100.0	3

Table S10. Statistically significant biological processes (Fisher's Exact test with False Discovery Rate (FDR) threshold > 0.05) of drought and water network modules based on primary transcript data. ID: numerical identifier of modules

ID	drought network	water network
1	rRNA export from nucleus; endonucleolytic cleavage to generate mature 3'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA); assembly of large subunit precursor of preribosome; exocyst localization; formation of translation preinitiation complex; endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); lipid oxidation; peptidyl-arginine modification; ribosomal large subunit assembly; regulation of translational initiation; ribosomal small subunit assembly; iron-sulfur cluster assembly; chromatin remodeling; response to endoplasmic reticulum stress; glycerophospholipid biosynthetic process; RNA secondary structure unwinding; mRNA splicing, via spliceosome; ER to Golgi vesicle-mediated transport; protein localization to membrane; protein targeting; nuclear-transcribed mRNA catabolic process; regulation of localization; protein folding; regulation of cellular component organization; detoxification; cell wall organization; defense response; secondary metabolic process; plant-type cell wall organization or biogenesis; regulation of hormone levels	spliceosomal complex assembly; cellular protein localization; protein transport; organelle organization
2	adenylate cyclase-modulating G-protein coupled receptor signaling pathway; tryptophan biosynthetic process; protein N-linked glycosylation; calcium ion transport; endocytosis; tricarboxylic acid cycle; vesicle organization; dicarboxylic acid metabolic process; membrane organization; Golgi vesicle transport; proton transmembrane transport; intracellular protein transport; nucleobase-containing small molecule metabolic process; cofactor metabolic process; protein phosphorylation; regulation of primary metabolic process; regulation of cellular metabolic process; RNA metabolic process; regulation of gene expression; nucleic acid phosphodiester bond hydrolysis	mitochondrial electron transport, cytochrome c to oxygen; regulation of vesicle targeting, to, from or within Golgi; response to zinc ion; regulation of sequestering of zinc ion; protein localization to endoplasmic reticulum exit site; purine nucleotide-sugar transmembrane transport; protein folding in endoplasmic reticulum; intra-Golgi vesicle-mediated transport; retrograde vesicle-mediated transport, Golgi to ER; signal peptide processing; ER to Golgi vesicle-mediated transport; tricarboxylic acid cycle; protein N-linked glycosylation; ubiquitin-dependent ERAD pathway; ATP synthesis coupled proton transport; regulation of protein complex assembly; carboxylic acid catabolic process; ATP hydrolysis coupled cation transmembrane transport; positive regulation of GTPase activity; protein targeting to mitochondrion; protein transmembrane transport; mitochondrial transmembrane transport; purine ribonucleoside diphosphate metabolic process; nucleotide catabolic process; nucleotide phosphorylation; aromatic amino acid family metabolic process; sulfur compound biosynthetic process; nicotinamide

		nucleotide biosynthetic process; fatty acid biosynthetic process; cell redox homeostasis; alpha-amino acid biosynthetic process; regulation of transcription, DNA-templated; transcription, DNA-templated; nucleic acid phosphodiester bond hydrolysis; RNA modification
3	Golgi vesicle transport	spliceosomal conformational changes to generate catalytic conformation; N-terminal protein amino acid modification; establishment of protein localization to organelle
4	photosystem II stabilization; CDP-diacylglycerol biosynthetic process; photosynthesis, light harvesting in photosystem I; protein-chromophore linkage; pentose-phosphate shunt; glucose metabolic process; porphyrin-containing compound biosynthetic process; response to reactive oxygen species; pigment biosynthetic process; response to light stimulus; small molecule biosynthetic process; carboxylic acid metabolic process; oxidation-reduction process; proteolysis; cellular macromolecule catabolic process	nitrogen compound transport
5	xylan biosynthetic process; tricarboxylic acid cycle; purine nucleotide biosynthetic process; purine ribonucleoside triphosphate metabolic process; purine ribonucleoside monophosphate metabolic process; purine ribonucleotide metabolic process; protein transport; intracellular transport	rRNA export from nucleus; box C/D snoRNP assembly; histone exchange; endonucleolytic cleavage to generate mature 3'-end of SSU-rRNA from (SSU-rRNA, 5.8S rRNA, LSU-rRNA); assembly of large subunit precursor of preribosome; endonucleolytic cleavage in ITS1 to separate SSU-rRNA from 5.8S rRNA and LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); regulation of translational elongation; maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); formation of translation preinitiation complex; ribosomal large subunit assembly; ribosomal small subunit assembly; regulation of mitotic metaphase/anaphase transition; rRNA methylation; Golgi vesicle transport; mRNA splicing, via spliceosome; signal transduction
6	tRNA N2-guanine methylation; ribosomal large subunit export from nucleus; cytoplasmic translation; mitochondrial gene expression; transcription by RNA polymerase I; mitochondrial respiratory chain complex assembly; protein import into mitochondrial matrix; tRNA aminoacylation for protein translation; translational initiation; ribosomal small subunit biogenesis; ribosomal large subunit biogenesis; rRNA processing; ribonucleoprotein complex assembly; transcription by RNA polymerase II; mRNA splicing, via spliceosome; protein phosphorylation	No statistically significant results

7	No statistically significant results	monocarboxylic acid catabolic process; hexose transmembrane transport; glucose import; amino acid transport; glycoprotein metabolic process; proton transmembrane transport; protein transport; protein phosphorylation; nucleic acid metabolic process
8	No statistically significant results	plant-type primary cell wall biogenesis; plant-type secondary cell wall biogenesis; cellulose biosynthetic process; xylan biosynthetic process; lignin metabolic process; phenylpropanoid biosynthetic process; microtubule-based process; cytoskeleton organization; ribonucleoside triphosphate biosynthetic process; cell wall organization; carbohydrate derivative biosynthetic process; protein modification by small protein conjugation or removal; RNA processing; RNA modification
9	No statistically significant results	L-serine biosynthetic process; tryptophan biosynthetic process; toxin catabolic process; response to salt stress; glutathione metabolic process; drug transport; nicotinamide nucleotide metabolic process
10	cold acclimation	menaquinone biosynthetic process; positive regulation of superoxide dismutase activity; glutaminyl-tRNAGln biosynthesis via transamidation; chaperone cofactor-dependent protein refolding; plastid translation; DNA-templated transcription, termination; chlorophyll biosynthetic process; lysine biosynthetic process via diaminopimelate; response to unfolded protein; tRNA aminoacylation for protein translation; serine family amino acid metabolic process; RNA methylation; photosynthesis; regulation of transcription, DNA-templated; protein phosphorylation; proteolysis involved in cellular protein catabolic process; signal transduction
11	response to heat; protein folding; protein phosphorylation	carboxylic acid metabolic process
12	trehalose biosynthetic process; regulation of response to stress; dephosphorylation; response to stimulus	positive regulation of response to salt stress; regulation of defense response; stress-activated protein kinase signaling cascade; activation of protein kinase activity; signal transduction by protein phosphorylation; protein dephosphorylation; transcription, DNA-templated; regulation of transcription, DNA-templated
13	positive regulation of superoxide dismutase activity; menaquinone biosynthetic process; chaperone cofactor-dependent protein refolding; response to unfolded protein; tRNA aminoacylation for protein translation; plastid organization; cellular response to stimulus	box H/ACA snoRNA 3'-end processing; box C/D snoRNA 3'-end processing; histone glutamine methylation; mRNA pseudouridine synthesis; snRNA pseudouridine synthesis; peptidyl-arginine methylation, to asymmetrical-dimethyl arginine; histone arginine methylation; formation of translation preinitiation complex; rRNA modification;

		regulation of translational initiation; ribosomal large subunit assembly; protein import into nucleus; maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA); RNA export from nucleus; tRNA modification; oxidation-reduction process;protein phosphorylation
14	oxidation-reduction process	glycine catabolic process; glycerol catabolic process; photosynthetic electron transport in photosystem I; glyceraldehyde-3-phosphate biosynthetic process; gluconeogenesis; response to reactive oxygen species; cellular response to oxidative stress; cellular metabolic compound salvage; hydrogen peroxide catabolic process; ribonucleoside monophosphate biosynthetic process; cellular oxidant detoxification; ribonucleotide metabolic process; RNA metabolic process; regulation of nucleic acid-templated transcription; regulation of cellular macromolecule biosynthetic process
15	response to karrikin; flavonoid biosynthetic process; plant-type primary cell wall biogenesis; cellulose biosynthetic process; phenylpropanoid biosynthetic process; alpha-amino acid biosynthetic process; cell wall organization; drug metabolic process; carbohydrate derivative metabolic process; nucleic acid metabolic process	No statistically significant results
16	No statistically significant results	No statistically significant results
17	No statistically significant results	No statistically significant results
18	No statistically significant results	No statistically significant results
19	No statistically significant results	response to heat; protein folding
20	plant-type primary cell wall biogenesis; plant-type secondary cell wall biogenesis; cellulose biosynthetic process; cell wall organization	No statistically significant results
21	regulation of transcription, DNA-templated	photosystem II stabilization; photosynthesis, light harvesting in photosystem I; protein-chromophore linkage; porphyrin-containing compound biosynthetic process; pigment biosynthetic process; response to light stimulus; lipid biosynthetic process; cellular lipid metabolic process
22	No statistically significant results	No statistically significant results
23	No statistically significant results	No statistically significant results
24	No statistically significant results	No statistically significant results
25	cellular response to phosphate starvation; cellular response to cold	No statistically significant results

Table S11. P-values of statistics computed for drought and water modules based on primary transcripts data set by the permutation test. ID: numerical identifier of modules. Avg. weight (the average magnitude of edge weights in the water (test) dataset: or how connected nodes in the module are to each other on average); coherence (the proportion of variance in the module data explained by the module's summary profile vector in the water (test) dataset); cor.cor (concordance of the correlation structure); cor.degree (concordance of the weighted degree of nodes between the two datasets, drought (discovery) and water (test) dataset); cor.contrib (concordance of the node contribution between the two dataset); avg.cor (average magnitude of the correlation coefficients of the module in the water (test) dataset); avg.contrib (average magnitude of the node contribution in the water (test) dataset). P-value > 0.01 are indicated in bold.

ID	avg.weight	coherence	cor.cor	cor.degree	cor.contrib	avg.cor	avg.contrib
1	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
2	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
3	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
4	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
5	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
6	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
7	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
8	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
9	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
10	**0.0628**	**0.1348**	1E-04	1E-04	1E-04	1E-04	1E-04
11	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
12	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
13	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
14	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
15	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
16	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
17	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
18	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
19	**0.0273**	**0.0296**	1E-04	1E-04	1E-04	1E-04	1E-04
20	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
21	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
22	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
23	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
24	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04	1E-04
25	**0.7869**	**0.9184**	**0.0026**	**0.7296**	**0.9712**	**0.0956**	**0.4872**

Table S12. Comparative analyses between differentially expressed (DE) transcripts and genes and co-expressed modules of all isoforms data set. ID: numerical identifier of modules.

	DE drought transcripts		DE water transcripts	
ID	**transcripts**	**genes**	**transcripts**	**genes**
0	712	562	535	417
1	267	191	743	601
2	253	202	256	194
3	602	503	297	204
4	499	359	358	296
5	133	112	369	296
6	200	148	211	176
7	99	85	157	119
8	339	270	442	366
9	333	252	372	301
10	172	141	102	78
11	243	211	130	85
12	46	46	73	58
13	46	41	85	70
14	74	57	126	91
15	101	81	64	46
16	93	55	170	144
17	57	34	80	48
18	40	36	47	41
19	56	47	28	22
20	56	40	50	42
21	76	63	45	31
22	41	33	30	25
23	78	66	39	32
24	90	75	42	32
25	43	25	22	18
26	38	31	11	10
27	16	16	27	24
28	11	8	16	11
29	27	21	1*	2*
30	15	10	13	5
31	4	4	-	-
32	14	9	-	-
33	23	18	-	-
34	25	21	-	-
35	0*	1*	-	-
36	5	4	-	-
37	7	2	-	-
38	7	5	-	-

*Some transcripts can belong to multiple modules, and therefore one gene can match to and be count in multiple modules. Usually the number of transcript is higher than the number of genes; however in some instances (module 29 in the water network; module 35 in the drought network) the gene count is higher than the transcripts count as a consequence of comparing different isoforms (transcript) from the same gene and the counting system for each case. All matches of the one gene to different modules will be counted.

Table S13. In-depth analyses of the most differentially expressed (DE) top-50 genes. KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) annotations; module identifiers of DE genes found within drought (D id) and water (W id) co-expression networks, and occupancy (H). Bold entries indicate cases related to water stress responses. NA (not available data).

rank	transcript id	gene id	KEGG/ec	enzyme KEGG	GO	Name and definition GO term	D id	W id	H
1	Bradi3g45080.1	Bradi3g45080	NA	NA	GO:0004857	enzyme inhibitor activity; Binds to and stops, prevents or reduces the activity of an enzyme	9	8	32
2	Bradi1g65780.1	Bradi1g65780	NA	NA	GO:0016021	integral component of membrane; The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane	24	8	33
3	Bradi2g18090.1	Bradi2g18090	NA	NA	NA	NA	9	0	32
4	Bradi5g09200.1	Bradi5g09200	NA	NA	GO:0033926	glycopeptide alpha-N-acetylgalactosaminidase activity; Catalysis of the reaction: D-galactosyl-3-(N-acetyl-alpha-D-galactosaminyl)-L-serine + H2O = D-galactosyl-3-N-acetyl-alpha-D-galactosamine + L-serine in mucin-type glycoproteins	9	4	32
5	Bradi4g39520.1	Bradi4g39520	NA	NA	NA	NA	24	8	31
6	Bradi4g40850.2	Bradi4g40850	NA	NA	NA	NA	24	15	32
7	Bradi5g09610.1	Bradi5g09610	NA	NA	NA	NA	9	8	31
8	Bradi3g37150.3	Bradi3g37150	NA	NA	NA	NA	9	8	30
9	Bradi3g50220.1	Bradi3g50220	NA	NA	GO:0043565 GO:0006355 GO:0003700 GO:0003677	sequence-specific DNA binding; Interacting selectively and non-covalently with DNA of a specific nucleotide composition, e.g. GC-rich DNA binding, or with a specific sequence motif or type of DNA e.g. promotor binding or rDNA binding regulation of transcription, DNA-templated; Any process that modulates the frequency, rate or extent of cellular DNA-templated transcription DNA binding transcription factor activity; Interacting selectively and non-covalently with a specific DNA sequence (sometimes referred to as a motif) within the regulatory region of a gene in order to modulate transcription DNA binding; Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid)	24	8	32
10	Bradi1g12117.1 Bradi1g12117.2	Bradi1g12117	3.6.3.6	H+-exporting ATPase	GO:0046872 GO:0000166	metal ion binding; Interacting selectively and non-covalently with any metal ion nucleotide binding; Interacting selectively and non-covalently with a nucleotide, any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the ribose or deoxyribose	9	8	33
11	**Bradi1g37410.1**	**Bradi1g37410**	**NA**	**NA**	**GO:0009415** **GO:0006950**	**response to water stimulus; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus reflecting the presence, absence, or concentration of water.** **response to stress; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a disturbance in organismal or cellular homeostasis, usually, but not necessarily, exogenous (e.g. temperature, humidity, ionizing radiation).**	**9**	**15**	**32**
12	Bradi2g45050.1	Bradi2g45050	NA	NA	NA	NA	24	15	32
13	Bradi3g43577.3	Bradi3g43577	2.4.1.255	protein O-GlcNAc transferase	NA	NA	3	8	33
14	Bradi1g07441.1	Bradi1g07441	NA	NA	NA	NA	9	8	33
15	**Bradi3g43870.1**	**Bradi3g43870**	**NA**	**NA**	**GO:0009415** **GO:0006950**	**response to water stimulus; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus reflecting the presence, absence, or concentration of water** **response to stress; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a disturbance in organismal or cellular homeostasis, usually, but not necessarily, exogenous (e.g. temperature, humidity, ionizing radiation)**	**9**	**15**	**33**
16	Bradi2g33170.1	Bradi2g33170	NA	NA	GO:0009790	embryo development; The process whose specific outcome is the progression of an embryo from its formation until the end of its embryonic life stage. The end of the embryonic stage is organism-specific. For example, for plant vegetative embryos, this would be from the initial determination of the cell or group of cells to form an embryo until the point when the embryo becomes independent of the parent plant	24	15	33
17	Bradi3g00910.1	Bradi3g00910	3.2.1.26 2.4.1.99	beta-fructofuranosidase sucrose:sucrose fructosyl transferase	GO:0004575 GO:0004564	sucrose alpha-glucosidase activity; Catalysis of the reaction: sucrose + H2O = alpha-D-glucose + beta-D-fructose beta-fructofuranosidase activity; Catalysis of the reaction: a fructofuranosylated fructofuranosyl acceptor + H2O = a non fructofuranosylated fructofuranosyl acceptor + a beta-D-fructofuranoside	3	8	30

18	Bradi4g30380.6	Bradi4g30380	2.7.11.1	non-specific serine/threonine protein kinase	GO:0007165	signal transduction; The cellular process in which a signal is conveyed to trigger a change in the activity or state of a cell. Signal transduction begins with reception of a signal (e.g. a ligand binding to a receptor or receptor activation by a stimulus such as light), or for signal transduction in the absence of ligand, signal-withdrawal or the activity of a constitutively active receptor. Signal transduction ends with regulation of a downstream cellular process, e.g. regulation of transcription or regulation of a metabolic process. Signal transduction covers signaling from receptors located on the surface of the cell and signaling via molecules located within the cell. For signaling between cells, signal transduction is restricted to events at and within the receiving cell	24	0	32
19	Bradi1g51800.1	Bradi1g51800	NA	NA	GO:0009790	embryo development; The process whose specific outcome is the progression of an embryo from its formation until the end of its embryonic life stage. The end of the embryonic stage is organism-specific. For example, for plant vegetative embryos, this would be from the initial determination of the cell or group of cells to form an embryo until the point when the embryo becomes independent of the parent plant	9	4	31
20	Bradi3g37130.4 Bradi3g37130.5	Bradi3g37130	3.1.27.1	ribonuclease T2	GO:0033897 GO:0003723	ribonuclease T2 activity; Catalysis of the two-stage endonucleolytic cleavage to nucleoside 3'-phosphates and 3'-phosphooligonucleotides with 2',3'-cyclic phosphate intermediates RNA binding; Interacting selectively and non-covalently with an RNA molecule or a portion thereof.	9	8	33
21	Bradi4g08240.1	Bradi4g08240	NA	NA	NA	NA	9	9	24
22	Bradi2g56750.1	Bradi2g56750	2.7.11.1	non-specific serine/threonine protein kinase	GO:0005515 GO:0006468 GO:0005524 GO:0004672	protein binding; nteracting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules). protein phosphorylation; The process of introducing a phosphate group on to a protein ATP binding; Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator protein kinase activity; Catalysis of the phosphorylation of an amino acid residue in a protein, usually according to the reaction: a protein + ATP = a phosphoprotein + ADP	4	0	33
23	Bradi3g14970.1	Bradi3g14970	NA	NA	GO:0009790	embryo development; The process whose specific outcome is the progression of an embryo from its formation until the end of its embryonic life stage. The end of the embryonic stage is organism-specific. For example, for plant vegetative embryos, this would	24	4	30
24	Bradi2g25460.1 Bradi2g25460.2 Bradi2g25460.3	Bradi2g25460	NA	NA	GO:0055085 GO:0016021 GO:0016020 GO:0006810 GO:0005215	transmembrane transport; Transmembrane transport requires transport of a solute across a lipid bilayer. Note that transport through the nuclear pore complex is not transmembrane because the nuclear membrane is a double membrane and is not traversed. For transport through the nuclear pore, consider instead the term 'nucleocytoplasmic transport; GO:0006913' and its children. Note also that this term is not intended for use in annotating lateral movement within membranes integral component of membrane; The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane membrane; A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it transport; The directed movement of substances (such as macromolecules, small molecules, ions) or cellular components (such as complexes and organelles) into, out of or within a cell, or between cells, or within a multicellular organism by means of some agent such as a transporter, pore or motor protein transporter activity; Enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of or within a cell, or between cells	9	8	33
25	Bradi4g40870.2	Bradi4g40870	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)	GO:0016624 GO:0008152	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor; Catalysis of an oxidation-reduction (redox) reaction in which an aldehyde or ketone (oxo) group acts as a hydrogen or electron donor and reduces a disulfide metabolic process; The chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation	9	8	33
26	**Bradi3g51200.1**	**Bradi3g51200**	**NA**	**NA**	**GO:0009415** **GO:0006950**	**response to water; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus reflecting the presence, absence, or concentration of water** **response to stress; Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a disturbance in organismal or cellular homeostasis, usually, exogenous (e.g. temperature, humidity, ionizing radiation)**	**9**	**9**	**33**
27	Bradi1g19713.1	Bradi1g19713	3.1.1.1	carboxyl esterase	NA	NA	9	4	32

28	Bradi2g23507.2	Bradi2g23507	1.2.1.41 2.7.2.11	glutamate-5-semialdehyde dehydrogenase glutamate 5-kinase	GO:0055114 GO:0016491 GO:0008152	oxidation-reduction process; A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons oxidoreductase activity; Catalysis of an oxidation-reduction (redox) reaction, a reversible chemical reaction in which the oxidation state of an atom or atoms within a molecule is altered. One substrate acts as a hydrogen or electron donor and becomes oxidized, while the other acts as hydrogen or electron acceptor and becomes reduced metabolic process; The chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as DNA repair and replication, and protein synthesis and degradation	9	8	33
29	Bradi5g13047.2 Bradi5g13047.3 Bradi5g13047.4	Bradi5g13047	NA	NA	NA	NA	9	28	7
30	Bradi5g10450.1	Bradi5g10450	NA	NA	NA	NA	9	0	33
31	Bradi1g63816.1	Bradi1g63816	NA	NA	NA	NA	9	9	33
32	Bradi5g09000.1	Bradi5g09000	2.3.1.75	long-chain-alcohol O-fatty-acyltransferase	NA	NA	9	15	33
33	Bradi4g17200.1	Bradi4g17200	NA	NA	NA	NA	9	9	33
34	Bradi1g47570.1 Bradi1g47570.2 Bradi1g47570.3	Bradi1g47570	2.7.11.1	non-specific serine/threonine protein kinase	GO:0005515 GO:0006468 GO:0004672	protein binding; nteracting selectively and non-covalently with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules). protein phosphorylation; The process of introducing a phosphate group on to a protein protein kinase activity; Catalysis of the phosphorylation of an amino acid residue in a protein, usually according to the reaction: a protein + ATP = a phosphoprotein + ADP	9	8	33
35	Bradi1g10310.1	Bradi1g10310	NA	NA	NA	NA	9	15	32
36	Bradi2g33270.2	Bradi2g33270	NA	NA	GO:0046872	metal ion binding; Interacting selectively and non-covalently with any metal ion	3	8	33
37	Bradi2g51480.1	Bradi2g51480	NA	NA	GO:0015979 GO:0009523 GO:0009507	photosynthesis; The synthesis by organisms of organic chemical compounds, especially carbohydrates, from carbon dioxide (CO2) using energy obtained from light rather than from the oxidation of chemical compounds photosystem II; A photosystem that contains a pheophytin-quinone reaction center with associated accessory pigments and electron carriers. In cyanobacteria and chloroplasts, in the presence of light, PSII functions as a water-plastoquinone oxidoreductase, transferring electrons from water to plastoquinone, whereas other photosynthetic bacteria carry out anoxygenic photosynthesis and oxidize other compounds to re-reduce the photoreaction center chloroplast; A chlorophyll-containing plastid with thylakoids organized into grana and frets, or stroma thylakoids, and embedded in a stroma	4	14	31
38	Bradi3g36407.1	Bradi3g36407	NA	NA	NA	NA	9	4	30
39	Bradi5g08290.1	Bradi5g08290	1.1.1.146	11beta-hydroxysteroid dehydrogenase	NA	NA	9	0	32
40	Bradi2g17550.1	Bradi2g17550	NA	NA	NA	NA	8	8	31
41	Bradi1g34647.1	Bradi1g34647	2.4.1.255	protein O-GlcNAc transferase	GO:0016757	transferase activity, transferring glycosyl groups; Catalysis of the transfer of a glycosyl group from one compound (donor) to another (acceptor)	3	8	33
42	Bradi5g17170.2	Bradi5g17170	NA	NA	GO:0043565 GO:0006355 GO:0003700 GO:0003677	sequence-specific DNA binding; Interacting selectively and non-covalently with DNA of a specific nucleotide composition, e.g. GC-rich DNA binding, or with a specific sequence motif or type of DNA e.g. promotor binding or rDNA binding regulation of transcription, DNA-templated; Any process that modulates the frequency, rate or extent of cellular DNA-templated transcription DNA binding transcription factor activity; Interacting selectively and non-covalently with a specific DNA sequence (sometimes referred to as a motif) within the regulatory region of a gene in order to modulate transcription DNA binding; Any molecular function by which a gene product interacts selectively and non-covalently with DNA (deoxyribonucleic acid)	24	8	32
43	Bradi1g44480.1	Bradi1g44480	1.2.4.1	pyruvate dehydrogenase (acetyl-transferring)	GO:0016624 GO:0008152	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor; Catalysis of an oxidation-reduction (redox) reaction in which an aldehyde or ketone (oxo) group acts as a hydrogen or electron donor and reduces a disulfide metabolic process; The chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as	9	8	33
44	Bradi2g07480.1	Bradi2g07480	NA	NA	NA	NA	24	15	31

45	Bradi1g62957.1 Bradi1g62957.2 Bradi1g62957.3	Bradi1g62957	2.4.1.13	sucrose synthase	GO:0016157 GO:0005985	sucrose synthase activity; Catalysis of the reaction: UDP-glucose + D-fructose = UDP + sucrose sucrose metabolic process; The chemical reactions and pathways involving sucrose, the disaccharide fructofuranosyl-glucopyranoside	9	1	33
46	Bradi2g54810.2 Bradi2g54810.3 Bradi2g54810.4	Bradi2g54810	3.1.3.16	protein-serine/threonine phosphatase	GO:0003824 GO:0006470 GO:0004722	catalytic activity; Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzymes possess specific binding sites for substrates, and are usually composed wholly or largely of protein, but RNA that has catalytic activity (ribozyme) is often also regarded as enzymatic protein dephosphorylation; The process of removing one or more phosphoric residues from a protein protein serine/threonine phosphatase activity; Catalysis of the reaction: protein serine phosphate + H2O = protein serine + phosphate, and protein threonine phosphate + H2O = protein threonine + phosphate	24	8	33
47	Bradi4g38960.1 /Bradi4g38960.2	Bradi4g38960	3.6.1.3	adenosinetriphosphatase	GO:0016887 GO:0005524	ATPase activity; Catalysis of the reaction: ATP + H2O = ADP + phosphate + 2 H+. May or may not be coupled to another reaction ATP binding; Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator	9	1	31
48	Bradi2g41950.1	Bradi2g41950	3.1.3.16	protein-serine/threonine phosphatase	GO:0003824 GO:0006470 GO:0004722	catalytic activity; Catalysis of a biochemical reaction at physiological temperatures. In biologically catalyzed reactions, the reactants are known as substrates, and the catalysts are naturally occurring macromolecular substances known as enzymes. Enzyme protein dephosphorylation; The process of removing one or more phosphoric residues from a protein protein serine/threonine phosphatase activity; Catalysis of the reaction: protein serine phosphate + H2O = protein serine + phosphate, and protein threonine phosphate + H2O = protein threonine + phosphate	9	1	32
49	Bradi4g36370.1	Bradi4g36370	3.1.4.11 4.6.1.13	phosphoinositide phospholipase C phosphatidylinositol diacylglycerol-lyase	NA	NA	8	8	32
50	Bradi2g54920.1	Bradi2g54920	1.2.1.41 2.7.2.11	glutamate-5-semialdehyde dehydrogenase glutamate 5-kinase	GO:0055114 GO:0016491 GO:0008152	oxidation-reduction process; A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons oxidoreductase activity; Catalysis of an oxidation-reduction (redox) reaction, a reversible chemical reaction in which the oxidation state of an atom or atoms within a molecule is altered. One substrate acts as a hydrogen or electron donor and becomes oxi metabolic process; The chemical reactions and pathways, including anabolism and catabolism, by which living organisms transform chemical substances. Metabolic processes typically transform small molecules, but also include macromolecular processes such as	9	8	33

Supporting Figures

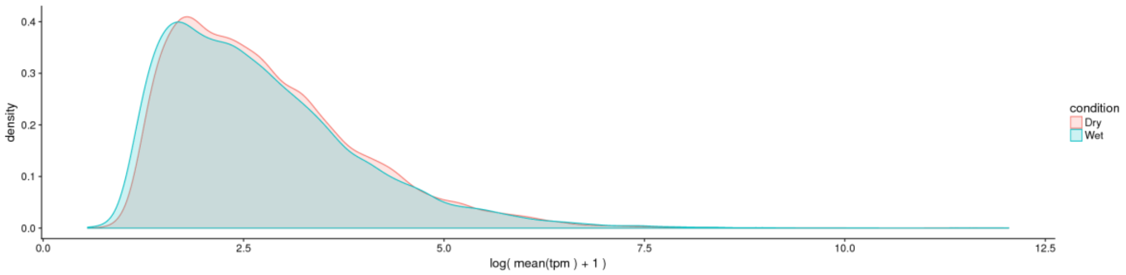


Figure S1. Density plots of filtered and normalized transcripts expression data in B. distachyon. Drought and water data are represented by orange and green projections, respectively.

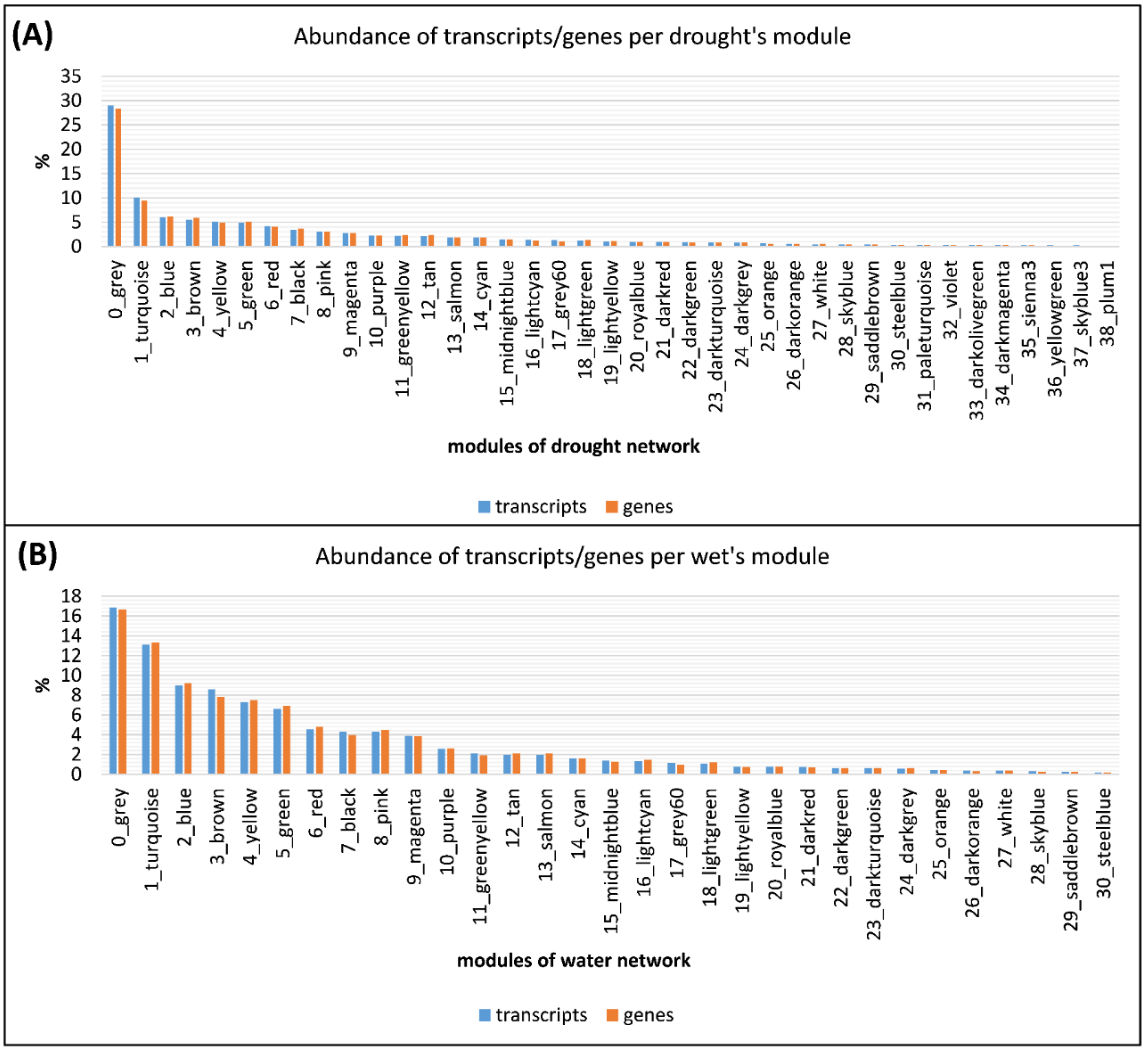


Figure S2. Histograms showing percentages of transcripts (blue) and genes (orange) found, respectively, in the 38 and 30 modules retrieved in the drought **(A)** and water **(B)** experiments of the studied *Brachypodium distachyon* accession.

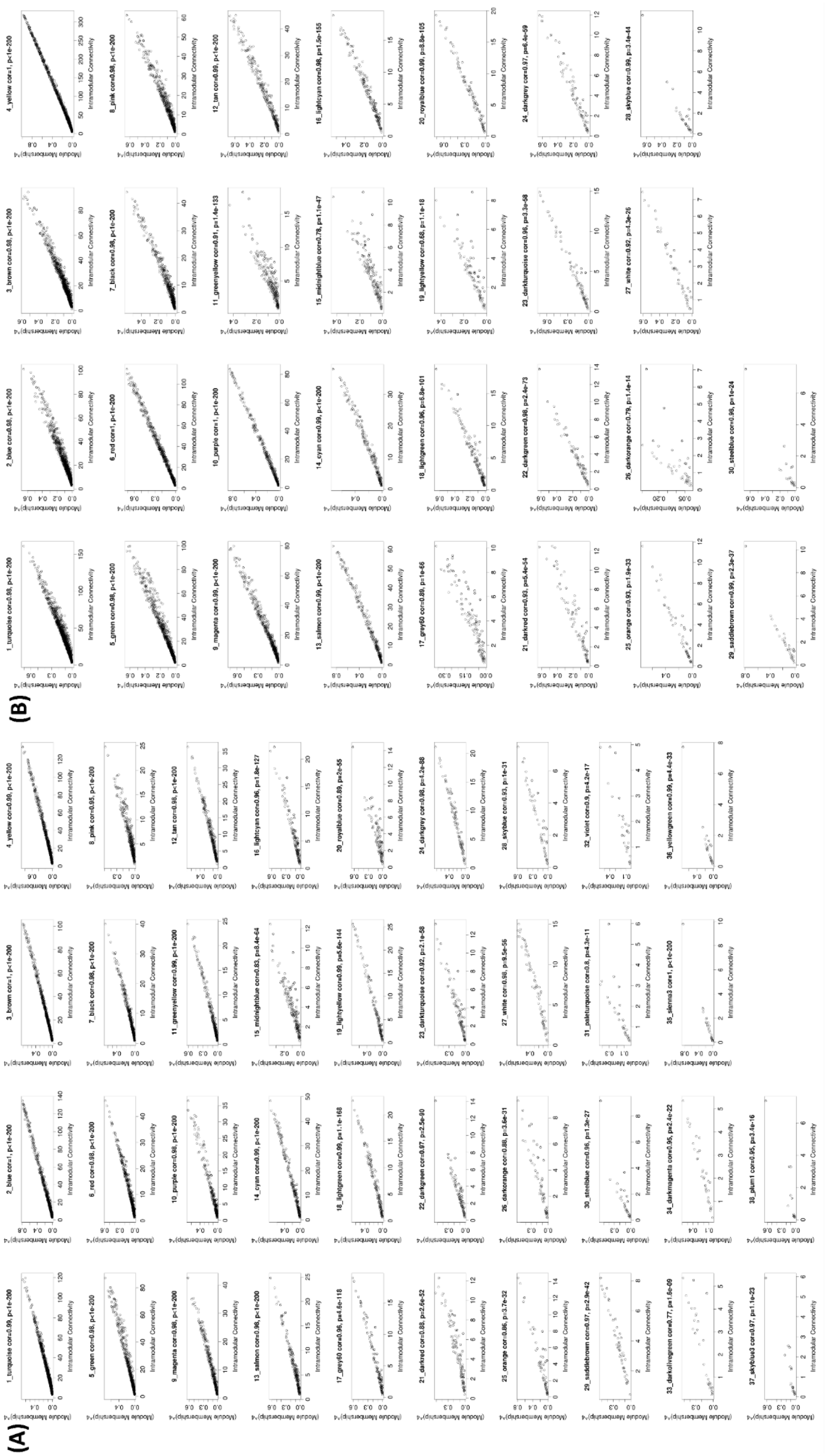


Figure S3. Plots of correlations between module membership (MM) and intra-modular connectivity for each module in the drought (**A**, 38 modules) and water (**B**, 30 modules) gene networks of the studied *Brachypodium distachyon* accessions.

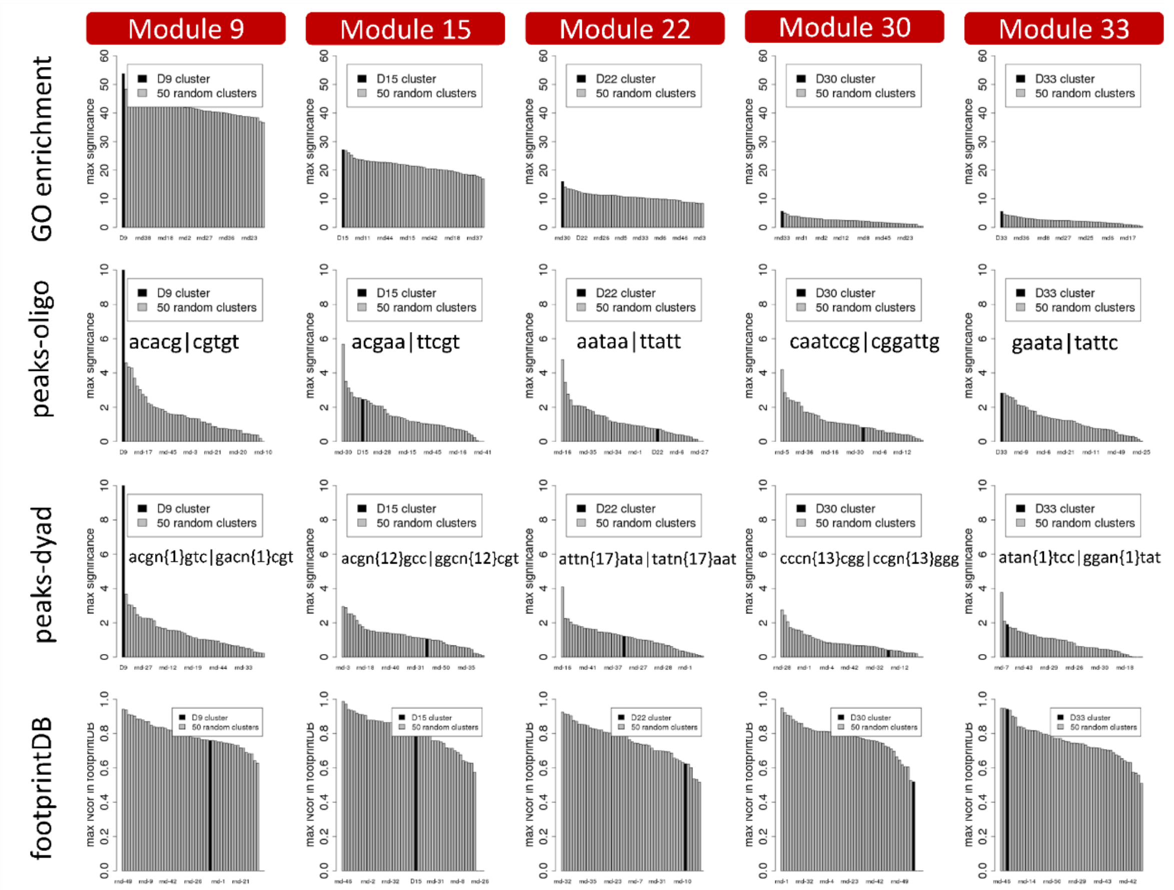


Figure S4. Discovered motifs in exclusive genes of drought network modules 9, 15, 22, 30 and 33 using 50 negative controls of equal size showing significance of target module (drought module) compared to random modules. GO enrichment, peaks-oligo, peaks-dyad and footprintDB analyses were used in the analysis.

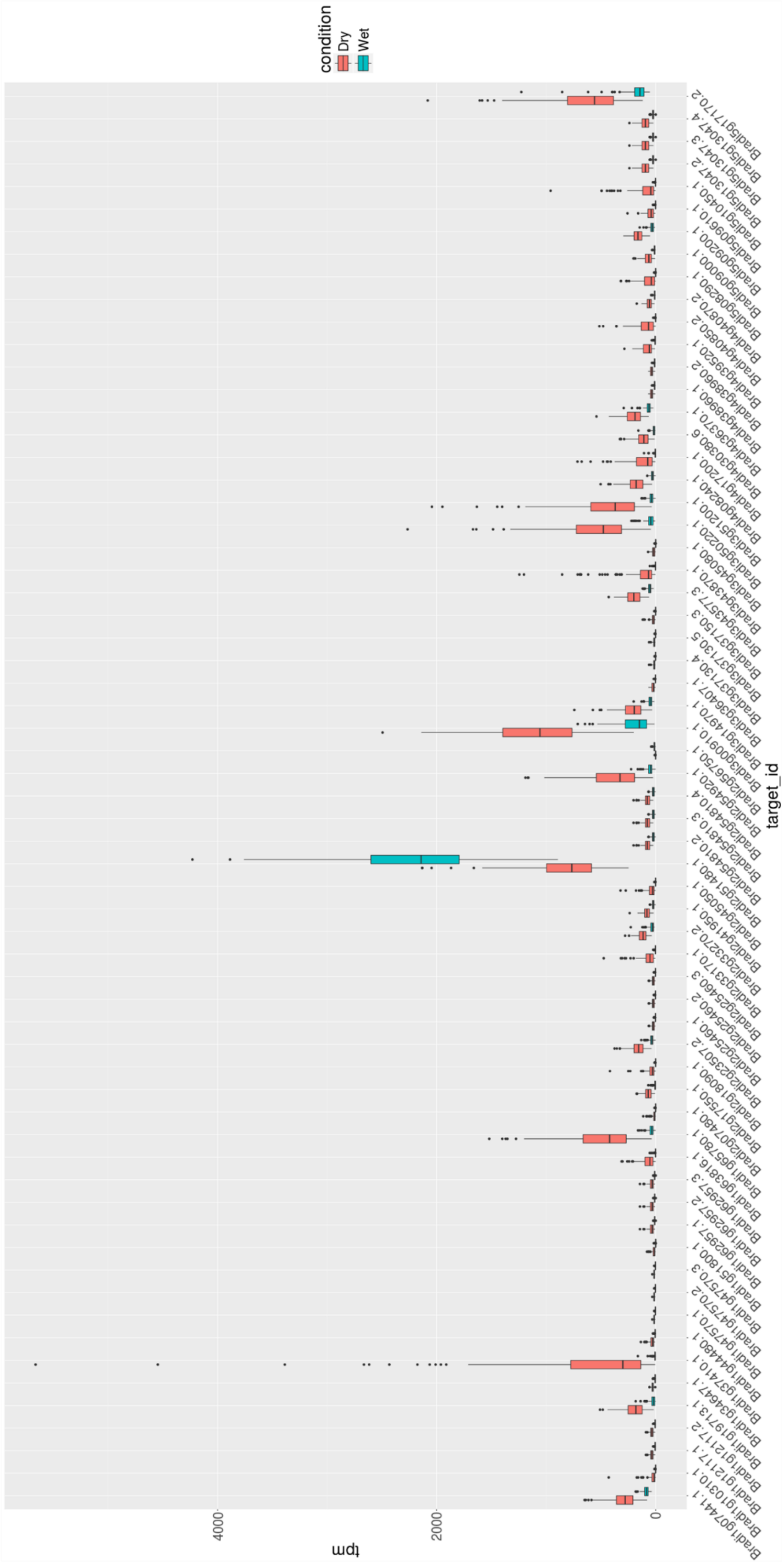


Figure S5. Boxplots of the top 50 most differentially expressed genes (DEs) in the drought (red) and water (blue) conditions. Target transcript identity (id) correspond to those of *B. distachyon* Bd21 v.3.1. TPM (Transcripts per million).

PUBLICATIONS OF THE PhD THESIS

Chapter 1: Article in press in Molecular, Phylogenetics and Evolution.

Reference: Antonio Díaz-Pérez, Diana López-Álvarez, **Rubén Sancho**, Pilar Catalán. (2018). Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches. DOI: 10.1016/j.ympev.2018.06.003.

Chapter 2: Article under final revisions by authors.

Title: Reference-genome syntenic mapping and multigene-based phylogenomics reveal the ancestry of homeologous subgenomes in grass *Brachypodium* allopolyploids

Authors: **Rubén Sancho**, Luis A. Inda, David L. Des Marais, Sean Gordon, John Vogel, Bruno Contreras-Moreira, Pilar Catalán

Chapter 3: Article published in New Phytologist.

Reference: **Sancho, R.**, Cantalapiedra, C.P., López-Alvarez, D., Gordon, S.P., Vogel, J.P., Catalán, P., & Contreras-Moreira, B. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium*: Flowering time signatures, introgression and recombination in recently diverged ecotypes. New Phytologist, 218, 1631–1644.

Chapter 4: Article under final revisions by authors.

Title: Co-expression network features and differentially expressed genes explain drought-response patterns in the model grass *Brachypodium distachyon*

Authors: **Rubén Sancho**, Pilar Catalán, Bruno Contreras-Moreira, David L. Des Marais

Other publications contributed to by the PhD thesis:

Reference: Catalán, P., López-Alvarez, D., Díaz-Pérez, A., **Sancho, R.**, & López-Herranz, M.L. (2016). Phylogeny and Evolution of the Genus *Brachypodium*. Genetics and genomics of *Brachypodium*. Plant Genetics and Genomics: Crops Models (ed. by J.P. Vogel), pp. 9–38. Springer.