

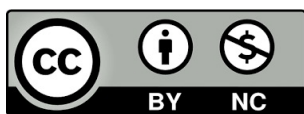
María Angeles Decena Rodríguez

"Evolution of allopolyploidy in  
species of the model genus  
***Brachypodium*** (Poaceae)  
using genomic, transcriptomic,  
cytogenetic and phenotypic  
approaches"

Director/es

Contreras Moreira, Bruno  
Hernández Molina, Pilar  
Catalán Rodríguez, María Pilar

<http://zaguan.unizar.es/collection/Tesis>



Universidad de Zaragoza  
Servicio de Publicaciones

ISSN 2254-7606

Tesis Doctoral

"EVOLUTION OF ALLOPOLYPLOIDY IN SPECIES  
OF THE MODEL GENUS **BRACHYPODIUM**  
(POACEAE) USING GENOMIC, TRANSCRIPTOMIC,  
CYTOGENETIC AND PHENOTYPIC APPROACHES"

Autor

María Angeles Decena Rodríguez

Director/es

Contreras Moreira, Bruno  
Hernández Molina, Pilar  
Catalán Rodríguez, María Pilar

**UNIVERSIDAD DE ZARAGOZA**  
**Escuela de Doctorado**

Programa de Doctorado en Ciencias Agrarias y del Medio Natural

2023





**Tesis doctoral**

**Doctorado internacional**

**"EVOLUTION OF ALLOPOLYPLOIDY IN SPECIES OF THE MODEL  
GENUS *BRACHYPODIUM* (POACEAE) USING GENOMIC,  
TRANSCRIPTOMIC, CYTOGENETIC AND PHENOTYPIC APPROACHES".**

Autora:

María de los Ángeles Decena Rodríguez

Directores:

Dra. Pilar Catalán Rodríguez

Dr. Bruno Contreras Moreira

Dra. Pilar Hernández Molina

Huesca, 31 de julio de 2023

Escuela Politécnica Superior de Huesca

Departamento de Ciencias Agrarias y del Medio Natural

Universidad de Zaragoza

**ESCUELA POLITÉCNICA SUPERIOR DE HUESCA**

**DEPARTAMENTO DE CIENCIAS AGRARIAS Y DEL MEDIO NATURAL**

**UNIVERSIDAD DE ZARAGOZA**

**EVOLUTION OF ALLOPOLYPLOIDY IN SPECIES OF THE  
MODEL GENUS *BRACHYPODIUM* (POACEAE) USING  
GENOMIC, TRANSCRIPTOMIC, CYTOGENETIC AND  
PHENOTYPIC APPROACHES.**

Memoria presentada por:

María de los Ángeles Decena Rodríguez

Directores de tesis:

Dra. Pilar Catalán Rodríguez

Dr. Bruno Contreras Moreira

Dra. Pilar Hernández Molina

Huesca, 31 de julio de 2023

La presente Tesis Doctoral se ha llevado a cabo bajo la dirección y tutela de la Dra. Pilar Catalán Rodríguez, y las codirecciones del Dr. Bruno Contreras Moreira y de la Dra. Pilar Hernández Molina en los laboratorios del grupo de investigación Bioflora de la Escuela Politécnica Superior de Huesca (Departamento de Ciencias Agrarias y del Medio Natural), Universidad de Zaragoza. Los análisis bioinformáticos se desarrollaron mediante tutorías virtuales y bajo la dirección del Dr. Bruno Contreras Moreira con los recursos informáticos del grupo de investigación Bioflora. La identificación y el filtrado *in silico* de genes codificadores de dehidrinas se realizaron en los laboratorios de la Escuela Técnica Superior de Ingeniería Informática de la Universidad de Málaga bajo la tutela de la Dra. Pilar Hernández Molina y la supervisión del Dr. Sergio Galvez. La introducción a los análisis citogenéticos de barcoding cromosómico y de la evolución del cariotipo en *Brachypodium* se llevaron a cabo durante la estancia de investigación virtual en la Universidad de Silesia en Katowice (Polonia) bajo las supervisiones del Prof. Robert Hasterok y la Dra. Joanna Lusinska. El resto de los análisis de la tesis, morfoanatómicos, taxonómicos, citogenéticos, bioinformáticos, genómicos y filogenómicos, fueron llevados a cabo en la Escuela Politécnica Superior de Huesca.

Esta tesis Doctoral ha sido financiada por un Contrato predoctoral para la formación de investigadores (BES-2017-080228) del Ministerio de Ciencia e Innovación de España a la doctoranda María de los Ángeles Decena Rodríguez. Los estudios de la tesis doctoral se han desarrollado dentro del marco de los proyectos de investigación CGL2016-79790-P (“Evolución de caracteres biológicos y procesos de especiación en el género modelo *Brachypodium* (Poaceae) mediante análisis de genómica comparada y funcional”) y PID2019-108195GB-I00 (“Genómica evolutiva integrada del sistema modelo poliploide *Brachypodium*: bases del éxito de la poliploidía en las angiospermas”) financiados por el Ministerio de Ciencia e Innovación de España.

*Porque el amor supremo es entregarse entero,  
porque el amor más noble no pone condición*

A mi familia.

A mi hermana Esperanza.

A mis padres Manuel y M<sup>a</sup> Ángeles.

## INDEX

<b>AGRADECIMIENTOS (Acknowledgements)</b> .....	<b>12</b>
<b>Summary</b> .....	<b>15</b>
<b>Resumen</b> .....	<b>17</b>
<b>Introduction</b> .....	<b>19</b>
Historical and evolutionary context of the grass model genus <i>Brachypodium</i> .....	19
<i>Brachypodium</i> taxa as model organisms for monocots.....	21
The large genomes of cereals, which are rich in repetitive DNA, the domestication syndrome, and the complex evolutionary history of the Poaceae were challenging before <i>Brachypodium</i> emerged as a model organism.....	22
Systematics and phylogenetics of <i>Brachypodium</i> .....	23
The need of new genomic and evolutionary studies of the perennial <i>Brachypodium</i> species.....	25
References.....	28
<b>OBJETIVOS</b> .....	<b>36</b>
<b>OBJECTIVES</b> .....	<b>37</b>
<b>ESTRUCTURA DE LA TESIS</b> .....	<b>38</b>
<b>PHD STRUCTURE</b> .....	<b>39</b>
<b>Chapter 1: Phylogenetics of the Palearctic model grass <i>Brachypodium sylvaticum</i> uncovers two divergent oriental and occidental micro-taxa lineages</b> .....	<b>40</b>
Material and Methods.....	43
Sampling and taxonomic identification.....	43
Genomic data and phylogenetic analysis .....	44
Results.....	48
Plastome and nuclear rDNA 35S phylogenies.....	48
Bayesian dated tree.....	50
Discussion.....	50
Literature cited.....	53
<b>Chapter 2: Cyto-phylogenomics of the <i>Brachypodium</i> grass species-complex unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends</b> .....	<b>55</b>
Introduction.....	57
Results .....	62
Cytogenetic study of <i>Brachypodium</i> .....	62
<i>Brachypodium</i> plastome analysis .....	64
<i>Brachypodium</i> nuclear rDNA 35S and 5S analysis .....	70
Supertree reconstruction and dating analysis of the <i>Brachypodium</i> phylogeny .....	75
Discussion.....	80

Untangling the genomic diversity and evolution of cryptic perennial diploid <i>Brachypodium</i> taxa and cytotypes .....	80
Uncovering the origins of arcane perennial <i>Brachypodium</i> polyploid taxa and cytotypes.....	83
Material and Methods.....	88
Sampling .....	88
Cytogenetic and karyotypic study .....	88
DNA isolation, genome skimming, plastome and rDNA 35S and 5S sequence assembling and multiple sequence alignments.....	89
Phylogenomics, supertree computation, and dating analyses.....	90
References:.....	93
<b>Chapter 3: Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in grass <i>Brachypodium</i> polyploids.....</b>	<b>101</b>
Introduction.....	102
Methods .....	105
Sampling, ploidy levels and genome skimming sequencing .....	105
Correlations of repeat amounts and genome size variation in <i>Brachypodium</i> .....	110
Landscape genomic diversity analysis of repeat types in <i>Brachypodium</i> .....	111
Repeatome phylogenomic network of <i>Brachypodium</i> .....	111
Results .....	112
Characterization and quantification of the <i>Brachypodium</i> repeatome .....	112
Global variability and genome landscape of the <i>Brachypodium</i> repeatome.....	120
The <i>Brachypodium</i> repeatome phylogenetic network and 5S rDNA graph-clusters.....	124
Discussion.....	131
Delineation of the <i>Brachypodium</i> repeatome and its impact on the striking genome size diversification of its lineages.....	131
Alternative evolutionary responses to the polyploidy genome shock hypothesis by different <i>Brachypodium</i> allopolyploids.....	135
Repeatome-based phylogenomics of <i>Brachypodium</i> and concordance between rDNA 5S graphs and subgenomes.....	136
References.....	140
<b>Chapter 4: Morphoanatomical analysis of cryptic and evolutionary divergent polyploid cytotypes of model Mediterranean <i>Brachypodium retusum</i> and <i>B. phoenicoides</i> grasses .....</b>	<b>145</b>
Introduction.....	146
Materials and Methods .....	149
Sampling.....	149
Phenotypic analysis .....	150
Statistical analyses.....	151
Results .....	152
Morphological differentiation within <i>Brachypodium retusum</i> and <i>B. phoenicoides</i> tetraploid and hexaploid cytotypes .....	152

Multivariate PCA and FAMD analyses of <i>Brachypodium retusum</i> and <i>B. phoenicoides</i> samples.	159
Discussion.....	166
Phenotypic variation within the <i>Brachypodium retusum</i> and <i>B. phoenicoides</i> polyploid complexes is partially concordant with cytotypic, geographical, environmental and phyletic variation .....	166
Gene flow and taxonomic differentiation of cryptic taxa within the <i>Brachypodium retusum</i> and <i>B. phoenicoides</i> polyploid complexes .....	169
References.....	172
<b>Chapter 5: Comparative Genomics, Evolution, and Drought-Induced Expression of Dehydrin Genes in Model <i>Brachypodium</i> Grasses.....</b>	<b>157</b>
Introduction.....	158
Results .....	162
Dehydrin Genes of <i>Brachypodium</i> Species and Outgroup Grasses .....	162
Cis-Regulatory Elements of Bdhn Genes.....	164
The <i>Brachypodium</i> Dehydrin Tree.....	166
Chromosome Distributions and Selection Analysis of Duplicated Bdhn Genes .....	166
Dehydrin Gene Clusters, Phylogenetics, and Climate Niche Variation in <i>B. distachyon</i> .....	166
Differential Expression of Bdhn Genes in <i>Brachypodium distachyon</i> Ecotypes under Drought and Temperature Stress Conditions.....	166
Effects of Drought on Dehydrin Gene Expression and Drought-Response Phenotypic Traits .....	170
Phylogenetic Signal of Dehydrin Expression, Phenotypic Trait Changes and Climate Variation in the <i>Brachypodium distachyon</i> Bdhn Tree.....	172
Discussion.....	172
The Dehydrin Gene Family in <i>Brachypodium</i> .....	172
Dehydrin Expression Induction in <i>Brachypodium distachyon</i> .....	173
Correlated Dehydrin and Phenotypic Drought Response and Phylogenetic Signal in <i>Brachypodium distachyon</i> .....	175
Material and Methods .....	176
Identification of Dehydrin Sequences .....	176
Structural Analysis, Conserved Motifs, and Cis-Regulatory Elements (CREs) .....	176
Multiple Alignments and Phylogenetic Analysis .....	177
Chromosomal Location, Gene Duplication, and Selection Analysis .....	177
Expression Analysis of Dehydrin Genes in <i>Brachypodium distachyon</i> .....	178
Drought-Induced Changes in Dehydrin Expressions, Phenotypic, Physiological, and Climatic Niche Traits, and Phylogenetic Signal in <i>Brachypodium distachyon</i> .....	179
Conclusions.....	180
References.....	183
<b>CONCLUSIONES.....</b>	<b>187</b>
<b>CONCLUSIONS .....</b>	<b>188</b>
<b>Supplementary data Supplementary figures and tables.....</b>	<b>190</b>

Supplementary figures and tables Chapter 2.....	190
Supplementary figures and tables Chapter 3.....	237
Supplementary figures and tables Chapter 4 .....	249
Supplementary figures and tables Chapter 5.....	268
<b>Appendices.....</b>	<b>288</b>
<b>Publications of the PhD thesis.....</b>	<b>289</b>



## AGRADECIMIENTOS (ACKNOWLEDGEMENTS)

Gracias, del latín *gratias*, nos llega desde la expresión latina “*gratia agere*”, “*dar las gracias*” y que ya en su raíz indoeuropea recoge el significado de “alabar en voz alta”. Que mejor ocasión, que la de dar comienzo a esta tesis, para elevar la voz y agradecer cuánto y a cuantos, en estos casi seis años, me han enseñado, acompañado y apoyado, en mi desarrollo como investigadora y persona a la vez.

En primer lugar, agradecer a Pilar Catalán, mi directora y tutora de tesis, por enviar ese correo que creí que nunca llegaría y, porque desde mi incorporación al grupo Bioflora, ha contribuido en todos los aspectos de mi formación. Desde la enseñanza de la taxonomía clásica a los más innovadores análisis y enfoques de la botánica, en general, y del género *Brachypodium* en particular. Agradecerle la guía, las inestimadas revisiones de resultados, análisis, interpretaciones, y manuscritos que hoy componen los capítulos de esta tesis.

A mi director, Bruno Contreras, con quien he desarrollado toda una nueva faceta como bióloga al formarme en el campo de la bioinformática. Agradecerle las horas al otro lado de la pantalla programando, las reuniones siempre disponibles, el apoyo y la guía en todo momento.

A mi directora, Pilar Hernández, por su contribución en el estudio de genes asociados a estrés hídricos en mis días en el IAS-CSIC, y al Dr. Sergio Gálvez, con quien desarrollé gran parte del estudio de genes dehidrinas en *Brachypodium* durante mi estancia en la ETSII de la Universidad de Málaga.

Al profesor Robert Hasterok y a la Dra. Joanna Lusinka, con quienes he tenido oportunidad de ampliar mis conocimientos en citogenética de las especies aloploidoides de *Brachypodium*, tanto durante mi estancia virtual en la Universidad de Silesia en Katowice como en las estancias de Joanna en la EPS de Huesca.

A los doctores Luis Ángel Inda, por enseñarme sobre la citometría de flujo y su colaboración en estudios cariológicos y morfoanatómicos que forman parte de esta tesis, y a Ernesto Pérez, con quien empecé extrayendo ADN y acabé hablando de escalada. Gracias a ambos por todo el apoyo, las risas y el cariño derrochado. A Antonio Díaz, quien me inició en las técnicas de *genomme skimming* y en el ensamblaje de plastomas y me contagió de su amor por la estadística, la docencia y la etimología.

A la Dra. Diana López por aceptar ser parte del tribunal de esta tesis, así como de revisar la misma. Y, por este y por muchos otros motivos, gracias al Dr. Juan Viruel, por darme ese valor

que necesitaba cuando nos conocimos, por guiarme antes de llegar a la EPS, y por ser siempre un ejemplo a seguir.

Gracias, gracias desde lo más profundo a Gemma Sausán, por la infinita paciencia y por hacer de la burocracia algo amable.

Como a Gemma, agradecer a todo el personal de administración, conserjería, dirección, docentes y técnicos, servicios de limpieza y seguridad de la EPS, porque ellos hacen de esta Escuela un hogar.

Nada de esto habría sido igual sin el apoyo del resto de componentes del grupo Bioflora. Gracias a mi compañera de todo, María Fernanda Moreno, por lo aprendido contigo y de ti, por ser mi trébol de cuatro hojas. A Miguel Campos, quien llegó para quedarse, siendo apoyo y revolución a la vez. A Rubén Sancho, por ser fuente de esperanza, de guía y por su implicación en varios trabajos de esta tesis. A las últimas incorporaciones, Alba, Samira, y Yana, que han traído fuerzas renovadas. A cuantos nos han visitado en nuestros laboratorios, Joanna, Evgeny, Diana y María.

A mis compañeros y amigos de los otros laboratorios de la EPS, Gianlu, Daniel, David, Lilia, las nuevas incorporaciones de Álvaro y Silvia; y, como no, a mis OpenHuerto y compañía, sin quienes el camino habría sido más aburrido. A Daniel y su entusiasmo, la dulzura de María Videgain, la música de Darío y a las muchas horas de escucha de Elisa aquí y allá, a Marta, Andoni, Julia, Lucía, Adrián, Diego, y muy especialmente a mis mejores compañeros de viaje, Christian y Miguel por los kilómetros que aún nos faltan por vivir.

A todas las profesoras que por el camino me encontré, a las hermanas de la Compasión y quienes forman parte de esta familia. A mis profesoras a lo largo de mi formación básica y muy especialmente, durante mis años de carrera, a la doctora Zoila Díaz, quien sin conocerme me abrió las puertas de su laboratorio y de sus narcisos, dándome la oportunidad de descubrir otro lado de la botánica, que me ha enseñado y acompañado durante esos años y todos los que han venido después. Gracias a los mejores biólogos que encontré en mi camino y siguen en él, a mis Salchipapas, Débora, Jesús, Pili, David y Laura. Junto a vosotros descubrí lo lejos que podía llegar la amistad.

Quienes me conocen saben que tengo vértigo, pero aquí en Huesca he aprendido a escalar. Y si venzo a ese miedo es porque sé que, bajo mis pies, se extiende la red más segura, el arnés más resistente, mi familia. Mi familia *sensu lato*, desde María, con quien empiezan a acumularse los recuerdos, M<sup>a</sup> José, mi mejor amiga, mi amiga del alma y toda la paciencia del ser humano concentrada; mi abuela María que me regaló mi primera gerbera, mis tíos que siempre me miran como quien mira algo importante, a mis primos, a Carlos y su música, a la voz serena de Antonio, el abrazo que se siente en el corazón de Inda y a Javi, quien no solo es mi más leal compañero, sino que también ha estado implicado en el proceso de desarrollo de varios scripts de esta tesis, sin importar tiempo u horarios, ha estado a mi lado enseñándome y apoyándome siempre.

A todos ellos porque han comprendido y perdonado mis ausencias, han apoyado aun sin terminar de entender qué es lo que hago.

A mi familia *sensu stricto*, a mi hermana Esperanza, que es mi ancla y mis alas al mismo tiempo y a mis padres, que nunca se cansan de responder a las mil y una preguntas de la niña que se asoma al balcón a preguntar, por enésima vez, por aquel árbol de enfrente, porque sin vosotros nada de esto tendría sentido.

## SUMMARY

The *Brachypodium* taxa fit the definition of model organisms and, for this reason, since the publication of the *B. distachyon* reference genome, they have become workhorses for biological research, specifically for the biology and evolution of grasses (family Poaceae). The recent development of massive genome sequencing techniques has become a crucial tool for research on evolution, systematics, and ecology of plants and for the generation of new model systems for temperate grass breeding.

The main evolutionary mechanism of speciation in angiosperm, which are thought to be the descendants of paleopolyploids, involved successive rounds of hybridization and whole genome duplication (WGD) (allopolyploidization) and subsequent diploidizations. *Brachypodium* provides an exceptional environment for studying genome duplication and allopolyploidy as well as the effects of these occurrences. Perenniality is also an important topic in plant biology, and some research aims to understand the mechanisms underlying the perennial-annual transition and the metabolic changes that cause it. With the exception of *B. mexicanum* and *B. boissieri*, all perennial members of the genus *Brachypodium* belong to the recently evolved core perennial clade. *Brachypodium* exhibits two perennial-to-annual transitions across its ancestral lineages which later experienced hybridization and WGD, generating a new allopolyploid annual species. As a result of it, the genus contains three annual and about 17 perennial species.

Despite all the previous advances in the study of this model genus, precise evolutionary, genomic, and taxonomic studies for the perennial species, are still lacking for *Brachypodium*, with a focus on micro-taxa and cryptic species. The research done in this PhD thesis has untapped the largest and most updated phylogeny of the genus in its broadests sense (20 species, 23 cytotypes) demonstrating also the need for further investigation of the perennial taxa.

Our cytogenetic and phylogenetic studies have revealed the presence of new hexaploid cytotypes for the allopolyploids *B. retusum* and *B. phoenicoides* and the contribution of different ancestral and intermediate evolved subgenomes in these high allopolyploids. The morphological variation within these cytotypes revealed overall significant smaller sizes of vegetative and reproductive traits for tetraploids than for hexaploids in both taxa. The discovery of two lineages within the *B. sylvaticum* *s.l.* complex, considered to be a monophyletic and homogeneous lineage so far, has established the distinction of two divergent oriental and occidental clades, and the close relationships of the former with African and S Asian endemics. Our *Brachypodium*

repeatome analysis has demonstrated the impact of the repetitive elements on the genome sizes of its species and in the genome expansions and contractions experienced by some polyploid lineages, and its phylogenetic value. This study has also corroborated the use of the 5S gene graph clusters to characterize the subgenomes of polyploid cytotypes. The comparative genomic and phylogenomic study of protein coding genes involved in the drought stress response (dehydrins) uncovered the existence of ten genes (Bdhn) in the annual *Brachypodium* species and in the perennial model species *B. sylvaticum*, which are under purifying selection, and the differential expression of four of them across the 54 *B. distachyon* ecotypes studied, supporting the potential role of *Brachypodium* as model for drought stress response mechanism within the Poaceae.

## RESUMEN

Los taxones de *Brachypodium* encajan en la definición de organismo modelo y, por esta razón, desde la publicación del genoma de referencia de *B. distachyon*, se han convertido en caballos de batalla para la investigación biológica, específicamente para la biología y la evolución de las gramíneas (familia Poaceae). Las técnicas de secuenciación masiva del genoma recientemente desarrolladas son ahora herramientas cruciales para la investigación sobre la evolución, la sistemática y la ecología de las plantas y para la generación de nuevos sistemas modelo para la mejora de pastos templados.

El principal mecanismo evolutivo de la especiación en las angiospermas, que se cree que son descendientes de los paleopoliploides, involucró rondas sucesivas de hibridación y duplicación del genoma completo (WGD) (alopoliploidización) y diploidizaciones posteriores. *Brachypodium* proporciona un entorno excepcional para estudiar la duplicación del genoma y la aopoliploidía, así como los efectos de estos sucesos. La perennidad es otro tema importante en la investigación en biología vegetal, y su objetivo es comprender los mecanismos que subyacen a la transición perenne-anual y los cambios metabólicos que la provocan. Con la excepción de *B. mexicanum* y *B. boissieri*, todos los miembros perennes del género *Brachypodium* pertenecen al clado "core perennials" recientemente evolucionado. *Brachypodium* exhibe dos transiciones perenne-anual en sus linajes ancestrales, seguidos de hibridación y WGD lo que originó una tercera especie aloploiploide anual. Como resultado, el género contiene tres especies anuales y unas 17 perennes.

A pesar de los avances en el estudio de este género modelo, todavía son necesarios estudios evolutivos, genómicos y taxonómicos más precisos para las especies perennes de *Brachypodium*, con un enfoque en micro-taxones y especies crípticas. La investigación de esta tesis doctoral ha computado la filogenia más grande y actualizada del género (20 especies, 23 citotipos), y ha puesto de manifiesto la necesidad de una mayor investigación de los taxones perennes.

Nuestros estudios citogenéticos y filogenéticos han revelado la presencia de nuevos citotipos hexaploides para los aloploiploides *B. retusum* y *B. phoenicoides* y la contribución de diferentes subgenomas ancestrales y de evolución intermedia en estos aloploiploides de elevado nivel de ploidía. La variación morfológica en estos citotipos reveló tamaños significativamente más pequeños en general, tanto para caracteres vegetativos como reproductivos, para los tetraploides en comparación con los hexaploides en ambos taxones. El descubrimiento de dos

linajes dentro del complejo *B. sylvaticum* s.l., considerado hasta ahora un linaje monofilético y homogéneo, ha establecido la distinción de dos clados divergentes oriental y occidental, y las estrechas relaciones del primero con especies endémicas africanas y sudasiáticas. Nuestro análisis del repeteoma de *Brachypodium* ha demostrado el impacto de los elementos repetitivos en el tamaño del genoma de las especies y en las expansiones y contracciones del genoma experimentadas por algunos linajes poliploides, y su valor filogenético, así como el uso de los grupos de gráficos de genes 5S para caracterizar los subgenomas de citotipos poliploides. El estudio genómico y filogenómico comparado de los genes codificadores de proteínas implicadas en la respuesta al estrés por sequía (dehidrinas) demostró la presencia de diez genes específicos (Bdhn) en las especies anuales de *Brachypodium* y la especie perenne modelo *B. sylvaticum*, que se hallan bajo presión selectiva y muestran distintas divergencias. La expresión de cuatro de ellos en los 54 ecotipos de *B. distachyon* estudiados y su sobreexpresión en condiciones de sequía, en especial en ecotipos más tolerantes a este estrés demuestra el papel potencial de *Brachypodium* como grupo modelo para analizar los mecanismos de respuesta al estrés por sequía dentro de las Poaceae.

## INTRODUCTION

### Historical and evolutionary context of the grass model genus *Brachypodium*

Currently, the genomes of staple cereals that feed the world, like rice, maize, sorghum, barley and wheat, have been sequenced and annotated and are used to improve their varieties (IWGSC et al., 2018; Kumar et al., 2009; Mascher et al., 2017). However, the genomes of other grass crops and pascicolous species (e. g., *Avena*, *Agrostis*, *Festuca*, *Lolium*, *Poa*) have not been sequenced yet or only partially (Bolger et al., 2014; Fox et al., 2014; Mayer KF et al., 2012; Pfeifer et al., 2013) due to the large size and complexity of their genomes. The future improvement of this key group of temperate grasses is limited, to some extent, by the absence of useful genome resources (Febrer et al., 2010), and by the difficulty of completing the sequence of their chromosomes. The use of a model system group of plants evolutionarily close to them is of great interest, since genomic and transcriptomic characteristics can be transferred from the model plants to these cereals and temperate pasture species, accelerating the process of discovery and expression of new genes and regulatory sequences and the development of bioinformatics tools that are essential for the success of a functional genomics program (Scholthof et al., 2018; Vogel, 2016).

In this context, the genus *Brachypodium* has received considerable attention since its annual species *Brachypodium distachyon* was selected as a functional model of temperate cereals and biofuel grasses (IBI et al., 2010; Mur et al., 2011; Scholthof et al., 2018), due to its peculiar biological and genomic characteristics (annual plant, with small and compact genome) (Draper et al., 2001; IBI et al., 2010; Vogel, 2016). This species and others in the genus are evolutionarily closer to temperate climate cereals (*Triticeae*, *Aveneae*) and fodder grasses (*Poeae s.l.*, *Bromeae*) than *Oryza sativa*, the subtropical model grass (Mur et al., 2011; Opanowicz et al., 2008), thus facilitating the development of comparative genomic analyses of *Brachypodium* species with wheat, barley and other temperate pasture and forage species (Catalan et al., 2014; Hasterok et al., 2022; IBI et al., 2010).

Molecular phylogenetic studies have shown that *Brachypodium*, the only genus of the monotypic tribe *Brachypodieae*, constitutes one of the divergent intermediate lineages of the Pooideae (Catalán et al., 1997; Schneider et al., 2011). *Brachypodium* diverged from its ancestor before the radiation of the clade of the most recently evolved pooideae, the so-called “core pooids” group that harbours temperate cereals and forage grasses of high ecological and economic value (Catalán et al., 1997; Catalan et al., 2016). The tribal arrangement of the



Pooideae has varied widely over the last centuries. The most recent classification indicated that twelve subtribes belong to the Poeae-type plastid DNA clade and seven tribes to the Aveneae-type plastid DNA clade (Soreng et al., 2015), and all of them have been classified within supertribe Poodae. Different studies focused on some particular subtribes such as the Airinae, Loliinae, Poinae and Aveninae and suggested that further changes to the taxonomy of the supertribe Poodae would be necessary (Pimentel et al., 2017). A supertribe Triticodae was also proposed including three tribes, Bromeae, Triticeae (encompassing subtribes Triticinae and Hordeinae) and Littledaleae (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, 2015).

According to the most recent molecular time divergence studies, the *Pooideae* stem and crown splits dated to 70.65 to 62.07 Ma. The divergence of the BOP and PACMAD clades has been estimated to occur during the Maastrichtian age (~70.65 Ma) and before that a whole genome duplication (WGD) was experienced by the protograin ancestor (90-70 Ma). However, the evolution of widespread grasses in the Miocene (25-15 Ma) was based on fossil grass evidence and paleofaunal studies (Jacobs et al., 1999; Strömberg, 2011). Interestingly, this increase in diversification of the pooids occurred before the divergence and diversification of the ungulate families Bovidae and Cervidae in moist Eurasian regions, which took place in the Late Oligocene (Bouchenak-Khelladi et al., 2009; Matthee & Davis, 2001). The diversification of the Pooideae during the Oligocene continued during the Miocene and the Pliocene (Pimentel et al., 2017) and developed into primary temperate grasslands in both hemispheres (Bouchenak-Khelladi et al., 2009; Edwards et al., 2010; Strömberg, 2011). 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  $x=11$ ; Lygeae  $x=10$ ; Nardeae  $x=13$ ; Phaenospermatae  $x=12$ ; Meliceae  $x=10, 9, 8$ ; Stipeae  $x=12, 11, 10$ ; Diarrheneae  $x=10$ ), through the intermediate ones of Brachypodieae ( $x=10, 9, 8, 5$ ) (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; Luo et al., 2009; Salse et al., 2008), although  $x=6, 5, 4, 2$  occasionally occur in Aveneae (Poodae) (Catalán et al., 2016).

*Brachypodium* is a relatively small genus that contains ca. 20 species distributed worldwide (Catalán et al., 2012, 2016; Catalan & Olmstead, 2000; Schippmann, 1991).

According to the most recent taxonomic updating (Catalán et al., 2012; Díaz-Pérez et al., 2018; Scholthof et al., 2018), three of them are annual species and 17 are perennial taxa. It has been recently demonstrated that the three annuals have a large distribution in their native circumMediterranean region (*B. distachyon*, *B. stacei*, *B. hybridum*) (Catalán et al., 2012; López-Alvarez et al., 2012, 2015, 2017). Among the perennials, few species show a large native Eurasian (*B. sylvaticum*, *B. pinnatum*, *B. rupestre*) or Mediterranean (*B. retusum*) distribution, whereas the rest have a restrict disjunct distributions in their respective native ranges [W Mediterranean (*B. phoenicoides*), C Mediterranean (*B. genuense*), E Mediterranean – SW Asia (*B. glaucovirens*), S Spain (*B. boissieri*), Canary isles (*B. arbuscula*), South Africa (*B. bolussi*), tropical and South Africa (*B. flexum*), Madagascar (*B. madagascariense*), Taiwan (*B. kawakamii*), SE Asia – New Guinea (*B. sylvaticum* var. *pseudodistachyon*), and America (*B. mexicanum*)] (Catalán et al., 2016; Schippmann, 1991). Since 1812 two segregated genera were erected, *Trachynia* Link, to cover the annual species, and *Brevipodium* Lovë & Lovë, to accommodate *B. sylvaticum*; however, in almost all modern works neither of these two segregates were recognized (Catalán et al., 1995), and all the newly described species have been subsumed within *Brachypodium* (Catalán et al., 2016; Schippmann, 1991). *Brachypodium* shows a remarkable descendant dysploidy trend from ancestral  $x=10$  karyotypes to intermediately-to-recently evolved  $x=9$ ,  $x=8$  and  $x=5$  karyotypes (Lusinska et al., 2019; Sancho et al., 2022a; Decena et al., 2023). Phylogenetic subgenome detection algorithms applied to transcriptome data and karyotype barcoding analysis further identified seven diploid subgenomes in the studied *Brachypodium* polyploids, three of them present in current diploid progenitor species and four orphan (only detected in polyploid species) (Sancho et al., 2022; Decena et al., 2023).

### ***Brachypodium* taxa as model organisms for monocots**

Early in the last XX century, advances in biology required the use of model or experimental organisms as a manner of simplifying the study object and making possible the transference to more complex systems. An experimental organism is related to a non-model organism, with enough economic importance to become a study object use to solve a specific question (Leonelli & Ankeny, 2013; Scholthof et al., 2018). Due to inherent domesticated-crop genetic erosion, lengthy seed-to-seed life cycles, and the requirement for substantial growth facilities, these plants have limits (Scholthof et al., 2018). Nonetheless, model organisms are chosen considering their tractability in the lab work, short life cycle, high fertility rates and ease of breeding, small genome size and organism size, and availability of standardised strains or their evolutionary position

(Ankeny & Leonelli, 2020). However, as pointed out by these authors, experimental organisms and model organisms differ, although both are essential for advances in biology. Model organisms are systems with deep resources for large-scale biology, ecology, evolution, genetics, cell biology, and availability of a plethora of diverse lines, 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 excel as a tool, or are interesting organisms or objects of scientific curiosity (Leonelli & Ankeny, 2013). From this perspective, *Arabidopsis* (for dicots) and *Brachypodium* (for monocots) can be defined as model organisms for plant biology (Hasterok et al., 2022; IBI et al., 2010; Scholthof et al., 2018)

**The large genomes of cereals, which are rich in repetitive DNA, the domestication syndrome, and the complex evolutionary history of the Poaceae were challenging before *Brachypodium* emerged as a model organism.**

*Brachypodium distachyon* became the first member of the Pooideae subfamily to be sequenced (IBI et al., 2010). Since the publication of the reference genome, a large number of studies have been carried out in different plant biology fields and many available resources, germoplasm collections, genetic markers, mutant collections, microarrays, and recently the published pangenome of 54 diverse inbred *B. distachyon* lines, and databases have been secured (Gordon et al., 2017) (<https://phytozome-next.jgi.doe.gov/brachypan/>, <https://phytozome-next.jgi.doe.gov/>).

After the separation of the three annual species, *B. distachyon* (2x), *B. stacei* (2x) and *B. hybridum* (4x), which were previously described as part of a single specie (*B. distachyon*) (Catalán et al., 2012), other *Brachypodium* species have become of interest. Nowadays, the interest has expanded to the whole genus (Hasterok et al., 2022; Scholthof et al., 2018). There are currently up to six reference genomes available, including the three annual species, the diploids *B. distachyon* and *B. stacei* and the allotetraploid *B. hybridum* (two accessions), which have been selected as a model system to elucidate the origins and evolution of allopolyploidy (Gordon et al. 2020, Scarlett et al. 2021), as well as three perennial species, the ancestral tetraploid *B. mexicanum* and the perennials diploids *B. arbuscula* and *B. sylvaticum* (with two accessions), the later selected as model system for grass perenniality (Gordon et al., 2016; Vogel, 2016). The constructions of the pangenomes of *B. stacei*, *B. hybridum* and *B. sylvaticum* represent ongoing works (Catalán & Vogel, 2020).

## Systematics and phylogenetics of *Brachypodium*

The temperate genus *Brachypodium* was described in 1812 by Palisot de Beauvois. *B. distachyon* (L.) P. Beauv. *sensu lato* (Palisot de Beauvois, 1812), was considered a single annual species and for more than three decades three cytotypes were recognised ( $2n = 10, 20, 30$ ) and considered part of an ascendant autopoliploidy series of  $x = 5$  (Talavera, 1978) until the description of the three independent annual species, the diploids *B. distachyon* ( $2n=10$ ) and *B. stacei* ( $2n= 20$ ) and their derived allotetraploid *B. hybridum* ( $2n =30$ ), in 2012 (Catalán et al., 2012). Despite the differences on the chromosome number, the diploid genome size was roughly similar (*B. stacei*, 0.564 pg/2C; *B. distachyon*, 0.631 pg/2C) whereas the genome size of the allotetraploid *B. hybridum* corresponded to the sum of the two progenitor genomes (1.265 pg/2C). Through molecular phylogenetic studies, it has been discovered that *B. stacei* is the oldest diploid lineage within the *Brachypodium* genus. It diverged from the common ancestor around 10 million years ago, following by the divergences of the *B. distachyon* lineage, which split off about 7 million years ago, and a clade of recent perennial lineages, the core perennial clade, which branched off approximately 3 million years ago. The allotetraploid *B. hybridum* species originated around 1 million years ago (Catalán et al., 2012; Díaz-Pérez et al., 2018; Sancho et al., 2022). This trio of annual species was proposed as a model complex for allo-polyploidy, due to the biological characteristics of its members (small stature, short life cycle, predominantly self-pollinating reproduction, small genome size, low amount of repetitive DNA, easy to transform, phylogenetically close to the temperate cereals)(Catalán et al., 2012), and because they can provide the cues for comparative functional genomics knowledge for the polyploid wheats and other pooid cereals and fodder grasses (Catalan et al., 2014; Gordon et al., 2016). Although the three species present morphological similarities, a comprehensive statistical analysis across a wide diversity of wild populations and inbred lines detected 12 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, leaf blade colour, leaf blade shape, softness, hairiness] and 434 tentatively annotated metabolite signals that significantly discriminated *B. distachyon*, *B. stacei* and *B. hybridum* (Catalán et al., 2012, 2016; López-Álvarez et al., 2017). Morphological significant variation correlated with genetic divergence in western Mediterranean populations of *B. stacei* and *B. distachyon* (López-Álvarez et al., 2017; Marques et al., 2017; Shiposha et al., 2016). Interestingly, populations of *B. hybridum* from the native circum-Mediterranean region that originated from different crossbreeds showed less noticeable differences in their phenotypical characteristics (López-Álvarez et al., 2017).

The reference genomes of the three annual species and their ongoing pangenomes have opened enormous possibilities for investigating the origins and consequences of plant polyploidy. The study of the reference genomes of the main annual accessions *B. distachyon* Bd21, *B. stacei* ABR114, recent *B. hybridum* ABR113 and ancestral *B. hybridum* Bd26 allowed researchers to define the divergence, hybridization and polyploidization events within the three species, and the multiple and bidirectional origins for the allotetraploid *B. hybridum* with ancestral [*B. distachyon*-type (D) plastotypes] and recent [*B. stacei*-type (S) plastotypes] allotetraploids, originating 1.4 and 0.4 Ma, respectively (Gordon et al., 2020). These two different hybridization and genome doubling events reflected a gradual polyploid genome evolution that could lead to a reproductive isolation of potentially cryptic species (Hasterok et al., 2022).

*Brachypodium* also contains ~17 perennial species (Catalán, et al., 2016; Schippmann, 1991), which share the typical sessile spikelet and exclusive embryo development, seed storage proteins, polysaccharides and globulins, stem, and leaf fructosans, small genome sizes and large dispolyploidy with the annual species (Catalán, al., 2016). Most of the perennial taxa present long-rhizomes and self-incompatibility, with the exception of the most ancestral perennial *B. mexicanum* and the perenniality model species *B. sylvaticum*, which are self-compatible and have slender rhizomes (Catalán et al., 1995; Khan & Stace, 1999; Steinwand et al., 2013).

The *Brachypodium* perennial species vary in both phenotype and origin. They range from the short-rhizomatous, self-fertile, ancestral American tetraploid *B. mexicanum*, which is biologically and genomically similar to the annual species, to the strong-rhizomatous recently evolved Eurasian and African diploid and allopolyploid species of the core-perennial clade. The core-perennial clade includes some Palaearctic species such as *B. sylvaticum* and *B. pinnatum*, along with more restrictively distributed endemic species (Catalán et al., 2016). Two xeric perennial allopolyploids *B. retusum* 4x and *B. boissieri* 6x, are characterized by their branched woody stems and short and strongly inrolled leaves whereas core perennial allopolyploids *B. phoenicoides* 4x-6x, *B. pinnatum* 2x, 4x and *B. rupestre* 2x, 4x, 6x are distinguished by their non-branched stems and long flat leaves. The long-rhizomatose *B. pinnatum*, *B. rupestre* and *B. phoenicoides* also show erect panicles. *B. phoenicoides*, adapted to dry places, is glabrous and presents partially inrolled leaves, semi-patent twisted spikelets and awnless lemmas, while the mesic *B. pinnatum* have short awns and bright green colored leaves, traits shared with *B. rupestre*, which differs from the former in its glabrous leaves and spikelets and in leaf epidermal traits (Catalán et al., 2016; Schippmann, 1991).

The diploid *B. sylvaticum*, selected as model species for perenniality (Gordon et al., 2016), is characterized by its densely hairy habit, nodding panicles and long awned lemma. This

species presents a large Eurosiberian distribution, ranging from the Canary Islands (West) to Japan and New Guinea (East) and from Scandinavia and Siberia (North) to N Africa and Malesia (South), though some of East Asian, and African populations probably correspond to different microtaxa (Catalán et al., 2016; Díaz-Pérez et al., 2018).

The origin of the tetraploids *B. phoenicoides* and *B. rupestre* ( $2n = 28$ ) seems to involve different hybridization events although resulting in the common possession of an E2 subgenome ( $x = 5$ ) and a G subgenome ( $x = 9$ ). In contrast, the Mediterranean tetraploid *B. retusum* shows a more divergent subgenomic origins having an ancestral A2 subgenome ( $x = 8$ ) and an intermediately evolved subgenome E1 ( $x = 8$ ).

Despite the advances in the studies of the *Brachypodium* polyploid origins, most of the perennial species and cytotypes described remains unclear. The disjunct distributions in their respective native ranges [Mediterranean (*B. genuense*), Mediterranean-SW Asia (*B. glaucovirens*), Central Asia (*B. spryginii*, *B. kawakamii*), East-Asia (*B. kurilense*, *B. miserum*), Madagascar (*B. madagascariense*), South Africa (*B. bolussi*, *B. flexum*)] and the morphological similarities and the lack of reference genomes difficult the disentanglement of their genomic origins.

The largest phylogenetic study on annual and perennial *Brachypodium* species corroborates the more ancestral divergences of the ancestral outcore species (*B. stacei*, *B. distachyon*, *B. hybridum*, *B. mexicanum* and *B. boissieri*) and the most recent branching off of the core perennial clade, with the first divergence of *B. arbuscula* followed by the successively split of the *B. sylvaticum*, *B. glaucovirens*, *B. pinnatum*, *B. phoenicoides* and *B. rupestre* clades.

### **The need of new genomic and evolutionary studies of the perennial *Brachypodium* species.**

The development of advanced genomic tools, through whole genome sequencing but also from low-pass genome sequencing and museomics, together with new taxonomic explorations, has set the scenario for a deeper study of the genomic diversity and evolution of the less studied *Brachypodium* species (Catalán et al., 2016).

Despite the large knowledge generated for the annual species of *Brachypodium* and the incipient study of some perennial species, the genus is still in lack of a thoroughly genomic, evolutionary and systematic review. The disjunct geographical distributions of minor taxa and the complex reticulate evolutionary history of most perennial species, and the low to high levels of polyploidy presented by these species, make these taxa an excellent case study to investigate



their evolutionary relationship, the identification of their genomes and subgenomes, the analysis of potential genomic rearrangements and expansions and contractions of their genomes, the variation of their morphological traits, and their adaptation to different stresses.

The aim of this PhD thesis is to uncover the gap of knowledge on the taxonomy, genomics, and evolution of several neglected *Brachypodium* perennial taxa and of potential cryptic species within the described allopolyploids. With this purpose, we have investigated the phylogeny of a large worldwide sampling of populations for the 20 *Brachypodium* species with special interest in the perennial species. Our study has targeted the following aims: 1) to conduct a phylogenomic study of the *B. sylvaticum* complex micro-taxa using whole plastid genome (plastome) and nuclear ribosomal DNA (rDNA) 35S gene sequences to elucidate their evolutionary relationships and to assess the systematic value of the morphological features that separate them; our study also aimed to infer the divergence ages of these lineages and to estimate their respective ancestries; 2) to scrutinize the potential cryptic cyto-diversity of perennial *Brachypodium* species through large population-level analysis of genome sizes, chromosome counts and rDNA 35S and 5S *in situ* hybridizations to estimate the number and distributions of cytotypes, and to reveal the existence of putative new *Brachypodium* cytotypes, analyze the structure and composition of plastomes in the perennial *Brachypodium* species and reconstruct a plastome-based phylogeny for all the *Brachypodium* taxa and cytotypes, analyze the sequences and number of ribotypic 35S and 5S families in the perennial *Brachypodium* species and reconstruct rDNA-based phylogenies for all the *Brachypodium* taxa and cytotypes, and detect all potential evolutionary lineages of *Brachypodium*, including new splits, and infer the maternal and paternal ancestors of polyploids through contrasted plastome and rDNA-based topological tests and estimate the nodal ages of a total-evidence *Brachypodium* supertree to elucidate the origins of new lineages within a timely evolutionary framework; 3) to characterize and quantify the repetitive elements of 44 representative samples of the main *Brachypodium* species and lineages identifying both shared and private repeats, analyze the expected correlation between genome size and abundance of the repeats; identify repeat types that could have contributed to the expansions or contractions of genomes, assess the phylogenetic value of repeats using phylogenetic reconstructions, and test the three alternative cases and responses to the “polyploid genome shock hypothesis” of polyploids and the putative paleo-polyploid origin of some *Brachypodium* diploids with large genome sizes using mobile and tandem repeat data analysis; 4) to select phenotypic traits that significantly discriminate the different tetraploid and hexaploid cytotypes of *B. retusum* and *B. phoenicoides* allopolyploids, identify other potential infraspecific morphotypic groups within each species and the phenotypic traits that discriminate them, establish the potential relationships

of infraspecific morphotypic groups to cytotypic, environmental or geographical variables, and uncover cryptic diversity within the *B. retusum* and *B. phoenicoides* polyploid complexes if present; and 5) to analyze, describe and explore the expression of *Brachypodium* dehydrin genes related to drought stress in the three annual *Brachypodium* species and the perennial diploid *B. sylvaticum*, and along a representative sampling of the *B. distachyon* pangenome. Our result would contribute to enlightening the knowledge on the genomic and evolutionary relationship of the ancestral outcore and recently evolved core perennial *Brachypodium* species, corroborate the presence of ancestral and recent subgenomes within the perennial *Brachypodium* allopolyploids, disentangle the phylogeny and origins of the Eurasian diploid *B. sylvaticum* complex, provide discriminant morphoanatomical traits to describe potential cryptic species within the Mediterranean allopolyploid *B. retusum* and *B. phoenicoides* complexes, and characterize the structure and evolution of dehydrin genes in *Brachypodium* and their expressions in *B. distachyon* ecotypes under drought and watered conditions

Our studies comprise the 20 described species of *Brachypodium* and an exhaustive sampling of 96 world populations, which has allowed us to analyze their phylogenetic relationship through complete sequences of plastome and rDNA 35S and 5S regions using genome skimming approaches. The study of their 5S rDNA regions facilitated the detection of two ribotypic subfamilies, a short 5S sequence subfamily associated to ancestral *Brachypodium* outcore genomes and a long 5S sequence subfamily corresponding to recently evolved *Brachypodium* core perennial clade genomes. Furthermore, the detected 5S loci match well the subgenomes described by Sancho et al. (2022) using transcriptome data of nuclear single copy genes. Our phylogenomic analysis detected twelve main evolutionary clades within the *Brachypodium* supertree (species tree) and also discovered two different lineages within the Palearctic *B. sylvaticum*, a more ancestral oriental clade of diploid species (*B. sylvaticum* vars. *breviglume*, *B. kurilense*, *B. miserum*) and African and S Asian endemics (*B. flexum*, *B. bolussi*, *B. madagascariense*, *B. kawakamii*) and a more recent occidental clade of *B. sylvaticum* s.s., *B. glaucovirens* and *B. sprynginii* taxa. Our genomic sequence analysis detected 10 types of dehydrin genes (*Bdhn*) across the four studied *Brachypodium* species which were constrained by purifying selection; only four genes were expressed in mature leaves of *B. distachyon*, all of them were more highly expressed in plants under drought conditions, and their expression was significantly correlated with drought-response phenotypic traits being more pronounced in drought-tolerant ecotypes. As a model genus, *Brachypodium* may be key to understand and develop strategies to face the climate change scenario.



## References

- Ankeny, R., & Leonelli, S. (2020). *Model Organisms*. Cambridge University Press. <https://doi.org/10.1017/9781108593014>
- Bolger, M. E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B., & Mayer, K. F. (2014). Plant genome sequencing — applications for crop improvement. *Current Opinion in Biotechnology*, 26, 31–37. <https://doi.org/10.1016/j.copbio.2013.08.019>
- Bouchenak-Khelladi, Y., Verboom, G. A., Hodkinson, T. R., Salamin, N., Francois, O., N?? Chonghaile, G., & Savolainen, V. (2009). The origins and diversification of C4 grasses and savanna-adapted ungulates. *Global Change Biology*, 15(10), 2397–2417. <https://doi.org/10.1111/j.1365-2486.2009.01860.x>
- 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(7), 414–418. <https://doi.org/10.1016/j.tplants.2014.05.002>
- 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(2), 150–166. <https://doi.org/10.1006/mpev.1997.0416>
- Catalán, P., López-Álvarez, D., Bellosta, C., & Villar, L. (2016). Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei*, and *B. hybridum* (Poaceae). *Anales Del Jardín Botánico de Madrid*, 73(1), e028. <https://doi.org/10.3989/ajbm.2428>
- Catalán, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., & López-Herránz, M. L. (2016). Phylogeny and Evolution of the Genus *Brachypodium*. In *Plant Genetics and Genomics: Crops and Models* (Issue October, pp. 9–38). [https://doi.org/10.1007/7397\\_2015\\_17](https://doi.org/10.1007/7397_2015_17)
- Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L. A. J., Langdon, T., Betekhtin, A., Siwinska, D., Pimentel, M., & López-Alvarez, D. (2012). Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany*, 109(2), 385–405. <https://doi.org/10.1093/aob/mcr294>
- 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(1–2), 1–19. <https://doi.org/10.1007/BF00985367>

- Catalán, P., Shi, Y., Armstrong, L., Draper, J., & Stace, C.A. (1995). Molecular phylogeny of the grass genus *Brachypodium* P. Beauv. based on RFLP and RAPD analysis. *Botanical Journal of the Linnean Society*, 117(4), 263–280. <https://doi.org/10.1111/j.1095-8339.1995.tb02590.x>
- Catalán, P., & Vogel, J. (2020). Advances on genomics, biology, ecology and evolution of *Brachypodium*, a bridging model grass system for cereals and biofuel grasses. *New Phytologist*, 227(6), 1587–1590. <https://doi.org/10.1111/NPH.16831>
- Decena, M. A., Sancho, R., Lusinska, J., Hasterok, R., Gorgojo, R., Montes, B., Perez-Collazos, E., Inda, L. A., & Catalan. (2023). Cyto-phylogenomics of the grass model genus *Brachypodium* unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends within this largely reticulate diploid-polyploid species-complex. *Frontiers in Plant Science*.
- Díaz-Pérez, A., López-Álvarez, D., Sancho, R., & Catalán, P. (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. *Molecular Phylogenetics and Evolution*, 127(May), 256–271. <https://doi.org/10.1016/j.ympev.2018.06.003>
- 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(4), 1539–1555. <https://doi.org/10.1104/PP.010196>
- Edwards, E. J., Osborne, C. P., Stromberg, C. A. E., Smith, S. A., Bond, W. J., Christin, P. A., Cousins, A. B., Duvall, M. R., Fox, D. L., Freckleton, R. P., Ghannoum, O., Hartwell, J., Huang, Y., Janis, C. M., Keeley, J. E., Kellogg, E. A., Knapp, A. K., Leakey, A. D. B., Nelson, D. M., ... Tipple, B. (2010). The Origins of C4 Grasslands: Integrating Evolutionary and Ecosystem Science. *Science*, 328(5978), 587–591. <https://doi.org/10.1126/science.1177216>
- Febrer, M., Goicoechea, J. L., Wright, J., McKenzie, N., Song, X., Lin, J., Collura, K., Wissotski, M., Yu, Y., Ammiraju, J. S. S., Wolny, E., Idziak, D., Betekhtin, A., Kudrna, D., Hasterok, R., Wing, R. A., & Bevan, M. W. (2010). An integrated physical, genetic and cytogenetic map of *Brachypodium distachyon*, a model system for grass research. *PLoS ONE*, 5(10). <https://doi.org/10.1371/journal.pone.0013461>

- Fox, S. E., Geniza, M., Hanumappa, M., Naithani, S., Sullivan, C., Preece, J., Tiwari, V. K., Elser, J., Leonard, J. M., Sage, A., Gresham, C., Kerhornou, A., Bolser, D., McCarthy, F., Kersey, P., Lazo, G. R., & Jaiswal, P. (2014). De Novo Transcriptome Assembly and Analyses of Gene Expression during Photomorphogenesis in Diploid Wheat *Triticum monococcum*. *PLoS ONE*, 9(5), e96855. <https://doi.org/10.1371/journal.pone.0096855>
- Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., Session, A., Martin, J., Cartwright, A., Katz, A., Singan, V. R., Goltsman, E., Barry, K., Dinh-Thi, V. H., Chalhoub, B., Diaz-Perez, A., Sancho, R., Lusinska, J., Wolny, E., ... Vogel, J. P. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nature Communications*, 11(1), 3670. <https://doi.org/10.1038/s41467-020-17302-5>
- Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., Stritt, C., Roulin, A. C., Schackwitz, W., Tyler, L., Martin, J., Lipzen, A., Dochy, N., Phillips, J., Barry, K., Geuten, K., Budak, H., Juenger, T. E., Amasino, R., ... Vogel, J. P. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, 8(1). <https://doi.org/10.1038/s41467-017-02292-8>
- Gordon, S. P., Liu, L., & Vogel, J. P. (2016). *The Genus Brachypodium as a Model for Perenniality and Polyploidy* (pp. 313–325). [https://doi.org/10.1007/7397\\_2015\\_19](https://doi.org/10.1007/7397_2015_19)
- Hasterok, R., Catalan, P., Hazen, S. P., Roulin, A. C., Vogel, J. P., Wang, K., & Mur, L. A. J. (2022). *Brachypodium*: 20 years as a grass biology model system; the way forward? *Trends in Plant Science*, 27(10), 1002–1016. <https://doi.org/10.1016/j.tplants.2022.04.008>

- 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(3–4), 389–398. <https://doi.org/10.1007/BF00221981>
- IBI, Dvorak, J., Wright, J., Febrer, M., Idziak, D., Hasterok, R., Lindquist, E., Wang, M., Fox, S. E., Priest, H. D., Filichkin, S. A., Givan, S. A., Bryant, D. W., Chang, J. H., Wu, H., Wu, W., Hsia, A. P., Schnable, P. S., Kalyanaraman, A., ... IBI. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 463(7282), 763–768. <https://doi.org/10.1038/nature08747>
- IWGSC, Appels, R., Eversole, K., Stein, N., Feuillet, C., Keller, B., Rogers, J., Pozniak, C. J., Choulet, F., Distelfeld, A., Poland, J., Ronen, G., Sharpe, A. G., Barad, O., Baruch, K., Keeble-Gagnère, G., Mascher, M., Ben-Zvi, G., Josselin, A.-A., ... Wang, L. (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403). <https://doi.org/10.1126/science.aar7191>
- Jacobs, B. F., Kingston, J. D., & Jacobs, L. L. (1999). The Origin of Grass-Dominated Ecosystems. *Annals of the Missouri Botanical Garden*, 86(2), 590. <https://doi.org/10.2307/2666186>
- Kellogg, E. A. (2015). *The Families and Genera of Vascular Plants. Vol. XIII. Flowering Plants. Monocots. Poaceae* (K. Kubitzki, Ed.). Springer.
- Khan, M. A., & Stace, C. A. (1999). Breeding relationships in the genus *Brachypodium* (Poaceae: Pooideae). *Nordic Journal of Botany*, 19(3), 257–269. <https://doi.org/10.1111/J.1756-1051.1999.TB01108.X>
- Kumar, S., Mohan, A., Balyan, H. S., & Gupta, P. K. (2009). Orthology between genomes of *Brachypodium*, wheat and rice. *BMC Research Notes*, 2(1), 93. <https://doi.org/10.1186/1756-0500-2-93>
- Leonelli, S., & Ankeny, R. A. (2013). What makes a model organism? *Endeavour*, 37(4), 209–212. <https://doi.org/10.1016/J.ENDEAVOUR.2013.06.001>

- 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(12). <https://doi.org/10.1371/journal.pone.0051058>
- 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., & Catalán, 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(7), 1073–1088. <https://doi.org/10.3732/ajb.1500128>
- 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(4), 545–561. <https://doi.org/10.1093/aob/mcw239>
- Luo, M. C., Deal, K. R., Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., Blake, N., Clegg, M. T., Coleman-Derr, D., Conley, E. J., Crossman, C. C., Dubcovsky, J., Gill, B. S., Gu, Y. Q., Hadam, J., Heo, H. Y., Huo, N., Lazo, G., Ma, Y., ... Dvorak, J. (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(37), 15780–15785. <https://doi.org/10.1073/pnas.0908195106>
- Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., & Hasterok, R. (2019). Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *International Journal of Molecular Sciences*, 20(22), 1–19. <https://doi.org/10.3390/ijms20225557>
- 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), 1–14. <https://doi.org/10.1186/s12862-017-0996-x>

- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P. E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X. Q., Zhang, Q., Barrero, R. A., Li, L., Taudien, S., ... Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature* 2017 544:7651, 544(7651), 427–433. <https://doi.org/10.1038/nature22043>
- 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(7), 1220–1230.
- Mayer KF;, Waugh R;, Langridge P;, Close TJ;, Wise RP;, Graner A;, Matsumoto T;, Sato K;, Schulman A;, Muehlbauer GJ;, Stein N;, Ariyadasa R;, Schulte D;, Poursarebani N;, Zhou R;, Steuernagel B;, Mascher M;, Scholz U;, Shi B;, ... Stein N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491(7426), 711–716. <https://doi.org/10.1038/nature11543>
- 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(2), 334–347. <https://doi.org/10.1111/j.1469-8137.2011.03748.x>
- Opanowicz, M., Vain, P., Draper, J., Parker, D., & Doonan, J. H. (2008). *Brachypodium distachyon*: making hay with a wild grass. *Trends in Plant Science*, 13(4), 172–177. <https://doi.org/10.1016/J.TPLANTS.2008.01.007>
- Palisot de Beauvois, A.-M.-F.-J. (1812). *Essai d'une nouvelle agrostographie, ou, Nouveaux genres des graminées :avec figures représentant les caractères de tous les genres /par A.-M.-F.-J. Palisot de Beauvois*. Chez l'auteur,. <https://doi.org/10.5962/bhl.title.474>
- Pfeifer, M., Martis, M., Asp, T., Mayer, K. F. X., Lübberstedt, T., Byrne, S., Frei, U., & Studer, B. (2013). The Perennial Ryegrass GenomeZipper: Targeted Use of Genome Resources for Comparative Grass Genomics. *Plant Physiology*, 161(2), 571–582. <https://doi.org/10.1104/pp.112.207282>

- Pimentel, M., Escudero, M., Sahuquillo, E., Minaya, M. Á., & Catalán, P. (2017). Are diversification rates and chromosome evolution in the temperate grasses (Pooideae) associated with major environmental changes in the Oligocene-Miocene? *PeerJ*, 5, e3815. <https://doi.org/10.7717/peerj.3815>
- 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(January), 11–24. <https://doi.org/10.1105/tpc.107.056309>
- Sancho, R., Inda, L. A., Díaz-Pérez, A., Des Marais, D. L., Gordon, S., Vogel, J. P., Lusinska, J., Hasterok, R., Contreras-Moreira, B., & Catalán, P. (2022). Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant Journal*, 109(6), 1535–1558. <https://doi.org/10.1111/tbj.15650>
- Schippmann. (1991). Revision der euroäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera*, 45.
- Schneider, J., Winterfeld, G., Hoffmann, M. H., & Röser, M. (2011). Duthieae, 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(1), 27–44. <https://doi.org/10.1080/14772000.2010.544339>
- Scholthof, K. B. G., Irigoyen, S., Catalan, P., & Mandadi, K. K. (2018). *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell*, 30(8), 1673–1694. <https://doi.org/10.1105/tpc.18.00083>
- 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*, 2016(9). <https://doi.org/10.7717/peerj.2407>
- 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); A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53(2), 117–137. <https://doi.org/10.1111/jse.12150>

- 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(9), e75180. <https://doi.org/10.1371/journal.pone.0075180>
- Strömberg, C. A. E. (2011). Evolution of Grasses and Grassland Ecosystems. *Annual Review of Earth and Planetary Sciences*, 39(1), 517–544. <https://doi.org/10.1146/annurev-earth-040809-152402>
- Talavera, S. (1978). Aportacion al estudio cariologico de las gramineas españolas. *Lagascalia*, 7, 133–142. *Lagascalia*, 7, 133–142.
- Vogel, J. P. (2016). *The Rise of Brachypodium as a Model System* (pp. 1–7). [https://doi.org/10.1007/7397\\_2015\\_14](https://doi.org/10.1007/7397_2015_14)



## OBJETIVOS

El objetivo principal de esta tesis es desentrañar las relaciones de parentesco evolutivo entre las especies del género modelo *Brachypodium* para esclarecer el origen de las especies alopoliplodes mediante análisis morfoanatómicos, genómicos y filogenéticos de genomas plastídicos, regiones ribosomales nucleares y repeteoma, además del estudio de genes inducidos por estrés hídrico.

Este objetivo principal incluye los siguientes objetivos específicos:

1. Esclarecer el origen y la evolución de la alopoliploidía de las especies perennes del género modelo de gramíneas *Brachypodium*.
2. Contrastar la hipótesis del *shock* poliploide a través del estudio de la correlación entre el tamaño genómico y la proporción de repeteoma en especies representativas del género *Brachypodium* con distintos subgenomas y niveles de ploidía.
3. Dilucidar la sistemática de especies perennes de *Brachypodium*, mediante análisis macro- y micromorfológicos y taxonómicos en un marco evolutivo.
4. Analizar la evolución y la expresión génica diferencial de genes de tolerancia al estrés hídrico (familia DHN, dehidrininas) en especies del género modelo de gramíneas *Brachypodium*.

## OBJECTIVES

The main objective of this PhD thesis is to uncover the evolutionary relationships among the perennial species of the grass model genus *Brachypodium* with the aim of clarifying the origin of the allopolyploid species, through morphoanatomical, genomic, and phylogenetic analysis of plastid genomes, nuclear ribosomal regions, and the repeatome. Additionally, the study will also analyze the expression of water stress induced genes.

To comply with the main objectives the thesis includes the following specific topics:

1. To clarify the origin and the evolution of allopolyploidy among the perennial species of the model genus *Brachypodium*.
2. To test the polyploid shock hypothesis through the correlation of genome size and repeatome content in representative species of the model genus *Brachypodium* with different subgenomes and levels of ploidy.
3. To elucidate the systematics of perennial *Brachypodium* species through the study of macro and micro-morphological traits within an evolutionary framework.
4. To analyse the evolution and differential gene expression of drought-induced dehydrin genes in model *Brachypodium* grasses.

## ESTRUCTURA DE LA TESIS

La presente tesis doctoral se estructura en tres capítulos generales (Introducción, Objetivos y Conclusiones) y cinco capítulos específicos relacionados con la investigación realizada durante el desarrollo de la misma (Capítulos 1 a 5). Además, hay una sección de referencias enumeradas para los capítulos generales, y cada capítulo tiene su propia sección de referencias específicas. La sección de anexos incluye información complementaria para ciertos capítulos. La tesis finaliza con una lista de publicaciones obtenidas de la investigación doctoral.

Los capítulos están organizados de la siguiente manera:

- Introducción
- Objetivos
- Capítulos de investigación de los estudios realizados durante la tesis:
  - **Capítulo 1:** La filogenética de la gramínea modelo paleártica *Brachypodium sylvaticum* descubre dos linajes divergentes de micro-taxones orientales y occidentales.
  - **Capítulo 2:** La cito-filogenómica del complejo de especies de gramíneas *Brachypodium* revela una diversidad críptica muy divergente y diferentes tendencias de disploidía descendiente antes y después de la poliploidización
  - **Capítulo 3:** Las expansiones y contracciones de elementos de ADN repetitivos revelan distintas respuestas evolutivas a la hipótesis del shock del genoma poliploide en los poliploides del género de gramíneas *Brachypodium*.
  - **Capítulo 4:** Análisis morfoanatómico de citotipos poliploides crípticos y evolutivamente divergentes de las gramíneas mediterráneas modelo *Brachypodium retusum* y *B. phoenicoides*
  - **Capítulo 5:** Genómica comparada, evolución y expresión de genes de dehidrinas inducidos por sequía en gramíneas modelos de *Brachypodium*
- Conclusiones
- Apéndice
- Publicaciones de la tesis doctoral

## PHD STRUCTURE

The PhD thesis is structured into three general chapters (Introduction, Objectives, and Conclusions) and five specific chapters related to the research conducted during the PhD work (Chapters 1 to 5). Additionally, there is a section of references listed for the general chapters, and each chapter has its own specific reference section. The appendix section includes supplementary information for several chapters. The thesis ends with a list of publications obtained from the PhD research.

The chapters are organised as follows:

- Introduction
- Objectives
- Research chapters of the studies conducted during the thesis:
  - **Chapter 1:** Phylogenetics of the Palearctic model grass *Brachypodium sylvaticum* uncovers two divergent oriental and occidental micro-taxa lineages.
  - **Chapter 2:** Cyto-phylogenomics of the *Brachypodium* grass species-complex unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends.
  - **Chapter 3:** Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in grass *Brachypodium* polyploids.
  - **Chapter 4:** Morphoanatomical analysis of cryptic and evolutionary divergent polyploid cytotypes of model Mediterranean *Brachypodium retusum* and *B. phoenicoides* grasses
  - **Chapter 5:** Comparative genomics, evolution, and drought-induced expression of dehydrin genes in model *Brachypodium* grasses
- Conclusions
- Appendix
- Publications of the PhD thesis

## PHYLOGENETICS OF THE PALEARCTIC MODEL GRASS *BRACHYPODIUM SYLVATICUM* UNCOVERS TWO DIVERGENT ORIENTAL AND OCCIDENTAL MICRO-TAXALINEAGES

Pilar Catalán<sup>1, 2\*</sup>, María Ángeles Decena<sup>1, 2</sup>, Rubén Sancho<sup>1, 2</sup>, Juan Viruel<sup>3</sup>, Ernesto Pérez-Collazos<sup>1,2</sup>, Luís. A. Inda<sup>1, 4</sup> & Nina S. Probatova<sup>5†</sup>

<sup>1</sup> Departamento de Ciencias Agrarias y del Medio Natural. Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Huesca, España

<sup>2</sup> Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza, España

<sup>3</sup> Royal Botanic Gardens, Kew, Richmond, UK

<sup>4</sup> Instituto Agroalimentario de Aragón-IA2 (Universidad de Zaragoza-CITA), Aragón, España

<sup>5</sup> Federal Scientific Center of the East Asian Terrestrial Biodiversity FEB RAS, Vladivostok, Russia

\* corresponding author

### ABSTRACT

*Brachypodium sylvaticum* has been selected as a model for perennial grasses, and considerable genomic resources have been generated and a reference genome and several resequenced pangenome accessions are available for this species. Despite these genomic advances, the evolution and systematics of diploid *B. sylvaticum* s. l. is almost unknown. The *B. sylvaticum* complex is formed by up to seven taxonomically close micro-taxa which differentiate from typical *B. sylvaticum* s. s. based on a few morphological features. Moreover, some of them show some largely disjunct geo- graphic distributions on both sides of their native Palearctic region. In this study, we used a phylogenomic approach including representative populations from the oriental and occidental distribution range of *B. sylvaticum* micro-taxa to elucidate their evolutionary relationships and assess the systematic value of the morpho- logical features that separate them. A combined plastome and nuclear phylogenetic tree supports an early split and high divergence of the oriental lineage, showing the close relationship of the Himalayan *B. sylvaticum* var. *breviglume* lineages to the Pa- cific *B. miserum* / *B. kurilense* clade, and the contrasting large homogeneity and low divergence of the occidental European, N African and SW and C Asian lineage, with several *B. sylvaticum* s. s., *B. spryginii*, and *B. glaucovirens* samples showing identical or similar sequences. Divergence time estimate analysis suggests that the oriental lineage diverged from the common ancestor in the early Pleistocene (2.0 Ma), fol- lowed by subsequent colonization and isolations in the Himalayas (2.0 – 1.7 Ma) and the Far East (0.36 Ma) in more recent times, while the occidental lineage split in the Mid-Late Pleistocene (0.97 Ma), followed by rapid radiation and postglacial spread in the western Palearctic during the last thousand years.

**Keywords:** *Brachypodium sylvaticum* complex, diploid populations, discriminant morpho- logical traits, eastern and western Eurasian micro-taxa, phylogenetics, plastomes, *rDNA* 35S gene, systematics

## РЕЗЮМЕ

Каталан П., Ангелес Децена М., Санчо Р., Вируел Х., Перес-Колазос Э., Пробатова Н.С. Филогенетика палеарктического модельного злака *Brachypodium sylvaticum* показывает две дивергентные восточную и западную линии микротаксонов. *Brachypodium sylvaticum* был выбран в качестве модели для многолетних трав, и для этого вида были созданы значительные геномные ресурсы и доступны эталонный геном и несколько повторно секвенированных пангеномных доступов. Несмотря на эти геномные достижения, эволюция и систематика диплоидного *B. sylvaticum* s. l. практически неизвестна. Комплекс *B. sylvaticum* образуют до семи таксономически близких микро-таксонов, которые отличаются от типичного *B. sylvaticum* s. s. по нескольким морфологическим признакам. Более того, некоторые из них демонстрируют в значительной степени дизъюнктивное географическое распространение по обе стороны их родного Палеарктического региона. В данном исследовании мы использовали филогеномный подход, включая репрезентативные популяции из восточного и западного ареалов распространения микро-таксонов *B. sylvaticum*, чтобы выяснить их эволюционные отношения и оценить систематическое значение разделяющих их морфологических признаков. Объединенное филогенетическое древо, построенное на основе пластома и ядерной ДНК, подтверждает раннее разделение и высокую дивергенцию восточной линии, показывая близость гималайской линии *B. sylvaticum* var. *breviglume* с тихоокеанской линией *B. miserum* / *B. kurilense*, и контрастирующую большую однородность и низкую дивергенцию восточно-европейской, североафриканской и южно- и среднеазиатской линии, с несколькими видами *B. sylvaticum* s. s., *B. spryginii* и *B. glaucovirens* демонстрируют идентичные или похожие последовательности. Анализ временной оценки дивергенции позволяет предположить, что восточная линия отделилась от общего предка в раннем плейстоцене (2,0 млн лет назад), после чего последовали колонизация и изоляция в Гималаях (2,0–1,7 млн лет назад) и на Дальнем Востоке (0,36 млн лет назад) в более поздние времена, в то время как окцидентальная линия отделилась в среднем-позднем плейстоцене западной Палеарктике в течение последней тысячи лет.

**Ключевые слова:** комплекс *Brachypodium sylvaticum*, диплоидные популяции, дискриминантные морфологические признаки, микротаксоны восточной и западной Евразии, филогенетика, пластома, ген 35S рДНК, систематика

Since its proposal as a model perennial grass system (Steinwand et al. 2013, Gordon et al. 2016), *Brachypodium sylvaticum* has attracted the attention of the scientific community as a suitable model plant for understanding the molecular mechanisms that caused the transitions between perennial-annual plants (Hu et al. 2011, Scholthof et al. 2018, Friedman 2020), and to design strategies for improving perennial energy crops (Carroll & Somerville 2009, Dohleman & Long 2009) and develop perennial grain crops (Glover et al. 2010, Gordon et al. 2016). This conceptual framework has conveyed the generation of large genomic and biological resources (e.g., transformation and inbred lines, Steinwand et al. 2013; transcriptome atlases, Fox et al. 2013; salt tolerance responses, Sade et al. 2018) for *B. sylvaticum*, paralleling those of its well-known annual congener *B. distachyon* (Scholthof et al. 2018). A reference genome has been assembled for one *Brachypodium sylvaticum* accession (*B.sylvaticum*-Ain1; Phytozome [https://phytozome-next.jgi.doe.gov/info/BsylvaticumAin\\_1\\_v2\\_1](https://phytozome-next.jgi.doe.gov/info/BsylvaticumAin_1_v2_1)) and current research on the genomic resequencing of other *B. sylvaticum* lines aims to generate a pangenome for this model perennial grass (Joint Genome Institute, Community Science Program; <https://jgi.doe.gov/Brachypodium-model-grass-genus-bioenergy/>).

*Brachypodium sylvaticum* has also been the subject of large taxonomic, cytogenetic, evolutionary and ecological research. This perennial grass presents the broadest native

distribution of all worldwide *Brachypodium* species, covering the entire Palearctic region, from Macaronesia (West) to New Guinea (East) and from Scandinavia and Siberia (North) to northern Africa and Malaysia (South) (Catalán et al. 2016). It is also the only known invasive perennial *Brachypodium* taxon in the New World (Rosenthal et al. 2008, POWO

<https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:393196-1>, accessed 16 April 2023). Morphologically, *B. sylvaticum* separates from other *Brachypodium* congeners based on the possession of short and slender rhizomes, nodding panicle, densely hairy habit and long-awned lemma (Schippmann 1991). It is a self-compatible perennial species (Khan & Stace 1999, Steinwand et al. 2013), adapted to mesic and nemoral habitats of humid forest (Schippmann 1991). Some of its features are also shared by the tropical and South African *B. flexum* and the Malagasy *B. madagascariense*, though they differ from the former in their shorter panicles, spikelets and awns (Catalán et al. 2016). The *B. sylvaticum* complex includes up to seven micro-taxa; *B. sylvaticum* s. s., described from England, is the most widespread species in Europe, N Africa, and SW and C Asia (Schippmann 1991). The remaining six cryptic micro-taxa were described from different regions of eastern Europe, Asia, and Malesia. They were formerly synonymized with *B. sylvaticum*, although some of them were later disgregated from it (*B. sylvaticum* var. *breviglume*, *B. sylvaticum* var. *pseudodistachyon*, *B. kurilense*, *B. miserum*, *B. pubescens*, *B. spryginii*; Keng 1982, Tzvelev 1983, 2015, Probatova & Skolovskaya 1985, Tzvelev & Probatova 2019) based on features related to the length of glumes, plant pubescence and

height of the plant. All the *B. sylvaticum* complex taxa are further characterized by a constant diploid chromosome number of  $2n = 2x = 18$  and a chromosome base number of  $x = 9$  (Wolny & Hasterok 2009, Catalán et al. 2016, Tzvelev & Probatova 2019, Decena et al. 2023). A close taxon to the *B. sylvaticum* complex is the eastern Mediterranean – SW Asian endemic *B. glaucovirens*, which is also a diploid species but showing a smaller chromosome number and chromosome base number ( $2n = 2x = 16$ ;  $x = 8$ ) than *B. sylvaticum* s. s. (Wolny & Hasterok 2009, Catalán et al. 2016).

Phylogenetic studies of *Brachypodium* based on plastid and nuclear loci have consistently reconstructed *B. sylvaticum* as a main lineage of the recently evolved core-perennial *Brachypodium* clade (Catalán & Olmstead 2000, Catalán et al. 2016, Díaz-Pérez et al. 2018). Molecular dating analysis inferred a recent Mid-Late Pleistocene origin for the *B. sylvaticum* s. l. lineages (1.2–0.2 Ma; Díaz-Pérez et al. 2018). A phylogenomic survey based on transcriptome data also supported a recent split of *B. sylvaticum* s. s. from its close *B. pinnatum* sister lineage in the Pleistocene (1.4 Ma; Sancho et al. 2022). Fluorescent In Situ Hybridization (FISH)-based comparative chromosome barcoding (CCB) analysis characterized the karyotypic profiles of *B. sylvaticum* s. s. ( $x = 9$ ) and of *B. glaucovirens* ( $x = 8$ ) and hypothesized their derived origins from either ancestral  $x = 10$  or intermediate-ancestral  $x = 9$  *Brachypodium* karyotypes through descendant dispoloidy (Lusinska et al. 2019, Sancho et al. 2022); however, the *B. sylvaticum* s. s. barcoded karyotype could not be differentiated

from those of other close diploid core perennial  $x = 9$  species (*B. arbuscula*, *B. pinnatum*) probably due to their recent divergences from the common ancestor (Sancho et al. 2022).

Despite the enormous advances attained in the genomic study of the model perennial grass *B. sylvaticum* s. s., the systematics of the *B. sylvaticum* complex micro-taxa is almost unknown. Taxonomically, the morphological characters used to separate several of these micro-taxa are extremely plastic and variable, and have been considered of doubtful taxonomic relevance in some cases (Tzvelev 2015, Catalán et al. 2016). Nonetheless, some of these taxa show large disjunct geographic distributions in both extremes of Eurasia, which could have led to genetic isolation and evolutionary divergence. Moreover, no evolutionary study has included representative samples of oriental and occidental populations of *B. sylvaticum* s. l. yet. Therefore, the objectives of our study were to conduct a phylogenomic study of the *B. sylvaticum* complex micro-taxa using whole plastid genome (plastome) and nuclear ribosomal DNA (rDNA) 35S gene sequences to elucidate their evolutionary relationships and to assess the systematic value of the morphological features that separate them. Our study also aimed to infer the divergence ages of these lineages and to estimate their respective ancestries.

## MATERIAL AND METHODS

### Sampling and taxonomic identification

A total of 18 *Brachypodium sylvaticum* s. s. samples collected from different European, N African and



Asian populations, plus six samples of close *B. sylvaticum* complex taxa [*B. sylvaticum* var. *breviglume* (2), *B. kurilense* (1), *B. miserum* (1), *B. glaucovirens* (1), *B. spryginii* (1)], five samples from other congeners [*B. pinnatum* (2), *B. arbuscula* (1), *B. distachyon* (1), *B. stacei* (1)] and three outgroup grass samples (*Triticumaestivum*, *Oryza sativa*, *Sorghum bicolor*) were included in the evolutionary and systematic study (Fig. 1, Table 1). The taxonomic identification of the studied materials was based on previous taxonomic treatments of *Brachypodium* (Keng 1982, Schippmann 1991, Catalán et al. 2012, Tzvelev 2015, Tzvelev & Probatova 2019). Chromosome and ploidy level information was retrieved from these studies and from a broad cyto-phylogenomic study of *Brachypodium* perennial species performed by Decena et al. (2023, and unpub. data).

### Genomic data and phylogenetic analysis

Total genomic DNA isolation and genome skimming sequencing procedures are detailed in Moreno-Aguilar et al. (2020) and Decena et al. (2023). Genomic data for plastome reconstruction and nuclear rDNA 35S genes were retrieved from the respective reference genomes of *B. sylvaticum* Ain1, *B. distachyon* Bd21 and *B. stacei* ABR114 accessions (Sancho et al. 2018, 2022) and downloaded from Phytozome (<https://phytozome-next.jgi.doe.gov/>). Assembly of plastomes and rDNA 35S sequences of the remaining *Brachypodium* samples was performed using

Novoplasty v. 2.7.1 (Dierckxsens et al. 2017) and read-mapping strategies to the reference 35S sequences with Geneious Prime 2023, respectively. Multiple sequence alignments (MSA) were performed separately for entire plastomes and nuclear 35S sequences using MAFFT.v7.450 (Katoh & Standley 2013), and removal of low quality regions from each of the MSAs was performed with TrimAl v. 1.2rev59 (Capella-Gutiérrez et al. 2009) followed by manual curation.

Maximum likelihood (ML) phylogenetic trees were re-constructed for each separated data set with IQ-TREE 1.6. imposing the best-fit nucleotide substitution model, according to the Bayesian Information Criterion (BIC), and estimating 1000 ultrafast bootstrap replicates (BS) and SH-alrt test with 1000 replicates for the branch support of the best tree (Nguyen et al. 2015). Due to the overall congruence of the plastome and the rDNA topologies, both data sets were concatenated into a combined data matrix and used to compute a combined ML plastome+35S tree. Ancestral divergence times of the *Brachypodium* lineages were estimated with BEAST 2 (Bouckaert et al. 2014) imposing independent site substitution models for each partition, log-normal relaxed clock and Yule tree linked models, a broad uniform distribution prior for the uncorrelated lognormal distribution (ucln) mean (lower = 1.0E-6; upper = 0.1) and an exponential prior for ucln standard deviation (SD). Five nodes of the tree were calibrated using secondary age constraints for the crown node of grasses (BOP + PACMAD clade; normal prior mean = 55.0 Ma, SD = 0.5 Ma), the BOP clade

## Chapter 1

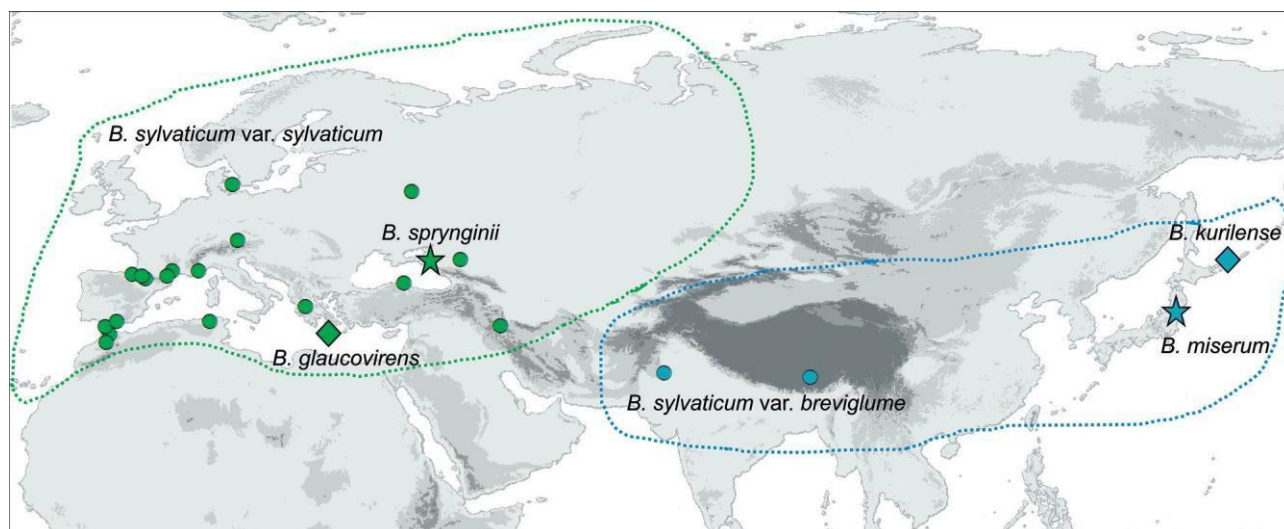
Botanica Pacifica. A journal of plant science and conservation. 2022. 12(1): 21–28 DOI: 10.17581/bp.2023.12119

(*Brachypodium* + *Oryza*; normal prior mean = 51.7 Ma, SD = 1.9 Ma), the Pooideae clade (*Brachypodium* + *Triticum*; normal prior mean = 33.2 Ma, SD = 9.52 Ma), the Brachypodium clade (normal prior mean = 32.2 Ma, SD = 9.5 Ma) and the *Brachypodium* core-perennial clade (normal prior mean = 2.95 Ma, SD = 1.49 Ma), following the grass-wide dating analysis of Sancho et al. (2018). We ran 600 million

Markov chain Monte Carlo (MCMC) generations in BEAST2 with a sampling frequency of 1,000 generations. The adequacy of parameters was checked using TRACER v. 1.66 with all the parameters showing Effective Sample Size (ESS) > 200. A Maximum clade credibility (MCC) tree was computed after discarding 10 % of the respective saved trees as burn-in.

**Table 1.** List of the *Brachypodium sylvaticum* complex samples (*B. sylvaticum* var. *sylvaticum*, *B. sylvaticum* var. *breviglume*, *B. glaucovirens*, *B. kurilense*, *B. miserum*, *B. spryginii*) and other congeners (*B. pinnatum*, *B. arbuscula*, *B. distachyon*, *B. stacei*) and outgroup taxa (*Triticum aestivum*, *Oryza sativa*, *Sorghum bicolor*) used in the study indicating taxon name and authority, population code, geographic location and sample source [INIA, Spanish National Institute of Agriculture accession; JGI, Joint Genome Institute accession; RIKEN, Japanese Center for Sustainable Resource Science accession; USDA, US Department of Agriculture accession; UZ, University of Zaragoza accession; Herbarium specimens: B (Berlin), M (München), LD (Lund), VLA (Vladivostok)], chromosome number (2n), ploidy level, and data source. All studied *Brachypodium* samples are diploids. Taxonomic identities are based in the taxonomic works of Keng (1982), Probatova (1985), Schippmann (1991), Catalán et al. (2012), Tzvelev (2015), Tzvelev & Probatova (2019).

Taxon	Code	Location, herbarium acronym	2n	Ploidy	Source
<i>B. sylvaticum</i> (Huds.) P. Beauv. var. <i>sylvaticum</i>	Bsylvaticum_Ain1	Tunisia: Ain Draham, JGI	18	2x	Sancho et al. 2022
	Bsylvaticum_Sin1	Turkey: Sinop, JGI	18	2x	Decena et al. 2023
	Bsylvaticum29	Greece : Evrytania, B100281411	18	2x	Decena et al. 2023
	Bsylvaticum30	Denmark: Zealand, B100566906	18	2x	Decena et al. 2023
	Bsylvaticum31	Germany: Ober Bayern, M0177011	18	2x	Decena et al. 2023
	Bsylvaticum32	Russia: Kaluga, M01769847	18	2x	Decena et al. 2023
	Bsylvaticum54	Morocco: Rif Mountains, UZ	18	2x	Decena et al. 2023
	Bsylvaticum63	Morocco: Mulay-Idriss, UZ	18	2x	Decena et al. 2023
	Bsylvaticum434	Ukraine: Krym: USDA PI639821	18	2x	Decena et al. 2023
	Bsylvaticum446	Iran: Ardebil, USDA PI380758	18	2x	Decena et al. 2023
	Bsylvaticum467	Spain: Huesca, Bessen, UZ	18	2x	Decena et al. 2023
	Bsylvaticum470	Spain: Guipuzcoa: San Sebastian, UZ	18	2x	Decena et al. 2023
	Bsylvaticum476	France: Hautes Pyrenees, Barbazan, UZ	18	2x	Decena et al. 2023
	Bsylvaticum501	France: Alpes Maritime. Roquefort les Pins, UZ	18	2x	Decena et al. 2023
	Bsylvaticum506	France: Herault, Saint Jean de Fos, UZ	18	2x	Decena et al. 2023
	Bsylvaticum508	France: Aude, Villesèque des Corbieres, UZ	18	2x	Decena et al. 2023
	Bsylvaticum552	Spain: Cadiz, Tarifa, UZ	18	2x	Decena et al. 2023
	Bsylvaticum554	Spain: Malaga, Canillas, UZ	18	2x	Decena et al. 2023
<i>B. sylvaticum</i> var. <i>breviglume</i> Keng	Bbreviglume33	Tibet: Gongbogyamda, LD135398	18	2x	Mo et al. 2023
	Bbreviglume34	Pakistan: Hazara, Paras-Shogran, M0175636	18	2x	Decena et al. 2023
<i>B. glaucovirens</i> (Murb.) T. Durand & B.D. Jacks	Bglaucovirens	Greece, Crete, B3151	16	2x	Decena et al. 2023
<i>B. kurilense</i> (Prob.) Prob.	Bkurilense9	Russia: Kuril Islands, Iturup, VLA 1625	18	2x	Tzvelev & Probatova 2019, Decena et al. 2023
<i>B. miserum</i> (Thunb.) Koidz.	Bmiserum67	Japan: Honshu, Aramakiob, RIKEN	18	2x	Tzvelev & Probatova 2019
<i>B. spryginii</i> Tzvelev	Bspryginii 29	Russia: Krasnodarskii Krai, VLA 11652	18	2x	Tzvelev & Probatova 2019, Decena et al. 2023
<i>B. pinnatum</i> (L.) P. Beauv.	Bpinnatum505	Norway: USDA PI345964	18	2x	Decena et al. 2023
	Bpinnatum515	Kazakhstan: USDA PI440176	18	2x	
<i>B. arbuscula</i> Gay ex Knoche	Barbuscula502	Spain: Canary Islands, La Gomera, INIA	18	2x	Sancho et al. 2022, Decena et al. 2023
<i>B. distachyon</i> (L.) P. Beauv.	Bdistachyon_Bd21	Irak: Salakkudin, JGI	10	2x	Catalán et al. 2012
<i>B. stacei</i> Catalán, Joch. Müll., L.A.J. Mur & T. Langdon	Bstacei_ABR114	Spain: Balearic islands, Formentera, JGI	20	2x	Catalán et al. 2012
<i>Triticum aestivum</i> L.		cultivar	42	6x	Genbank
<i>Oryza sativa</i> L.		cultivar	24	2x	Genbank
<i>Sorghum bicolor</i> (L.) Moench		cultivar	20	2x	Genbank



**Figure 1** Geographical distribution of the studied samples of the *Brachypodium sylvaticum* complex micro-taxa. Western taxa (green region): *B. sylvaticum* s. s. (green circles), *B. sprynginii* (green star), *B. glaucovirens* (green diamond); Eastern taxa (blue region): *B. sylvaticum* var. *breviglume* (blue circles), *B. miserum* (blue star), *B. kurilense* (blue diamond) (see Table 1 for more details)

## RESULTS

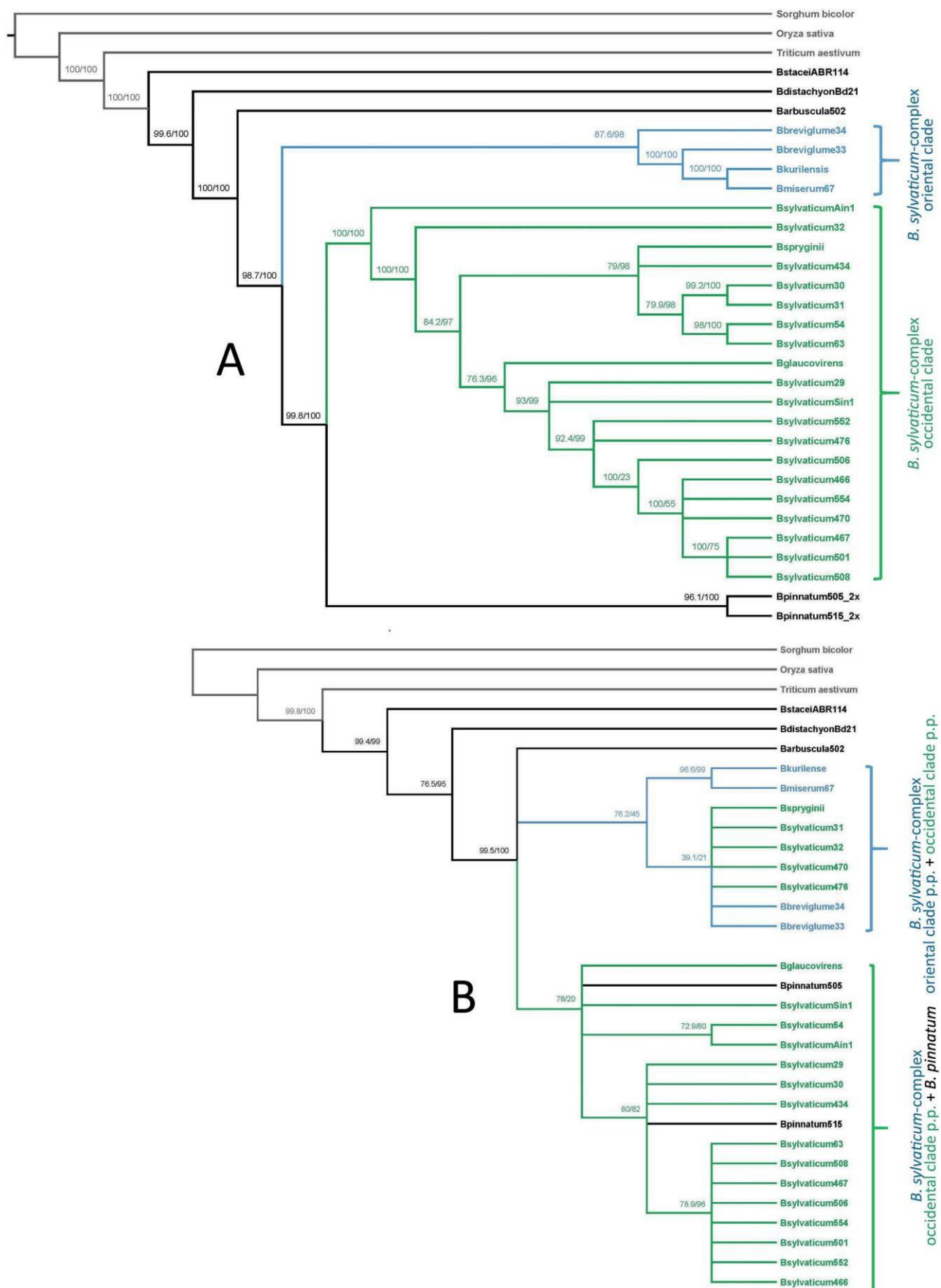
### Plastome and nuclear rDNA 35S phylogenies

The full *Brachypodium* plastome data set included 136,380 filtered positions and the 35S data set 6,237 positions. Whole plastome sequences of the *Brachypodium sylvaticum* complex species were highly conserved in terms of synteny and gene number: they contained a total of 133 genes (76 protein coding genes, 20 non-redundant *t*RNAs, four *r*RNAs in both inverted repeats, four pseudogenes, and two hypothetical open reading frames) and a structure and length similar to that of the *B. stacei* plastome (Sancho et al. 2018). The 35S showed a conserved structure and similar average lengths along its aligned transcriptional unit in all the samples studied (5'-external transcribed spacer (ETS) (~500 bp), 18S gene (~1,818 bp), internal transcribed spacers and 5.8S gene (ITS1-5.8S-ITS2) (~590 bp), 25S gene (~3,350 bp).

The best ML plastome tree retrieved a strong to well-supported topology for most *Brachypodium* lineages (100–79 % bootstrap support (BS), Fig. 2A). The earliest splits were those of the annual *B. stacei* and *B. distachyon* lineages, followed by the divergence of the *Brachypodium* core-perennial clade lineages. Within it, the Canarian *B. arbuscula* lineage branched off first, and the next divergences corresponded to the *B. sylvaticum*-complex oriental clade and then that of the sister *B. sylvaticum*-complex occidental

clade and *B. pinnatum* clade lineages. The oriental clade showed in turn the successive splits of the strongly supported Himalayan *B. sylvaticum* var. *breviglume* lineages and Pacific *B. miserum* / *B. kurilense* lineages, while the *B. sylvaticum* occidental clade reconstructed the nesting of *B. glaucovirens* and *B. spryginii* within this broadly homogeneous and non-geographically structured lineage, with several tips collapsing in polytomies (Fig. 2A).

The optimal ML 35S tree was less resolved than the plastome tree (Fig. 2B). This phylogeny also recovered the successive divergences of the early splitting *B. stacei*, *B. distachyon* and *B. arbuscula* lineages, followed by those of the more recently evolved core-perennials. Within the recentmost in-core group, all members of the *B. sylvaticum*-complex oriental lineages plus a few members of the occidental lineage (*B. spryginii*, and *B. sylvaticum* s. s. samples from eastern, central, and SW Europe) formed a moderately supported clade (76 % BS) with low internal resolution except for the strongly supported Pacific *B. miserum* / *B. kurilense* clade (96 % BS). The remaining members of the occidental clade collapsed in a series of nested polytomies in a moderately supported clade that also included the *B. pinnatum* samples (78 % BS); part of these lineages formed another moderately supported *B. sylvaticum* s. s. pro partim subclade (79 %) without clear geographic structure (but with most samples from the western Mediterranean region) (Fig. 2B).



**Figure 2** Maximum Likelihood phylogenomic trees of the *Brachypodium sylvaticum* complex taxa studied and other representative species of the genus. A – plastome tree cladogram. B – nuclear *rDNA* 35S gene tree cladogram. Ultrafast bootstrap / SH-aIrt support values are indicated above the branches. *Sorghum bicolor*, *Oryza sativa* and *Triticum aestivum* outgroups were used to root the trees. *Brachypodium sylvaticum* complex oriental and occidental lineages are shown in blue and green colors, respectively



## Bayesian dated tree.

We applied a total-evidence approach to retrieve a consensus phylogeny for this diploid *Brachypodium* tree and to infer the nodal ages of the *B. sylvaticum*-complex lineages (Fig. 3). The resolution of the combined plastome+35S tree reflected the matrilineal plastome topology, and the Bayesian dating analysis of its maximum clade credibility tree inferred a Mid-Miocene (9.04 Ma) origin for the MRCA of *Brachypodium* and a Late-Pliocene (2.71 Ma) origin for that of the core-perennial clade. The splits of the crown ancestors of the *B. sylvaticum*-oriental (2.0 Ma) and *B. sylvaticum*-occidental (0.97 Ma) ancestors were estimated to have occurred during the early Pleistocene, while the subsequent splits spanned the Late-Quaternary (Fig. 3). The dating analysis indicated that the *B. sylvaticum* var. *breviglume* ancestors were comparatively older (2.00–1.72 Ma) than that of the young Pacific *B. miserum* / *B. kurilense* group (0.36 Ma), while the western taxa showed recent but different ancestors (*B. spryginii* and close lineages 0.52 Ma; *B. glaucovirens* and close lineages 0.29 Ma) (Fig. 3).

## DISCUSSION

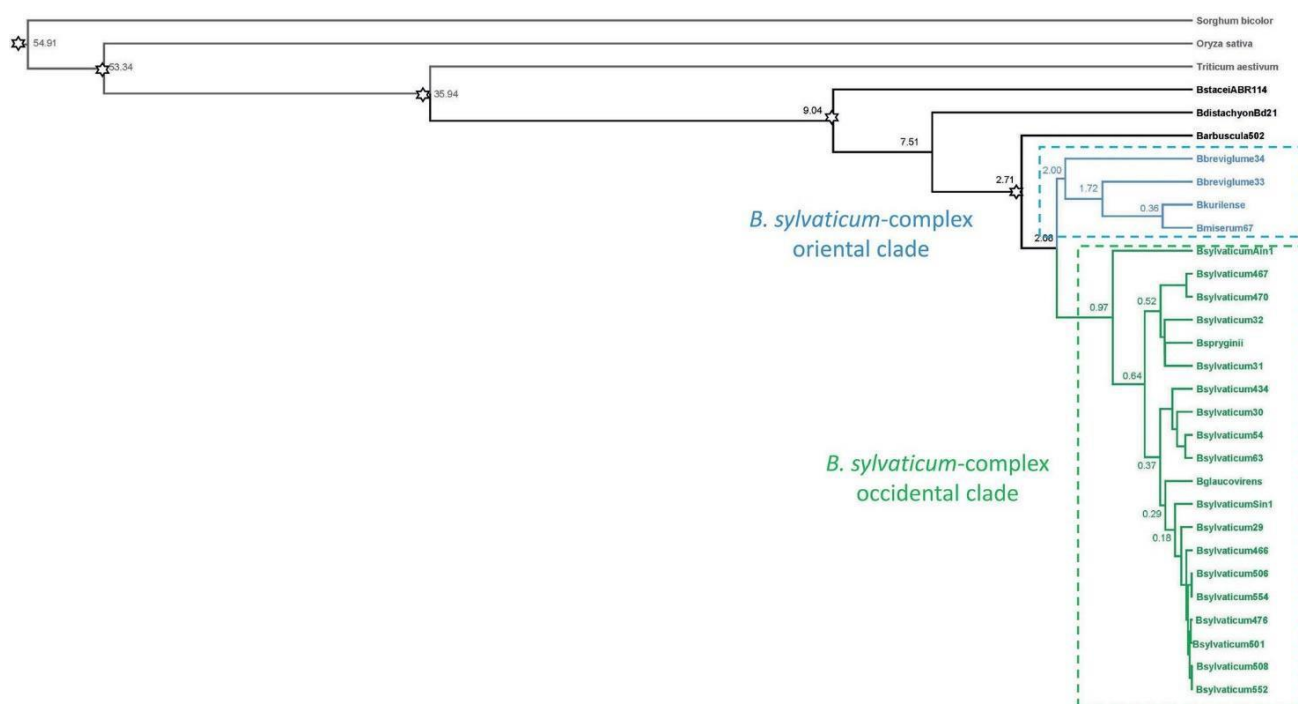
Our study has unveiled the existence of two main divergent lineages within the hitherto considered single monophyletic species *B. sylvaticum* or unclearly resolved *B. sylvaticum*-complex taxa (Figs 1, 2, 3; Catalan et al. 2016, Díaz-Pérez et al. 2018). Although *B. sylvaticum* s. l. has been

taxonomically split into several minor satellite taxa in its eastern and western distribution ranges (Tzvelev 1983, Scholz 2007, Tzvelev & Probatova 2019), the evolutionary relationships of both the oriental and occidental taxa and populations of *B. sylvaticum* s. l. were mostly unknown. Our analysis has demonstrated that this cytologically homogeneous diploid species complex is composed of two main lineages which are distributed in two largely disjunct Himalayas–Pacific and EuroMediterranean–Siberian regions (Figs 1, 2). Our plastome-based phylogeny clearly separated the two main *B. sylvaticum* s. l. diploid lineages, placing the oriental taxa within a first diverging clade, and the occidental taxa within a subsequently split *B. sylvaticum* s. s. clade (Fig. 2A), sister to its close *B. pinnatum* relative, a resolution mostly congruent with that of the rDNA 35S tree (Fig. 2B). The oriental clade showed the successive divergences of Himalayan *B. sylvaticum* subsp. *breviglume* lineages from their respective ancestors, assumed to have occurred in the Early Pleistocene (Gelasian, 2.0–1.7 Ma), while the two Pacific close species *B. miserum* and *B. kurilense* were inferred to have diverged from their common ancestor in the Middle Pleistocene (Chibabian, ~0.3 Ma; Fig. 3). The purported colonization ages of the Tibet and pan-Himalayan Pakistan mountains agreed with previous findings about the migrations of other cool grass lineages in South-Eastern Asia as well as components of the Himalayan flora (Shen et al. 2017) in the Late Pliocene–Early Pleistocene. Population genetic analyses detected high rates of genetic diversity for Tibetan population of *B.*

*sylvaticum* var. *breviglume* (Mo et al. 2013). The close relationship uncovered for the Far East *B. kurilense* and *B. miserum* lineages concurred with the main contribution of southern Japanese floristic elements to the northern Kurils' flora (Pietsch et al. 2003). The survival and adaptation of these diploid lineages to the cold climate conditions of NE Asia could have been favored by the existence of glacial refugia in western Beringia (Lozhkin et al. 2018). The highly diversifying history of the oriental *B. sylvaticum* complex group contrast with the large genomic homogeneity of the occidental lineage. The latter group was composed of European, Asian, and North African *B. sylvaticum* s. s. plus *B. glaucovirens* and *B. spryginii* lineages that did not reveal a strong geographic pattern in any of the phylogenetic trees reconstructed (Fig. 2). Several samples shared the same plastome sequence and collapsed in large polytomies in the plastome-based phylogenetic tree (Fig. 2A), while they split into independent but mostly invariable and unresolved lineages, sometimes mixed with oriental Himalayan and *B. pinnatum* lineages, in the 35S-based phylogenetic trees (Fig. 2B). Their origins from the crown ancestor were dated to more recent times than those of the oriental lineages, having presumably occurred in the last five hundred thousand years (Fig. 3). These results point towards a rapid spread of genomically homogeneous occidental *B. sylvaticum* s. s. lineages across the western Palearctic in late glacial and postglacial times with occasional crosses with oriental lineages (Figs 2, 3).

Moreover, some of the characterized *B. sylvaticum* s. s. invasive genotypes in northwestern North America were apparently introduced from occidental European ancestors (Rosenthal et al. 2008). The taxonomic implications of the phylogenomic results obtained in this study support the oriental micro-taxa *B. sylvaticum* var. *breviglume*, *B. kurilense*, and *B. miserum* as distinct to typical *B. sylvaticum* s. s. based on their short lower glume (more than 1.5 times shorter than the upper glume), which is shorter in *B. sylvaticum* var. *breviglume* (3–5 mm) than in the Pacific taxa (4.5–6 mm) (Keng 1982, Tzvelev 2015). *Brachypodium kurilense*, also subordinated to *B. miserum* in some taxonomic treatments, departs from the latter taxon based on the short prickles and hairiness of its spikelet pedicel and axis (Probatova & Skolovskaya 1985, Tzvelev 2015). These characters, although potentially variable, are in general fixed in the populations and constitute therefore phylogenetic signal traits of systematic value. The studied occidental micro-taxa (*B. glaucovirens*, *B. spryginii*) have distribution ranges that overlap with that of *B. sylvaticum* s. s. in the eastern Mediterranean and SW Asia, and in Ciscaucasia, respectively (Fig 1, Tzvelev 2015). *Brachypodium glaucovirens*, formerly synonymized to *B. sylvaticum*, was later recognized as a separate species (Scholz 2007). Morphologically it shows intermediate features, resembling *B. sylvaticum* in its short rhizome and long awn, and *B. pinnatum* in its bright green leaf color, broad leaf ribs, and erect panicle, and thus being a taxon of purported hybrid origin between the two species (Schippmann 1991).





**Figure 3** Bayesian maximum clade credibility dated chronogram of *Brachypodium sylvaticum* complex taxa and other congeners constructed with BEAST2 using plastome and *rDNA* 35S data showing estimated nodal divergence times (medians, in Ma) above branches. Stars indicate secondary nodal calibration priors (means  $\pm$  SD, in Ma) for the crown nodes of the Poaceae, BOP, *Brachypodium* + core pooids, *Brachypodium*, and *Brachypodium* core-perennial clades. *Brachypodium sylvaticum* complex oriental and occidental lineages are shown in blue and green colors, respectively

Our phylogenies corroborate this hypothesis and support a *B. sylvaticum* s. s.-type maternal progenitor species for *B. glaucoirens* (Fig. 2). *Brachypodium spryginii* differentiates from *B. sylvaticum* s. s. based on its more abundant plant pubescence and longer hairs (Tzvelev 2015). However, these variable characters are probably plastic and of low systematic value.

## ACKNOWLEDGEMENTS

This article is dedicated to the memory of our colleague Prof. Nina Sergeyevna Probatova, an excellent expert on the taxonomy of Eurasian grasses, who stimulated us to investigate the phylogeny and systematics of *Brachypodium kurilense*, a species described by her. We thank the B, LD, M, and VLA herbaria for facilitating the study of the loaned *Brachypodium* specimens and the INIA and USDA germplasm

banks, and the JGI and RIKEN institutes for providing seeds for our analyses. All the cytogenetic, bioinformatic and evolutionary analyses were performed in the Bioflora laboratory of the Escuela Politécnica Superior de Huesca (University of Zaragoza, Spain).

**Author contributions:** PC designed the study. PC, LAI, EP, and NP collected samples. MD, RS, LAI, and NP developed the experimental work. MD, RS, LAI, EP, JV, and PC analyzed the data and interpreted the results. PC wrote the manuscript. All authors contributed to the article and approved the submitted version.

**Funding:** This study was funded by the Spanish Ministry of Science and Innovation PID2019-108195GB-I00 and TED2021-131073B-I00 and by the Spanish Aragon Government Bioflora LMP82-21

research grants. MD was supported by a Spanish Ministry of Science predoctoral FPI fellowship. RS was supported by a Bioflora postdoctoral fellowship.

**Data Availability Statement:** Data generated for this study are available at Github (<https://github.com/Bioflora/BsylvaticumBotPacifica> (accessed on 21 April 2023)).

**Conflict of interest:** The authors declare no conflict of interest.

## LITERATURE CITED

- Bouckaert, R., J. Heled, D.K. Hnert, T. Vaughan & C.H. Wu 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* 10:1003537.
- Capella-Gutiérrez, S., J. M. Silla-Martínez & T. Gabaldón 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Carroll, A. & C. Somerville 2009. Cellulosic biofuels. *Annual Review of Plant Biology* 60:165–182.
- Catalán, P., D. López-Álvarez, A. Díaz-Pérez, R. Sancho & M.L. López-Herránz 2016. Phylogeny and evolution of the genus *Brachypodium*. In: *Genetics and genomics of Brachypodium* (J.P. Vogel, ed.), pp. 9–38, Springer International Publishing.
- Catalán, P., J. Müller, R. Hasterok, G. Jenkins, L.A.J. Mur, T. Langdon, A. Betekhtin, D. Siwinska, M. Pimentel & D. López-Álvarez 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany* 109:385–405.
- Catalán, P. & R.G. Olmstead 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.
- Decena, M.A., R. Sancho, J. Lusinska, R. Hasterok, R. Gorgojo, B. Montes, et al. 2023. Cyto-phylogenomics of the grass model genus *Brachypodium* unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends within this largely reticulate diploid-polyploid species-complex (under review).
- Díaz-Pérez, A., D. López-Álvarez, R. Sancho & P. 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. *Molecular Phylogenetics and Evolution* 127:256–271.
- Dierckxsens, N., P. Mardulyn & G. Smits 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45(4):e18.
- Dohleman, F.G. & S.P. Long 2009. More productive than maize in the Midwest: How does *Miscanthus* do it? *Plant Physiology* 150:2104–2115.
- Fox, S.E., J. Preece, J.A. Kimbrel, G.L. Marchini, A. Sage, K. Youens-Clark, M.B. Cruzan & P. Jaiswal 2013. Sequencing and De Novo Transcriptome Assembly of *Brachypodium sylvaticum* (Poaceae). *Applications in Plant Sciences* 1:1200011.
- Friedman, J. 2020. The evolution of annual and perennial plant life histories: Ecological correlates and genetic mechanisms. *Annual Review of Ecology, Evolution, and Systematics* 51:461–481.
- Glover, J.D., J.P. Reganold, L.W. Bell, J. Borevitz, E.C. Brummer, E. S. Buckler, C.M. Cox, T.S. Cox, T.E. Crews, S.W. Culman, ... & Y. Xu 2010. Increased food and ecosystem security via perennial grains. *Science* 328:1638–1639.
- Gordon, S.P., L. Liu & J.P. Vogel 2016. The genus *Brachypodium* as a model for perenniality and polyploidy. In: *Genetics and Genomics of Brachypodium*, (J.P. Vogel, ed.), pp. 313–325, Springer International Publishing.
- Hu, F., D. Wang, X. Zhao, T. Zhang, H. Sun, L. Zhu, F. Zhang, L. Li, Q. Li, D. Tao, B. Fu & Z. Li 2011. Identification of rhizome-specific genes by genome-wide differential expression analysis in *Oryza longistaminata*. *BMC Plant Biology* 11:18.
- Katoh, K. & D.M. Standley 2013. MAFFT Multiple sequence alignment software, Version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.
- Moreno-Aguilar, M.F., L.A. Inda, A. Sánchez-Rodríguez & P. Catalán 2022. Phylogenomics and systematics of overlooked Mesoamerican and South American polyploid broad-leaved *Festuca* grasses differentiate *F.* sects. *Glabriarcae* and *Ruprechtia* and *F.* subgen. *Asperifolia*, *Erosiflorae*, *Mallopetalon* and *Coironhuecu* (subgen. nov.). *Plants* 11(17):2303.
- Nguyen, L.-T., H.A. Schmidt, A. von Haeseler & B.Q. Minh 2015. IQ-TREE: a fast and effective stochastic

## Chapter 1

Botanica Pacifica. A journal of plant science and conservation. 2022. 12(1): 21–28 DOI: 10.17581/bp.2023.12119

- algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274.
- Pietsch, T.W., V.V. Bogatov, K. Amaoka, Y.N. Zhuravlev, V.Y. Barkalov, S. Gage, H. Takahashi, A.S. Lelej, S.Y. Storozhenko, N. Minakawa, ... & E.L. MacDonald 2003. Biodiversity and biogeography of the islands of the Kuril Archipelago. *Journal of Biogeography* 30(9):1297–1310.
- Probatova, N.S. 1985. Grass family – Poaceae Barnh. (Gramineae Juss.) In: *Vascular plants of the Soviet Far East, vol. 1* (S.S. Kharkevich, ed.), pp. 89–382, Nauka, Leningrad (in Russian). [Пробатова Н.С. 1985. Мятликовые, или злаки – Poaceae Barnh. (Gramineae Juss.) // Сосудистые растения советского Дальнего Востока / отв. ред. С.С. Харкевич. Л.: Наука. Т. 1. С. 89–382].
- Rosenthal, D.M., A.P. Ramakrishnan & M.B. Cruzan 2008. Evidence for multiple sources of invasion and intra-specific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Molecular Ecology* 17:4657–4669.
- Sade, N., M. Del Mar Rubio Wilhelmi, X. Ke, Y. Brotman, M. Wright, I. Khan, W. De Souza, E. Bassil, C.M. Tobias, R. Thilmony, ... & E. Blumwald 2018. Salt tolerance of two perennial grass *Brachypodium sylvaticum* accessions. *Plant Molecular Biology* 96:305–314.
- Sancho, R., C.P. Cantalapiedra, D. López-Alvarez, S.P. Gordon, J.P. Vogel, P. Catalán & B. Contreras-Moreira 2018a. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist* 218:1631–1644.
- Sancho, R., L.A. Inda, A. Díaz-Pérez, D.L. Des Marais, S. Gordon, J.P. Vogel, J. Lusinska, R. Hasterok, B. Contreras-Moreira & P. Catalán 2022. Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *The Plant Journal* 109:1535–1558.
- Schippmann, U. 1991. Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45:1–249.
- Scholthof, K.B.G., S. Irigoyen, P. Catalán & K.K. Mandadi 2018. *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30:1673–1694.
- Scholz, H. 2007. On the identity of *Brachypodium firmifolium* (Poaceae) from Cyprus. *Willdenowia* 37:215–220.
- Shen, Z., M. Yang, J. Feng, X. Li, P. Peng & Z. Zheng 2017. Geographic patterns of alpine flora in China in relation to environmental and spatial factors. *Biodiversity Science* 25:182–194.
- Steinwand, M.A., H.A. Young, J.N. Bragg, C.M. Tobias & J.P. Vogel 2013. *Brachypodium sylvaticum*, a model for perennial grasses: Transformation and inbred line development. *PLoS One* 8:1–11.
- Tzvelev N. 1983. *Brachypodium* P. Beauv. In: *Grasses of the Soviet Union* (N. Tzvelev, ed.), pp. 141–145, Academy of Sciences of the USSR, Smithsonian Institution, New Delhi.
- Tzvelev, N. 2015. On the genus *Brachypodium* P. Beauv. (Poaceae) in Russia. *Novosti Sistematiki Vysshikh Rastenii* 46:91–97 (in Russian). [Цвелев Н.Н. 2015. О роде *Brachypodium* Р. Беаув. (Poaceae) в России // Новости систематики высших растений. Т. 46. С. 91–97].
- Tzvelev, N.N. & N.S. Probatova 2019. *Brachypodium* P. Beauv. In: Tzvelev, N.N. & N.S. Probatova. *Grasses of Russia*, pp. 48–50, KMK Scientific Press, Moscow (in Russian). [Цвелев Н.Н., Пробатова Н.С. 2019. Род *Brachypodium* Р. Беаув. // Цвелев Н.Н., Пробатова Н.С. Злаки России. М.: КМК. С. 48–50].
- Wolny, E. & R. Hasterok 2009. Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Annals of Botany* 104:873–881.

## Cyto-phylogenomics of the *Brachypodium* grass species-complex unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends.

María Ángeles Decena<sup>1, 2, \*</sup>, Rubén Sancho<sup>1, 2, \*</sup>, Joanna Lusinska<sup>3</sup>, Robert Hasterok<sup>3</sup>, Rubén Gorgojo<sup>1</sup>, Beatriz Montes<sup>1</sup>, Ernesto Pérez-Collazos<sup>1, 2</sup>, Luís. A. Inda<sup>1, 4</sup>, Pilar Catalán<sup>1, 2</sup>

<sup>1</sup>Departamento de Ciencias Agrarias y del Medio Natural. Escuela Politécnica Superior de Huesca. Universidad de Zaragoza. C/ Carretera de Cuarte Km 1. E-22071 Huesca. Spain.

<sup>2</sup>Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, E-50059 Zaragoza. Spain.

<sup>3</sup>Plant Cytogenetics and Molecular Biology Group, Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, Katowice 40-032, Poland.

<sup>4</sup>Instituto Agroalimentario de Aragón-IA2 (Universidad de Zaragoza-CITA), Aragón, Spain.

\*Both authors corresponded equally as first coauthors

**Corresponding authors:** Luis Angel Inda. Departamento de Ciencias Pilar Catalán. Departamento de Ciencias Agrarias y del Medio Natural. Escuela Politécnica Superior de Huesca. Universidad de Zaragoza. C/ Carretera de Cuarte Km 1. E-22071 Huesca. Spain, [lainda@unizar.es](mailto:lainda@unizar.es). Pilar Catalán. Departamento de Ciencias Agrarias y del Medio Natural. Escuela Politécnica Superior de Huesca. Universidad de Zaragoza. C/ Carretera de Cuarte Km 1. E-22071 Huesca. Spain. [pcatalan@unizar.es](mailto:pcatalan@unizar.es).

## Abstract

Genomic diversity of plants usually explodes in reticulate species complexes formed by historically interbreeding lineages with different allo-autopolyploid series. Descendant dysploidy has been considered a mechanism that converts polyploids into functional diploids; however, reduced dysploid karyotypes could be also present in neopolyploids. The model grass genus *Brachypodium* contains more than twenty recognized species showing a remarkable descendant dysploidy and various ploidy levels. In contrast to the three well-known annual species, the remaining *Brachypodium* perennial species have been poorly investigated to date. We have performed an exhaustive cytogenetic and phylogenomic study of all its species (21 species, 26 cytotypes), and more specifically of 18 perennial species and minor satellite taxa. Genome size, chromosome, rDNA 35S and 5S FISH, and comparative chromosome painting (CCP) analyses across 147 populations have revealed previously unknown or neglected cytotypes for several species and have characterized the karyotypic profiles of their constituent genomes. Pan-plastomic analysis has demonstrated that all perennial taxa bear a large *B. stacei*-type plastome. Phylogenomics of plastome and nuclear rDNA 35S and 5S sequences evidenced the existence of two divergent lineages within diploid *B. sylvaticum*, a highly diverse oriental lineage consisting of micro-taxa closely related to the South Asian and Tropical and South African endemics, and a largely homogeneous occidental lineage, which may have arisen from post-polyploid diploidization. Short (S) and long (L) 5S haplomes were present in ancestral and recent perennial *Brachypodium* lineages and concurred in the intermediately evolved *B. retusum* and *B. phoenicoides*. A *Brachypodium* supertree revealed the participation of ancestral progenitor genomes in American *B. mexicanum*-4x, ancestral and recent progenitor genomes (with different chromosome base numbers and karyotypic profiles, and at different dosages) in aridic-to-mesic Mediterranean and Western European lineages (*B. retusum*-4x, 6x, *B. boissieri*-6x, *B. phoenicoides*-4x, 6x, *B. rupestre*-4x, 6x), and intermediate-recent progenitor genomes in currently diploid core-perennial mesic lineages (*B. arbuscula*-2x, *B. sylvaticum* s. l.-2x, *B. pinnatum* s. l.-2x). Phylogenetic inference suggests rapid evolution of distinct descendant dysploidy trends from ancestral or intermediate karyotypes ( $x = 10$ ,  $x = 9$ ) to different sorts of derived karyotypes ( $x = 8$ ,  $x = 5$ ), sometimes occurring in parallel in closely diverging lineages. Our hypothesized scenarios indicate that some of the descendant dysploidies predated the last polyploidizations, as these unique karyotypes correspond to orphan diploid progenitor genomes of current perennial *Brachypodium* polyploids, while others developed very recently and only affected distinct types of diploids.



**Keywords:** allo- and autopolyploidizations, *Brachypodium* perennial grasses, cytogenetics, descendant dysploidy, diploid-polyploid species complex, phylogenomics, plastomes, rapid karyotype evolution, rDNA 35S and 5S genes.

## Introduction

Cytotypic diversity has been a source of evolutionary novelty and speciation in most angiosperms (Soltis et al., 2014; Melichárková et al., 2020). Advanced cytogenetic and evolutionary research has demonstrated that ascendant series of auto or allopolyploids, which were considered in the past to be part of the same species, correspond in fact to independent species (Soltis et al., 2007; Soltis et al., 2009; Padilla-García et al., 2018). Most of these aggregates are constituted by closely related taxa that recently evolved from diploids to autopolyploids, or intercrossed, generating reticulate scenarios of homoploid hybrids or derived allopolyploids which returned to the diploid state, forming archetypal syngameons or diploid-polyploid species complexes (Soltis et al., 2016). The elements of many diploid-polyploid species complexes are often morphologically similar, hindering their recognition as separate independent species (Padilla-García et al., 2018). Although most of these hybridogenous complexes are constituted by recently evolved Pliocene-Pleistocene species, in some species-complexes the relatively young neopolyploids resulted from the cross of widely divergent diploid progenitor species (Melichárková et al., 2020). In general terms, the largest the genomic divergence of the parental species the highest the chromosomal stability of the nascent allopolyploid, and the fastest its reproductive isolation from the progenitor species and its speciation (Gordon et al., 2020). However, in many recent diploid-polyploid species complexes, occasional or continuous gene flow blurs the reproductive barriers between species. It creates new hybridogenous lineages, introgressed with progenitor or close species, which may adapt to different environments or colonize new areas but could still intercross with other polyploids or with the diploids, breaking their genetic isolation and obscuring the identification of newly arising species (Baduel et al., 2018; Melichárková et al., 2020). Descendant dysploidy, or polyploid drop, is a common phenomenon in grasses (Murat et al., 2010) and other angiosperms families, and has been viewed as a mechanism that converts polyploids into functional diploids (Mandáková and Lysak, 2018). However, reduced dysploid karyotypes could have been formed within diploids (Winterfeld et al., 2018), and may have participated in the origin of new meso- and neopolyploids (Mandáková et al., 2017). Angiosperm groups with multiple post-polyploid diploidizations (PPD) present multiple chromosome base numbers; these dysploidies could have

led to reproductive isolation among the PPD descendants contributing to diversification and speciation (Mandáková and Lysak, 2018).

*Brachypodium* is recognized today as a model genus for functional genomics of grasses and monocots and is characterized by its remarkable dysploidy (Hasterok et al., 2022; Scholthof et al., 2018). The three exhaustively studied annual species (diploids *B. distachyon* and *B. stacei*, and their derived allotetraploid *B. hybridum*) are considered a model system for allopolyploidy (Gordon et al., 2020), and one perennial species, *B. sylvaticum*, has been proposed as model plant for perenniality (Gordon et al., 2016; Scholthof et al., 2018). Systematic, cytogenetic and evolutionary studies demonstrated that *Brachypodium* constitutes an isolated lineage (monotypic tribe Brachypodieae), intermediately evolved within the Pooideae and sister to the main core pooids (Catalan et al., 2016), that exhibits a large dysploidy, consisting of plants with different chromosome base number ( $x = 10, 9, 8, 5$ ) (Betekhtin et al., 2014; Lusinska et al., 2019). Most of its 17 recognized species and cytotypes were corroborated or inferred to be polyploids (Schippmann, 1991; Catalan et al., 2016; Díaz-Pérez et al., 2018). The *Brachypodium* species show different native geographic distribution ranges and ecological preferences. Although the three annual species are broadly distributed in the circumMediterranean region, *B. distachyon* grows in more mesic, *B. stacei* in more aridic and *B. hybridum* in intermediate but also aridic habitats (López-Alvarez et al., 2015). Among the perennials, the most largely distributed taxa are the Eurosiberian moist-forest dwelling *B. sylvaticum* and continental *B. pinnatum*, and the Mediterranean aridic *B. retusum*, whereas the remaining taxa show more restricted distributions and narrower ecological niches in their respective native ranges (W European mesic *B. rupestre*, W Mediterranean wet-edaphic *B. phoenicoides*, C Mediterranean mesic *B. genuense*, E Mediterranean—SW Asia mesic *B. glaucovirens*, S Spain dolomitic *B. boissieri*, Canary isles Macaronesian *B. arbuscula*, South African afroalpine *B. bolussi*, moist-mesic Tropical and South African *B. flexum* and Madagascar *B. madagascariense*, Taiwan alpine *B. kawakamii*, and America Madrean and paramo mountain *B. mexicanum*) (Catalan et al., 2016). Current delimitations of some *Brachypodium* species, based in a taxonomic species concept, have classified within the same taxon several cytotypes showing different ploidy levels but similar morphological features (e. g., *B. pinnatum* 2x, 4x, *B. rupestre* 2x, 4x; *B. retusum* 4x, 6x) (Schippmann, 1991; Catalan et al., 2016). Recent comparative chromosome painting (CCP) and transcriptomic analyses have further characterized the karyotypic patterns of genomes and have identified the known and unknown (ghost) diploid progenitor genomes present in some *Brachypodium* polyploids (Sancho et al., 2022b). The annual

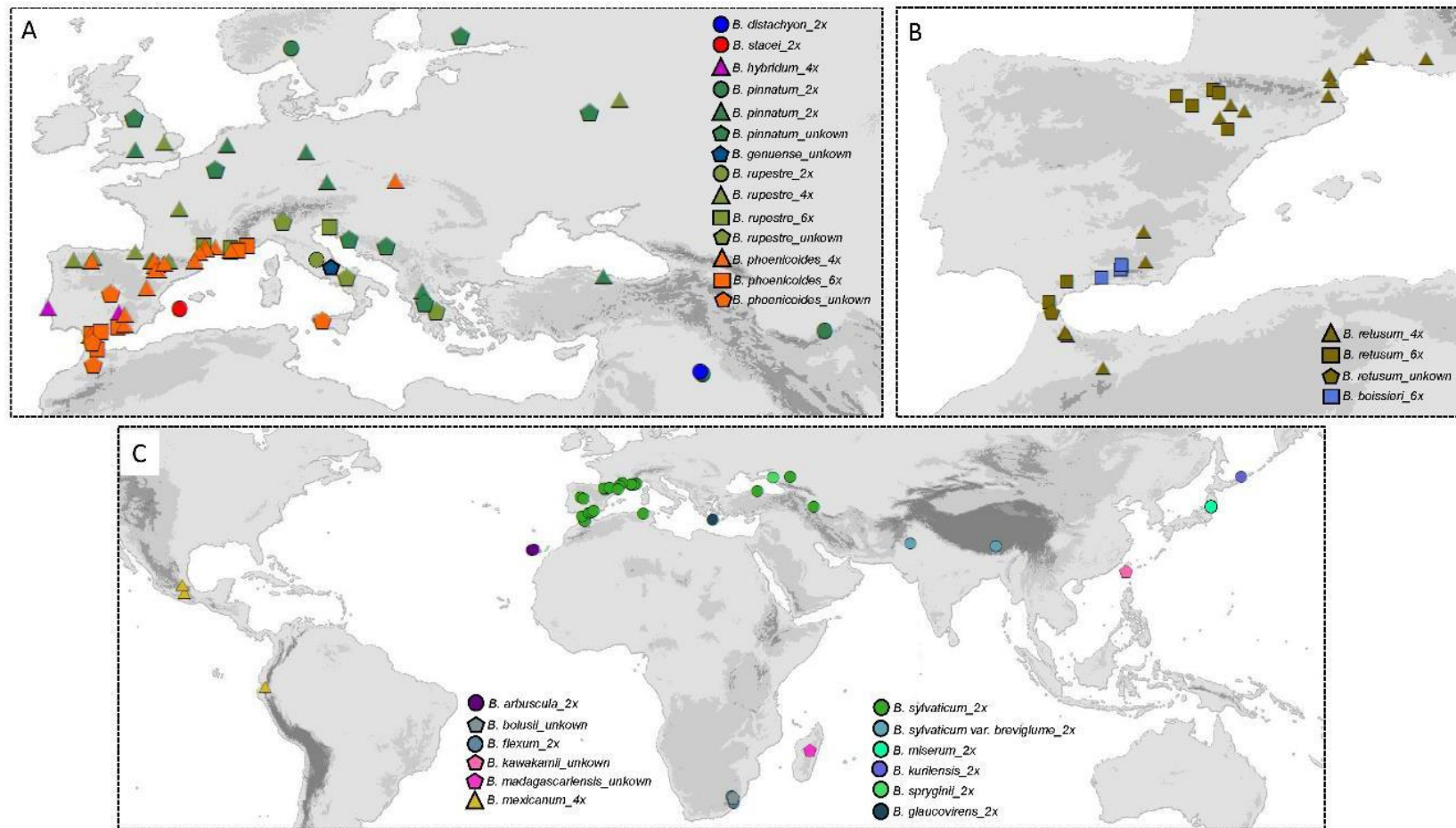
*B. hybridum* is the only allopolyploid species with extant diploid progenitor species, while the studied perennial polyploids have one or all their (sub)genomes inherited from unknown (probably extinct), diploid progenitor species. The transcriptomic and karyotypic CCP data revealed that most of the intermediately evolved (*B. retusum* 4x) and recently evolved (*B. rupestre* 4x, *B. phoenicoides* 4x) perennial polyploids were allopolyploids, while the ancestral polyploids showed signatures of segmental allopolyploidization (*B. mexicanum* 4x) or autopolyploidization (*B. boissieri* 6x) (Sancho et al., 2022b).

Although genetic and genomic resources are abundant for the model system *B. distachyon* (Gordon et al., 2017; Stritt et al., 2021) and for other annual species of its allopolyploid complex (*B. stacei*, *B. hybridum*) (Gordon et al., 2020b; Scarlett et al., 2022), fewer resources are available for the perennial species, except *B. sylvaticum* (Steinwand et al., 2013; Gordon et al., 2016). A phylogenetic and biogeographical study of *Brachypodium* using a few nuclear and plastid loci inferred a Mid-Miocene origin of its common ancestor in the Holarctic region and a Late-Miocene origin for its modern annual diploids in the Mediterranean region followed by a profusion of splits, range expansions, and different genome mergings for the perennial diploid genomes in the Mediterranean and Eurasian regions, with sporadic colonizations and further mergings in other continents during the Quaternary (Díaz-Pérez et al., 2018). However, more than half of the currently recognized *Brachypodium* species have not been analyzed genomically yet, and many of their infraspecific taxa or cytotypes have never been investigated molecularly. Cryptic diversity could be also associated to glacial refugia, where microendemisms could have evolved, and to poorly studied areas, where populations of broadly distributed species could have diverged (Fi et al., 2018; Struck et al., 2018). Plastid gene phylogenies of the annual *Brachypodium* species served to identify the maternal progenitors of various polyploids (Lopez-Alvarez et al., 2012; Díaz-Pérez et al., 2018; Sancho et al., 2022b), while plastome-based population studies corroborated the recurrent and bidirectional origin of allotetraploid *B. hybridum* and multiple chloroplast capture events between previously diverged *B. distachyon* lineages (Sancho et al., 2018; Gordon et al., 2020b). Nuclear rDNA genes (35S, 5S) have been profusely used as cytogenetic landmarks for inferring the allopolyploid nature of many angiosperms (Garcia et al., 2020; Rosselló et al., 2022), including the annual *Brachypodium hybridum* (Catalán et al., 2012), and to infer taxonomic relationships in *Brachypodium* (Shi et al., 1993; Wolny and Hasterok, 2009). In contrast to the predominant concerted evolution of 35S genes in angiosperms, the 5S genes are less prone to it, although allopolyploidization also tend to trigger the elimination of 5S families in some grasses as a



consequence of the genome doubling (Baum and Feldman, 2010). Both rDNA 35S and 5S regions have been used for reconstructing the evolutionary relationships of several grass groups (Baum et al., 2009; Mahelka et al., 2013; Moreno-Aguilar et al., 2022a, 2022b). Moreover, the co-occurrence of several 5S ribotypic families within the same plant has evidenced past hybridizations and recent allopolyploidizations and has been also used to detect progenitor haplomes in plant allopolyploids (Baum and Johnson, 2008; Baum and Feldman, 2010; Garcia et al., 2020; Vozárová et al., 2021). With its presumable diversity of rDNA 5S ribotypes along its abundant allopolyploid species and cytotypes, *Brachypodium* has yet to be investigated for 5S gene content and phylogenetic signal. Comparative chromosome barcoding (CCB), which enables the specific painting of whole chromosomes or specific regions, has proved to be effective in tracking the structural and evolutionary trajectories of individual chromosomes and whole karyotypes in *Brachypodium* (Lusinska et al., 2019; Sancho et al., 2022b). However, several *Brachypodium* cytotypes have not been explored through CCB approaches yet.

Therefore, the objectives of our study were to: i) scrutinize the potential cryptic cyto-diversity of perennial *Brachypodium* species through large population-level analysis of genome sizes, chromosome counts and rDNA 35S and 5S *in situ* hybridizations to estimate the number and distributions of cytotypes, and to reveal the existence of putative new *Brachypodium* cytotypes; ii) analyse the structure and composition of plastomes in the perennial *Brachypodium* species and reconstruct a plastome-based phylogeny for all the *Brachypodium* taxa and cytotypes; (iii) analyse the sequences and number of ribotypic 35S and 5S families in the perennial *Brachypodium* species and reconstruct rDNA-based phylogenies for all the *Brachypodium* taxa and cytotypes; (iv) detect all potential evolutionary lineages of *Brachypodium*, including new splits, and infer the maternal and paternal ancestors of polyploids through contrasted plastome and rDNA-based topological tests and estimate the nodal ages of a total-evidence *Brachypodium* supertree to elucidate the origins of new lineages within a timely evolutionary framework.



**Figure 1.** Geographical distribution of the studied populations of 21 *Brachypodium* taxa (26 cytotypes). (A) Annual species (*B. distachyon*, *B. stacei*, *B. hybridum*) and perennial taxa of the *B. pinnatum* complex (*B. pinnatum*, *B. rupestre*, *B. phoenicoides*, *B. genuense*). (B) Perennial *B. retusum* and *B. boissieri* species. (C) Perennial taxa of the Eurasian *B. sylvaticum* complex (*B. sylvaticum*, *B. sylvaticum* var. *breviglume*, *B. glaucovirens*, *B. kurilense*, *B. spryginii*) and other world-wide distributed perennial species (*B. arbuscula*, *B. bolusii*, *B. flexum*, *B. kawakamii*, *B. madagascariensis*, *B. mexicanum*). Colour codes for taxa and symbol codes for ploidy level (diploid: circle, tetraploid: triangle, hexaploid: square, unknown: polygon) are indicated in the corresponding charts.

## Results

### Cytogenetic study of *Brachypodium*

A total of 90 populations of perennial *Brachypodium* species (22 *B. sylvaticum* s. l., 22 *B. retusum*, 20 *B. phoenicoides*, 8 *B. pinnatum*, 13 *B. rupestre*, 5 *B. boissieri*), mostly from the highly diverse western Mediterranean region, allowed us to perform a comprehensive cytogenetic analysis of these *Brachypodium* taxa and to uncover new cytotypes (Figs. 1, S1; Table S1A). In contrast to annual species, which may present mixed populations containing individuals from more than one taxon or from different ancestries (Shiposha et al. 2020), the studied perennial populations were homogeneous with all individuals showing the same cytogenetic features (Table S1A). All the 22 surveyed samples of European, North African and Western (Iran), Central (Russian, Siberia) and Eastern (Japan; *B. miserum*) Asian populations of *B. sylvaticum sensu lato* showed  $2n = 2x = 18$  chromosomes, with a chromosome base number of  $x = 9$ , and a mean genome size (GS) of  $0.92 \pm 0.02$  2C/pg (Tables S1A), supporting the known diploid status of this species along its wide geographical range (Fig. 1). This ploidy level is the same as that found in other taxa described within the *B. sylvaticum* complex, showing the same  $x = 9$  (*B. sylvaticum* var. *breviglume*, *B. kurilense*, *B. miserum*, *B. spryginii*; Keng, 1982; Tzvelev, 1983; Tzvelev and Probatova, 2019) or different  $x = 8$  (*B. glaucovirens*; Wolny and Hasterok, 2009) chromosome base numbers. Three different ploidy levels ( $2x$ ,  $4x$ ,  $6x$ ) were detected in populations of the four species of the *B. pinnatum* complex (*B. pinnatum*, *B. rupestre*, *B. phoenicoides*, *B. genuense*) (Table S1A). Among the 8 analysed samples of *B. pinnatum sensu stricto*, populations from Western Asia (Iraq) contained diploids with  $2n = 2x = 16$  chromosomes (Lusinska et al., 2019), showing a different chromosome base number of  $x = 8$ , and a mean GS of  $0.97 \pm 0.01$ , while those from Northern Europe (Norway) and W and C Asia (Iran, Kazakhstan) contained diploids with  $2n = 2x = 18$  chromosomes ( $x = 9$ ) and a slightly lower GS (mean  $0.84 \pm 0.01$ ), and those of C Europe (Netherlands, Germany, Czech Republic) and E Mediterranean (Turkey, Greece) consisted of tetraploids with  $2x = 4x = 28$  and a mean GS of  $1.48 \pm 0.02$  (Figs. 1, S1A; Table S1A). Analysis of *B. genuense* (*B. rupestre sensu lato*) individuals indicated that one population from S Europe (Italy) contained diploids with  $2n = 2x = 18$  chromosomes ( $x = 9$ ). Most surveyed populations of *B. rupestre* from its native W European range (Spain, France, UK), plus one population from Russia, consisted of

tetraploids with  $2n = 4x = 28$  and GS values ranging from  $1.43 \pm 0.02$  to  $1.5 \pm 0.02$ ; noticeably, we discovered a new cytotype for this species in populations from southern France and Croatia, which contained hexaploids with  $2n = 6x = 38$  and GS mean value of  $2.27 \pm 0.02$  (Figs. 1, S1A; Table S1A). Of the 27 surveyed samples of *B. phoenicoides*, most populations from its native W Mediterranean region (Spain, France) consisted of tetraploids with  $2n = 4x = 28$  and a mean GS value of  $1.46 \pm 0.03$  (Fig. 1, Table S1A). Interestingly, we identified two different types of *B. phoenicoides* hexaploids; individuals from S Spain and NW Africa (Morocco) populations exhibited higher chromosome number ( $2n = 6x = 38$ ) and GS values ( $2.15 \pm 0.02$  to  $2.18 \pm 0.01$ ) than those from S France populations ( $2n = 6x = 36$ ;  $1.85 \pm 0.03$  to  $1.99 \pm 0.02$ ) (Fig. 1, Table S1A). Our cytogenetic survey within *B. retusum* detected two ploidy levels (4x, 6x) for this species; most populations from its native Mediterranean region (Spain, France, Morocco) contained tetraploids with  $2n = 4x = 32$  and GS values of  $1.73 \pm 0.4$  while some populations from northern and southern Spain consisted of hexaploids with  $2n = 6x = 42$  and GS values of  $2.37 \pm 0.17$  (Figs. 1, S1A; Table S1A). *Brachypodium retusum* individuals from Moroccan populations with mean GS values of  $1.86 \pm 0.2$  and those from S Spain populations with mean GS values of  $2.11 \pm 0.02$  were inferred to be, respectively, tetraploids and hexaploids. All analyzed individuals from the five surveyed populations of the southern Spain endemic species *B. boissieri* consistently detected hexaploids with  $2n = 6x = 48$  and a mean GS of  $3.15 \pm 0.05$  (Figs. 1, S1A; Table S1A). Within the perennial *Brachypodium* species, the lowest genome sizes correspond to diploids of the Canarian endemic *B. arbuscula* ( $2n = 2x = 18$ ,  $x = 9$ ; GS  $0.71 \pm 0.01$ ), and the highest to tetraploids of the American endemic *B. mexicanum* ( $2n = 4x = 40$ ,  $x = 10$ ; GS  $3.77 \pm 0.03$ ; Sancho et al. 2022) while the annual *Brachypodium* species present the smallest genomes of the genus (diploids *B. stacei*  $2n = 2x = 20$ ,  $x = 10$ , GS 0.56, and *B. distachyon*  $2n = 2x = 10$ ,  $x = 5$ , GS 0.63, and allotetraploid *B. hybridum*  $2n = 4x = 30$ ,  $x = 10 + 5$ , GS 1.26; Catalán et al., 2012) (Fig. 1; Table S1A).

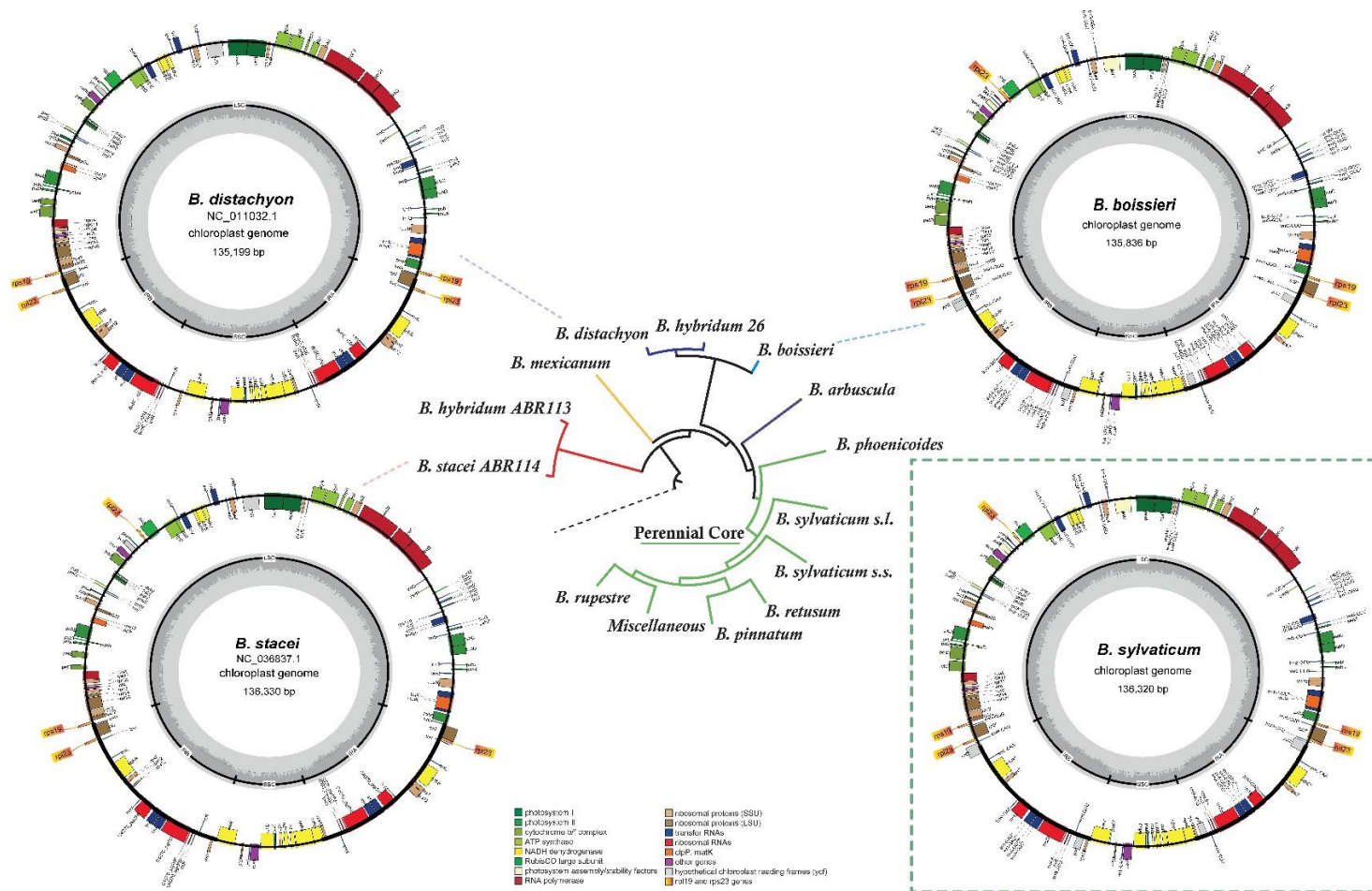
In situ hybridizations of rDNA genes detected different numbers of 35S and 5S loci in the surveyed individuals and studied species (Fig. S1B; Tables S1A; Wolny and Hasterok, 2009). In most cases these figures corresponded to the presumed number of loci, according to the inferred ploidy levels of the individuals; however, some individuals presented unexpected rDNA loci. For the more variable 35S markers, some diploids presented more than 2 loci (*B. pinnatum* 3-4, *B. genuense* 4, *B. glaucovirens* 4,

*B. sylvaticum* 4-6) and some hexaploids less than 6 loci (*B. rupestre* 5, *B. retusum* 5) (Fig. S1B; Tables S1A; Wolny and Hasterok, 2009). By contrast, the number of 5S loci detected in the studied individuals were in general highly consistent with their ploidy levels except for the *B. boissieri* hexaploids, which presented more loci than expected (8-9-10) (Fig. S1B; Tables S1A).

### ***Brachypodium* plastome analysis**

Newly assembled *Brachypodium* plastomes were deposited in ENA under project number PRJEB64465. Most plastid assemblies from whole-sequenced data and from genome skimming data produced a single plastome contig with a deep coverage of >584 per sample that contained its two inverted repeat regions (IRa, IRb), but 24 assemblies had to be constructed through a read mapping approach using the *B. sylvaticum*-Ain1 plastome as reference sequence (Table S2). Fully annotated *Brachypodium* plastomes were compared for gene composition and structural variation analysis (Fig. 2). Whole plastome sequences of perennial *Brachypodium* species were highly conserved in terms of synteny and gene number; the plastome lengths ranged from 135,241 bp (*B. madagascariense*) to 137,151 bp (*B. mexicanum* 504). All the perennial *Brachypodium* species plastomes shared a ~1,143 bp insertion between *psaI* and *rbcL* in the LSC region, corresponding to a *rpl23* pseudogene, and a ~301 bp corresponding to a duplicated *rps19* gene between *psbA* and *trnH* in the IRb repeat (Fig. 2). Sharing the first indel with the ancestral plastome of *B. stacei* and the second indel with the more recently evolved plastome of *B. distachyon* suggests that both were probably present in the most recent common ancestor (MRCA) of the genus and the loss of each of these fragments occurred exclusively in the respective lineages of the annuals *B. stacei* and *B. distachyon* (Sancho et al., 2018).



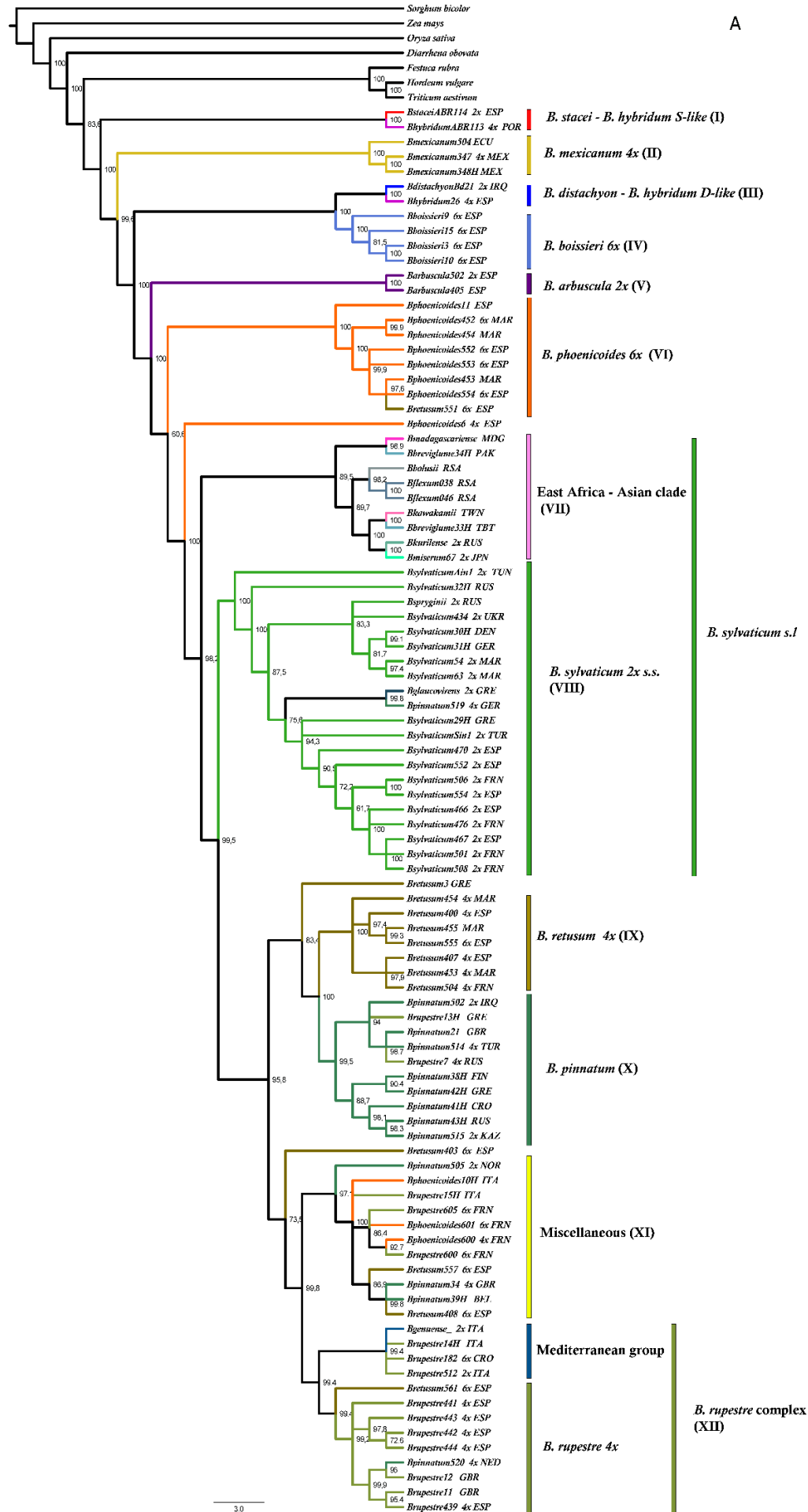


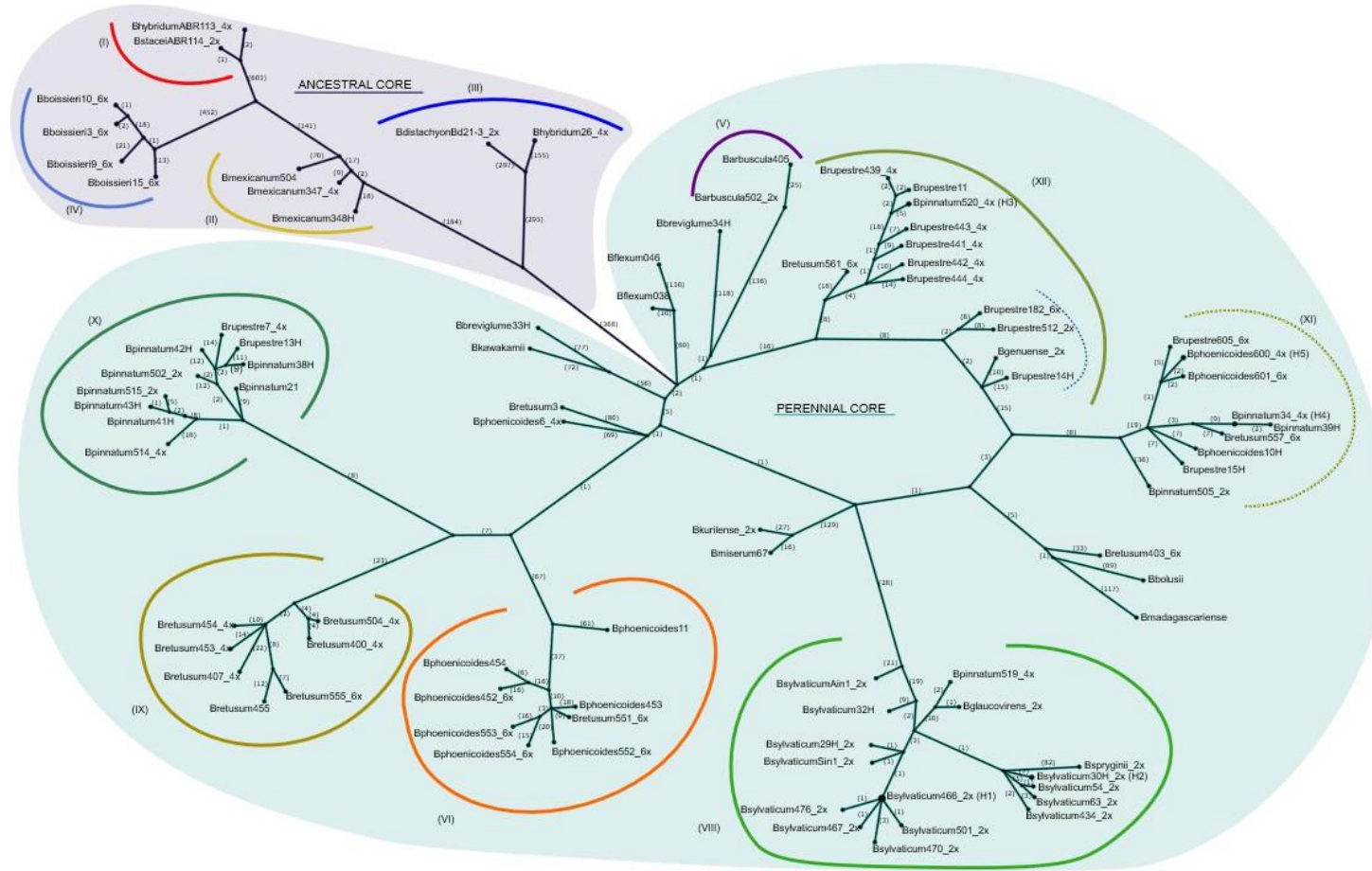
**Figure 2.** Plasmome maps of four representatives *Brachypodium* plastotypes: *B. distachyon* ABR6 (1161 bp deletion, rps19 locus duplication), *B. stacei* ABR114 (1161 bp insertion, rps19 locus deletion), *B. boissieri* 3 (1161 bp insertion, rps19 locus duplication), *B. sylvaticum* 467 (1161 bp insertion, rps19 locus duplication) (see text). Smaller inner circles correspond to a map of plasmome regions (LSC, SSC, IRA and IRB) and large outer circles to a map of plasmid genes and landmarks; colour codes of genes are indicated in the chart, landmarks (rps19 and rpl23 genes) are highlighted in orange. In the center of the figure a summarized plasmome-based ML tree showing the phylogenetic positions of the four plastotypes.

The full *Brachypodium* plastome data set included 136,380 filtered positions of which 13,708 were variable and 8,368 potentially informative. The best Maximum likelihood (ML) phylogenetic tree retrieved a strong to well-supported topology for most of the 12 *Brachypodium* lineages (I-XII; Fig. 3A). This tree uncovered new lineages corresponding to specific cytotypes (ploidy levels) of some perennial species and unveiled the maternal origin of the East African – S & E Asian endemic *Brachypodium* species. The earliest splits were those of the three annual *Brachypodium* and two ancestral perennial (*B. mexicanum*, *B. boissieri*) out-core lineages, which preceded the subsequent divergences of the more recently evolved core perennial clade lineages (Fig. 3A). Within the out-core group, the annual Mediterranean diploid *B. stacei* plus its derived allotetraploid *B. hybridum* S-plastotype lineage (I) split first, followed by the tetraploid *B. mexicanum* (II), which showed the sister relationship of its Mexican and S American lineages (Fig. 3A). The next divergence was that of the annual Mediterranean diploid *B. distachyon* plus its derived allotetraploid *B. hybridum* D-plastotype lineage (III), which formed a clade sister to the S Spain endemic hexaploid *B. boissieri* lineage (IV). Within the core perennial clade, the Canarian diploid *B. arbuscula* lineage (V) diverged first, followed by a poorly supported split of a strong *B. phoenicoides* 6x clade (VI) (including all hexaploid representatives but one plus two tetraploids from this species, and one *B. retusum*-4x sample) (Fig. 3A). The next divergences corresponded to two strongly supported lineages of the diploid Eurasian *B. sylvaticum sensu lato* complex species and close taxa. The strong relationships recovered for all branches of the first diverging East African – S & E Asian clade (VII) indicated that the Malagasy endemic *B. madagascariense* lineage shared its maternal ancestry with a diploid *B. sylvaticum* var. *breviglume* lineage from Pakistan and this clade was in turn sister to a clade of two sister African and Asian groups. The first Tropical and South African subclade showed the split of the montane diploid *B. flexum* and the afroalpine *B. bolussi* lineages, while the second Asian subclade retrieved the sister relationship of a Pacific sublineage formed by the sister diploids *B. miserum* from Japan and *B. kurilense* from the Kurils islands, and a South Asian sublineage composed by the sister diploid *B. sylvaticum* var. *breviglume* from Tibet and the alpine *B. kawakamii* from Taiwan (Fig. 3A). The second highly supported *B. sylvaticum sensu stricto* clade (VIII) included widely distributed representative samples from Europe, N Africa and SW Asia of *B. sylvaticum* subsp. *sylvaticum* plus close species. The different lineages did not show a clear geographic structure and some of them diverged separately [(i) a relatively well-supported clade included E, C & N Europe

(Russia, Ukraine, Germany, Denmark) and NW African (Morocco) lineages plus one Caucasian *B. spryginii* lineage; (ii) a strong clade was formed by E Mediterranean (Greece, Turkey) and W Mediterranean (N and S Spain, S France) lineages, many of which were highly homogeneous and collapsed in a large polytomy; (iii) a highly supported clade was composed of the Cretan diploid *B. glaucovirens* and its sister C European (Germany) tetraploid *B. pinnatum* lineage] (Fig. 3A). The most recently evolved branches of the *Brachypodium* plastome tree showed strong relationships for the *B. pinnatum* species complex and *B. retusum* lineages. The common ancestor of this group split into two clades; within the first clade a tetraploid *B. retusum* subclade (IX) that constitutes a new cytotypic-specific lineage was sister to a *B. pinnatum* subclade (X) which showed the divergences of a *B. rupestre* complex subclade (XI) and a miscellaneous subclade containing representatives from four species (XII) (Fig. 3A). The *B. pinnatum* clade (X) included diploid ( $x = 9$ ,  $x = 8$ ) and tetraploid representatives of this species and two *B. rupestre* samples; the group diverged into different lineages that did not show a clear geographic pattern [a subclade composed by E Europe and Siberia (Croatia, Russia, Kazakhstan) samples, and its sister subclade containing W, E & N Europe (UK, Bosnia, Greece, Finland) and SW & C Asia (Turkey, Irak) samples plus *B. rupestre* from S Europe (Greece)] (Fig. 3A). The *B. retusum*-4x clade (IX) comprised all detected tetraploids of this species; the common ancestor split into different lineages, recovering strong but separated relationships of different Moroccan lineages with, respectively, S Spain and N Spain plus S France lineages (Fig. 3A). The miscellaneous clade (XI) included representative *B. pinnatum* 2x and 4x, *B. retusum* 6x, *B. phoenicoides* 4x and 6x, and *B. rupestre* 6x samples from Europe. In this group, the hexaploid *B. retusum* samples from N and S Spain were closely related to tetraploid *B. pinnatum* samples from W Europe (UK, Belgium) while the hexaploid *B. rupestre* samples from S France were to tetraploid and hexaploid *B. phoenicoides* samples from the same area, and both groups to *B. phoenicoides* and *B. rupestre* samples from Italy. Other independent *B. retusum* 6x and *B. phoenicoides* 4x lineages from N Spain were found to be closest to, respectively, the *B. pinnatum* complex clade and the *B. phoenicoides* 6x clade (Fig. 3A). The *B. rupestre* complex clade (XII) showed two strong sister subclades; one included diploid *B. genuense* samples from Italy plus the hexaploid *B. rupestre* representative from Croatia (Mediterranean clade), and the other tetraploid *B. rupestre* samples from W Europe (N Spain, UK, Netherland) (Fig. 3A).







**Figure 3.** (A) Maximum Likelihood plastome tree of 95 *Brachypodium* accessions representing all known species and cytotypes. Numbers on branches indicate UltraFast Bootstrap supports (BS). Branches with likelihood bootstrap support values <30% were collapsed into polytomies. Ploidy levels and geographic origins of studied samples are indicated on terminals. *Sorghum bicolor*, *Zea mays*, *Oryza sativa*, *Diarrhena obovata*, *Festuca rubra*, *Triticum aestivum* and *Hordeum vulgare* were used to root the trees. Color codes of *Brachypodium* accessions correspond to those indicated in Figure 1. *Brachypodium* lineages and groups are shown in the right part of the figure. Scale bar: number of mutations per site. Country codes correspond to those indicated in Supplementary Table S1B. (B) Haplotypic parsimony network constructed with 4,070 syntenic SNPs of *Brachypodium* plastomes using the TCS algorithm. The number of mutation steps are indicated on branches. Ancestral out-core and recent core-perennial evolutionary lineages (I-XII) correspond to the main haplotypic groups.

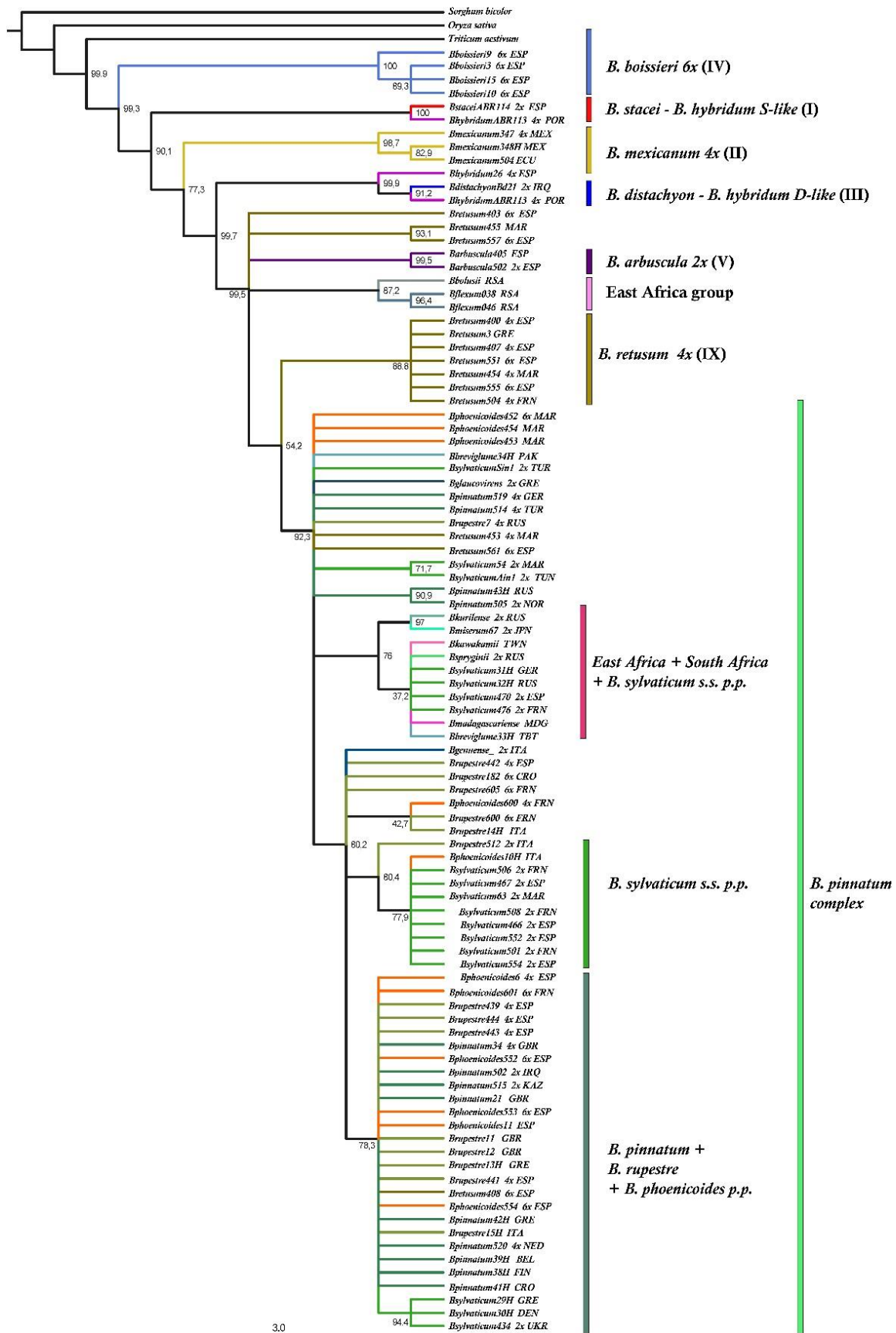
The plastome haplotypic network based on 4,070 segregating sites (SNPs), of which 2,421 were parsimony-informative, reconstructed a similar evolutionary scenario for the studied *Brachypodium* samples and allowed us to identify 87 distinct plastohaplotypes (Fig. 3B), of which two haplotypes were taxon- or cytotype-specific [H1 (*B. sylvaticum*466\_2x, *B. sylvaticum*506\_2x, *B. sylvaticum*508\_2x, *B. sylvaticum*552\_2x, *B. sylvaticum*554\_2x), H2 (*B. sylvaticum*30H\_2x, *B. sylvaticum*31H)], three were shared by different taxa [H3 (*B. pinnatum*520\_4x, *B. rupestre*12), H4 (*B. pinnatum*34\_4x, *B. retusum*408\_6x), H5 (*B. phoenicoides*600\_4x, *B. rupestre*600\_6x)], and 82 were sample-specific (Fig. 3B). The haplotypic network was fully resolved except for one internal loop and showed the relationships of the most ancestral groups to the recent core perennial groups (Fig. 3B). The out-core ancestral haplotypes separated by 368 step mutation from the core-perennial group. The *Brachypodium* plastome genomic diversity was overall high within the genus [number of segregating sites ( $S$ ) = 4,070; haplotypes ( $h$ ) = 87; Haplotype diversity ( $H_d$ ) = 0.9969; Nucleotide diversity ( $\pi$ ) = 0.0685209]. The samples clustered into 12 main groups which were highly concordant to those of the inferred plastome phylogeny (Figs. 4A, 4B; Table S3). In terms of number of variable sites, our analysis indicated that the *B. distachyon* + *B. hybridum*-D (III) (452) and the *B. stacei* + *B. hybridum*-S (I) (3) were, respectively, the most and the less variable groups. The ancestral *B. mexicanum* group (II) (116) showed more than twice variable sites than the *B. boissieri* group (IV) (56). Within the recent core perennial group, *B. phoenicoides* (VI) (227) and *B. arbuscula* (V) (25) had, respectively, the highest and the smallest number of variable sites. The lowest haplotype diversity index ( $H_d$ ) was found in the *B. sylvaticum* (VIII) group (0.942) followed by the miscellaneous (XI) (0.963) and *B. rupestre*-4x (XII) (0.987) groups. The largely sampled *B. sylvaticum* group (VIII) was the only group that showed a statistically significant ( $P < 0.05$ ) negative Tajima's  $D$  value, suggesting that the *B. sylvaticum* populations might have been under selective sweep or population expansion after a recent bottleneck (Table S3).

### ***Brachypodium* nuclear rDNA 35S and 5S analysis**

Comparative genomics and phylogenomic analyses for the same *Brachypodium* samples used in the plastome analysis were also performed using nuclear ribosomal DNA 35S and 5S genes (Figs. 5, 6, S2; Tables S1B, S4). The newly sequenced rDNA 35S cistrons were deposited in ENA with project number PRJEB64129 (Table S1B). We

obtained a single contig for the rDNA 35S cistron of the studied samples, which ranged from 8,135 bp (*B. rupestre* 14H) to 6,948 bp (*B. boissieri* 15), with a coverage depth relatively constant across all the sequences (>250). The aligned *Brachypodium* 35S data set included 6,197 positions of which 621 were variable and 288 potentially informative. The 35S showed a conserved structure and similar average lengths along its aligned transcriptional unit in all the samples studied [5'-external transcribed spacer (ETS) (~456 bp), 18S gene (1,810 bp), internal transcribed spacers and 5.8S gene (ITS1-5.8S-ITS2) (595 bp), 25S gene (3,344 bp)].

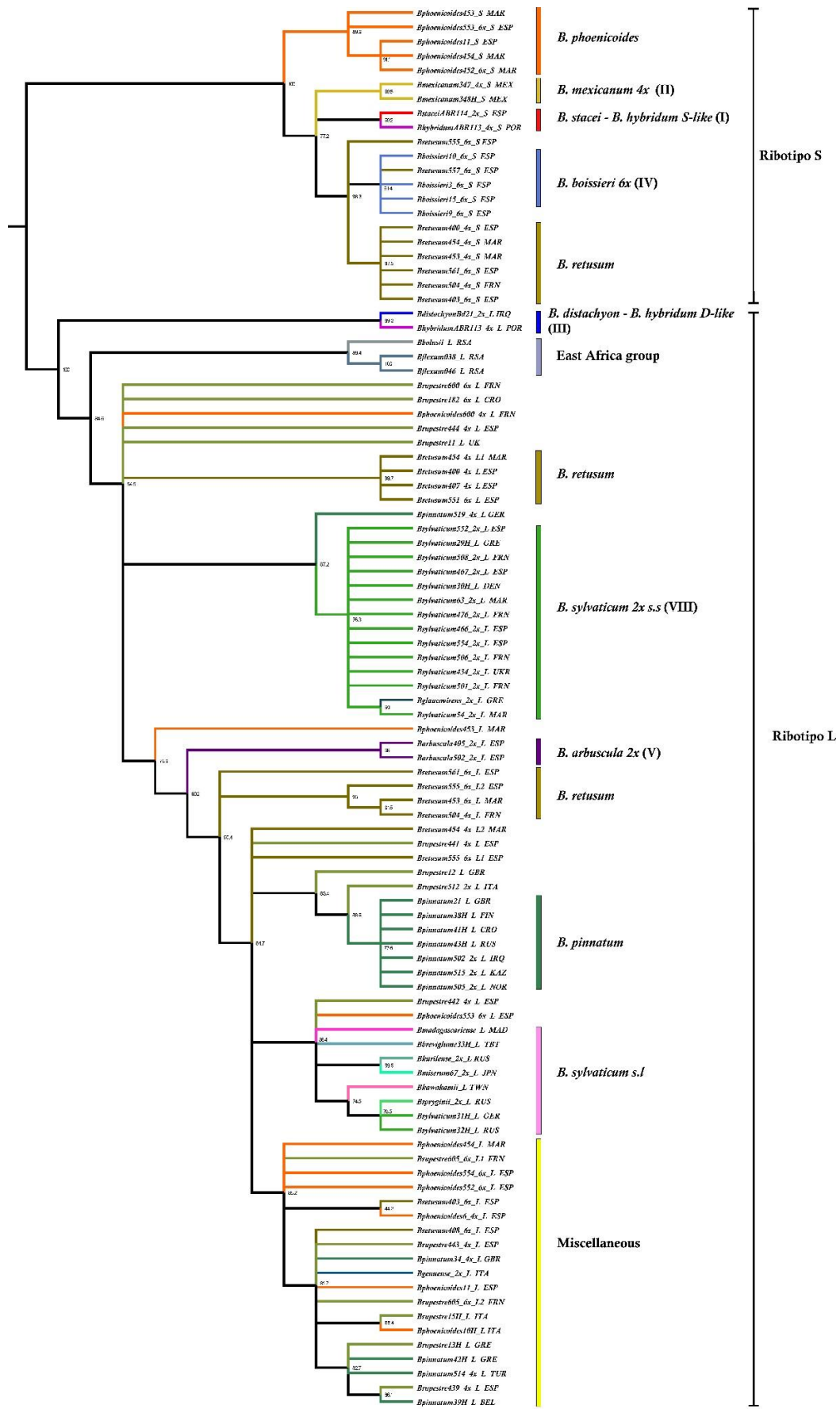
The best ML 35S phylogenetic tree was less resolved than the plastome-based tree, retrieving only 8 of its 12 lineages (Fig. 4); the out-core lineages were well supported in the 35S topology while most of the core-perennial lineages were weakly supported and collapsed in polytomies. The earliest split in the 35S phylogeny corresponded to the ancestral *B. boissieri* lineage, followed by *B. stacei* plus its derived allotetraploid *B. hybridum* lineage (S-plastotype, with *B. stacei*-type BhS subgenome) (Fig. 4). The next divergences were those of the ancestral *B. mexicanum* and of *B. distachyon* plus its derived allotetraploid *B. hybridum* lineages (D- and S-plastotypes, with *B. distachyon*-type BhD subgenomes) (Fig. 4). Within the core-perennial clade, a large polytomy included the strongly to relatively well supported *B. arbuscula*, East Africa (*B. flexum* / *B. bolussi*), *B. retusum*-4x lineages and a clade of remaining perennials (Fig. 4). The last group was largely unresolved except for the moderately supported Madagascar – S & E Asian *p. p.* (*pro partim*) and *B. pinnatum* complex *p. p.* clades. The African-Asian group included a strong Pacific lineage of the sister *B. miserum* and *B. kurilense*, plus *B. madagascariense*, *B. kawakamii*, and a *B. sylvaticum* var. *breviglume* Himalayan lineage; it also incorporated Caucasian *B. spryginii* and some Russian and European *B. sylvaticum* s. *l.* lineages (Fig. 4). The *B. pinnatum* complex group was largely unresolved in a polytomy that included *B. genuense*, all *B. rupestre* 4x and 6x, most *B. phoenicoides* 4x and 6x, some *B. pinnatum* 2x and 4x, one *B. retusum* 6x, and two *B. sylvaticum* s. *s.* lineages (SW Europe and NW Africa). *B. glaucovirens*, and some *B. pinnatum*, *B. phoenicoides*, and *B. retusum* (4x, 6x) lineages showed different nestings within the core-perennial clade (Fig. 4).



**Figure 4.** ML 35S gene tree of 95 *Brachypodium* accessions representing all studied taxa and cytotypes. Numbers on branches indicate UltraFast Bootstrap supports (BS). Branches with likelihood bootstrap support values <30% were collapsed into polytomies. Ploidy levels and geographical origins of studied samples are indicated on terminals. *Sorghum bicolor*, *Oryza sativa* and *Triticum aestivum* were used to root the tree. Color codes of *Brachypodium* species correspond to those indicated in Figure 1. *Brachypodium* lineages and groups are shown in the right part of the figure. Country codes correspond to those indicated in Supplementary Table S1B. Scale bar: number of mutations per site.

The newly assembled sequences of the nuclear rDNA 5S region ranged from 463 bp (*B. flexum*) to 270 bp (*B. stacei*) (Fig. S2; Table S4). The sequences were deposited in ENA with project number PRJEB60266 (Table S1B). In contrast to *Triticeae* grasses where diploid species present two paralogous 5S ribotypes (short and long 5S sequences) and allopolyploids less copies than expected according to ploidy level (Baum and Johnson, 2008; Baum et al., 2009; Baum and Feldman, 2010), *Brachypodium* diploids only showed one copy, short in the ancestral lineages and long in the recent lineages, while some intermediately evolved *B. retusum* and *B. phoenicoides* allopolyploids presented the two 5S ribotypic families (short and long). The 5S region consisted of a conserved 5S gene (120 bp in all species) and a variable intergenic spacer (IGS) of 243 bp in the whole 5S aligned data set. In contrast to the homogeneous 35S region, the 5S region showed different types of structural IGS length families within *Brachypodium* (Fig. 5; Tables S4A, S4B). The short 5S haplome (5S-S) of 272 bp (270 to 273 bp) was shared by the ancestral outcore *B. mexicanum*, *B. stacei* (plus the BhS subgenomes of *B. hybridum* S and D plastotypes) and *B. boissieri* species. *Brachypodium distachyon* and the core-perennial species shared the long 5S haplome (5S-L), having a more variable IGS region of 383bp (371- to 463 bp) (Table S4B). The *B. phoenicoides* and *B. retusum* 4x and 6x cytotypes had both short and long IGS families, except for some samples in which only one ribotype was found (5S-S: *B. phoenicoides* 452 and *B. retusum* 557; 5S-L: *B. retusum* 407, 408 and 551). Moreover, *B. retusum* 454-4x and *B. retusum* 555-6x showed three ribotypes, one 5S-S copy and two functional 5S-L copies with nearly identical sequences (77% pairwise identity), while *B. rupestre* 605, a recently evolved core perennial lineage, presented two paralogous 5S-L copies with 98% of pairwise identity (Fig 5; Table S4B).





**Figure 5.** ML 5S gene tree of 84 *Brachypodium* accessions showing the resolution of short (5S-S) and long (5S-L) 5S sequences. Numbers on branches indicate UltraFast Bootstrap supports (BS). Branches with likelihood bootstrap support values <30% were collapsed into polytomies. Ploidy levels and geographic origins of studied samples are indicated on terminals. ML tree was rooted at mid-point. Color codes of *Brachypodium* species correspond to those indicated in Figure 1. *Brachypodium* lineages and groups are shown in the right part of the figure. Country codes correspond to those indicated in Supplementary Table S1B. Scale bar: number of mutations per site.

The optimal ML 5S phylogeny (Fig. 5) was better resolved than the 35S-based tree (Fig. 4). The 5S topology reflected the divergence between the short-IGS and long-IGS 5S families and retrieved 10 of the 12 plastome-based lineages (Fig. 5). The first split separated the fully supported 5S-S and 5S-L lineages. Within the first group, two sister and relatively well supported clades showed an ancestral group of *B. stacei* (plus its derived *B. hybridum* S-plastotype BhS subgenome), *B. mexicanum*, and *B. boissieri* + *B. retusum* 4x- 6x strongly-supported lineages collapsed in a polytomy and a clade of *B. phoenicoides* 6x lineages (Fig 5). Within the second group, the first split separated the well supported annual *B. distachyon* and its derived *B. hybridum* BhD subgenome clade, followed by the divergence of the East Africa clade and a polytomy of remaining core perennial lineages. Relatively well-supported *B. arbuscula*, *B. retusum* 4x, *B. retusum* 6x, *B. sylvaticum* s.s. *p. p.*, *B. sylvaticum* complex *p.p.*, *B. pinnatum* *p. p.*, and *B. pinnatum* complex *p. p.* clades fell within the last group (Fig 5).

### Supertree reconstruction and dating analysis of the *Brachypodium* phylogeny.

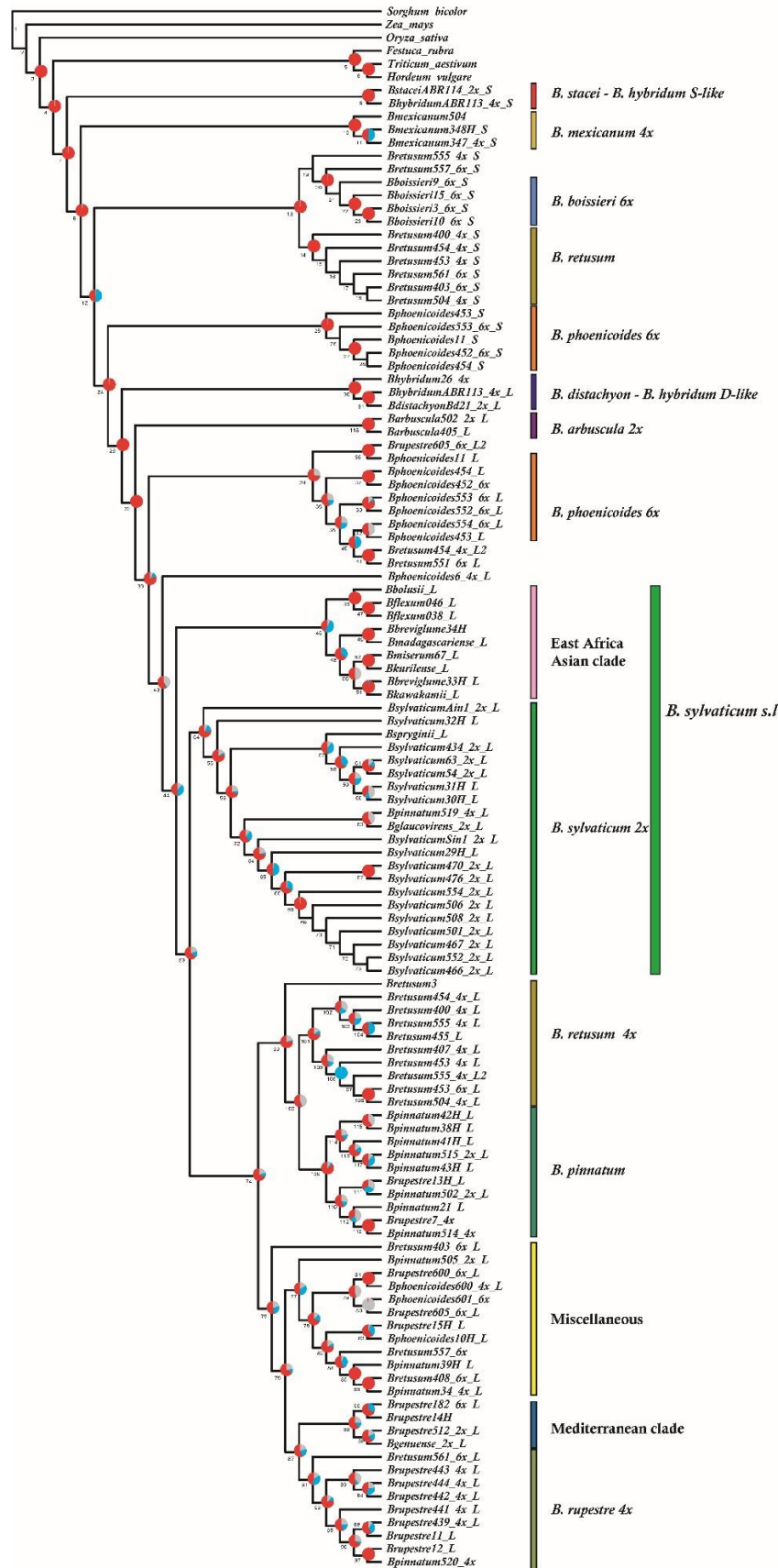
To avoid the overriding weight of the plastid data over the nuclear data and to uncover the potential diploid genomes that contributed to the origins of the different polyploid cytotypes, we adopted a multispecies coalescent (MSC) phylogenetic reconstruction approach that incorporated the resolution of the biparentally inherited 5S (and 35S)-based trees with the maternally inherited plastome-based trees to yield the final species-tree phylogeny within a supertree framework (Bininda-Emonds, 2014). The *Brachypodium* phylogenetic trees retrieved from the MSC approaches recovered topologies that were relatively highly congruent with each other as shown in the computed

*Brachypodium* ASTRAL supertree (Fig. 6). Estimation of the proportion of gene tree quartets that agree with the ASTRAL species tree through normalized quartet scores indicated considerable intragenomic congruence for the ancestral *Brachypodium* lineages and increasing levels of incongruence for the more recently evolved core-perennial

lineages (Fig. 6, Tables S5A, S5B). The proportion of gene tree quartets concordant with the species tree was 74.5%. The MSC analysis revealed that nodal branches obtained values between 3.7-100% for the main topology (*q1*: 35 nodes out of 118 with 100%

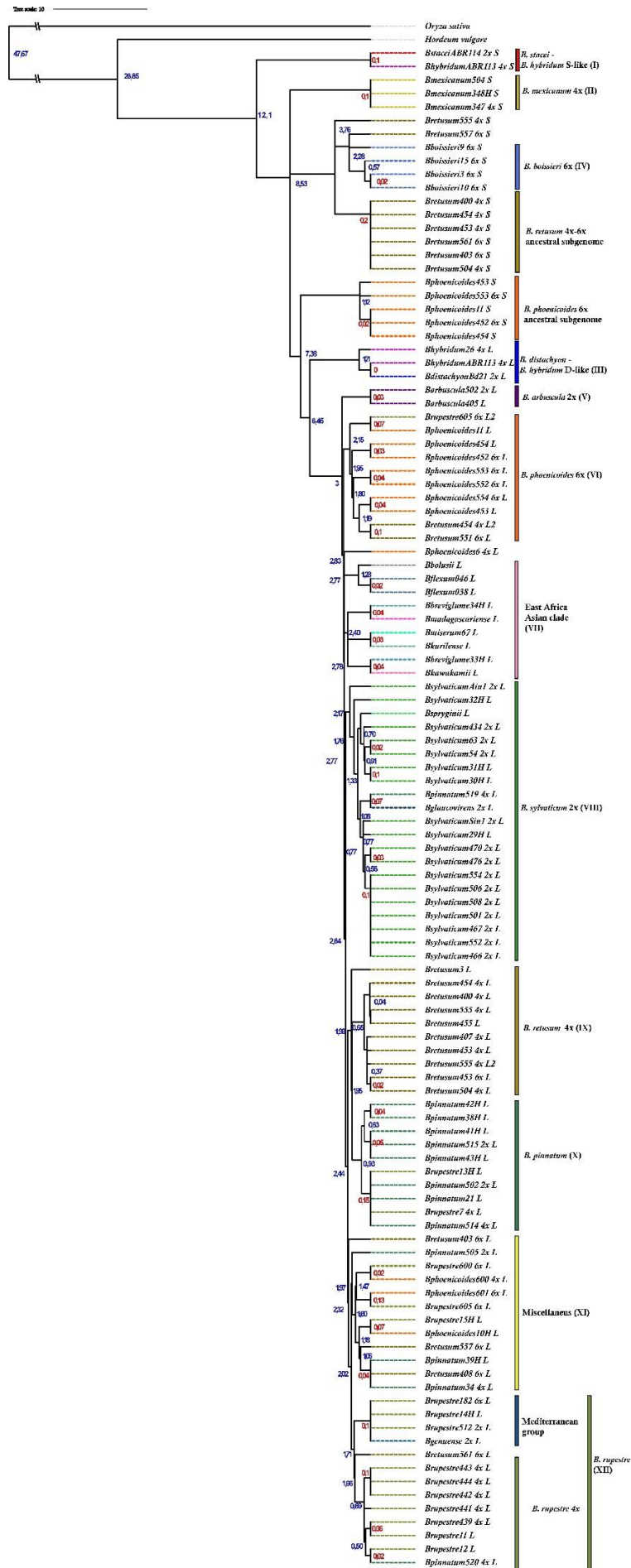


support), 0-100 % for the first alternative topology (q2: one node with 100%), and 0-96.3 % for the second alternative topology (q3: zero nodes with 100%) (Table S5B). Genomic discordance between nuclear rDNA gene data and the plastome data was low for the outcore *Brachypodium* lineages and high for the intermediately evolved *B. retusum* and some of the most recently diverged *B. pinnatum* complex (*B. phoenicoides*, *B. pinnatum*, *B. rupestre*) lineages (Fig. 6). The *Brachypodium* supertree showed the successive ancestral divergences of the out-core lineages *B. stacei* (+ *B. hybridum*-BhS), *B. mexicanum*, the *B. boissieri* + *B. retusum*-4x-6x 5S-S clade, *B. phoenicoides*-6x 5S-S, and *B. distachyon* (+ *B. hybridum*-BhD), and the recent divergences of core perennial lineages. Within the latter group, the resolution of the clade lineages was similar to that of the plastome-based tree (Fig. 6).



**Figure 6.** ASTRAL species tree (supertree) of *Brachypodium* inferred from whole plastome and 35S and 5S ML gene trees. The nodes of the tree show consecutive numbering. Pie diagrams at nodes correspond to quartet support values for the alternative branch topologies [q1 (red), q2 (blue), q3 (gray)] (see Supplementary Table S4). *Sorghum bicolor*, *Zea mays*, *Oryza sativa*, *Festuca rubra*, *Triticum aestivum* and *Hordeum vulgare* were used to root the trees. *Brachypodium* lineages and groups are shown in the right part of the figure.

A divergence time estimation for the *Brachypodium* lineages through calibrated penalized likelihood approaches retrieved a TreePL topology (Fig. 7) highly congruent with that of the *Brachypodium* supertree (Fig. 6). The TreePL search inferred a Mid-Late-Miocene origin (12.1 Ma) for the MRCA of *Brachypodium*, an age concordant with that of previous estimations (Sancho et al., 2022). The split of the ancestral *B. stacei* lineage (12.1 Ma) was followed by those of other out-core clades, which were dated to the Late-Miocene (*B. mexicanum* and *B. boissieri* + *B. retusum*-4x-6x-5S-S, 8.53 Ma; *B. phoenicoides*-6x-5S-S, 7.38 Ma; *B. distachyon* lineage, 6.45 Ma), while that of the core-perennial clade was dated to the Late-Pliocene (3 Ma) (Fig. 7). The divergence of the *B. boissieri* and *B. retusum*-5S-S lineages probably took part in the Early-Pliocene (3.76 – 2.28 Ma) and the inferred origin of the ancestral *B. hybridum* BhD lineage (1.21 Ma) agreed with previous records (Gordon et al., 2020). The divergences of the core perennial lineages were estimated to have occurred very recently, since the Late-Pliocene through the Quaternary (Fig. 7). The oldest splits from their respective crown ancestors corresponded to *B. arbuscula* (3 Ma), *B. phoenicoides*-6x-5S-L (2.83 Ma), and the East Africa – S & E Asian group (2.77 Ma) in the Pliocene, and the *B. sylvaticum* s. s. (2.17 Ma), *B. retusum*-4x-5S-L (1.99 Ma) and the *B. pinnatum* complex lineages ('miscellaneous' 1.97 Ma, *B. rupestre* 1.71, *B. pinnatum* 0.93 Ma) in the Pleistocene. Intraclade divergences, or intraspecific or intracytotypic radiations were estimated to have occurred very recently in the Late-Quaternary, except for splits of the *B. phoenicoides*-6x-5S-L *p.p.* (1.95 Ma), *B. sylvaticum* s. s. *p. p.* (1.33 Ma), and East African *B. flexum* and *B. bolussi* lineages (1.28 Ma), dated to the Early Quaternary (Fig. 7). Our time-calibrated inferences indicated that the speciations of *B. madagascariense* in Madagascar and of *B. kawakamii* in Taiwan from Himalayan *B. sylvaticum* var. *breviglumis* ancestors likely occurred 45 kya and 38 kya, respectively. The youngest divergences were dated to only several thousand years ago (e. g., *B. arbuscula*, 33 kya; *B. miserum* / *B. kurilense*, 28 kya; Fig. 7).



**Figure 7.** Divergence time estimations of the *Brachypodium* supertree calculated with TreePL showing node mean ages. Values above branches represent nodes ages in millions years (Ma, blue) or thousand years (Kya, red). Ploidy levels of studied samples are indicated on terminals. *Oryza sativa* and *Hordeum vulgare* were used to root the tree. Color codes of *Brachypodium* species correspond to those indicated in Figure 1. *Brachypodium* lineages and groups are shown in the right part of the figure.

## Discussion

### Untangling the genomic diversity and evolution of cryptic perennial diploid *Brachypodium* taxa and cytotypes.

Our cytogenetic and phylogenetic analyses of largely sampled world *Brachypodium* species have uncovered new cytotypes and cryptic taxa with different ploidy levels, revealing their origins and evolutionary relationships within a solid phylogenomic framework (Figs. 1-7; Tables S1A, S1B). Our study has corroborated the existence of unique diploid cytotypes for *B. arbuscula*, *B. genuense*, *B. flexum*, *B. glaucovirens*, and the large *B. sylvaticum* complex composed by *B. sylvaticum* s. s. and its satellite taxa (*B. sylvaticum* var. *breviglume*, *B. spryginii*, *B. glaucovirens*, *B. miserum*, *B. kurilense*). It has also confirmed the existence of diploid populations within the *B. pinnatum* s. s. complex (Fig. 1; Table S1A). The individuals of the majority of these taxa have a constant chromosome base number of  $x = 9$  ( $2n = 2x = 18$ ), while *B. glaucovirens*, a lineage of the *B. sylvaticum* aggregate, and those of one population of the *B. pinnatum* aggregate (*Bpinnatum*502) show a karyotype of  $x = 8$  ( $2n = 2x = 16$ ) (Fig. 1, Table S1A; Wolny and Hasterok, 2009). Comparative chromosome painting (CCP) analysis performed in some of these species (*B. sylvaticum* s.s, *B. pinnatum*, *B. arbuscula*) demonstrated the uniformity of the recent  $x = 9$  karyotype, characterized by a shared banding profile across their nine chromosomes (Bp1-Bp9), and a potential reduction from  $x = 9$  to the recent  $x=8$  karyotype in *B. glaucovirens* and *B. pinnatum* (*pro partim*), by descendant dysploidy via end-to-end fusion of chromosomes Bp3 and Bp6 (Lusinska et al., 2019; Sancho et al., 2022). The homogeneity of the  $x = 9$  karyotype for phylogenetically divergent core-perennial diploid lineages such as *B. arbuscula*, *B. sylvaticum* and *B. pinnatum* (Figs. 4-6; Lusinska et al., 2019; Sancho et al., 2022) as well as for their derived  $x = 8$  karyotypes of the also divergent *B. glaucovirens* and *B. pinnatum* p.p. (Figs. 4-6; Lusinska et al., 2019) reinforces the overall similarity of the core perennial genomes (Scholthof et al., 2018) and the limited resolution of the selected cytomolecular probes to separate the karyotypes of these closely related genomes (Sancho et al., 2022). However, our plastome and rDNA data have demonstrated the parallel

descendant dysploidy trend along the two diverging *B. sylvaticum* s. s. (*B. glaucovirens*) and *B. pinnatum* s. s. (*Bpinnatum502*) lineages (Figs. 4A, 4B, 5, 6), which apparently occurred within the last 72 - 66 kya (Fig. 7), suggesting the existence of recent chromosomal rearrangements and reductions within the diploid *Brachypodium* perennials.

Furthermore, our study has untapped the existence of two main divergent lineages within the previously considered monophyletic *B. sylvaticum* species (Figs. 1, 4-8; Díaz-Pérez et al., 2018; Catalán et al., 2023). Selected as a model grass for perenniality, *B. sylvaticum* accumulates the highest genomic resources of all perennial *Brachypodium* taxa (Steinwand et al., 2013; Gordon et al., 2016; Scholthof et al., 2018) and covers the largest native geographic distribution of all *Brachypodium* species, with a broad Holarctic range expanding from the Atlantic Macaronesian archipelagos through Eurasia to the Pacific Japan and Malesia archipelagos (Fig. 1; Catalán et al., 2023). It is also the only invasive perennial *Brachypodium* species present in the New World (Rosenthal et al., 2008; Marchini et al., 2018). In addition, southern hemisphere tropical mountain forest species have been considered morphologically close to *B. sylvaticum* (e. g., East Africa *B. flexum*, Madagascar *B. madagascariense*; Catalan et al., 2016). Although *B. sylvaticum* s. l. has been taxonomically split into several minor satellite taxa in its eastern and western distribution ranges (e. g., *B. sylvaticum* var. *breviglume* (Himalayas), *B. miserum* (Japan), *B. kurilense* (Kuril Islands), *B. glaucovirens* (Eastern Mediterranean, SW Asia), *B. spryginii* (Eastern Europe, Caucasus); Tzvelev, 1983; Scholz, 2007; Tzvelev and Probatova, 2019), the evolutionary relationships of the oriental and occidental taxa and populations of *B. sylvaticum* s. l. were only recently untapped (Catalán et al., 2023). Our plastome-based phylogeny clearly separated the two main *B. sylvaticum* s. l. diploid lineages, placing the oriental taxa within a first diverging East African – S & E Asian clade, and the occidental taxa within a subsequently diverging *B. sylvaticum* s. s. clade (Fig. 3a, b), a resolution broadly congruent with those of the rDNA 35S and 5S trees (Figs. 4, 5). The first clade showed the potential contributions of diploid Himalayan *B. sylvaticum* subsp. *breviglume*-type ancestors to the independent origins of Malagasy *B. madagascariense* and Taiwan *B. kawakamii*, assumed to have occurred in the Mid-Late Holocene (Northgrippian- Meghalayan, 45 – 38 kya), while the two Pacifican close diploid species *B. miseum* and *B. kurilense* were inferred to have split from their common ancestor in the Late Holocene (Meghalayan, 28 kya (Fig. 7). An oriental *B. sylvaticum*-



type ancestor could also have participated in the origin of the Early-Mid Pleistocene (Calabrian; 1.27 Mya) Tropical East Africa *B. flexum*/*B. bolussi* lineages, with diploid *B. flexum* populations radiating in the recent Late-Holocene (17 kya) (Fig. 7). The purported trans-Indic ocean and trans-China sea colonizations agreed with previous findings about the migrations of other cool grass lineages from South-Eastern Asia to Madagascar and to Eastern-Southern Africa (Pimentel et al., 2013), and of the alpine Himalaya flora to Taiwan (Shen et al., 2017) in the Late Pliocene and throughout the Pleistocene. The close relationship of *B. kurilense* to *B. miserum* concurred with the main contribution of southern source floristic elements (Japan) to the non-relict Kurils' flora (Pietsch et al., 2003).

The highly diversifying history of the oriental *B. sylvaticum* s. l. group and its closely related African-Asian endemics contrast with the large genomic and taxonomic homogeneity of the occidental *B. sylvaticum* s. s. lineage. The latter group was composed of European, Asian, and North African *B. sylvaticum* s. s. lineages that did not show clear geographic patterns in any phylogeny (Figs. 3, 4, 5), plus *B. glaucovirens* and *B. spryginii*. Several samples shared the same plastome haplotypes (Fig. 3B, Table S3) and collapsed in large polytomies in the plastome-based tree (Fig. 3A), while they split into independent but mostly invariable and unresolved lineages, sometimes mixed with some oriental Himalayan lineages, in the 35S and 5S-based trees (Fig. 4, 6). These results point towards to a rapid spread of genomically homogeneous occidental *B. sylvaticum* s. s. lineages across Europe, N Africa and western Asia in late glacial and postglacial times with occasional crosses with oriental lineages (Figs. 3-7). Noticeably, the large genomic uniformity of the European *B. sylvaticum* s. s. populations was mirrored by that of their fungal *Epichloë sylvatica* endophytes (Leuchtmann and Schardl, 2022), vertically transmitted through seeds and systemically present in these grasses, suggesting that the postglacial spread affected similarly to close *B. sylvaticum* plants that contained a single endophyte lineage. The microtaxon *B. spryginii* and the morphologically distinct *B. glaucovirens* diverged from their close *B. sylvaticum* s. s. relatives in the Middle Pleistocene (Fig. 7); *B. glaucovirens*, which shows intermediate features between *B. sylvaticum* and *B. pinnaum*, was resolved as a strong sister taxon of a single *B. pinnaum* lineage from C Europe in the plastome but not in the nuclear rDNA phylogenies (Figs. 4, 5), suggesting a maternal *B. sylvaticum* s. s.-type progenitor species for *B. glaucovirens* and some *B. pinnaum* lineages. Surprisingly, the genomically

homogeneous occidental *B. sylvaticum* s. s. individuals showed a larger genome size ( $2C = 0.88 - 0.96$ ) than those of the other core perennial *Brachypodium* diploids (*B. arbuscula*, 0.71; *B. pinnatum* 0.82) and a higher number of 35S loci (3-5) than those expected for diploids (Table S1A), suggesting a potential paleo-polyploid origin followed by a recent diploidization for the genomic constitution of this perennial model species.

### Uncovering the origins of arcane perennial *Brachypodium* polyploid taxa and cytotypes.

Our analyses have confirmed the existence of unique cytotypes for the old out-core tetraploid *B. mexicanum* ( $2n = 4x = 40$ ), characterized by an ancestral karyotype of  $x=10$  composed by two similar chromosomal sets in this putative segmental allotetraploid (Bm1-Bm10, subgenomes A1 and A1'; Lusinska et al., 2019), and hexaploid *B. boissieri* ( $2n = 6x = 48$ ), characterized by a reduced ancestral karyotype of  $x=8$  present by triplicate in this putative autohexaploid (Bb1-Bb8, subgenome A2; Sancho et al., 2022), while the other perennial species surveyed showed more than one cytotype (Fig. 1, Table S1A). *Brachypodium mexicanum* was resolved as the next diverging lineage after the ancestralmost *B. stacei* lineage in the *Brachypodium* supertree (Fig. 6), with both species showing highly-collinear but different  $x = 10$  karyotypic profiles (Sancho et al., 2022). *B. mexicanum* presents the highest genome size ( $2C = 3.7$ ) of all *Brachypodium* species, being more than twice the mean size of that of other annual and tetraploid perennials (Table S1A), suggesting a unique increasing genome size trend along this old lineage within a genus with, otherwise, small and compact genomes (Table S1A; Scholthof et al., 2018). Intraclade sister relationships were recovered for Mexican and Ecuadorean *B. mexicanum* lineages (Figs. 3a, 3b, 4, 6), paralleling those of Mexican and Venezuelan lineages in the study of Díaz-Pérez et al. (2018), suggesting the existence of a divergent northern South-American paramo lineage from the North-American Madrean lineage within this species. The narrow S Spain endemic *B. boissieri* showed its ancestral divergence in the nuclear rDNA trees but more recent sister relationships to the annual *B. distachyon* lineage in the plastome tree (Figs. 4-7). Interestingly, this species displays the second highest genome size of the genus ( $2C = 3.1-3.2$ ) and several 5S loci (8 - 10) higher than expected for an hexaploid (Table S1A); however, its 5S sequences are highly homogeneous and resolve *B. boissieri* close to its morphologically like *B. retusum* (Fig. 5). The collected data indicate that old out-core *B. boissieri* has probably gone through



consecutive rounds of allo- and autopolyploidizations; the nuclear 35S and 5S tree support the ancestry of the progenitor paternal diploid genomes while the plastome tree points to a more recently evolved maternal diploid genome (Figs. 3, 4, 5), confirming previous findings from transcriptome and plastid data (Sancho et al. 2020). The hypothesized evolutionary scenario of initial allopolyploidizations for *B. boissieri*, suggested by its contrasting nuclear and plastome resolutions, was probably followed by a uniformizing autohexaploidy, corroborated by its triplicate  $x = 8$  ancestral karyotype (Sancho et al. 2020) and a remarkable convergent evolution of its multiple 5S loci into similar sequences (Fig. 5, Table S1A). The dolomitic stenocious *B. boissieri* likely arose in the Early Pliocene (2.27 Ma; Fig. 7) in the western Mediterranean region from unknown and probably extinct diploid progenitor species.

Our study indicated that the intermediately evolved Mediterranean *B. retusum* presents two cytotypes, with tetraploid ( $2n = 4x = 32$ ) populations being widely distributed in the western Mediterranean region (Morocco, Spain, S France) and hexaploid ( $2n = 6x = 42$ ) populations present in S & N Spain (Fig. 1, Tables S1A, S1B). Our searches could not confirm the existence of other purported tetraploid individuals with a lower chromosome number ( $2n = 4x = 28$ ) reported for this species (Schippmann, 1991). CCP analysis indicated that *B. retusum* allotetraploids bear an ancestral  $x = 8$  karyotype (Bb1-Bb8), shared with its morphologically close *B. boissieri*, and an intermediately evolved  $x = 8$  karyotype (Br1 - Br8, subgenome E1) (Sancho et al., 2022). Although the *B. retusum* 6x karyotype has not been published yet, preliminary CCP analysis indicate that it corresponds to *B. retusum* allohexaploids composed of the same ancestral  $x=8$  (Bb1 - Bb8) and intermediate  $x=8$  (Br1 - Br8) subgenomes, plus a third chromosomal set of another intermediately evolved but more reduced  $x = 5$  karyotype (Bph1 - Bph5), the latter also present in *B. phoenicoides* and *B. rupestre* allotetraploids (Sancho et al., 2022, and unpub. data). Our phylogenomic analyses have detected both out-core ancestral 5S-S and recent core perennial *B. pinnatum*-complex-type 5S-L allotetraploid and allohexaploid *B. retusum* lineages (Fig. 5), a recent *B. pinnatum*-type *B. retusum* lineage in the plastome tree (Fig. 3A), and core-perennial-type *B. retusum* sequences in the poorly resolved 35S tree (Fig. 4). The recruited data indicate that the ancestral *B. retusum*-4x and *B. retusum*-6x 5S lineages, sister to those of *B. boissieri*, would probably correspond to their ancestral A2 paternal subgenome, while their recent *B. pinnatum*-type 5S-L and plastome lineages could correspond to their maternal E1

subgenome (Figs. 5-7). The ‘intermediately evolved’ position of the *B. retusum* E1 subgenome was hypothesized based on the grafting positions of its homeologs in the *Brachypodium* diploid skeleton tree, although the plastid data also suggested a more recent core-perennial clade maternal origin (Sancho et al., 2022). Our current results allow us to infer a different scenario for a more recent origin of the *B. retusum* E1 subgenome. In consonance with our phylogenies (Figs. 3-7), it probably derived from a *B. pinnatum*-complex-type  $x = 9$  karyotype (Bp1 - Bp9) that experienced a distinct dysploidy descendant trend of  $x=9 \rightarrow x = 8$  than those of *B. glaucovirens* and *B. pinnatum*502 (end-to-end Bp3+Bp6 chromosomal fusion; Lusinska et al., 2019), here involving two nested centromeric chromosomal fusions (Bp4 + Bp6, Bp5 + Bp7) and one centromeric chromosomal fission (Bp5 + Bp7) which ended in the current  $x = 8$  E1 karyotype (Sancho et al., 2022). According to this scenario, the recent E1 genome also contributed as the maternal plastome donor of the *B. retusum* allotetraploids (Fig. 3). The deconvolution of the reticulate origins of the *B. retusum* allohexaploids is more complicated. Although the 5S data detected a similar ancestral  $x = 8$  A2 paternal subgenome than the *B. retusum* allotetraploids (Fig. 5) and the plastome data a recent *B. pinnatum*-complex-type  $x = 8$  E1 maternal subgenome (Fig. 3), there was no consistent evidence for the contribution of the third  $x = 5$  E2 subgenome, except for the nestings of some *B. retusum*-6x lineages within the recent *B. rupestre*-4x and *B. phoenicoides*-4x clades in the plastome, 35S and 5S trees (Figs. 3-5), which may indicate alternative recent maternal or paternal origins from this E2 genome (or potential convergent evolution of its rDNA genes into the E1 ribotypes). The *B. retusum* 4x and 6x cytotypes may constitute cryptic species although its taxonomic separation remains elusive; unlike its close stenocious congener *B. boissieri*, the euryoicous *B. retusum* is adapted to variable soil types and xeric to sub-mesic conditions across the Mediterranean region (Fig. 1), probably resulting from different expressions of ancestral A2 and recently evolved E1 (and E2) genes.

Our evolutionary analyses inferred different lineages for various taxa and ploidy levels within the *B. pinnatum*-complex (Fig. 3-7; Table S1A). Diploid *B. pinnatum* s. s. and *B. genuense* presented the typical core-perennial  $x = 9$  karyotype but nested in separate *B. pinnatum*-2x and *B. rupestre*-type clades in the plastome and 5S trees (Figs. 3,5). Individuals of all studied tetraploid *B. pinnatum*, *B. rupestre*, and *B. phoenicoides* populations showed a constant genome size and chromosome number ( $2C = 1.4$  pg,

2n = 28; Table S1A). CPP analysis of representative samples of these species indicated that they were allopolyploids formed by a recent  $x = 9$  (Bp1 - Bp9) core-perennial subgenome and an 'intermediately evolved' and reduced  $x = 5$  E2 (Bph1 - Bph5) subgenome (Lusinska et al., 2019; Sancho et al., 2022). However, our phylogenies showed that the morphologically close and mesic *B. pinnatum* and *B. rupestre* tetraploids nested preferentially in closely related but separate recent clades in the plastome tree (Fig. 3) and in a recent *B. rupestre*-type 5S-L clade in the 5S tree (Fig. 5), while the morphologically distinct and xeric-wet edaphic *B. phoenicoides* tetraploids nested in intermediately-evolved positions or within a recent *B. phoenicoides* (miscellaneous) clade in the plastome tree (Fig. 3a) but in two divergent ancestral 5S-S and recently evolved 5S-L clades in the 5S tree (Fig. 5). All these lineages collapsed, however, in a largely polytomic core-perennial clade in the 35S tree (Fig. 4). The new plastome and 5S phylogenies suggest that, although the *B. phoenicoides* and the *B. pinnatum* – *B. rupestre* allotetraploids show similar cytogenetic features and karyotypes, they had different origins. *B. phoenicoides* shows intermediacy of morphological features between xeric ancestral *B. boissieri* – *B. retusum* and recent mesic *B. pinnatum* – *B. rupestre* species (Schippmann, 1991; Catalan et al., 2016). It shares 5S families with both groups (Fig. 5), suggesting a hybrid origin from one out-core (or intermediately-evolved)-type ancestor and one core perennial *B. pinnatum*-complex-type ancestor (either maternal or paternal, and either  $x = 9$  or  $x = 5$  for any of them; Figs. 4, 5, 6). The *B. retusum*-4x out-core or intermediately-evolved  $x = 5$  E2 karyotype could have resulted from a contraction of an ancestral ( $x = 10$ ) or an intermediate ( $x = 9$ , Bp1 - Bp9) karyotype to an  $x = 5$  (Bph1 - Bph5) karyotype through five or four nested centromeric chromosome fusions (e. g., Bp2 + Bp3 = Bph1, Bp4 + Bp6 = Bph2, Bp5 + Bp7 = Bph3, Bp8 + Bp9 = Bph4; Lusinska et al., 2019; Sancho et al., 2022), respectively. Conversely, the phylogenomic data indicate that the  $x = 5$  E2 karyotype present in the *B. pinnatum* and *B. rupestre* allotetraploids may have had a recent origin as these taxa did not present signatures of out-core ancestral or intermediately evolved sequences in any of the reconstructed phylogenies (Fig. 3-6). Therefore, we could hypothesize that parallel descendant dysploidy trends from  $x = 9$  (or  $x = 10$ ) karyotypes to the  $x = 5$  E2 karyotype could have occurred in different *Brachypodium* lineages and at different times (e.g., ~2.1 Ma, in *B. phoenicoides*-4x;  $\leq 0.52$  Ma, in *B. pinnatum* – *B. rupestre*-4x; Fig. 7). Some *B. pinnatum* tetraploids likely inherited a maternal *B. pinnatum*  $x = 9$  progenitor genome and a paternal  $x = 5$  *B. rupestre*-type progenitor genome, and some *B. rupestre* tetraploids a maternal *B. genuense*

*B. rupestre*-type  $x = 5$  progenitor genome and a paternal  $x = 9$  *B. sylvaticum*-type progenitor genome (Fig 3, 4), although other lineages departed from these scenarios.

The newly discovered hexaploid cytotype of *B. rupestre* in S Europe and the hexaploid cytotype of *B. phoenicoides* from S Spain and Morocco also showed a constant genome size and chromosome number ( $2C = 2.2\text{--}2.4$  pg,  $2n = 6x = 38$ ), while another hexaploid cytotype of *B. phoenicoides* from S France presented slightly smaller values ( $2C = 1.8\text{--}1.9$  pg,  $2n = 6x = 36$ ) (Fig. 1, Table S1A). Preliminary CPP analysis of representative *B. rupestre* and *B. phoenicoides*  $2n = 38$  hexaploids indicated that they consist of three subgenomes formed by one  $x=9$  (Bp1 - Bp9) and two  $x = 5$  (Bph1 - Bph5) chromosomal sets (Lusinska et al., 2019; Sancho et al., 2022, and unpub. data). We could then hypothesize two alternative evolutionary scenarios for a second round of hybridization and whole genome duplication originating these allohexaploid cytotypes, either from the cross of an  $x = 9 + 5$  allotetraploid with another  $x = 5$  progenitor diploid genome, or of an  $x = 5 + 5$  auto(or allo)-tetraploid with a  $x = 9$  diploid progenitor genome (Figs. 3-6). Our phylogenomic data showed similar resolutions in the plastome and 5S trees for the hexaploid than for the tetraploid *B. rupestre* and *B. phoenicoides* lineages (Figs. 3-5), supporting again the participation of an out-core ancestral or intermediately-evolved ancestor and a recent core ancestor in *B. phoenicoides* hexaploids and only recent ancestors in *B. rupestre* hexaploids (Fig. 6). Interestingly, the maternal ancestor of the *B. phoenicoides*- $6x$   $2n = 6x = 38$  lineages from S Spain and Morocco was more ancestral (intermediately evolved) than that of the *B. phoenicoides*- $6x$   $2n = 6x = 36$  lineage from S France (recently evolved, miscellaneous clade) (Fig. 4-6), suggesting the potential existence of different ongoing micro-speciation processes with potential chromosomal losses in the last lineage, while all the studied *B. rupestre*- $6x$  lineages from S France showed recent maternal *B. sylvaticum*-type and paternal *B. genuense*/*B. rupestre*-type ancestors (Figs. 3-6).

Our analyses have considerably enlarged and contributed to deciphering the highly reticulate evolutionary framework of *Brachypodium*, which constitutes a large agamic complex of different allo- and autopolyploid taxa that resulted from ancestral and recent hybridizations and polyploidizations of both divergent and closely related lineages (Figs. 3-7). Although the characterization of all these potential crypto-species and the elucidation of their genomic changes and polyploidization mechanisms are still elusive, our data have elucidated the old origins of *B. stacei*- $2x$  and *B. mexicanum*- $4x$ , the mixed

origins of *B. boissieri*-6x, *B. retusum*-4x, 6x and *B. phoenicoides*-4x, 6x from both ancestral and recent ancestors, the relatively recent origin of *B. distachyon*-2x from intermediately evolved ancestors, and the young origins of core-perennial *B. sylvaticum* s. l. complex-2x (*B. sylvaticum* s. s. and satellite taxa) and *B. pinnatum*-complex-2x, 4x, 6x (*B. pinnatum*, *B. genuense*, *B. rupestre*) from recent ancestors.

## Material and Methods

### Sampling

A comprehensive genus-wide cyto-phylogenomic analysis of *Brachypodium* was performed in this study. Individuals from a total of 147 populations from all recognized 21 *Brachypodium* taxa, and their 26 cytotypes, sampled across their respective native geographical ranges, were used in our analyses; of them, 138 were newly studied (Fig. 1, Tables S1A, S1B). 89 samples from the *Brachypodium* perennial species were employed in the plastome and rDNA 35S and 5S phylogenomic analyses (Table S1B). In addition, reference genome sequence data from the three annual *Brachypodium* species (*B. distachyon*, *B. stacei*, *B. hybridum*) and from perennial *B. sylvaticum*, plus six close outgroups (*Diarrhena obovate*, *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Festuca rubra*, *Hordeum vulgare* and *Triticum aestivum*) were retrieved from NCBI and incorporated to the phylogenomic study. Samples from 90 populations of 8 *Brachypodium* taxa were used in the cytogenetic study, and chromosome counting was performed in 61 populations (Fig. 1, Table S1A). Diploid populations collected in the Caucasus, previously attributed to *B. rupestre* (Wolny and Hasterok, 2009), were found to belong to *B. sylvaticum* (Table S1).

### Cytogenetic and karyotypic study

Chromosome counting (2n values) analysis was performed on DAPI-stained meristematic root cells following the protocols of Jenkins and Hasterok (2007) and using a Motic BA410 fluorescence microscope. Chromosome numbers of studied *Brachypodium* taxa and cytotypes obtained from previous authors were incorporated to the study (Table S1A). Chromosome records indicate that *B. flexum* is a diploid species (2n=2x=18; Hoshino and Davidse, 1988) and not a polyploid (Díaz-Pérez et al., 2018). Genome sizes (GS, 2C values) were estimated from fresh leaf tissue using Propidium Iodide staining of cell nuclei and flow cytometry measurements (Sysmex Ploidy Analyser) following the protocol of Doležal et al. (2007). The final DNA content for each

*Brachypodium* population was calculated based on six independent measurements. For each individual measurement, at least 5000 nuclei were assessed, with peaks having a coefficient of variation  $< 3$ , and using *Solanum lycopersicum* L. “Stupické polní rané” (1.96 pg/2C; Doležel et al., 1992), *Raphanus sativus* L. “Saxa” (1.11 pg/2C; Doležel et al., 1992), and *Brachypodium rupestre* (Host) Roem. & Schult “Brup 435” (0.82 pg/2C; Sancho et al., 2022) as references.

The 25S and 5S FISH procedure followed the protocol of Jenkins and Hasterok (2007). A 5S rDNA probe was obtained from *Triticum aestivum* clone pTa794 (Gerlach and Dyer, 1980) by PCR amplification and labelling with digoxigenin-11-dUTP (Roche) and a 25S rDNA probe was generated by nick translation of a 2.3-kb ClaI subclone of the 25S rDNA coding region of *Arabidopsis thaliana*. This probe was labelled with tetramethyl-rhodamine-5-dUTP (Roche) and used to visualize 25S rDNA loci containing the genes coding for 18S, 5.8S and 25S rRNA (Hasterok et al., 2002). All images were observed in a Motic BA410 fluorescence microscope.

### **DNA isolation, genome skimming, plastome and rDNA 35S and 5S sequence assembling and multiple sequence alignments.**

Total genomic DNA was isolated from silica gel dried samples or herbarium specimens using a modified CTAB protocol (Doyle and Doyle, 1987) using ~20mg of tissue. DNA concentration (100-200ng/ul) and absorbance (260/230 nm of 1.8 to 2.1 and 280/260 nm of 1.8 to 2.0) were estimated using Qubit ® 3.0 (Life Technologies, Grand Island, NY) and Biodrop µLITE (Harvard Biosciences), respectively. Genome skimming sequencing was performed from PCR-free libraries through the Illumina technology at the Spanish Centro Nacional de Análisis Genómicos (CNAG, Barcelona) and Macrogen (Madrid), and the pair-end (PE) reads were processed following the procedures described in Moreno-Aguilar et al., (2022).

Assembled plastomes for 70 out of 92 newly sequenced samples were obtained with Novoplasty v. 2.7.1 (Dierckxsens et al., 2017) using the previously assembled *B. sylvaticum*-Ain1 plastome and other genome sequences retrieved from Phytozome (<https://phytozome-next.jgi.doe.gov/>) as references, and standardized parameters (k-mer: 39, insert size: ~92–329 bp). The plastomes of another 22 samples were assembled using a read-mapping strategy to the same *B. sylvaticum* Ain-1 using Geneious Prime v.2023.1.1



(<https://www.geneious.com/>) (Table S2). Plastome gene annotation was performed using GeSeq (Tillich et., 2017).

The nuclear rDNA 35S region (transcribed cistron 5'-ETS-18S gene- ITS1-5.8S gene- ITS2-25S gene) of the 92 *Brachypodium* samples studied was assembled using a similar read-mapping strategy to the *B. sylvaticum*-Ain1 35S reference sequence. 35S sequences from *B. distachyon* Bd21, *B. stacei* ABR114, *B. hybridum* ABR113 (recent S-plastotype) and Bhyb26 (ancient D-plastotype) reference genomes were retrieved from Phytozome (<https://phytozome-next.jgi.doe.gov/brachypan>) and incorporated to the analysis.

A total of 98 nuclear rDNA 5S sequences (5S gene + IGS long and short sequences) from 84 *Brachypodium* samples were assembled using a customized assembly protocol (Fig. S2). After a quality check of genome skimming paired-end (PE) reads (FastQC\_v.0.11.9; <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), they were mapped onto two separated 5S reference data sets (5S-long, 5S-short) that included eight outgroup grasses and six *Brachypodium* references genomes available at NCBI and retrieved through Blastn (Table S4). After a first mapping step (minimap\_v.2.1; Li et al., 2018), unmapped reads were removed (samtools\_v.1.16.1; Danecek et al., 2021) and putative 5S reads (PE and single-end (SE)) were sorted according to their matching to either the 5S-long or 5S-short data sets (split\_pairs\_v.0.5; [https://github.com/eead-csic-compbio/split\\_pairs](https://github.com/eead-csic-compbio/split_pairs)). Each set of reads was assembled separately to retrieve preliminary 5S long (L) and short (S) sequences using SPAdes (SPAdes\_v.3.15.5; Prjibelski, et al., 2020). The resulting scaffolds were assembled *de novo* with Geneious Prime v.2023.1.1 using the default parameters and manually curated, removing poorly assembled scaffolds. The final 5S-L and 5S-S filtered data sets were used for downstream analyses (Fig. S2).

Multiple sequence alignments (MSA) were performed separately for entire plastomes, nuclear 35S and 5S sequences using MAFT.v7.450 (Kato et al., 2002). TrimAl v. 1.2rev59 (Capella-Gutiérrez et al., 2009) was used to remove low quality regions from each of the MSAs by imposing the *-automated1* parameter followed by manual curation.

### **Phylogenomics, supertree computation, and dating analyses.**

Maximum likelihood (ML) phylogenetic trees were constructed for each separated data set with IQ-tree 1.6.12. imposing the best-fit nucleotide substitution model,

according to the Bayesian Information Criterion (BIC), and estimating 1000 ultrafast bootstrap replicates (BS) for the branch support of the best tree (Nguyen et al., 2015; Chernomor et al., 2016; Kalyaanamoorthy et al., 2017). Branches with a low bootstrap support ( $<30$ ) were collapse using newick utils (Junier et al., 2010).

The parsimony haplotypic network based on 4,070 segregating sites (SNPs) of 95 *Brachypodium* plastomes was computed using TCS algorithm (Clement et al., 2002) implemented in PopART v.1.7 (Leigh and Briant, 2015). This tool was used to calculate the number of segregating sites (S), haplotypes (h), Nucleotide diversity ( $\pi$ ), and the Tajima's D statistics for the complete network and the main haplo-groups (Table S4). The number of variable sites and haplotype diversity index (Hd) among the *Brachypodium* plastome groups retrieved from the phylogenomic analysis (see Results) were estimated with DNASP v.6 (Rozas et al., 2017).

The plastome, 35S and 5S ML trees were used to build a species-tree supertree under the multispecies coalescent using partially resolved gene trees analysis with ASTRAL-III by imposing a  $-t2$  flag parameter search (Zhang et al., 2018). The intragenomic discordance in the dataset were estimated through the analysis of normalized quartets scores for the main ( $q1$ ), first alternative ( $q2$ ) and secondary alternative ( $q3$ ) topologies generated by ASTRAL (Sayyari and Mirarab, 2016). The estimated intragenomic discordance in the trees was computed using the `plot_Astral_trees_v2.R` script (<https://github.com/sidonieB/scripts>) in RStudio\_2022.12.0. The quartet score value was proportionally inverse to incongruence (values of 1 indicate the absence of gene tree discordance) (Pérez-Escobar et al., 2021).

Ancestral divergence ages of the *Brachypodium* samples under study included in the ASTRAL supertree were estimated through the penalty likelihood method implemented in TreePL (Smith and O'Meara, 2012). Four nodes of the tree were calibrated using secondary age constrains according to the estimates of Sancho et al. (2022) for the BOP (max= 55.08 Ma, min=47.67 Ma), Pooideae (max= 47.35 Ma, min= 26.85 Ma), *Brachypodium* (max (set to mean)= 12.1 Ma, min= 8.47 Ma), and *Brachypodium* core-perennial (max (set to mean)= 3 Ma, min= 2 Ma) clades. A first priming analysis was run to establish subsequent adjustments, a second analysis (with adjusted cross-validation values) to establish the smoothing parameter, and a third analysis to assess node age estimates using the optimized settings for the ASTRAL supertree. Cross validation analysis indicated 0.1 as the best smoothing value.



**Author contributions:** PC, LAI, and EP designed the study. PC, LAI, and EP collected samples. MD, JL, RS, RG, BM and LAI developed the experimental work. MD, JL, RS, LAI, EP, RH and PC analyzed the data and interpreted the results. PC, MD, RS and LAI wrote the manuscript. All authors contributed to the article and approved the submitted version.

**Funding:** This study was funded by the Spanish Ministry of Science and Innovation PID2019-108195GB-I00 and TED2021-131073B-I00 and by the Spanish Aragon Government Bioflora LMP82-21 research grants. MD was supported by a Spanish Ministry of Science predoctoral FPI fellowship. RS was supported by a Bioflora postdoctoral fellowship.

**Data Availability Statement:** Input and output data, and Supplementary information are available at Github (<https://github.com/Bioflora/BrachypodiumPhylogenomics> (accessed on 22 June 2023)).

**Acknowledgements:** We thank Nina Probatova, Iztiar Arnelas, Michael Wilcox, the INIA, JGI, RIKEN and USDA germplasm banks, and the B, FI, G, LD, M and VLA herbaria for facilitating samples of some of the *Brachypodium* taxa studied. Genome skimming data from the studied *Brachypodium* samples were generated at the Centro Nacional de Análisis Genómicos (CNAG) and MacroGen laboratories (Spain). The cytogenetic, bioinformatic and evolutionary analyses were performed in the Bioflora laboratory of the Escuela Politécnica Superior de Huesca (EPSHU, Universidad de Zaragoza, Spain). The karyotypic CCB analyses were performed in the Plant cytogenetics laboratory of the University of Silesia in Katowice (Poland).

**Conflict of interest:** The authors declare no conflict of interest.

## References:

- Baduel, P., Bray, S., Vallejo-Marin, M., Kolář, F., and Yant, L. (2018). The “Polyploid Hop”: Shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* 6, 1–19. doi:10.3389/fevo.2018.00117.
- Baum, B., and Feldman, M. (2010). Elimination of 5S DNA unit classes in newly formed allopolyploids of the genera *Aegilops* and *Triticum*. doi:10.1139/g10-017.
- Baum, B. R., Edwards, T., and Johnson, D. A. (2009). Molecular Phylogenetics and Evolution Phylogenetic relationships among diploid *Aegilops* species inferred from 5S rDNA units. *Mol. Phylogenet. Evol.* 53, 34–44. doi:10.1016/j.ympev.2009.06.005.
- Baum, B. R., and Johnson, D. A. (2008). Molecular confirmation of the genomic constitution of *Douglasdeweya* (Triticeae : Poaceae ): demonstration of the utility of the 5S rDNA sequence as a tool for haplome identification. 621–628. doi:10.1007/s00438-008-0338-1.
- Betekhtin, A., Jenkins, G., and Hasterok, R. (2014). Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. *PLoS One* 9, 1–26. doi:10.1371/journal.pone.0115108.
- Bininda-Emonds (2014). “An Introduction to Supertree Construction (and Partitioned Phylogenetic Analyses) with a View Toward the Distinction Between Gene Trees and Species Trees,” in *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*, 1–552. doi:10.1007/978-3-662-43550-2.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi:10.1093/bioinformatics/btp348.
- Catalán, P., Ángeles Decena, M., Sancho, R., Viruel, J., Pérez-Collazos, E., Inda, L. A., et al. (2023). Phylogenetics of the Palearctic model grass *Brachypodium sylvaticum* uncovers two divergent oriental and occidental micro-taxa lineages. *Bot. Pacifica* 12. doi:10.17581/bp.2023.12119.

- Catalan, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., and López-Herránz, M. L. (2016). Phylogeny and Evolution of the Genus *Brachypodium*. 9–38. doi:10.1007/7397\_2015\_17.
- Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L. A. J., Langdon, T., et al. (2012). Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann. Bot.* 109, 385–405. doi:10.1093/aob/mcr294.
- Chernomor, O., von Haeseler, A., and Minh, B. Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Syst. Biol.* 65, 997–1008. doi:10.1093/sysbio/syw037.
- Díaz-Pérez, A., López-Álvarez, D., Sancho, R., and Catalán, P. (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. *Mol. Phylogenet. Evol.* 127, 256–271. doi:10.1016/j.ympev.2018.06.003.
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45. doi:10.1093/nar/gkw955.
- Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2, 2233–2244. doi:10.1038/nprot.2007.310.
- Doležel, J., Sgorbati, S., and Lucretti, S. (1992). Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol. Plant.* 85, 625–631. doi:10.1111/j.1399-3054.1992.tb04764.x.
- Doyle, J., and Doyle, J. (1987). A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* 19, 11–15.
- Fi, C., Robinson, C. T., and Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. 613–635. doi:10.1111/mec.14486.
- Garcia, S., Wendel, J. F., Borowska-Zuchowska, N., Ainouche, M., Kuderova, A., and Kovarik, A. (2020). The Utility of Graph Clustering of 5S Ribosomal DNA Homoeologs in Plant Allopolyploids, Homoploid Hybrids, and Cryptic Introgressants. *Front. Plant Sci.* 11, 41. doi:10.3389/FPLS.2020.00041/BIBTEX.

- Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., et al. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.* 11, 1–16. doi:10.1038/s41467-020-17302-5.
- Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., et al. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 8. doi:10.1038/s41467-017-02292-8.
- Gordon, S. P., Liu, L., and Vogel, J. P. (2016). “The Genus *Brachypodium* as a Model for Perenniality and Polyploidy,” in *Genetics and Genomics of Brachypodium*, ed. J. P. Vogel (Cham: Springer International Publishing), 313–325. doi:10.1007/7397\_2015\_19.
- Hasterok, R., Catalan, P., Hazen, S. P., Roulin, A. C., Vogel, J. P., Wang, K., et al. (2022). *Brachypodium*: 20 years as a grass biology model system; the way forward? *Plant Sci.* 2022. doi:10.1016/j.tplants.2022.04.008.
- Hoshino, T., and Davidse, G. (1988). Chromosome Numbers of Grasses (Poaceae) From Southern Africa. I. *Ann. Missouri Bot. Gard.* 75, 866–873. doi:10.2307/2399375.
- Jenkins, G., and Hasterok, R. (2007). BAC “landing” on chromosomes of *Brachypodium distachyon* for comparative genome alignment. *Nat. Protoc.* 2, 88–98. doi:10.1038/nprot.2006.490.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi:10.1038/nmeth.4285.
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi:10.1093/nar/gkf436.

- Keng, Y. (1982). *Brachypodium sylvaticum* var. *breviglume* Keng ex Keng f. *Acta Bot. Yunnanica* 4, 277–278.
- Leuchtmann, A., and Schardl, C. L. (2022). Genetic Diversity of *Epichloë* Endophytes Associated with *Brachypodium* and *Calamagrostis* Host Grass Genera including Two New Species.
- Lopez-Alvarez, D., Lopez-Herranz, Maria Luisa Betekhtin, A., and Catalan, P. (2012). A DNA Barcoding Method to Discriminate between the Model Plant *Brachypodium distachyon* and Its Close Relatives *B. stacei* and *B. hybridum* (Poaceae). 7. doi:10.1371/journal.pone.0051058.
- López-Alvarez, D., Manzaneda, A. J., Rey, P. J., Giraldo, P., Benavente, E., Allainguillaume, J., et al. (2015). Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. *Am. J. Bot.* 102, 1073–1088.
- Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., et al. (2019). Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *Int. J. Mol. Sci.* 20, 1–19. doi:10.3390/ijms20225557.
- Mahelka, V., Kopecky, D., and Baum, B. R. (2013). Contrasting Patterns of Evolution of 45S and 5S rDNA Families Uncover New Aspects in the Genome Constitution of the Agronomically Important Grass *Thinopyrum intermedium* (Triticeae). 30, 2065–2086. doi:10.1093/molbev/mst106.
- Mandáková, T., Li, Z., Barker, M. S., and Lysak, M. A. (2017). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 91, 3–21. doi:10.1111/tbj.13553.
- Mandáková, T., and Lysak, M. A. (2018). Post-polyploid diploidization and diversification through dysploid changes. *Curr. Opin. Plant Biol.* 42, 55–65. doi:10.1016/j.pbi.2018.03.001.
- Marchini, G. L., Arredondo, T. M., and Cruzan, M. B. (2018). Selective differentiation during the colonization and establishment of a newly invasive species. 31, 1689–1703. doi:10.1111/jeb.13369.

- Melichárková, A., Šlenker, M., Zozomová-Lihová, J., Skokanová, K., Šingliarová, B., Kačmárová, T., et al. (2020). So Closely Related and Yet So Different: Strong Contrasts Between the Evolutionary Histories of Species of the *Cardamine pratensis* Polyploid Complex in Central Europe. *Front. Plant Sci.* 11. doi:10.3389/fpls.2020.588856.
- Moreno-aguilar, M. F., Inda, L. A., and Sánchez-rodríguez, A. (2022a). Evolutionary Dynamics of the Repeatome Explains Contrasting Differences in Genome Sizes and Hybrid and Polyploid Origins of Grass Loliinae Lineages. 13. doi:10.3389/fpls.2022.901733.
- Moreno-aguilar, M. F., Inda, L. A., Sánchez-rodríguez, A., and Catalán, P. (2022b). Phylogenomics and Systematics of Overlooked Mesoamerican and South American Polyploid Broad-Leaved Festuca Grasses Differentiate *F. sects. Glabricarpae* and *Ruprechtia* and *F. subgen. Asperifolia*, *Erosiflorae*, *Mallopetalon* and. 1–27.
- Murat, F., Xu, J. H., Tannier, E., Abrouk, M., Guilhot, N., Pont, C., et al. (2010). Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res.* 20, 1545–1557. doi:10.1101/gr.109744.110.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi:10.1093/molbev/msu300.
- Padilla-García, N., Rojas-Andrés, B. M., López-González, N., Castro, M., Castro, S., Loureiro, J., et al. (2018). The challenge of species delimitation in the diploid-polyploid complex *Veronica* subsection *Pentasepalae*. *Mol. Phylogenet. Evol.* 119, 196–209. doi:10.1016/j.ympev.2017.11.007.
- Pérez-Escobar, O. A., Dodsworth, S., Bogarín, D., Bellot, S., Balbuena, J. A., Schley, R. J., et al. (2021). Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *Am. J. Bot.* 108, 1166–1180. doi:10.1002/ajb2.1702.
- Pietsch, T. W., Bogatov, V. V., Amaoka, K., Zhuravlev, Y. N., Barkalov, V. Y., Gage, S., et al. (2003). Biodiversity and biogeography of the islands of the Kuril Archipelago. 1297–1310.

- Pimentel, M., Sahuquillo, E., Torrecilla, Z., Popp, M., Catalán, P., and Brochmann, C. (2013). Hybridization and long-distance colonization at different time scales: towards resolution of long-term controversies in the sweet vernal grasses (*Anthoxanthum*). *Ann. Bot.* 112, 1015–1030. doi:10.1093/aob/mct170.
- Rosenthal, D. M., Ramakrishnan, A. P., and Cruzan, M. B. (2008). Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Mol. Ecol.* 17, 4657–4669. doi:10.1111/j.1365-294X.2008.03844.x.
- Rosselló, J. A., Maravilla, A. J., and Rosato, M. (2022). The Nuclear 35S rDNA World in Plant Systematics and Evolution : A Primer of Cautions and Common Misconceptions in Cytogenetic Studies. 13, 1–15. doi:10.3389/fpls.2022.788911.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302.
- Sancho, R., Cantalapiedra, C. P., López-Alvarez, D., Gordon, S. P., Vogel, J. P., Catalán, P., et al. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* 218, 1631–1644. doi:10.1111/NPH.14926.
- Sancho, R., Inda, L. A., Díaz-Pérez, A., Des Marais, D. L., Gordon, S., Vogel, J. P., et al. (2022). Tracking the ancestry of known and ‘ghost’homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J.* 109, 1535–1558.
- Scarlett, V. T., Lovell, J. T., Shao, M., Phillips, J., Shu, S., Goodstein, D. M., et al. (2022). Multiple origins , one evolutionary trajectory : gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*.
- Schippmann (1991). Revision der euroäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45.
- Scholthof, K. B. G., Irigoyen, S., Catalan, P., and Mandadi, K. K. (2018a). *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30, 1673–1694. doi:10.1105/tpc.18.00083.
- Scholz, H. (2007). On the identity of *Brachypodium firmifolium* ( Poaceae ) from Cyprus. 215–220. doi:10.3372/wi.37.37111.

- Shen, Z., Yang, M., Feng, J., Li, X., Peng, P., and Zheng, Z. (2017). Geographic patterns of alpine flora in China in relation to environmental and spatial factors. *Biodiversity Science* 25:182–194. doi:10.17520/biods.2017014.
- Shi, Y., Draper, J., and Stace, C. (1993). Ribosomal DNA variation and its phylogenetic implication in the genus *Brachypodium* (Poaceae). *Plant Syst. Evol.* 188, 125–138. doi:10.1007/BF00937726/METRICS.
- Smith, S. A., and O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689–2690. doi:10.1093/bioinformatics/bts492.
- Soltis, Douglas, E., Soltis, Pamela, S., Schemske, Douglas, W., Hancock, James, F., Thompson, John, N., Husband, Brian, C., et al. (2007). Autoploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56, 13–30. Available at: [https://is.muni.cz/el/1431/podzim2007/Bi9020/um/4567165/4567187/Soltis\\_autoployploidy\\_plants\\_Taxon\\_2007.pdf](https://is.muni.cz/el/1431/podzim2007/Bi9020/um/4567165/4567187/Soltis_autoployploidy_plants_Taxon_2007.pdf).
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi:10.3732/ajb.0800079.
- Soltis, D. E., Visger, C. J., Blaine Marchant, D., and Soltis, P. S. (2016). Polyploidy: Pitfalls and paths to a paradigm. *Am. J. Bot.* 103, 1146–1166. doi:10.3732/ajb.1500501.
- Soltis, P. S., Liu, X., Marchant, D. B., Visger, C. J., and Soltis, D. E. (2014). Polyploidy and novelty: Gottlieb's legacy. *Philos. Trans. R. Soc. B Biol. Sci.* 369, 20130351. doi:10.1098/rstb.2013.0351.
- Steinwand, M. A., Young, H. A., Bragg, J. N., Tobias, C. M., and Vogel, J. P. (2013). *Brachypodium sylvaticum*, a Model for Perennial Grasses: Transformation and Inbred Line Development. *PLoS One* 8, 1–11. doi:10.1371/journal.pone.0075180.
- Stritt, C., Gimmi, E. L., Wyler, M., Bakali, A. H., Skalska, A., Hasterok, R., et al. (2021). Migration without interbreeding : Evolutionary history of a highly selfing Mediterranean grass inferred from whole genomes. 1–16. doi:10.1111/mec.16207.
- Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., et al.



- (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends Ecol. Evol.* 33, 153–163. doi:10.1016/j.tree.2017.11.007.
- Tzvelev, N. (1983). *Brachypodium* P. Beauv. In: *Grasses of the Soviet Union* (N. Tzvelev, ed.), pp. 141–145, Academy of Sciences of the USSR, Smithsonian Institution, New Delhi.
- Tzvelev, N., and Probatova, N. (2019). *Brachypodium* P. Beauv. In: Tzvelev, N.N. & N.S. Probatova. *Grasses of Russia*, pp. 48–50, KMK Scientific Press, Moscow
- Vozárová, R., Herklotz, V., Kovařík, A., Tynkevich, Y. O., Volkov, R. A., Ritz, C. M., et al. (2021). Ancient Origin of Two 5S rDNA Families Dominating in the Genus *Rosa* and Their Behavior in the Canina-Type Meiosis . *Front. Plant Sci.* 12. Available at: <https://www.frontiersin.org/article/10.3389/fpls.2021.643548>.
- Winterfeld, G., Becher, H., Voshell, S., Hilu, K., and Röser, M. (2018). Karyotype evolution in *Phalaris* (Poaceae): The role of reductional dysploidy, polyploidy and chromosome alteration in a wide-spread and diverse genus. *PLoS One* 13, 1–19. doi:10.1371/journal.pone.0192869.
- Wolny, E., and Hasterok, R. (2009). Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Ann. Bot.* 104, 873–881. doi:10.1093/aob/mcp179.

## Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in grass *Brachypodium* polyploids.

María Ángeles Decena<sup>1,2,♦</sup>, Rubén Sancho<sup>1,2,♦</sup>, Luis A. Inda<sup>1,3</sup>, Ernesto Pérez-Collazos<sup>1,2</sup>, Pilar Catalán<sup>1,2,\*</sup>

<sup>1</sup>Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Huesca, España.

<sup>2</sup>Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, España.

<sup>3</sup>Instituto Agroalimentario de Aragón-IA2 (Universidad de Zaragoza-CITA), Aragón, España

♦Both authors contributed equally as first coauthors

\* **Correspondence:** Pilar Catalán (pcatalán@unizar.es)

### Abstract

The *Brachypodium* grass species have been selected as model plants for functional genomics of grass crops, and for elucidating the origins of allopolyploidy and perenniality in monocots due to their small genome sizes and feasibility of cultivation. However, genome sizes differed enormously among either *Brachypodium* diploid or polyploid lineages. We have used low-pass genome sequencing data to uncover the composition, abundance, and phylogenetic value of repetitive elements across 44 representatives of the main *Brachypodium* lineages. We also aimed to test the potential mechanisms and consequences of the “polyploid genome shock hypothesis” (PGSH) in three different evolutionary scenarios of variation of repeats and genome sizes of *Brachypodium* allopolyploids. Our data indicated that the proportion of the genome covered by the repeatome in the *Brachypodium* species showed a 3.3-fold difference between the highest content of *B. mexicanum*-4x (67.97%) and the lowest of *B. stacei*-2x (20.77%), and that changes in their genome sizes were a consequence of gains or losses in their repeat elements. LTR-Retand and Tekay retrotransposons were the most frequent repeat elements in the *Brachypodium* genomes, while the Ogre retrotransposons were exclusively found in *B. mexicanum*. The repeatome phylogenetic networks showed a high topological congruence with the plastome and nuclear rDNA and transcriptome trees, differentiating the ancestral outcore lineages from the recently evolved core-perennial lineages. The rDNA 5S graph topologies had a great match with the ploidy levels and subgenomes of the *Brachypodium* polyploids. The core-perennial *B. sylvaticum* presents a large repeatome and features of a post-polyploid diploidized origin. Our study evidenced that expansions and contractions in the repeatome were responsible for the

three contrasting responses to the PGSH; the exacerbated genome expansion of the ancestral allotetraploid *B. mexicanum* was a consequence of chromosome-wide proliferation of TEs and not of WGD, the additive repeatome pattern of young allotetraploid *B. hybridum* of stabilized post-WGD genome evolution, and the genome shrinkage of recent core perennials polyploids (*B. pinnatum*, *B. phoenicoides*) of repeat losses through recombination. Our analyses have contributed to disentangling the evolution of the repeatome and the genome size variation of model *Brachypodium* grasses.

**Keywords:** *Brachypodium*, evolution, genome size diversification, polyploid genome shock, repeatome, transposable elements, 5S loci.

## INTRODUCTION

The “polyploid genome shock hypothesis” (PGSH), first proposed by McClintock (1984), postulates rapid genome restructuring after hybridization and whole genome duplication (WGD). PGSH is assumed to be a response to the sudden combination of different genomes in a single nucleus, and the necessary genetic and epigenetic regulatory adjustments to keep pace between them (Bird et al., 2018; Edger et al., 2019), coupled with the potential alteration of genome integrity induced by WGD (Scarlett et al., 2022). It is less clear, however, how this genomic re-patterning may occur and affect, in the short and long term, the new polyploid genome. Research conducted in some wild and synthetic allopolyploid plants indicates that genomic reshuffling is common in the first generations after WGD, while the allopolyploid genome tends to become stabilized and diploidized over time (De Storme and Mason, 2014; Wendel et al., 2018; Mason and Wendel, 2020; Deb et al., 2023). However, other plant allopolyploids have evidenced either a long-term genomic instability, perpetuated over generations (Chalhoub et al., 2014; Mason and Wendel, 2020), or a total absence of genomic restructuring, with the immediate creation of the amphidiploid (Scarlett et al., 2022; Deb et al., 2023). One of the genome’s constituents more severely impacted by the potential genome restructuring is the repetitive DNA fraction, or repeatome (mobile elements (retrotransposons and transposons), and tandem ribosomal DNA and satellite repeats), which is largely present in the nuclear genome of most plants (Macas et al., 2015; Hidalgo et al., 2017; Pellicer et al., 2018). Three contrasting evolutionary scenarios have been proposed to explain the potential consequences of PGSH in polyploid repeatome turnovers. In some angiosperms, a rapid increase of repeats has been detected in the genomes after rounds of

polyploidizations (Chen et al., 2020). In others, the polyploid genomes show equivalent repeatome sizes to those of their diploid progenitor species (McCann et al., 2018). And in other groups, high-level polyploids exhibit a considerable reduction of their repeatome relative to that of their diploid and low-level polyploid relatives (Chen, 2007; Parisod et al., 2010; Moreno-Aguilar et al., 2022). The ability of centromeric retrotransposon families to proliferate has been interpreted as the potential mechanism for the increased repeatome of the first group of plants, while stabilized post-WGD genome evolution would explain the additive patterns of the second group, and the tendency of other repeat families to recombine and lose repeats may have caused the repeatome shrinkage in the third group (Michael, 2014; Chen et al., 2020; Scarlett et al., 2022). Although the proliferation or removal of the repetitive elements from the genomes could have resulted from recombination processes or through double-strand break repair (Hawkins et al., 2009; Vu et al., 2017), the driven forces that balance the expansions and contractions of the repeatome are poorly known (Fedoroff, 2012; Drouin et al., 2021). The analysis of the three alternative PGSH evolutionary scenarios has been also hampered by the lack of a proper specific group for testing all its cases.

The importance and impact that the dynamics of the repetitive elements have had in the variation of the genome size and its evolution across the angiosperms have been corroborated in several studies (Dodsworth et al., 2015; Hidalgo et al., 2017; Pellicer et al., 2018). In plants with available reference genomes, the dynamics of transposable elements (TEs) insertions have been also related to the expression of some core or dispensable genes, which are differentially regulated (Gordon et al., 2017a), and to epigenetic effects (Chen, 2007; Fedoroff, 2012; Negi et al., 2016). However, the analysis of the repetitive families in the majority of angiosperms that lack assembled and annotated genomes has been performed using genome skimming data and repeatome graph-topology methods (Weiss-Schneeweiss et al., 2015; Garcia et al., 2020; Vitales et al., 2020; Moreno-Aguilar et al., 2022). The quantification and annotation of repeats in the plant genomes is based upon the fact that similarity-based clustering of low coverage genome sequencing reads, confidentially representing 0.01-0.50x of the total haploid genome coverage, is proportional to the genomic abundance and longitude of the corresponding repeat-types and could be used to quantify them (Macas et al., 2015; Pellicer et al., 2018; Novák et al., 2020). Moreover, comparative analysis of shared repeat clusters has confirmed the phylogenetic information of the repeatome in several

angiosperm groups (Vitales et al., 2020a, 2020b; Herklotz et al., 2021) and its utility to infer ancestral and recent polyploidization and diploidization events (Moreno-Aguilar et al., 2022). In addition, 5S rDNA graph-based clustering approaches have corroborated the identity of the ancestral progenitor genomes of several polyploid plants (Garcia et al., 2020) and have also uncovered the paleopolyploid nature of current diploidized plant species (Vozárová et al., 2021; Moreno-Aguilar et al., 2022).

The cool seasonal genus *Brachypodium*, which consists of approximately 23 taxa (Catalan et al., 2016; Decena et al., 2023), has been selected as a model functional system for cereal and biofuel crops and for investigating the evolution of polyploidy in grasses. Annotated reference genomes and considerable genomic resources have been produced for its three annual species (*B. distachyon*, *B. stacei*, *B. hybridum*) (Scholthof et al., 2018; Hasterok et al., 2022). Comparative genomic studies of these annual species evidenced that the recurrently originated allotetraploid *B. hybridum* did not experience major genomic restructurings, showing equivalently inherited parental transposon contents in its two subgenomes (Gordon et al., 2020; Scarlett et al., 2022; Mu et al. 2023). However, cytogenetic analysis of the less investigated perennial taxa detected considerable differences between the large genome sizes of ancestral polyploids (*B. mexicanum*-4x, 3.7 pg (2C); *B. boissieri*-6x, 3.1 pg) and the relatively small sizes of recently evolved polyploids (e. g., *B. pinnatum*-4x, *B. rupestre*-4x, *B. phoenicoides*-4x, 1.4 pg; *B. phoenicoides*-6x, 2.1 pg) (Decena et al., 2023). *Brachypodium* shows a remarkable descendant dysploidy trend from ancestral x=10 karyotypes to intermediately-to-recently evolved x=9, x=8 and x=5 karyotypes (Lusinska et al., 2019; Sancho et al., 2022; Decena et al., 2023). Phylogenetic subgenome detection algorithms applied to transcriptome data, 5S rDNA phylogenetic reconstruction and karyotype barcoding analysis further identified seven diploid subgenomes in the studied *Brachypodium* polyploids, three of them present in current diploid progenitor species and four orphan (only detected in polyploid species) (Sancho et al., 2022; Decena et al., 2023). Except for the thoroughly investigated *B. distachyon* annual species, where transposon landscape analysis revealed a high transposable activity of LTR-Copia Angela elements and a large contribution to genome size of highly methylated LTR-Gypsy Retand elements (Stritt et al., 2020a), and the identification of centromeric CRBd retrotransposons elements in six *Brachypodium* species (Li et al., 2018), no other study has exhaustively explored the composition and dynamics of repetitive elements in a large representation of the *Brachypodium* taxa.

We were particularly interested in using *Brachypodium* as a test-bed case study for the three alternative PGSH evolutionary scenarios. These *Brachypodium* samples constitute exemplary case studies for investigating the putative role of the repeat elements' dynamics in the evolution of these genomes and for testing the potential mechanisms and consequences of the “polyploid genome shock hypothesis” in three different evolutionary scenarios of proliferation, maintenance, and reduction of repeats and genome sizes of allopolyploids occurring within this genus. We also aimed to comparatively analyze the repeatome variations in diploid *Brachypodium* species which show substantial differences in genome sizes (Decena et al., 2023). The repeatome analysis was also used to assess the potential phylogenetic value of the repeat elements in the monotypic Brachypodieae tribe. The objectives of our study were: (i) to characterize and quantify the repetitive elements of 44 representative samples of the main *Brachypodium* species and lineages identifying both shared and private repeats; (ii) to analyze the expected correlation between genome size and abundance of the repeats; (iii) to identify repeat types that could have contributed to the expansions or contractions of genomes; (iv) to assess the phylogenetic value of repeats using phylogenetic reconstructions; and (v) to assess the three alternative cases and responses to the “polyploid genome shock hypothesis” of polyploids and the putative paleo-polyploid origin of some *Brachypodium* diploids with large genome sizes using mobile and tandem repeat data analysis.

## METHODS

### Sampling, ploidy levels and genome skimming sequencing

Genomic sequences from 44 accessions of 11 *Brachypodium* species, representing its main out-core and core-perennial lineages were incorporated to the study (**Table 1** and **Supplementary Table S1**; **Supplementary Figure S1**). 41 perennial samples were sequenced *de novo* while genome data from three annual samples was downloaded from NCBI (*B. distachyon* Bd21-3 (SRR4236817); *B. stacei* ABR114 (SRR3944701); *B. hybridum* ABR113 (SRR3945056; SRR3945058)). Samples were classified into groups based on previous phylogenomic studies (Sancho et al., 2022; Decena et al., 2023). We obtained a large cytotypic representation of *Brachypodium* through the study of 15 different cytotypes found within these taxa, including diploids, tetraploids and hexaploids (**Table 1** and **Supplementary Table S1**). The 44 selected samples represent 11 out of the

12 evolutionary lineages currently recognized within *Brachypodium* (Decena et al., 2023). They include all the main diploid and polyploid annual and perennial *Brachypodium* lineages, formed at different evolutionary times and spanning the last 10 Ma (Catalan et al., 2016; Sancho et al., 2022; Decena et al., 2023).

Cytogenetic and karyotypic knowledge of *Brachypodium* has been recently expanded in the recent studies of Sancho et al., (2022) and Decena et al., (2023), where up to 26 different taxonomic cytotypes were identified through chromosome counting and genome size (GS) analysis and up to seven diploid progenitor genomes were characterized in the polyploids through phylogenetic inference and karyotypic barcode profiles. Ploidy levels of the *Brachypodium* samples under study were inferred from chromosome and GS estimations performed in the same accessions in Decena et al., (2023) and from the repeat data obtained in this study (see **Table 1** and **Supplementary Table S1**). Total DNA for 41 (perennial *Brachypodium* samples) out of the 44 samples studied was extracted from fresh and silica gel dried leaf tissues of plants growing in the University of Zaragoza – High Polytechnic School of Huesca common garden and from herbarium specimens (*B. mexicanum* 348H, Herbarium B) (**Supplementary Table 1**). Isolation of DNA was accomplished using a modified CTAB protocol (Doyle and Doyle, 1987) using ~20mg of tissue. DNA concentration (100-200ng/ul) and absorbance (260/230 nm of 1.8 to 2.1 and 280/260 nm of 1.8 to 2.0) were estimated using Qubit ® 3.0 (Life Technologies, Grand Island, NY) and Biodrop µLITE (Harvard Biosciences), respectively. Genome skimming sequencing was performed from a PCR-free multiplexed pool of KAPA libraries through the Illumina technology in paired-end mode (2 x 100 bp) at the Spanish Centro Nacional de Análisis Genómicos (CNAG, Barcelona). Illumina paired-end (PE, 101 bp) reads were checked using FastQC\_v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and the adapters and low quality sequences were trimmed and removed using Trimmomatic-0.39 (Bolger et al., 2014). The 41 *Brachypodium* genomic samples used in downstream analysis contained between 10.09 – 30.22 million reads (average 16.7 million reads) with insert sizes ranging between 123 – 329 bp (**Supplementary Table S2**). In addition, sequence data from the three annual *Brachypodium* samples was retrieved from NCBI. The downloaded sequences were filtered using Trimmomatic with the following parameters: SLIDINGWINDOW:15:28 (window of bases: quality threshold) and CROP, HEADCROP and MINLEN according to the per-base sequence content and the length of



101 bp, according to the GS length of other samples, and 1.4-3 million reads of their respective Illumina sequencing data were used for the repeatome analysis (**Supplementary Table S2**).

### **Repeat clustering and annotation, and 5S rDNA graph-clustering analysis**

The repeatome of the *Brachypodium* samples under study was analyzed using RepeatExplorer2 (RE2), a computational pipeline that uses similarity graph-based clustering of filtered PE reads for the identification of the composition and proportion of repetitive elements (Novák et al., 2010, 2013, 2020). Two clustering analyses were performed. First, each sample was analyzed independently using RE2 through the Galaxy platform (<https://repeatexplorer-elixir.cerit-sc.cz>; Galaxy Version 2.3.8.1) following the protocol of Novák et al. (2020). The clustering analysis of individual samples was fed with variable amount of PE reads per sample in order to attain the recommended genome coverage (0.01–0.5×) of each taxon (**Supplementary Table S2**). The clustering was conducted employing default RE2 settings (90% similarity, minimum overlap = 55; cluster size threshold = 0.01 %). Automated RE2 annotation of clusters was used to quantify the clusters and to calculate the proportions of repetitive elements in each sample. Sequence data of the three annual *Brachypodium* samples, enriched with organelle sequences, were pre-processed to remove them using the DUK software (Li et al., 2011) with k-mer size of 24, a cut-off threshold for 2 matched reads, and the respective plastomes of *B. distachyon* (LT558596.1), *B. stacei* (NC\_036837.1; Sancho et al. 2018) and *B. hybridum* ABR113 (NC\_036836.1; S-plastotype), as references. Pre-processing steps were applied to format the sequences according to the requirements for the downstream analyses. Thus split\_pairs.v0.5 ([https://github.com/eead-csic-compbio/split\\_pairs](https://github.com/eead-csic-compbio/split_pairs); Contreras-Moreira et al. (2014) was used to obtain the interleaved paired input format, and seqtk.v.1.3-r117 (<https://github.com/lh3/seqtk>) to convert fastq to fasta format. The headers of the sequences were formatted following the specifications required by the downstream analysis. This independent RE2 analysis resulted in the automatic repeat annotation and quantification of the studied repeatomes.

**Table 1.** Samples included in the repeatome analysis of *Brachypodium*. Species, sample's code, chromosome number (2n), inferred ploidy level (nx), genome size (2C, pg), monoploid genome size (1Cx, pg) and locality of origin. Ploidy levels inferred from chromosome counting, genome sizes, or repeat content. Asterisks indicate inferred 2n values according to those of the same taxonomic cytotypes (this study and previous records). See Supplementary Table S1 for additional information on taxon authorship, detailed localities and vouchers, and cytogenetic and genomic data sources.

Species	Code	Chromosome number (2n)	Genome Size (2C/pg)	Ploidy (nx)	Monoploid genome(1Cx/pg)	Locality
<i>B.distachyon</i>	Bdis_Bd21-3	10	0.631±0	2	0.316	Iraq: Salakudin
<i>B.stacei</i>	Bsta_ABR114	20	0.564±0	2	0.282	Spain: Balearic isles
<i>B. hybridum</i>	BhybABR113	30	1.265±0	4	0.633	Portugal: Lisbon
<i>B.arbuscula</i>	Barb502	18	0.713±0.004	2	0.357	Spain: Canary Isles
	Bboi3	48	3.236±0.072	6	0.539	Spain: Granada
<i>B.boissieri</i>	Bboi10	48	3.152±0.04	6	0.525	Spain: Granada
	Bboi15	48	3.149±0.032	6	0.525	Spain: Granada
	Bmex347	40	3.774	4	0.944	Mexico: Hidalgo
<i>B.mexicanum</i>	Bmex348H	40*	3.774*	4*	0.944*	Mexico: Puebla
	Bmex504	40*	3.774*	4*	0.944*	Ecuador: Loja
	Bpho422	28	1.469±0.012	4	0.367	Slovakia: Ružomberok
	Bpho6-1R	28	1.443±0.019	4	0.361	Spain: Huesca
<i>B.phoenicoides</i>	Bpho452	38	2.176±0.017	6	0.363	Morocco: Rif Mountains
	Bpho552	38	2.204±0.039	6	0.367	Spain: Cádiz
	Bpho553	38	2.183±0.013	6	0.364	Spain: Málaga
	Bpho554-1	38	2.155±0.02	6	0.359	Spain: Granada
	Bpin505	18	0.822±0.009	2	0.411	Norway: Oslo
<i>B.pinnatum</i>	Bpin34	28	1.449±0.018	4	0.362	Great Britain: North Wiltshire
	Bpin514	28	1.537±0.012	4	0.384	Turkey: Samsun
	Bpin520	28	1.499±0.014	4	0.375	Netherlands: Scherpenzeel
<i>B.retusum</i>	Bret400	32	1704±0.024	4	426,000	Spain: Huesca

### Chapter 3

	Bret407	32	1.715±0.017	4	0.429	Spain: Huesca
	Bret453-4	32*	1.855±0.1	4	0.464	Morocco: Rif Mountains
	Bret454	32*	1.836±0.115	4	0.459	Morocco: Tazza-Bou Idir
	Bret504	32	1.669±0.026	4	0.417	France: Hérault
	Bret555	32	1.715±0.017	4	0.429	Spain: Granada
	Bret403	42	2.373±0.958	6	0.396	Spain: Huesca
	Bret408	42	2.431±0.033	6	0.405	Spain: Navarra
	Bret551	42*	2.109±0.025	6	0.352	Spain: Málaga
	Bret557	42	2.464±0.026	6	0.411	Spain: Cádiz
	Bret561	42	2.362±0.046	6	0.394	Spain: Zaragoza
<i>B. rupestre</i>	Brup7	28	1.562±0.016	4	0.391	Russia: Moscow
	Brup439-1	28	1.55±0.022	4	0.388	Spain: Huesca
	Brup441	28	1.483±0.008	4	0.371	Spain: Leon
	Brup442	28	1.56±0.03	4	0.39	Spain: Huesca
	Brup443	28	1.498±0.012	4	0.375	Spain: Guipuzcoa
	Brup444	28	1.492±0.021	4	0.373	Spain: Lugo
	Brup182	38	2.258±0.026	6	0.376	Croatia: Istria
	Brup600	38	2.216±0.013	6	0.369	France: Nans les Pins
	Brup605	38	2.265±0.013	6	0.378	France: Pourrieres
<i>B. sylvaticum</i>	Bsyl54-1	18*	0.888±0.008	2	0.444	Morocco: Rif Mountains
	Bsyl466-6	18*	0.928±0.013	2	0.464	Spain: Huesca
	Bsyl477-1	18*	0.932±0.017	2	0.466	Spain: Lérida
	Bsyl501-6	18*	0.947±0.01	2	0.474	France: Alpes Maritimes

Secondly, the comparative analysis of all the *Brachypodium* samples under study was conducted using the RE2 program installed in our local server (command `repex_tarean/seqclust`) using the following parameters: `/repex_tarean/seqclust -p -l Brachy_clustering.log -c 0 -P 2 -v ./Brachy_clustering Brachy_RE.fasta -C -tax VIRIDIPLANTAE3.0 -opt ILLUMINA`. This comparative clustering analysis was performed employing the same RE2 configuration used for the individual analyses. Organelle clusters and/or clusters with missing data were removed. The resulting clusters were used for phylogenetic downstream analysis

The retrieved 5S data were used to feed the 5S clustering graph analysis performed with the TAREAN option of RE2 (Novák et al., 2010; Garcia et al., 2020) using the same input showed previously for the individualized RE2 analysis (**Table 1** and **Supplementary Table S1** and **S2**). The 5S rDNA clusters' shapes were characterized by a connected component index parameter (C) and their k-mer score was calculated as the sum of frequencies of all k-mers used for consensus sequence reconstruction (García et al. 2020). The 5S rDNA cluster graph topologies were visually inspected and classified into three graph groups (type 1, simple circular-shaped graph; type 2, complex graph with two loops; type 3, complex graph with three loops); in the complex graphs the interconnected loops represent IGS spacers (Garcia et al., 2020). The 5S graphs were inspected to detect potential variation of 5S rDNA loci (ribotypic 5S families) and to identify presumable hybrids and allopolyploids.

### **Correlations of repeat amounts and genome size variation in *Brachypodium***

To analyze the potential contribution of the different repeat types and the repeatome to the variation in monoploid genome size (1Cx) observed between and within the studied *Brachypodium* lineages and samples, we performed a test search using the data from the individual analysis and linear regression model analyses (Pearson correlation coefficient) with the *ggscatter* function from the *ggpubr* package (Kassambara 2023) in R. The respective contributions of repeats to pairwise differences in genome sizes were estimated following Macas et al. (2015) and Moreno-Aguilar et al. (2022) using absolute amounts (Mbp) of repeats calculated for individual species (**Table 2** and **Supplementary Table S3**). We also tested whether there were significant differences in repeat amount for different repeat families obtained from the individual analysis through Kruskal–Wallis rank tests using the *PMCMRplus* package (Pohlert 2023) in R.

### **Landscape genomic diversity analysis of repeat types in *Brachypodium***

To investigate the levels of conservatism or diversity of the repeat types that most contributed to genome size variation in *Brachypodium* (44 studied samples) we performed a genome landscape search for the global variability of these individual repeat types across the *Brachypodium* genomes following Macas et al. (2015) and Moreno-Aguilar et al. (2022). We pooled the pairwise similarity values of reads, retrieved from the RE2 outputs (hitsort files) of the global comparative analysis (all samples together), for each sample and repeat type in a separate dataset and evaluated their similarities with respect to similarities of reads from the same repeat in all other samples. We then calculated intraspecific versus interspecific similarity hit ratios (Hs/Ho ratios) considering that conservative sequence repeats will produce similarity hits with about the same frequency for Hs and Ho, while diversified sequence repeats will generate similarity hits with different frequencies (Macas et al., 2015). We also calculated similarity hit ratios for the 5S tandem-repeat rDNA to compare its gene-conserved vs IGS-variable Hs/Ho ratios with those obtained from the other repeat elements analyzed (Moreno-Aguilar et al., 2022).

### **Repeatome phylogenomic network of *Brachypodium***

We performed phylogenomic analyses with the repeat data obtained from the comparative clustering of the *Brachypodium* repetitive elements. The repeatome super-network was inferred following the steps described by Vitales et al. (2020). The most abundant repeats (342 top clusters), defined as possessing more than 0.01% of the total input reads in the dataset, were employed for phylogenetic analyses. Organelle (plastid/mitochondrial) clusters were removed prior to the phylogenetic inference. For each cluster, the starting data set consisted in the matrices of the observed/expected number of edges between species, which is a measure of the pairwise similarity among the reads of the species. These matrices were extracted from the RE results folder using the Perl script `parse_re2_matrix` (Sánchez-Rodríguez and Moreno-Aguilar, unpub. data). Incomplete matrices lacking pairwise similarity or zero values were excluded. These similarity matrices were transformed to distance matrices by calculating the inverse of the values. The function `NJ` from package `ape` v.5.4-1 (Paradis and Schliep 2019) in R v.4.0.5 (R Core Team 2021) was used to build the neighbor-joining tree for each of the 55 surviving clusters. The super-network was performed using the default parameters in

SplitsTree4 v.4.17.0 (Huson and Bryant 2006). The potential phylogenetic information of the repeatome data set was assessed through topological comparisons of the repeatome network with phylogenomic trees of *Brachypodium* retrieved from transcriptome (Sancho et al., 2022) and whole plastome and nuclear rDNA 35S and 5S data (Decena et al., 2023).

## RESULTS

### Characterization and quantification of the *Brachypodium* repeatome

The annotated repeats recovered by RE2 in the individual analysis showed remarkable differences in repeat types and contents among the 44 *Brachypodium* samples studied (**Table 2** and **Supplementary Table S3; Figure 1**). The proportion of the monoploid genome occupied with repeats ranged from 67.97% (*B. mexicanum*-4x) to 20.77% (*B. stacei*-2x), with a mean across the genus of 28.65% (**Table 2** and **Supplementary Table S3, Figure 1**). The amount and 1Cx-percent coverage of repetitive elements varied considerably within both *Brachypodium* polyploids and diploids. Among the polyploids, the highest percentages corresponded to the ancestral outcore perennial *B. mexicanum* tetraploid samples (56.6-67.9%, mean 60.81%), which outpassed those of all the remaining samples, followed by ancestral outcore perennial *B. boissieri* hexaploid samples (29.35-32.2%, 30.92%), and the lowest to the intermediately evolved *B. retusum* and the recently evolved core perennial *B. pinnatum*, *B. phoenicoides* and *B. rupestre* tetraploid and hexaploid samples (22.5-27.9%), and the annual *B. hybridum* tetraploid sample (22.05%). Within the intermediately-to-recently-evolved groups, some species showed higher and non-overlapping ranges of repeatome percentages in diploids and low polyploids than in high polyploids [e. g., *B. pinnatum* diploids (26.7%) vs tetraploids (22.5-24.9%, mean 23.9%), *B. phoenicoides* tetraploids (25.9-27.9%, 26.9%) vs hexaploids (22.7-24.3%; 23.2%)], while others showed overlapping ranges [e. g., *B. retusum* tetraploids (25.9-27.9; 27.2%) vs hexaploids (24.9-27.2%; 26.1%), *B. rupestre* tetraploids (23.4-25.4%, 24.6%) vs hexaploids (24.2-24.9%; 24.6%)] (**Table 2** and **Supplementary Table 3, Figure 1**). Among the diploids, the highest percentages corresponded to the recent core perennial *B. sylvaticum* samples (32.9-36.1%, mean 34.2%), which notably exceeded those of other core perennial diploids (*B. pinnatum* 26.7%, *B. arbuscula* 22.54%) and of ancestral outcore annual diploids (*B. distachyon* 22.75, *B. stacei* 20.77%). The annual allotetraploid *B. hybridum* sample showed a 1Cx-percent coverage of repeatome equivalent to the mean between those of its annual

progenitor diploid species *B. stacei* and *B. distachyon* (**Table 2** and **Supplementary Table S3, Figure 1**).



# Chapter 3

**Table 2** Genome proportion (percentage) of repeats estimated by RepeatExplorer2 for individual *Brachypodium* samples (estimations per monoploid genome, 1Cx)

Code	Class I/Ty1 Copia										Class I/Ty3 gypsy										Class I		Class II/TIR		Class II										TOTAL (%)
	Class I	Class I LTR	Ty1 copia	Ale	Alesia	Angela	Bianca	Ikeros	Ivana	SIRE	TAR	Tork	Ty3 gypsy	Athila	Ogre	Retand	CRV	Tekay	Reina	pararetrovirus	LINE	EnSpm CACTA	HAT	MuDR Mutator	P/F Harbinger	Heltron	mobile element	satellite	45S rDNA	18S rDNA	25S rDNA	5S rDNA	Unclassified repeat	Unclassified	
Bdis_Bd 21-3	0	0.25	0	0	0.09	0.92	0.34	0.11	0.11	1.17	0.4	0.02	0.45	0	0	6.83	0.01	2.62	0.02	0.11	0.01	0.76	0	0.36	0.02	0	0	0.01	2.15	0	0	0.09	1.3	4.6	22.7 5
Bsta_A BR114	0	0.97	0	0.03	0	1.53	0.07	0.15	0.04	1.06	0.26	0.01	0	1.69	0	1.18	0.64	3.28	0	0	0.07	0.32	0	0.15	0	0	0.04	0.12	0.49	0	0	0.07	2.74	5.86	20.7 7
Bhyb_A BR113	0	1.95	0	0.11	0	0.59	0.46	0.16	0	0.95	0.19	0.2	0	0.67	0	0.06	1.19	1.2	0	0.28	0.03	0.2	0.02	0.05	0.02	0	0	0.16	0	0.04	0.01	0.01	0	13.5	22.0 5
Barb50 2	0	0.04	0	0	0	0.2	0.25	0.27	0.03	1.88	0.22	0.25	0	0	0	4.02	4.22	1.1	0	0.18	0.17	0.34	0	1.49	0.17	0	0	0.06	0.97	0	0	0.28	0	6.4	22.5 4
Bboi3	0	1.06	0	0	0	0.18	0.19	0.19	0	1.45	0.5	0.26	0	0	0	10.4 6	0.57	1.42	0	0.15	0.16	0.55	0	0.4	0.21	0	0	1.15	0.66	0	0	0.1	4.85	4.84	29.3 5
Bboi10	0	0.98	0	0	0	0.17	0.1	0.16	0	1.22	0.47	0.2	0	0	0	11.7 7	0.45	1.47	0	0.07	0.14	0.62	0	0.43	0.21	0	0	0.22	1.7	0	0	0.16	6.55	4.13	31.2 2
Bboi15	0	1.12	0	0	0	0.17	0.12	0.21	0.01	1.23	0.47	0.25	0	0	0	13.1 9	0.46	1.28	0	0.04	0.16	0.63	0	0.42	0.21	0	0	1.52	0.79	0	0	0.12	5.42	4.37	32.1 9
Bmex34 7-2	0	1.4	0	0.2	0	0.72	0.2	0.53	0.2	2.22	0.87	0	0	0.28	1.93	20.7	0.81	27.0 3	0	0	0.02	2.11	0	3.39	0.03	0.01	0	1.67	0.76	0	0	0.18	0	2.71	67.9 7
Bmex34 8H	0	0.98	0	0.16	0	0.76	0.22	0.51	0.16	1.02	0.76	0.34	0	0.09	1.93	16.6 4	0.2	21.8 4	0	0.07	0.08	0.97	0.18	2.29	0.12	0.04	0.04	1.43	0.66	0	0	0.17	1.15	3.83	56.6 4
Bmex50 4	0	0.3	0	0.2	0	0	0.59	0.66	0.28	1.57	0.85	0	0	1.45	1.14	16.8 3	2.23	18.7 5	0.02	0.12	0.05	2.3	0.33	2.63	0.35	0.35	0.93	1.33	0.45	0	0	0.26	0.38	3.47	57.8 2

Chapter 3

Bpho6-1R	0	1.5	0	0.03	0	0.05	0.1	0.17	0.03	1.69	0.69	0.31	0	0	0	4.73	4.41	1.95	0	0.02	0.18	0.84	0	1.01	0.35	0	0	0.46	0	0	0	0.44	0.75	6.22	25.93
Bpho422	0	1.87	0	0	0	0.07	0	0.18	0	1.63	0.72	0.35	0	0	0	5.47	0.86	2.07	0	0	0.09	1.35	0	1.03	0.33	0	0	4.92	1.51	0	0	0.16	0.03	5.24	27.88
Bpho452	0	1.63	0	0.01	0	0.06	0.12	0.13	0.03	1.91	0.46	0.18	0	0	0	4.39	0.03	1.14	0	0.17	0.07	0.71	0.02	0.89	0.28	0	0	0.17	0	0	0	0.1	7.56	4.2	24.26
Bpho552	0	0.4	0	0.02	0	0.06	0.12	0.14	0.04	1.81	0.44	0.18	0	0.11	0	4.19	0.61	1.97	0	0.13	0.14	0.74	0	0.79	0.27	0	0	5.27	0.75	0	0	0.17	0	4.42	22.77
Bpho553	0	1.94	0	0.03	0	0.06	0.11	0.15	0.05	1.57	0.44	0	0	0.01	0	4.17	0.61	0.62	0	0.25	0.22	0.61	0	0.84	0.36	0	0	4.42	1.48	0	0	0.15	0	5.15	23.24
Bpho554-1	0	2.06	0	0.02	0	0.08	0.12	0.14	0	1.79	0.46	0.21	0	0.16	0	3.39	0.03	0.56	0	0.11	0.18	0.75	0	0.87	0.2	0	0	0.47	0	0	0	0.11	5.66	5.34	22.71
Bpin505	0	1.14	0	0.01	0	0.01	0.25	0.18	0.01	1.94	0.58	0.27	0	0.01	0	5.64	0.03	2.47	0	0	0.19	1.21	0	1.19	0.25	0.45	0	0.29	1.4	0	0	0.08	4.5	4.58	26.68
Bpin34	0	2.36	0	0	0	0	0.13	0.2	0	1.59	0.48	0.26	0	0	0	5.3	0.61	0.94	0	0	0.07	0.88	0	0.85	0.31	0	0	4.92	1.14	0	0	0.06	0	4.22	24.32
Bpin514	0	1.8	0	0	0	0.13	0.15	0.21	0	1.71	0.57	0.17	0	0	0	5.48	0.58	1.21	0	0	0	1.04	0	0.89	0.3	0	0	6.05	0.72	0	0	0.1	0	3.8	24.91
Bpin520	0	2.58	0	0.02	0	0.1	0.18	0.19	0	1.89	0.47	0.31	0	0.17	0	3.83	0	1.02	0	0	0.13	1	0	0.9	0.27	0	0	0.26	0.63	0	0	0.07	4.1	4.35	22.47
Bret400	0	1.34	0	0	0	0.17	0.19	0.21	0.06	1.64	0.44	0.24	0	0	0	6.59	0.67	2	0	0.15	0.21	0.63	0	0.7	0.22	0	0	0.29	0	0	0	0.08	4.82	5.25	25.9
Bret407	0	0.97	0	0	0	0.12	0.15	0.19	0.08	1.59	0.44	0.21	0	0	0	7.78	0.5	2.45	0	0.13	0.07	0.76	0.01	0.69	0.38	0	0	1.11	1.18	0	0	0.12	3.38	5.07	27.38
Bret453-4	0	2.07	0	0.01	0	0.14	0.16	0.2	0.08	1.76	0.48	0.23	0	0	0	7.75	0.54	1.77	0	0.18	0.14	0.99	0	0.64	0.25	0	0	5.11	0.89	0	0	0.07	0	4.48	27.94
Bret454	0	1.5	0	0.01	0	0.15	0.18	0.18	0.09	1.62	0.4	0.2	0	0	0	7.33	0.6	1.95	0	0.14	0.04	0.62	0	0.71	0.3	0	0	0.52	0.95	0	0	0.17	4.75	4.77	27.18

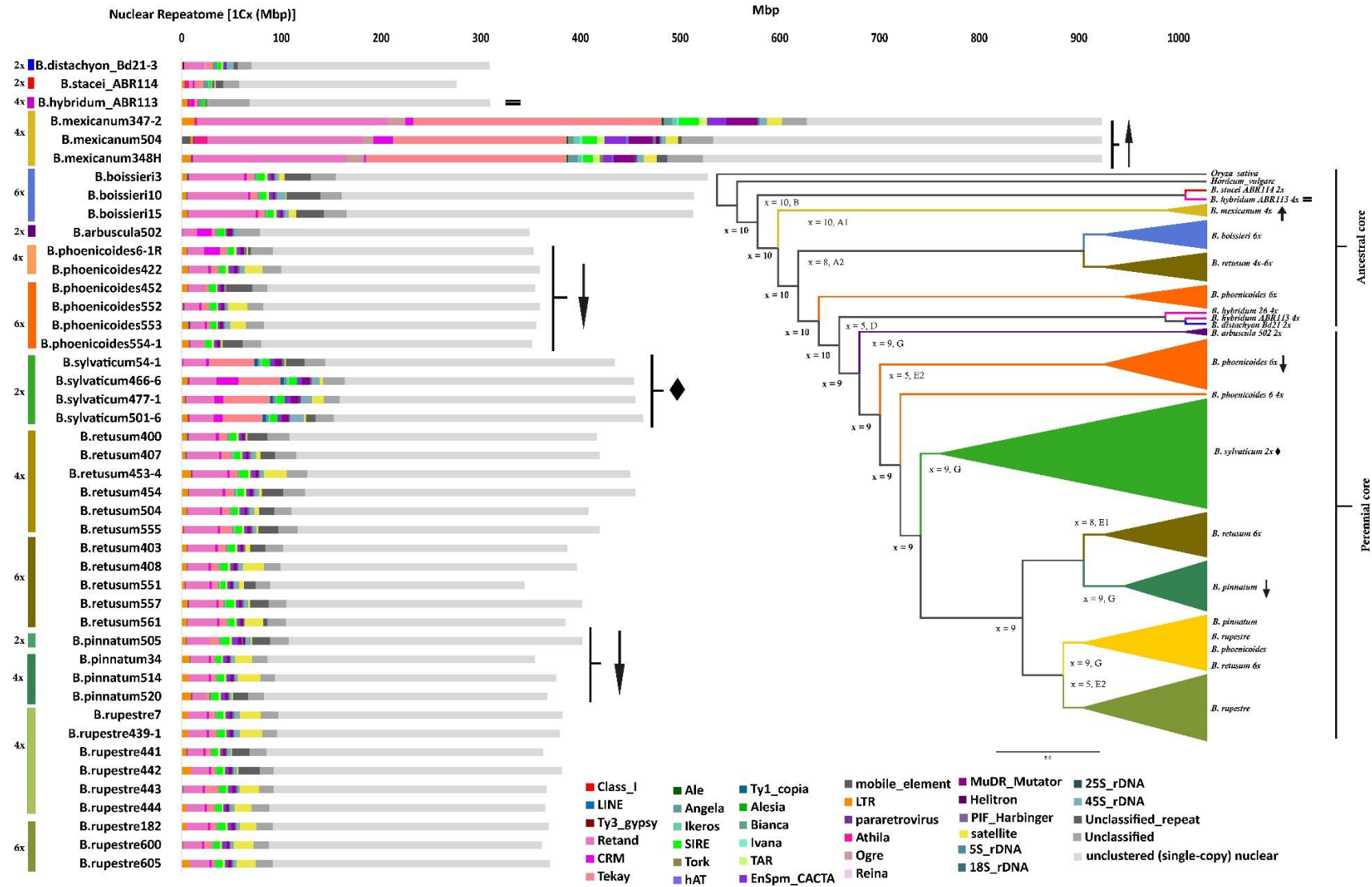
Chapter 3

Bret504	0	1.28	0	0	0	0.04	0.16	0.2	0.1	1.41	0.43	0.22	0	0	0	7.95	0.54	1.9	0	0.16	0.04	1.11	0	0.66	0.33	0	0	1.18	1.21	0	0	0.13	3.74	4.17	26.96
Bret555	0	0.37	0	0	0	0.16	0.13	0.19	0.09	1.62	0.45	0.19	0	0	0	8.17	0.48	3.01	0	0.15	0.03	0.75	0	0.73	0.34	0	0	0.45	0.94	0	0	0.13	4.77	4.58	27.73
Bret403	0	1.36	0	0.05	0	0.13	0.15	0.19	0.05	1.71	0.48	0.32	0	0	0	7.12	0.61	2.2	0	0.09	0.18	0.61	0	0.76	0.37	0	0	1.35	0	0	0	0.12	3.81	4.73	26.39
Bret408	0	1.39	0	0	0	0	0.16	0.21	0	1.87	0.5	0.26	0	0.04	0	5.27	0.63	2.05	0	0	0.1	0.7	0	0.93	0.27	0	0	5.29	1.1	0	0	0.06	0	4.14	24.97
Bret551	0	0.96	0	0	0	0.07	0.15	0.17	0.09	1.57	0.43	0	0	0	0	7.08	0.6	2	0	0.14	0.03	0.94	0	0.68	0.3	0	0	1.39	1.43	0	0	0.09	3.43	4.24	25.79
Bret557	0	1.44	0	0	0	0.1	0.14	0.2	0.07	2.06	0.44	0.17	0	0.02	0	6.93	0.5	1.31	0	0.12	0.21	0.68	0	0.75	0.33	0	0	0.53	1.08	0	0	0.14	4.47	4.44	26.13
Bret561	0	1.47	0	0.03	0	0.07	0.13	0.2	0.05	1.64	0.49	0.32	0	0	0	7.63	0.63	1.83	0	0.12	0.04	0.66	0	0.71	0.29	0	0	4.81	0	0	0	0.1	1.21	4.8	27.23
Brup7	0	1.53	0	0.05	0	0.04	0.13	0.21	0	1.69	0.57	0.21	0	0	0	5.06	0.57	1.65	0	0	0.04	0.74	0	0.89	0.35	0	0	5.45	1.53	0	0	0.11	0	4.59	25.41
Brup43 9-1	0	1.54	0	0	0	0	0.14	0.23	0	1.76	0.47	0.26	0	0.03	0	5.09	0.58	1.94	0	0	0.02	0.72	0	0.86	0.23	0	0	5.85	1.5	0	0	0.12	0	3.97	25.31
Brup44 1	0	1.35	0	0	0	0.07	0.13	0.15	0	1.54	0.53	0.28	0	0	0	4.33	0.63	1.62	0	0.2	0.09	0.55	0	0.75	0.34	0	0	0.34	0.9	0	0	0.13	4.83	4.66	23.42
Brup44 2	0	2.43	0	0	0	0.05	0.15	0.18	0	1.73	0.46	0.24	0	0.01	0	4.84	0.52	0.98	0	0	0	0.86	0	0.79	0.3	0	0	0.34	0.89	0	0	0.15	5.56	3.73	24.21
Brup44 3	0	0.33	0	0.01	0	0	0.15	0.19	0.01	1.76	0.47	0.25	0	0.01	0	5.35	0.58	3.66	0	0	0.12	1.02	0	0.98	0.39	0	0	5.18	0.61	0	0	0.1	0	4.04	25.21
Brup44 4	0	1.27	0	0	0	0.01	0.16	0.17	0	1.71	0.46	0	0	0.01	0	5.08	0.58	2.04	0	0	0.04	0.6	0	0.84	0.26	0	0	4.67	1.21	0	0	0.15	0	4.89	24.15

Chapter 3

Brup182	0	1.25	0	0	0	0.07	0.15	0.22	0	1.97	0.6	0.24	0	0.01	0	5.9	0.55	1.78	0	0	0.01	0.79	0	0.95	0.35	0	0	4.45	0.93	0	0	0.12	0	4.59	24.93
Brup600	0	0.42	0	0	0	0	0.13	0.17	0	1.71	0.54	0.2	0	0	0	5.3	0.54	2.31	0	0	0.02	1.12	0	0.78	0.21	0	0	5.52	0.98	0	0	0.09	0	4.2	24.24
Brup605	0	2.19	0	0	0	0.07	0.14	0.19	0	1.78	0.6	0.21	0	0	0	5.32	0.58	0.82	0	0.01	0.01	0.72	0	0.88	0.34	0	0	5.22	0.94	0	0	0.07	0	4.64	24.73
Bsyl54-1	0	0.03	0.72	0	0	0.37	0.19	0.38	0.11	1.82	0	0	0	0.29	0	5.24	0.6	10.55	0	0.06	0.06	1.15	0	1.61	0.35	0	0	0.38	0	0	0	0.22	4.3	4.68	33.11
Bsyl466-6	0	1.35	0.72	0	0	0.35	0.2	0.38	0.08	1.92	0	0	0	0.26	0	5.97	4.82	9.35	0	0.01	0.16	0.97	0.08	1.81	0.29	0.03	0	0.86	1.64	0	0	0.07	0	4.76	36.08
Bsyl477-1	0.1	0.69	0.66	0	0	0.39	0.16	0.33	0.03	1.68	0	0	0	0.11	0	6.23	2.04	10.26	0	0.02	0.02	1.38	0	1.68	0.28	0.02	0	2.46	2.55	0	0	0.14	0	3.49	34.82
Bsyl501-6	0	1.33	0.66	0	0	0.39	0.15	0.3	0.03	1.68	0	0	0	0.25	0	5.31	1.9	8.69	0	0.05	0.06	0.87	0.01	1.46	0.3	0.02	0	0.53	2.91	0	0	0.09	2.03	3.92	32.94
Mean	0.00	1.28	0.06	0.02	0.00	0.20	0.17	0.22	0.05	1.65	0.45	0.18	0.01	0.13	0.11	6.75	0.89	3.91	0.00	0.08	0.09	0.86	0.01	0.98	0.26	0.02	0.02	2.23	0.95	0.00	0.00	0.13	2.18	4.73	28.65

LTR-Gypsy retrotransposons represented the major fractions of the repeatome in all the studied *Brachypodium* genomes followed by Satellite repeats, LTR-Copia retrotransposons, and Class II TIR-transposons (**Table 2** and **Supplementary Table S3, Figure 1**). LTR-Gypsy Retand (mean 6.75%) and Tekay (3.91%) elements were the most represented repeats in all genomes. However, while the Retand repeats covered the highest percentages of genomes in almost all species and similar values in most core perennial clade samples (16.6-20.7% *B. mexicanum*, 10.5-13.2% *B. boissieri*, 5.2-8.2% *B. retusum*, 4-6.2% core perennials, 6.8% *B. distachyon*, 1.8-0.06% *B. stacei*, *B. hybridum*, **Table 2**; non-significant differences in Kruskal-Wallis tests, **Supplementary Table 3**), the Tekay repeats showed considerable percent differences among species, being highly abundant in the *B. mexicanum* (18.08-27%) and *B. sylvaticum* (8.7-10.6%) genomes (more frequent than the Retand elements), and less abundant in the annual (1-3%) and the remaining core perennial genomes (0.6-3.7%) (significant differences in Kruskal-Wallis tests). LTR-Copia SIRE elements (1-2.2; mean 1.65%) were relatively uniformly distributed across the *Brachypodium* genomes and showed similar coverage percentages than Tekay elements in most core perennial species (0.6-3.7%). Other types of repeats showed in general coverages <10% in most genomes with some exceptions in those of the *B. mexicanum* samples. ClassII-TIR Mutator (0.05-3.4; mean 0.98%) and CACTA (0.2-2.3; mean 0.86%) transposons were also evenly distributed across the *Brachypodium* genomes, although the former were more abundant in the *B. mexicanum* (Mutator: 2.8%; CACTA: 1.8) and *B. sylvaticum* (Mutator: 1.64%; CACTA: 1.09) than in the other species genomes.



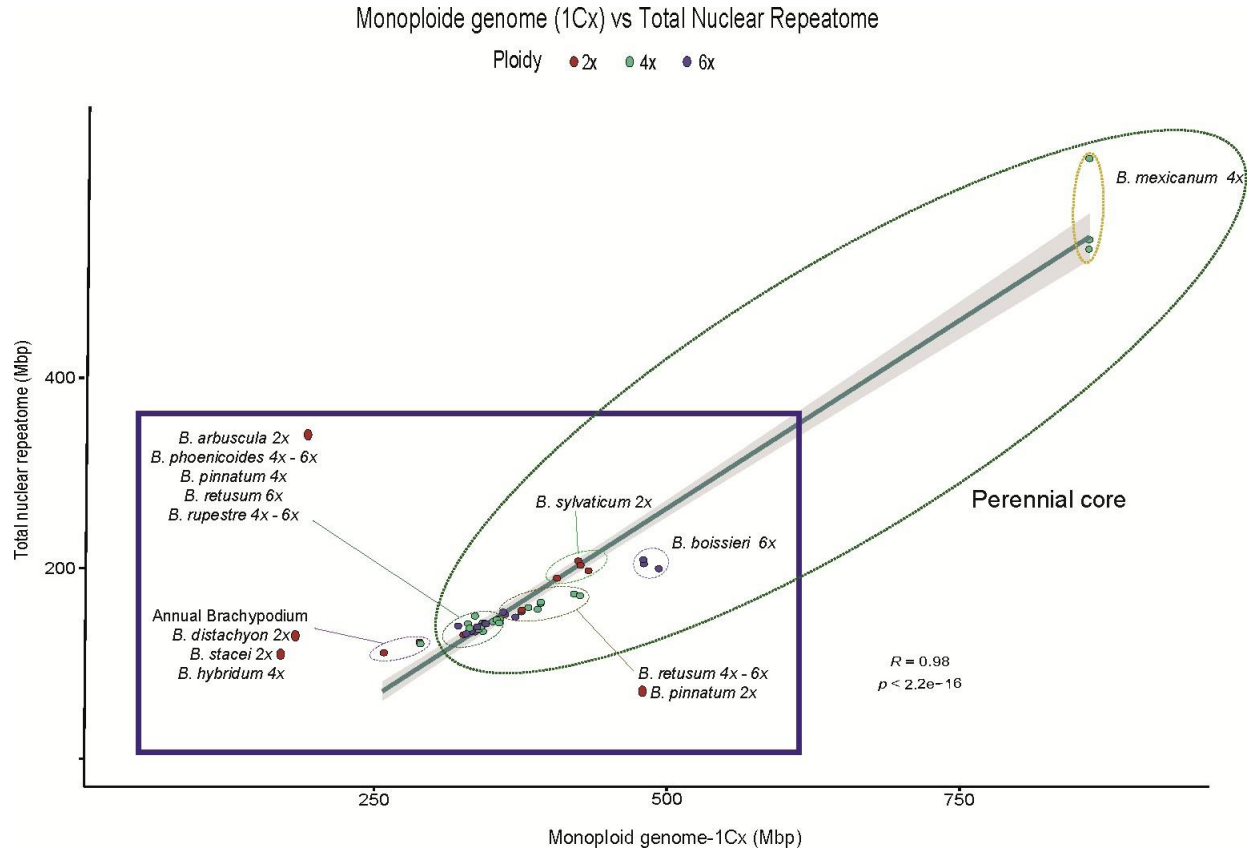
**Figure 1.** Stacked charts of repeat contents per holoploid genome (1Cx) retrieved from the individual RepeatExplorer2 analyses of the studied *Brachypodium* samples. Colour codes for species - cytotypes are indicated in the left part of the figure. Colour codes for repeat subfamilies are indicated in the chart. The inset shows a summarized Astral species tree (supertree of whole plastome and nuclear rDNA 35S and 5S trees) of the studied samples based on the broad phylogenomic study of Decena et al. (2023); karyotypes and subgenomes correspond to those retrieved in Sancho et al. (2022) and Decena et al. (2023), inferred karyotypes of ancestors are highlighted in bold. Scale bar: number of mutations per site. Symbols in the histograms and the summarized tree specify examples of the three alternative scenarios of repeatome size variation in response to the polyploid shock hypothesis (arrow up: *B. mexicanum*, exacerbated increase of repeatome; equality sign: *B. hybridum*, equivalent amount of repeatome to that of diploid progenitor species; arrow down: *B. phoenicoides*, *B. pinnatum*, considerable reduction of repeatome with increasing ploidy-level) and of potential diploidized paleopolyploidy (diamond: *B. sylvaticum*).

Similarly, LTR-Copia Ikeros (0.22%) and Angela (0.2%) elements were more frequent in the *B. mexicanum* (~0.5%, ~0.7%) and *B. sylvaticum* (~0.38%, ~0.4%), and Angela also in the annual species (0.6-1.5%) than in the remaining *Brachypodium* genomes (<0.3%). LTR-Copia TAR elements (0.45%) were present in all but *B. sylvaticum* genomes, while Tork elements (0.18%) were absent in *B. sylvaticum*, most *B. mexicanum* and a few core perennial genomes. LTR-Gypsy Ogre elements were exclusively found in the *B. mexicanum* genomes (1.14-1.93%). LTR-Copia Bianca (0.17%), Ivana (0.05%), Ale (0.02%) and Alesia (<0.01%), LTR-Gypsy Athila (0.13%) and Reina (<0.01%), and Class II Harbinger (0.26%), Helitron (0.02%) and hAT (0.01%) elements were only residually present in a few genomes. Nonspecific tandem satellite repeats (2.23%) were generally well to moderately or poorly represented in most *Brachypodium* genomes, although their frequencies were unevenly distributed across the different groups (**Table 2** and **Supplementary Table S3**). The obese *B. mexicanum* genomes presented the highest percentages of genome coverage for most repeat families (**Table 2** and **Supplementary Table S3**, **Figure 1**). Kruskal-Wallis rank tests performed for each of the *Brachypodium* repeat elements found significant differences for Tekay, Angela, Bianca, TAR, Tork, Helitron, and satellite repeats when examined in all samples (**Supplementary Table 3**).

### Global variability and genome landscape of the *Brachypodium* repeatome

Regression model analysis of repeat content and monoploid genome sizes differences among the *Brachypodium* samples showed a strong correlation when data from all main repeats were combined ( $R^2 = 0.98$ ,  $p < 2.2E-16$ ), accounting for 49.4% differences in genome size between species (**Table 3** and **Figure 2**). Most repetitive elements (22) presented high correlations. Among them, Retand had the highest correlation values ( $R^2 = 0.96$ ,  $p = 5.41E-30$ ), followed by Ikeros ( $R^2 = 0.91$ ,  $p = 9.07E-24$ ), Tekay ( $R^2 = 0.87$ ,  $p = 7.45E-20$ ), Mutator ( $R^2 = 0.84$ ,  $p = 2.08E-18$ ), Ogre ( $R^2 = 0.83$ ,  $p = 1.79E-17$ ), and others (**Table 3**). The repeat family accounting for the highest contribution to pairwise differences in genome sizes was Retand (21.7%), followed by Tekay (6.69%), while the contributions of the other repeats were <3% (**Table 3** and **Supplementary Figure S2**).

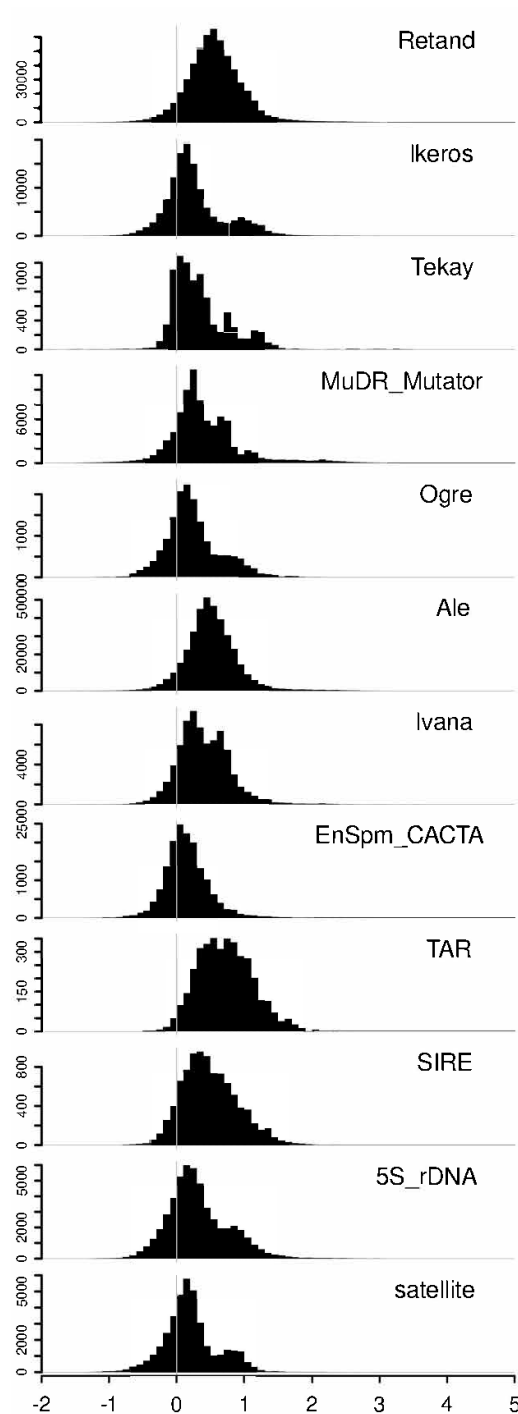




**Figure 2.** Correlation plot of repeat content and genome size variation (1Cx) for the studied *Brachypodium* samples. Summed abundance values of the most represented repeat types obtained from the individual RepeatExplorer2 analysis. Pearson correlation analysis ( $R = 0.98$ ,  $p < 2.2e-16$ ). Ellipses with dashed lines encircle the main *Brachypodium* groups. Colour codes for ploidy level are indicated in the chart.

The global variability analysis of the individual repeat types across the *Brachypodium* genomes showed different histogram profiles of Hs/Ho hit ratios (**Figure 3**). Using the histogram of 5S rDNA sequences as a reference, where a narrow major peak near zero on the  $\log(Hs/Ho)$  x-axis indicated that the ratios of intraspecific Hs to interspecific Ho hit frequencies were close to one, therefore reflecting the high sequence conservation of the 5S genes, while a wide right-hand tail of  $\log(Hs/Ho)$  hit values ranging from 0.1 to 3, indicated the high divergence of IGS sequences of 5S rDNA (Moreno-Aguilar et al., 2022), the histograms of the ten analyzed repeats showed contrasting patterns. Although most histograms had general Gaussian distributions for  $\log(Hs/Ho)$  hit values, a majority of them presented main peak values  $>0.5$  and a skewed distribution towards positive 1-3 values (Retand, Mutator, Ale, Ivana, TAR, SIRE, satellite), while the others had main peak values close to zero (Tekay, Ikeros, Ogre, CACTA) but also with skewed tails towards positive x-axis values (**Figure 3**). These results suggested an overall higher conservatism of the Tekay sequences and a higher diversification of the Retand sequences in the *Brachypodium* genome landscape

regarding these two types of major repeats, and similar dynamics for the other minor repeats (**Table 3**).



**Figure 3.** Global variability of the main repeat types and their sequence conservation across the *Brachypodium* genome. Histograms showing distributions of read similarity (Hs/Ho) hit ratios [frequencies of read similarity hits to reads from the same species (Hs) or to reads from all other species (Ho) (log scale, x-axis) and the number of reads (y-axis)]. Hs/Ho ratios close to one (0 on the logarithmic scale) indicate sequence conservatism while larger values indicate sequence diversification.

**Table 3.** Pearson linear correlation of repeat abundance with genome size variation (1Cx) in *Brachypodium* and contribution of individual repeats to the genome size differences between species. Only the most represented *Brachypodium* repeat types are shown. Significant values are highlighted in bold.

Repeat	Correlation to genome size		Abundance in analysed genomes [Mb/1Cx]		Average contribution to pairwise differences in genome sizes [%]
	R <sup>2</sup>	P-value	min.	max	
Retand	<b>0.955</b>	<b>5.41E-30</b>	0.19	191.01	<b>21.7</b>
Ikeros	<b>0.912</b>	<b>9.07E-24</b>	0.34	6.09	<b>0.45</b>
Tekay	<b>0.865</b>	<b>7.45E-20</b>	1.97	249.42	<b>6.69</b>
MuDR_Mutator	<b>0.842</b>	<b>2.08E-18</b>	0.15	31.28	<b>1.45</b>
Ogre	<b>0.825</b>	<b>1.79E-17</b>	0.00	17.81	<b>0</b>
Ale	<b>0.804</b>	<b>1.92E-16</b>	0.00	1.85	<b>0</b>
Ivana	<b>0.784</b>	<b>1.51E-15</b>	0.00	2.58	<b>0.112</b>
EnSpm_CACTA	<b>0.765</b>	<b>8.92E-15</b>	0.62	21.22	<b>1.24</b>
TAR	<b>0.744</b>	<b>5.40E-14</b>	0.00	8.03	<b>0.551</b>
SIRE	<b>0.698</b>	<b>1.75E-12</b>	2.92	20.48	<b>1.62</b>
Bianca	<b>0.578</b>	<b>2.18E-09</b>	0.00	5.44	<b>0.265</b>
hAT	<b>0.517</b>	<b>3.83E-08</b>	0.00	3.05	<b>0</b>
Angela	<b>0.344</b>	<b>2.90E-05</b>	0.00	7.01	<b>0.426</b>
Unclassified	<b>0.326</b>	<b>5.12E-05</b>	14.19	41.75	<b>2.86</b>
Athila	<b>0.298</b>	<b>0.000125</b>	0.00	13.38	<b>0</b>
mobile_element	<b>0.298</b>	<b>0.000126</b>	0.00	8.58	<b>0</b>
Helitron	<b>0.269</b>	<b>0.00031</b>	0.00	3.23	<b>0</b>
Reina	<b>0.213</b>	<b>0.00162</b>	0.00	0.18	<b>0</b>
rDNA	<b>0.169</b>	<b>0.00562</b>	0.19	13.89	<b>1.12</b>

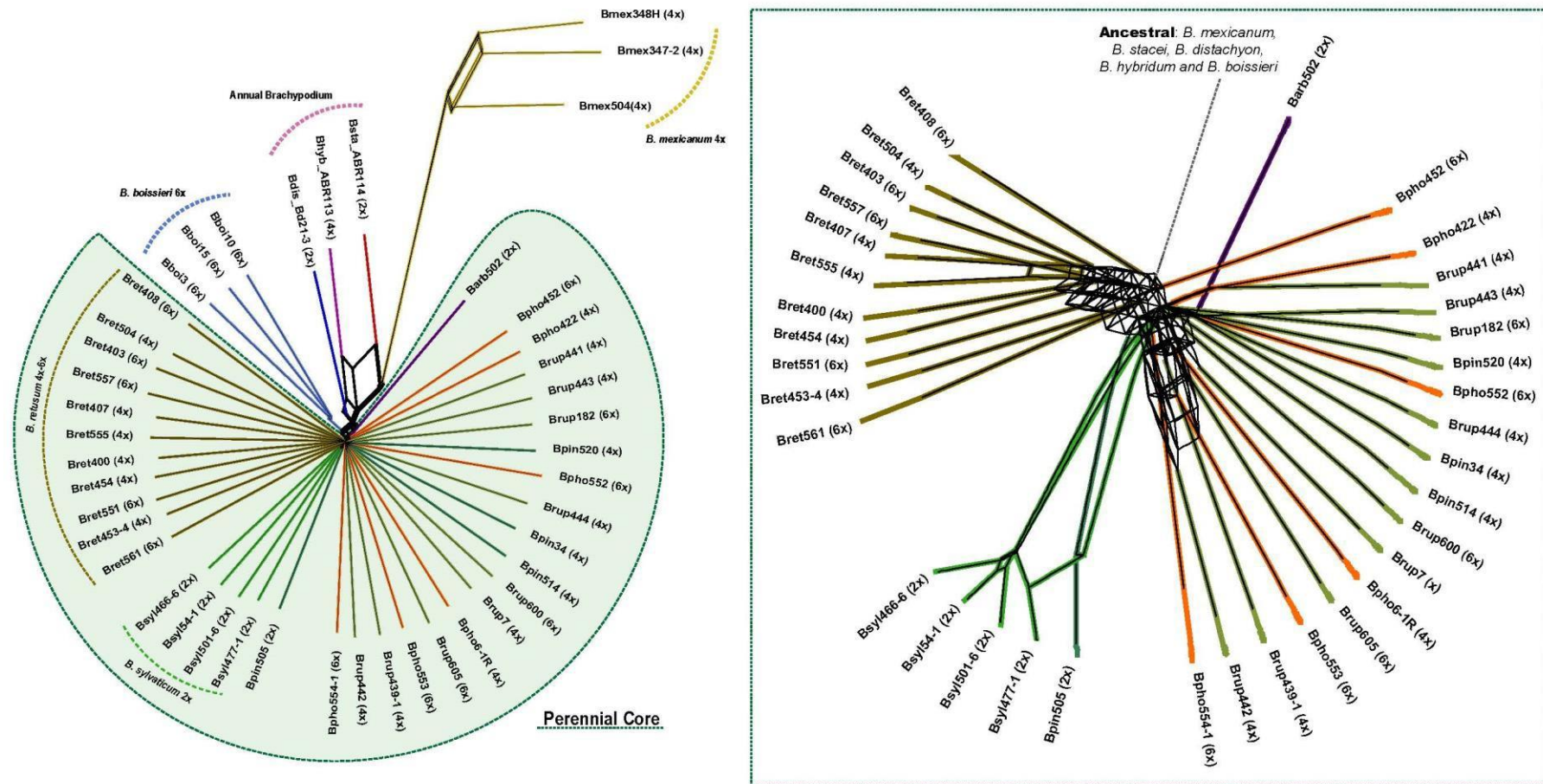
PIF_Harbinger	<b>0.13</b>	<b>0.0161</b>	0.00	3.23	<b>0.3</b>
CRM	<b>0.129</b>	<b>0.0166</b>	0.00	21.87	<b>0.516</b>
LTR	<b>0.106</b>	<b>0.0313</b>	0.13	12.92	<b>0.388</b>
pararetrovirus	0.0635	0.099	0.00	1.11	0.0109
LINE	0.0503	0.143	0.00	0.87	0.0724
Tork	0.0354	0.221	0.00	3.14	0.13
Alesia	0.016	0.413	0.00	0.28	0
Ty3_gypsy	0.016	0.413	0.00	1.39	0
satellite	0.0133	0.456	0.03	22.99	1.32
Unclassified repeat conflicting	0.004	0.683	0.00	33.65	0.618
Ty1_copia	0.00277	0.735	0.00	3.27	0
Class_I	0.000885	0.848	0.00	0.46	0
All	<b>0.961</b>	<b>3.97E-31</b>	57.76	627.19	<b>49.4</b>

### The *Brachypodium* repeatome phylogenetic network and 5S rDNA graph-clusters

The RE2 comparative analysis of repeats recovered different types and numbers of shared or species specific repetitive elements in each *Brachypodium* lineage (**Supplementary Tables S3 and S4**, and **Figure 1**). RE2 annotated different numbers of top clusters in the studied taxa (342 *Brachypodium* clusters; total number of reads 2914070 (48.1%); minimum number of reads 775 (0.03% of the top cluster reads) (**Supplementary Table S4**), representing presumably orthologous repeat families from different samples which grouped together due to their high repeat sequence similarity (Macas et al., 2015).

We reduced the number of top clusters used to build the NJ trees to 55 clusters after discarding clusters with NA or zero read values for some samples (**Supplementary Table S4**) and computed the respective NJ trees using the transformed similarity matrices to distance matrices by calculating the inverse of the values. The phylogenomic network constructed from the distance-based NJ trees revealed the clear divergences of the ancestral *Brachypodium* outcore lineages and

the less resolved relationships of the recent core perennial lineages (**Figure 4**). Among the former group, a *B. mexicanum* subcluster was highly isolated from the others, although it was more closely related to the also ancestral *B. stacei* lineage. The allotetraploid *B. hybridum* lineage nested between its two diploid progenitor species *B. stacei* and *B. distachyon* lineages, while the outcore *B. boissieri* subcluster fell close to the *B. distachyon* lineage (**Figure 4**). Within the intricate core perennial group, a slightly older subcluster included most *B. retusum* 4x and 6x samples, the diploid *B. arbuscula* lineage separated from the rest, all the diploid *B. sylvaticum* samples grouped together in an isolated subcluster, and a more recent subcluster included the representative samples of the *B. pinnatum* complex taxa (*B. pinnatum* 2x and 4x, *B. rupestre* 4x and 6x, *B. phoenicoides* 4x and 6x) (**Figure 4**).



**Figure 4.** Evolutionary network based on standardized repeat data set obtained from the comparative RepeatExplorer2 analysis of *Brachypodium*. A consensus network was constructed with SplitsTree from distance-based NJ trees computed with transformed similarity matrices to distance matrices by calculating the inverse of the values between samples (see text). Abbreviations of *Brachypodium* species and cytotype samples correspond to those indicated in Table 1, and colours codes to those indicated in Supplementary Figure S1. The inset shows a detailed view of the *Brachypodium* core-perennial subnetwork.

**Table 4.** Genomic pair-end read features of 5S rDNA loci and cluster graph parameters of the studied *Brachypodium* taxonomic cytotypes. Graph shape types (type 1, simple circular-shaped graph with one loop; types 2 and 3, complex graph with two and three loops, respectively, where the interconnected loops represent IGS spacers).

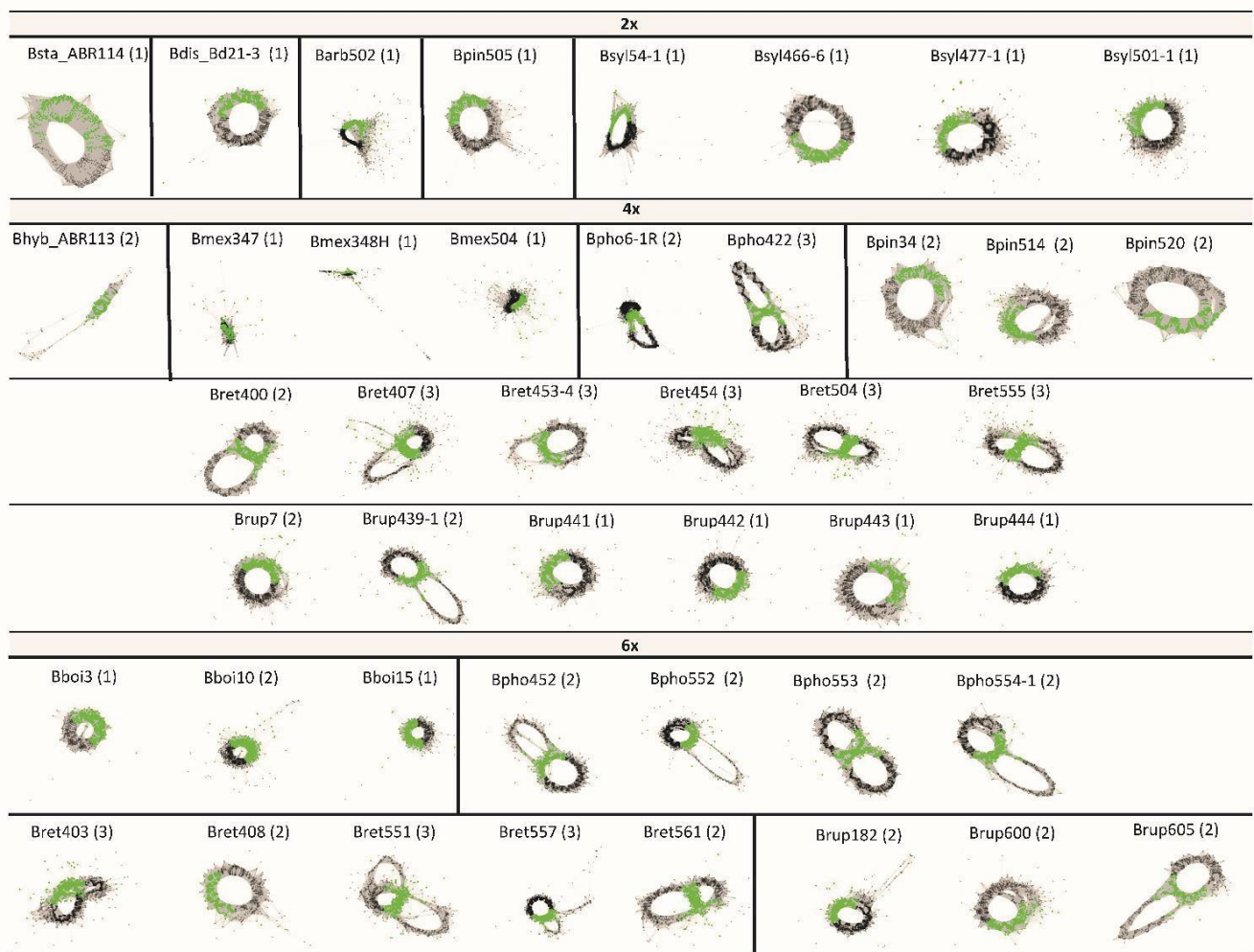
Taxon-Cytotype	Code	No.Reads in cluster	Genome proportion (%)	Consensus repeat length (bp)	k-mer coverage	Connected component index C	Graph shape (type)
<i>B.stacei</i> -2x	Bsta_ABR114	970	0.071	270	0.974	0.997	1
<i>B.distachyon</i> -2x	Bdis_Bd21-3	1450	0.095	370	0.844	0.995	1
<i>B.hybridum</i> -4x	Bhyb_ABR113	315	0.01	272	0.818	0.8	2
<i>B.arbuscula</i> -2x	Barb502	4887	0.28	400	0.801	0.983	1
	Bboi3	1153	0.1	271	0.82	0.989	1
<i>B.boissieri</i> -6x	Bboi10	2072	0.16	271	0.823	0.964	2
	Bboi15	1253	0.12	271	0.815	0.974	1
	Bmex347	5483	0.18	273	0.758	0.992	1
<i>B.mexicanum</i> -4x	Bmex348H	6625	0.17	260	0.719	0.979	1
	Bmex504	9701	0.26	339	0.767	0.981	1
<i>B.phoenicoides</i> -4x	Bpho6-1R	9781	0.44	375	0.825	0.998	2
	Bpho422	2766	0.16	427	0.614	0.996	3
	Bpho452	2048	0.1	430	0.712	0.976	2
<i>B.phoenicoides</i> -6x	Bpho552	3282	0.17	376	0.767	0.967	2
	Bpho553	3404	0.15	378	0.585	0.986	2
	Bpho554-1	2356	0.11	376	0.717	0.999	2
<i>B.pinnatum</i> -2x	Bpin505	1580	0.079	373	0.926	0.98	1
<i>B.pinnatum</i> -4x	Bpin34	1028	0.059	373	0.919	0.989	2
	Bpin514	1425	0.1	373	0.817	0.965	2



	Bpin520	1427	0.072	373	0.929	0.999	2
<i>B.retusum-4x</i>	Bret400	1486	0.081	273	0.616	0.99	2
	Bret407	2298	0.12	276	0.669	0.932	3
	Bret453-4	1081	0.069	408	0.762	0.981	3
	Bret454	2938	0.17	373	0.577	0.938	3
	Bret504	2248	0.13	373	0.647	0.944	3
	Bret555	2196	0.13	373	0.614	0.957	3
<i>B.retusum-6x</i>	Bret403	2737	0.12	375	0.705	0.946	3
	Bret408	969	0.059	373	0.818	0.971	2
	Bret551	1684	0.094	274	0.542	0.943	3
	Bret557	2692	0.14	430	0.816	0.974	3
	Bret561	1776	0.1	395	0.563	0.97	2
<i>B.rupestre-4x</i>	Brup7	1908	0.11	373	0.81	0.971	2
	Brup439-1	1812	0.12	373	0.731	0.973	2
	Brup441	2391	0.13	373	0.86	0.967	1
	Brup442	2481	0.15	388	0.811	0.969	1
	Brup443	1788	0.097	373	0.877	0.977	1
	Brup444	2646	0.15	397	0.912	0.98	1
<i>B.rupestre-6x</i>	Brup182	2283	0.12	365	0.814	0.977	2
	Brup600	1314	0.094	376	0.828	0.978	2
	Brup605	1067	0.07	373	0.668	0.993	2

The analysis of the 5S rDNA clusters of the 44 *Brachypodium* samples studied produced different types of simple and complex graphs (**Table 4** and **Figure 5**) which corresponded to the short (5S-S) and long (5S-L) 5S sequences detected in our previous study (Decena et al., 2023). Noticeably, in most cases the retrieved graphs matched the

expected types for their respective ploidy levels (**Table 1** and **Supplementary Table S1**), detected FISH 5S rDNA loci, nature of the polyploidy (auto- vs segmental- vs allopolyploidy) and number and identity of subgenomes (Sancho et al., 2022; Decena et al., 2023). Thus, graph topologies of diploid taxa corresponded to a simple circular type-1 graph which probably represents a single 5S gene family and locus (outcore *B. stacei* and *B. distachyon* and core perennial *B. arbuscula*, *B. pinnatum* and *B. sylvaticum* 2x samples). In contrast, all allotetraploid *B. hybridum*, *B. phoenicoides* and *B. retusum* samples had complex type-2 graphs showing two IGS loops interconnected by a junction section (coding region of the 5S gene), suggesting they may have two 5S ribotypes. The separation of the two IGS loops was less clear in the graphs of the *B. mexicanum*-4x and *B. pinnatum*-4x samples, while tetraploid *B. rupestre* samples showed both complex type-2 (Brup439-1, Brup7) and simple type-1 (remaining samples) graphs (**Figure 5**). Among the hexaploid taxonomic cytotypes, some *B. retusum*-6x samples (Bret551, Bret557) presented complex type-3 graphs with three interconnected IGS loops, indicating that they could have three 5S loci, and the other samples (Bret408, Bret561) type-2 graphs, while all the *B. phoenicoides*-6x and *B. rupestre*-6x samples presented type-2 graphs and all *B. boissieri*-6x samples simple type-1 graphs (**Figure 5**).



**Figure 5.** 5S clustering graph plots of *Brachypodium* samples generated by the individual RepeatExplorer2 analysis sorted by ploidy level. Diploids (2x) show graph type 1, while some tetraploids and hexaploids show graph types 2 and 3, respectively (see also Table 4).

## DISCUSSION

### Delineation of the *Brachypodium* repeatome and its impact on the striking genome size diversification of its lineages

Our comprehensive analysis of the *Brachypodium* repeatome has revealed the composition and frequency of the main repetitive DNA elements across the genome landscape of all its lineages (**Table 2** and **Supplementary Table S3, Figure 1**). Our data confirm the decisive contribution of the repeatome to the genome size diversification of the studied *Brachypodium* genomes. The repeatome represents a main or considerable percentage of the holoploid genome of the surveyed samples. One of the most compelling results is the enormous differences in genome size, and its correlated repeatome amount, detected among species and lineages (**Table 2** and **Supplementary Table S3, Figures 1 and 2**). For a genus selected as model system of monocots due to the small genome size of its flagship species *B. distachyon* (IBI, 2010; Gordon et al., 2017), the differences between the smallest genome sizes found within its annual species (*B. stacei*, 551 Mbp, holoploid genome; 275 Mbp monoploid genome), having the lowest repeatome contents (20%), and the largest genome sizes of the slender perennial *B. mexicanum* samples (3690 Mbp, holoploid genome; 922 Mbp monoploid genome), presenting the highest repeatome contents (67.9%), are 6.7-fold and 3.3-fold, respectively (**Tables 1 and 2** and **Supplementary Tables S1 and S3**). Although the majority of the analyzed *Brachypodium* species' genomes are small (annual species; 275-309 Mbp monoploid genome) or relatively small (most core perennial species;  $\leq 352$  Mbp) and their respective repeatome percentages are also consistently low (20.7-22.7% annuals;  $\leq 27.9\%$  core perennials), *B. mexicanum* plus the ancestral *B. boissieri*-6x (508 Mbp; 31%) and the recent core-perennial *B. sylvaticum*-2x (450 Mbp; 34%) lineages depart from this tendency, and the intermediately evolved *B. retusum*-4x-6x (410 Mbp; 26.7%) also differs slightly from it (**Table 2** and **Supplementary Table S3, Figures 1 and 2**).

Surprisingly, the main differences in genome sizes and repeatome amounts have been found among the most ancestral x=10 karyotype lineages, the smallest *B. stacei* (S karyotype) and the largest *B. mexicanum* (A1 karyotype) genomes (**Figure 1**, Sancho et al., 2022; Decena et al., 2023). Although the genome (and repeatome) contractions observed in *B. stacei* and in the also ephemeral outcore *B. distachyon* (intermediately evolved x=5 D karyotype; Sancho et al., 2022; Decena et al., 2023) and *B. hybridum* (S+D

karyotypes) lineages (**Figure 1**) is a general feature detected in other annual angiosperms (Suda et al., 2015; Pellicer et al., 2018; Hloušková et al., 2019), the gross genomes of the weakly-rhizomatose perennial *B. mexicanum* and the strongly-rhizomatose perennial *B. boissieri* (ancestral  $x=8$  A2 karyotype; Sancho et al., 2022; Decena et al., 2023) (**Figure 1**) points toward to a *Brachypodium* common ancestor with an expanded genome that preceded the diversification of its oldest outcore lineages. A similar evolutionary scenario has been hypothesized for the obese-genome ancestor of the *Hesperis* subclade (~1600 Mbp), within the otherwise overall small-genome Brassicaceae clade that includes the model dicot *Arabidopsis thaliana* with one of the smallest genome sizes of angiosperms (157 Mbp; Hloušková et al., 2019). Our repeatome data, together with the extremely high synteny (coding genes) detected between the *B. stacei* and *B. mexicanum* chromosomes (Sancho et al, unpub. data) and high collinearity and similar CCB karyotypes of the A1 and A1' subgenomes of *B. mexicanum* (Sancho et al., 2022, and unpub. data), suggest that the 3.3-fold differences in their monoploid genomes sizes (for the same number of chromosomes) were caused by expansions of LTR-Gypsy retrotransposons in the *B. mexicanum* chromosomes (probably coupled with some potential losses in the *B. stacei* chromosomes) and not by WGD (**Table 2** and **Supplementary Table S3, Figures 1 and 2**). The inflated genome of mesopolyploid *B. mexicanum* (10.4-8.6 Ma; Sancho et al., 2022; Decena et al., 2023) likely resulted from the proliferation of Tekay (22-27%) and Retand (17-21%) repeat families, and the enrichment in other less abundant elements (Mutator, 2-3%; Angela, 0.7%, Ogre, 1-2%). Interestingly, Ogre retrotransposons, frequent in dicot Fabaeae legume genomes (Macas et al., 2015) and also common in Brassicaceae genomes (Hloušková et al., 2019), were only residually present in some grass Loliinae genomes (Moreno-Aguilar et al., 2022), and have been exclusively found in *B. mexicanum* within our low-pass genomic survey of the genus (**Table 2** and **Supplementary Table S3, Figure 1**). The relatively large genome of the also ancestral *B. boissieri* (5.4-3.7 Ma, Sancho et al., 2022; Decena et al., 2023) probably resulted from the burst of Retand (10-13%) retrotransposons, which were also predominant but less enriched in the intermediately evolved strong-rhizomatous perennial *B. retusum* genome (5-8%). The overall decrease in the amounts of Retand ( $\leq 6\%$ ) and other repetitive elements in the core-perennial diploid and their derived neopolyploid lineages' genomes (**Table 2** and **Supplementary Table S3, Figure 1**) was likely a consequence of long-lasting post-paleopolyploid diploidizations and genome downsizings due to the shedding of the excess of repeats (Michael, 2014; Hloušková et al., 2019).

The large genome reductions observed in ancestral outcore *Brachypodium* annual species (**Table 2** and **Supplementary Table S3, Figure 1**) could be also related to the transition in life form. Evidence suggests that annuality has evolved convergently from perennality in different lineages of flowering plants, and that it could have been facilitated by evolutionary precursors (correlated developmental, physiological, and genomic traits) in the temperate pooid grasses, which also include *Brachypodium* (Hjertaas et al., 2023). It has been also demonstrated that plants with small genomes can grow in more diverse habitats and tend to be annuals while those with large genomes are restricted to narrow ecological niches and are perennials (Suda et al., 2015; Pellicer et al., 2018). Although the annual *Brachypodium* species share similar mesic and arid habitats and distribution areas that other Mediterranean perennial relatives (Catalan et al., 2016), they show shorter generation times and thus larger ability for long-distance dispersal and colonization of new niches and non-native continents than the perennials (Scholthof et al., 2018), probably facilitated by their extremely downsized genomes (**Table 2** and **Supplementary Table S3, Figure 1**). The great dysploid reduction from the  $x=10$  ancestral S karyotype of *B. stacei* to the intermediate  $x=5$  D karyotype of *B. distachyon* resulted from four nested chromosome centromeric fusions; however, the high collinearity of the two genomes was corroborated by their almost similar genomic sizes (Gordon et al., 2020). Our repeatome analysis further support the analogous genome coverage of repeatome in these diploids (*B. stacei* 20.7%; *B. distachyon* 22.7%), with the differences caused by a plausible recent proliferation of Retand elements in the youngest lineage (*B. stacei* 1.2%; *B. distachyon* 6.8%) while the Tekay and Angela elements were slightly higher in the oldest lineage (*B. stacei* 3.3% and 1.5% vs *B. distachyon* 2.6% and 0.9%) (**Table 2** and **Supplementary Table S3, Figure 1**). Our data reinforce the findings of Stritt et al. (2020) about the major contribution of Retand elements in the variation of genome sizes between *B. distachyon* accessions, extending it at the whole genus level. These authors also postulated a high dynamic activity and source of within-species polymorphisms of very young Angela elements in the *B. distachyon* genome landscape; however, our Hs/Ho ratios indicated a higher diversification of the Retand and to a lesser extent of the Tekay sequences across the *Brachypodium* genomes (**Figure 3**), probably due to the low contribution of the Angela repeats to the genome landscape at the genus level (**Tables 2 and 3**). Our analysis has also confirmed the low repeatome content the annual allotetraploid *B. hybridum* ( $x=10$  S+  $x=5$  D karyotype), which showed a balanced

percentage (22%) between those of its diploid progenitor species (**Table 2** and **Supplementary Table S3**).

The striking larger repeatome coverage of the recently evolved core perennial *B. sylvaticum* diploid genome (34.2%) with respect to other diploid (24.%) and polyploid (<26%) core-perennial genomes that share the same recently evolved  $x=9$  karyotype (Sancho et al., 2022; Decena et al., 2023) is not correlated with parallel differences in 1Cx genome sizes, which have similar values for *B. sylvaticum*-2x (456 Mbp) than for *B. pinnatum*-2x (401 Mbp) and other core-perennial 2x-4x-6x cytotypes (349-382 Mbp) (**Tables 1 and 2** and **Supplementary Tables S1 and S3**). This unexpected result could be a consequence of a relatively recent polyploidization and subsequent diploidization of the Late Pliocene - Early Pleistocene *B. sylvaticum* occidental lineage (2.78 – 2.17 Ma; **Figure 1**, Decena et al., 2023). The *B. sylvaticum* samples showed a proliferation of Tekay retrotransposons (10%) compared to the other core perennial lineages (1-2%), and also higher proportions of Mutator, Ikeros and Angela elements (**Table 2** and **Supplementary Table S3**). This finding, together with other cytotypic features, like a larger number of rDNA 25 loci than expected for a diploid (4-5; Decena et al., 2023), would suggest a likely post-polyploid diploidized origin. High repeatome coverages were also found in grass Loliinae diploidized paleo-polyploids (Moreno-Aguilar et al., 2022). Yet, another special characteristic of *B. sylvaticum* is its symbiotic relationship with fungal endophytes of the genus *Epichloë*, a long-term association that presumably contributed to the expansion and colonization success of mesic and humid forest biomes by this species (Saikkonen et al., 2016). In contrast to other core-perennial *Brachypodium* species (e. g., *B. pinnatum*, *B. phoenicoides*), which are only occasionally infected with *Epichloë typhina* or *E. sylvatica* strains, and to annual *Brachypodium* species that lack *Epichloë* symbionts, all *B. sylvaticum* individuals are ubiquitously infected with *E. sylvatica* (Leuchtmann and Schardl, 2022). Horizontal gene transfer (HGT) of fungal *Epichloë* genes have been detected in different pooid lineages (Shinozuka et al., 2020); genome screenings also indicated that transposable elements can be transferred among distantly related organisms, and that transferred TEs remained functional and experienced bursts (El Baidouri et al., 2014). Although the *B. sylvaticum* repeatome reads filtered by RE2 should correspond to Viridiplantae species, and the analysis of the potential symbiont-to-plant HGTs are out of the scope of this study, the intrinsic and long-lasting *B. sylvaticum*-*E. sylvatica* holobiont relationships should be re-examined within a large co-evolving



genomic scenario to assess the potential impact of the symbiont in the repeatome dynamics of its host plant.

### **Alternative evolutionary responses to the polyploidy genome shock hypothesis by different *Brachypodium* allopolyploids**

Our study has evidenced that expansions and contractions in the repeatome are responsible for the three contrasting responses to the PGSH in different *Brachypodium* allopolyploid lineages and that each response was caused by different biological, cytological and temporal scenarios.

Thus, the exacerbated genome expansion of the old allotetraploid *B. mexicanum* was not a consequence of WGD per se but of proliferations of Tekay and Retands retrotransposons and other repetitive DNA elements (**Table 2** and **Supplementary Table S3, Figure 1**). However, TE annotations in the reference genome of this species (Bmex347; Sancho et al. unpub. data) indicated that amplifications of the LTR and other transposable elements were not only circumscribed to (peri)centromeric regions but also to telomeres and chromosome arms, as TEs could intersperse with coding regions (Hloušková et al., 2019), therefore supporting chromosome-wide amplifications in all its  $x=10$  karyotype. Although it is surprising why the ancestral *B. mexicanum* genome has the propensity to tolerate or benefit from such repeatome bloatings and subsequent genome expansions, which were likely inherited from the common ancestor but not purged over time, the increased chromosome arm lengths were compensated by increased centromere size and copy number of centromeric tandem repeats (Sancho et al., unpub. data) which would guarantee a correct segregation of the obese chromosomes during cell division (Zhang and Dawe, 2012).

By contrast, the additive repeatome pattern exhibited by the annual allotetraploid *B. hybridum* with respect to those of its contracted-genome diploid progenitor species (**Table 2** and **Supplementary Table S3, Figure 1**) is a likely response to a stabilized post-WGD genome evolution. The three detected independent recurrent origins of this young neopolyploid, which spanned the last 1.4 Ma to 20 kya, ended in the same phenotypic allotetraploid which did not show evidences of homeologous recombinations, subgenomic dominance or pronounced TE activations (Gordon et al., 2020; Scarlett et al., 2022; Mu et al., 2023). It was probably caused by the high evolutionary and structural divergence of the *B. stacei* and *B. distachyon* progenitor genomes ( $x=10$  S and  $x=5$  D

karyotypes) which likely favored the non-recombinant integrity of the resulting subgenomes in the hybrid and the immediate creation of the amphidiploids (Mu et al., 2023).

The observed repeatome and genome size reductions with increasing ploidy level in the recently evolved core perennial *B. pinnatum* and *B. phoenicoides* lineages (**Table 2** and **Supplementary Table 3, Figure 1**) are likely the result of loss of repeats through recombination which resulted in repeatome shrinkage (Michael, 2014). Although the *B. phoenicoides* polyploid cytotypes share one recent subgenome with a diploid *B. pinnatum*-type karyotype ( $x=9$ ), a second intermediately subgenome with a reduced karyotype ( $x=5$  E2) is present once in the allotetraploids ( $x=9$  G +  $x=5$  E2) and twice in the allohexaploids ( $x=9$  G +  $x=5$  E2 +  $x=5$  E2) (Sancho et al., 2022; Decena et al., 2023), thus favoring more frequent recombinations between identical or very similar subgenomes and therefore more potential repeatome losses in the high polyploids. Paralleling the case of the highly hybridogenous high-polyploid Loliinae lineages, which experienced large genomic rearrangements causing massive repeatome and genome contractions (Moreno-Aguilar et al., 2022), the *B. pinnatum* and *B. phoenicoides* lineages of the core-perennial clade also evidenced high interspecific hybridization rates (Khan and Stace, 1999), thus favoring the reductions in their repetitive elements and genomes.

### Repeatome-based phylogenomics of *Brachypodium* and concordance between rDNA 5S graphs and subgenomes

As in previous studies of angiosperms (Dodsworth et al., 2015; McCann et al., 2018, 2020; Vitales et al., 2020; Herklotz et al., 2021; Moreno-Aguilar et al., 2022), the shared repeat clusters retrieved from the comparative RE2 analysis of *Brachypodium* have been observed to contain phylogenetic information for its main lineages (**Figures 1** and **4; Supplementary Tables S3** and **S4**). The topology of the repeatome network constructed from the independent distance-based NJ trees (**Figure 4**) is highly congruent with those of the *Brachypodium* Astral supertree based on whole plastome and rDNA 35S and 5S gene trees (**Figure 1**; Decena et al., 2023) and the *Brachypodium* transcriptome-based subgenomic tree (Sancho et al., 2022). The unrooted network showed the larger divergences of the ancestral outcore lineages and the recent separations of the core-perennial lineages (**Figure 4**). The network retrieved the high isolation of the *B. mexicanum* lineage from the other lineages; this large divergence resulted from the higher

amounts of repeats for the common elements of some repetitive families within the representatives of the genus (**Supplementary Table S4**). The *B. mexicanum* group included two close samples from Mexico and a less related sample from Ecuador (**Figure 4**), paralleling the results obtained from plastomes (Decena et al., 2023); although all the studied *B. mexicanum* samples show a similar repeatome composition (**Figure 1**), the divergence of the South American Andean sample from the North American Mexican samples, coincident with that observed in previous phylogenetic analyses (Díaz-Pérez et al., 2018), indicated the plausible existence of two geographically separated lineages. The closeness of the *B. stacei* lineage to the *B. mexicanum* cluster support the shared ancestry of the two  $x=10$  karyotypes, and the intermediate placement of *B. hybridum* between its two progenitor lineages reinforced its additive pattern. Interestingly, the ancestral *B. boissieri* cluster was resolved as closer to the *B. distachyon* lineage than to the more ancestral *B. mexicanum* and *B. stacei* outcore lineages, matching the relationships recovered in the plastome tree but diverging from those of the rDNA and transcriptome trees (**Figure 4**; Sancho et al., 2022c; Decena et al., 2023). Thus, the repeatome data indicate the additional contribution of a more recently evolved ancestor to the nuclear genome of this purported ancestral autohexaploid ( $x=8$  A2 karyotype) species. Within the recently evolved core-perennial group, the earlier divergence of the *B. retusum* subcluster from the rest supports its intermediate evolutionary position between the outcore *B. boissieri* and nuclear core-perennial lineages; however, its tighter closeness to the core group suggest an admixed pattern of ancestral ( $x=8$  A2) and more recent ( $x=8$  E1 (+  $x=5$  E2) repeat amounts or a convergent evolution towards the core repetitive elements (**Figure 4**), paralleling the plastome and rDNA findings (Sancho et al., 2022; Decena et al., 2023). The respective divergences of the *B. arbuscula* and *B. sylvaticum* lineages from the rest, and the closeness of the *B. pinnatum* complex taxa and cytotypes (*B. pinnatum*, *B. rupestre*, *B. phoenicoides*) likely resulted from their specific repeat compositions, particularly those of the highly differentiated *B. sylvaticum* group, and were coincident with those retrieved from plastomes and rDNA genes (**Figures 1 and 4**; Decena et al., 2023).

The rDNA 5S graph topologies (**Table 4** and **Figure 5**) showed a great match with the number and nature of the genomes and subgenomes of the studied *Brachypodium* samples (**Figure 1**; Sancho et al., 2022; Decena et al., 2023), corroborating their value to uncover ploidy levels, ancient rDNA families, and known and orphan subgenomes in

angiosperms (Garcia et al., 2020; Vozárová et al., 2021; Moreno-Aguilar et al., 2022). The single type-1 graphs of diploids corresponded to their respective monoploid extant genomes (*B. stacei*,  $x=10$  S; *B. distachyon*,  $x=5$  D; *B. arbuscula*, *B. pinnatum*, *B. sylvaticum*,  $x=9$  G), the type-2 graphs to the two different subgenomes of allotetraploids (*B. hybridum*,  $x=10$  S +  $x=5$  D (extant); *B. phoenicoides*,  $x=9$  G (extant) +  $x=5$  E2 (orphan); *B. retusum*,  $x=8$  A2 +  $x=8$  E1 (orphan)), and the type-3 graphs to the three different subgenomes of some allohexaploids (*B. retusum*,  $x=8$  A2 +  $x=8$  E1 +  $x=5$  E2 (orphans) (**Figure 1**; Decena et al., 2023). Interestingly, the poorly resolved type-2 graphs of *B. mexicanum*-4x may indicate that they correspond to close ribotypic 5S families, agreeing with the two similar A1 and A1'  $x=10$  orphan subgenomes of this putative ancestral segmental allotetraploid, the type-2 graphs of the *B. phoenicoides*-6x and *B. rupestre*-6x samples to their two different ribotypes and subgenomes  $x=9$  G (extant) +  $x=5$  E2 (orphan) plus a duplicated  $x=5$  E2 (orphan) copy in these allo-autohexaploids, and the type-1 graph of *B. boissieri*-6x samples to the triplicated  $x=8$  A2 subgenomes of this ancestral autohexaploid (**Figures 1 and 5**; Sancho et al., 2022; Decena et al., 2023). The maintenance of 5S rDNA loci in high allopolyploid *Brachypodium* species is consistent with their conserved patterns in other angiosperm allopolyploids (Garcia et al., 2017). The few cases of fewer 5S graph loops than expected according to ploidy level and distinct subgenomes (*B. rupestre*-6x) could be due to convergent evolution to another ribotypes or, more likely, to failure of the low-pass genome sequencing repeat reads in detecting the distinct 5S IGS sequences.

## DATA AVAILABILITY STATEMENT

Input and output data, and Supplementary information are available at Github ([https://github.com/Bioflora/Brachypodium\\_Repeatome](https://github.com/Bioflora/Brachypodium_Repeatome)).

## AUTHOR CONTRIBUTIONS

PC designed the study. EP, LAI and PC collected samples. MD, RS and LAI developed the experimental work. MD, RS and PC analyzed the data and interpreted the results. PC, MD and RS prepared the manuscript. All authors revised the manuscript. The authors declare no conflict of interest.

## FUNDING

This study was funded by the Spanish Ministry of Science and Innovation PID2019-108195GB-I00, TED2021-131073B-I00 and PDC2022-133712-I00, and by the Spanish Aragon Government LMP82-21 and Bioflora A01-23R research grants. MD was supported by a Spanish Ministry of Science predoctoral FPI fellowship. RS was supported by a Bioflora postdoctoral contract.

## ACKNOWLEDGEMENTS

We thank the INIA, JGI and USDA germplasm banks for facilitating seeds of some of the *Brachypodium* samples studied. Genome skimming data from the studied *Brachypodium* samples were generated at the Centro Nacional de Análisis Genómicos (CNAG). The repeatome and evolutionary analyses were performed in the Bioflora laboratory of the Escuela Politécnica Superior de Huesca (EPSHU, Universidad de Zaragoza, Spain).

## SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at [https://github.com/Bioflora/Brachypodium\\_Repeatome](https://github.com/Bioflora/Brachypodium_Repeatome)

## REFERENCES

- Bird, K. A., VanBuren, R., Puzey, J. R., and Edger, P. P. (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 220, 87–93. doi:10.1111/nph.15256.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.
- Cantalapiedra, C. P., Diana, L., Gordon, S. P., Vogel, J. P., and Contreras-moreira, B. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium* : flowering time signatures , introgression and recombination in recently diverged ecotypes Rub. 1631–1644. doi:10.1111/nph.14926.
- Catalan, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., and López-Herránz, M. L. (2016). Phylogeny and Evolution of the Genus *Brachypodium*. 9–38. doi:10.1007/7397\_2015\_17.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. *Science* (80-. ). 345, 950–953. doi:10.1126/science.1253435.
- Chen, Z. J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* 58, 377–406. doi:10.1146/annurev.arplant.58.032806.103835.
- Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., et al. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* 52, 525–533. doi:10.1038/s41588-020-0614-5.
- De Storme, N., and Mason, A. (2014). Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. *Curr. Plant Biol.* 1, 10–33. doi:10.1016/j.cpb.2014.09.002.
- Deb, S. K., Edger, P. P., Pires, J. C., and McKain, M. R. (2023). Patterns, mechanisms, and consequences of homoeologous exchange in allopolyploid angiosperms: a genomic and epigenomic perspective. *New Phytol.*, 2284–2304. doi:10.1111/nph.18927.
- Decena, M. A., Sancho, R., Lusinska, J., Hasterok, R., Gorgojo, R., Montes, B., et al. (2023). Cyto-phylogenomics of the grass model genus *Brachypodium* unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends within this largely reticulate diploid-polyploid species-complex. *Front. Plant Sci.*
- Díaz-Pérez, A., López-Álvarez, D., Sancho, R., and Catalán, P. (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. *Mol. Phylogenet. Evol.* 127, 256–271. doi:10.1016/j.ympev.2018.06.003.
- Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Novak, P., et al. (2015).

- Genomic repeat abundances contain phylogenetic signal. *Syst. Biol.* 64, 112–126. doi:10.1093/sysbio/syu080.
- Doyle, J., and Doyle, J. (1987). A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.* 19, 11–15.
- Drouin, M., Hénault, M., Hallin, J., and Landry, C. R. (2021). Testing the Genomic Shock Hypothesis Using Transposable Element Expression in Yeast Hybrids. *Front. Fungal Biol.* 2, 1–15. doi:10.3389/ffunb.2021.729264.
- Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.* 51, 541–547. doi:10.1038/s41588-019-0356-4.
- El Baidouri, M., Carpentier, M.-C., Cooke, R., Gao, D., Lasserre, E., Llauro, C., et al. (2014). Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res.* 24, 831–838. doi:10.1101/gr.164400.113.
- Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* (80-. ). 338, 758–767. doi:10.1126/science.338.6108.758.
- Garcia, S., Kovařík, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 89, 1020–1030. doi:10.1111/tpj.13442.
- Garcia, S., Wendel, J. F., Borowska-Zuchowska, N., Ainouche, M., Kuderova, A., and Kovarik, A. (2020). The Utility of Graph Clustering of 5S Ribosomal DNA Homoeologs in Plant Allopolyploids, Homoploid Hybrids, and Cryptic Introgressants. *Front. Plant Sci.* 11, 1–14. doi:10.3389/fpls.2020.00041.
- Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., et al. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.* 11, 1–16. doi:10.1038/s41467-020-17302-5.
- Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., et al. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* 8. doi:10.1038/s41467-017-02292-8.
- Hasterok, R., Catalan, P., Hazen, S. P., Roulin, A. C., Vogel, J. P., Wang, K., et al. (2022). *Brachypodium*: 20 years as a grass biology model system; the way forward? *Plant Sci.* 2022. doi:10.1016/j.tplants.2022.04.008.
- Hawkins, J. S., Proulx, S. R., Rapp, R. A., and Wendel, J. F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17811–17816. doi:10.1073/pnas.0904339106.
- Herklotz, V., Kovařík, A., Wissemann, V., Lunerová, J., Vozárová, R., Buschmann, S., et al. (2021). Power and Weakness of Repetition – Evaluating the Phylogenetic Signal From Repeatomes in the Family Rosaceae With Two Case Studies From Genera Prone to Polyploidy and Hybridization (*Rosa* and *Fragaria*). *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.738119.
- Hidalgo, O., Pellicer, J., Christenhusz, M., Schneider, H., Leitch, A. R., and Leitch, I. J. (2017). Is There an Upper Limit to Genome Size? *Trends Plant Sci.* 22, 567–573.



- doi:10.1016/j.tplants.2017.04.005.
- Hjertaas, A. C., Preston, J. C., Kainulainen, K., Humphreys, A. M., and Fjellheim, S. (2023). Convergent evolution of the annual life history syndrome from perennial ancestors. *Front. Plant Sci.* 13, 1–18. doi:10.3389/fpls.2022.1048656.
- Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., and Lysak, M. A. (2019). The large genome size variation in the Hesperis clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Ann. Bot.* 124, 103–120. doi:10.1093/aob/mcz036.
- IBI (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi:10.1038/nature08747.
- Khan, M. A., and Stace, C. A. (1999). Breeding relationships in the genus *Brachypodium* (Poaceae).
- Leuchtmann, A., and Schardl, C. L. (2022). Genetic Diversity of Epichloë Endophytes Associated with *Brachypodium* and Calamagrostis Host Grass Genera including Two New Species.
- Li, M., Copeland, A., and Han, J. (2011). DUK - A Fast and Efficient Kmer Based Sequence Matching Tool.
- Li, Y., Zuo, S., Zhang, Z., Li, Z., Han, J., Chu, Z., et al. (2018). Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *Plant J.* 93, 1088–1101. doi:10.1111/tpj.13832.
- Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., et al. (2019). Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *Int. J. Mol. Sci.* 20, 1–19. doi:10.3390/ijms20225557.
- Macas, J., Novak, P., Pellicer, J., Cizkova, J., Koblizkova, A., Neumann, P., et al. (2015). In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe fabaeae. *PLoS One* 10, 1–23. doi:10.1371/journal.pone.0143424.
- Mason, A. S., and Wendel, J. F. (2020). Homoeologous Exchanges, Segmental Allopolyploidy, and Polyploid Genome Evolution. *Front. Genet.* 11, 1–10. doi:10.3389/fgene.2020.01014.
- McCann, J., Jang, T. S., Macas, J., Schneeweiss, G. M., Matzke, N. J., Novák, P., et al. (2018). Dating the species network: Allopolyploidy and repetitive DNA evolution in American daisies (*Melampodium* sect. *Melampodium*, Asteraceae). *Syst. Biol.* 67, 1010–1024. doi:10.1093/sysbio/syy024.
- McCann, J., Macas, J., Novák, P., Stuessy, T. F., Villaseñor, J. L., and Weiss-Schneeweiss, H. (2020). Differential Genome Size and Repetitive DNA Evolution in Diploid Species of *Melampodium* sect. *Melampodium* (Asteraceae). *Front. Plant Sci.* 11, 1–14. doi:10.3389/fpls.2020.00362.
- McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* (80-. ). 226, 792–801.
- Michael, T. P. (2014). Plant genome size variation: Bloating and purging DNA. *Briefings*

- Funct. Genomics Proteomics* 13, 308–317. doi:10.1093/bfgp/elu005.
- Moreno-aguilar, M. F., Inda, L. A., and Sánchez-rodríguez, A. (2022). Evolutionary Dynamics of the Repeatome Explains Contrasting Differences in Genome Sizes and Hybrid and Polyploid Origins of Grass Loliinae Lineages. 13. doi:10.3389/fpls.2022.901733.
- Mu W, Li K, Yang Y, Breiman A, Yang J, Wu J, Zhu M, Wang S, Catalán P, Nevo E, L. J. (2023). Subgenomic stability of progenitor genomes during repeated allotetraploid origins of the same grass *Brachypodium hybridum*. *bioRxiv*.
- Negi, P., Rai, A. N., and Suprasanna, P. (2016). Moving through the stressed genome: Emerging regulatory roles for transposons in plant stress response. *Front. Plant Sci.* 7. doi:10.3389/fpls.2016.01448.
- Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. Available at: <http://www.biomedcentral.com/1471-2105/11/378>.
- Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi:10.1038/s41596-020-0400-y.
- Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and MacAs, J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792–793. doi:10.1093/bioinformatics/btt054.
- Parisod, C., Holderegger, R., and Brochmann, C. (2010). Evolutionary consequences of autopolyploidy. *New Phytol.* 186, 5–17. doi:10.1111/j.1469-8137.2009.03142.x.
- Pellicer, J., Hidalgo, O., Dodsworth, S., and Leitch, I. J. (2018). Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)*. 9. doi:10.3390/genes9020088.
- Saikkonen, K., Young, C. A., Helander, M., and Schardl, C. L. (2016). Endophytic *Epichloë* species and their grass hosts: from evolution to applications. *Plant Mol. Biol.* 90, 665–675. doi:10.1007/s11103-015-0399-6.
- Sancho, R., Inda, L. A., Díaz-Pérez, A., Des Marais, D. L., Gordon, S., Vogel, J. P., et al. (2022). Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J.* 109, 1535–1558. doi:10.1111/tpj.15650.
- Scarlett, V. T., Lovell, J. T., Shao, M., Phillips, J., Shu, S., Goodstein, D. M., et al. (2022). Multiple origins , one evolutionary trajectory : gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*.
- Scholthof, K. B. G., Irigoyen, S., Catalan, P., and Mandadi, K. K. (2018). *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30, 1673–1694. doi:10.1105/tpc.18.00083.
- Shinozuka, H., Shinozuka, M., de Vries, E. M., Sawbridge, T. I., Spangenberg, G. C., and Cocks, B. G. (2020). Fungus-originated genes in the genomes of cereal and pasture grasses acquired through ancient lateral transfer. *Sci. Rep.* 10, 19883. doi:10.1038/s41598-020-76478-4.

- Stritt, C., Wyler, M., Gimmi, E. L., Pippel, M., and Roulin, A. C. (2020). Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. *New Phytol.* 227, 1736–1748. doi:10.1111/nph.16308.
- Suda, J., Meyerson, L. A., Leitch, I. J., and Pyšek, P. (2015). The hidden side of plant invasions: The role of genome size. *New Phytol.* 205, 994–1007. doi:10.1111/nph.13107.
- Vitales, D., Álvarez, I., Garcia, S., Hidalgo, O., Feliner, G. N., Pellicer, J., et al. (2020a). Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). *Ann. Bot.* 125, 611–623. doi:10.1093/aob/mcz183.
- Vitales, D., Garcia, S., and Dodsworth, S. (2020b). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Mol. Phylogenet. Evol.* 147, 106766. doi:10.1016/j.ympev.2020.106766.
- Vozárová, R., Herklotz, V., Kovařík, A., Tynkevich, Y. O., Volkov, R. A., Ritz, C. M., et al. (2021). Ancient Origin of Two 5S rDNA Families Dominating in the Genus *Rosa* and Their Behavior in the Canina-Type Meiosis. *Front. Plant Sci.* 12. Available at: <https://www.frontiersin.org/article/10.3389/fpls.2021.643548>.
- Vu, G. T. H., Cao, H. X., Reiss, B., and Schubert, I. (2017). Deletion-bias in DNA double-strand break repair differentially contributes to plant genome shrinkage. *New Phytol.* 214, 1712–1721. doi:10.1111/nph.14490.
- Weiss-Schneeweiss, H., Leitch, A. R., McCann, J., Jang, T. S., and Macas, J. (2015). Employing next generation sequencing to explore the repeat landscape of the plant genome. in.
- Wendel, J. F., Lisch, D., Hu, G., and Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr. Opin. Genet. Dev.* 49, 1–7. doi:10.1016/j.gde.2018.01.004.
- Zhang, H., and Dawe, R. K. (2012). Total centromere size and genome size are strongly correlated in ten grass species. *Chromosom. Res.* 20, 403–412. doi:10.1007/s10577-012-9284-1.

## MORPHOANATOMICAL ANALYSIS OF CRYPTIC AND EVOLUTIONARY DIVERGENT POLYPLOID CYTOTYPES OF MODEL MEDITERRANEAN *BRACHYPODIUM RETUSUM* AND *B. PHOENICOIDES* GRASSES

María Ángeles Decena<sup>1,2</sup>, Luis A. Inda<sup>1,3</sup>, Ernesto Pérez-Collazos<sup>1,2</sup>, Pilar Catalán<sup>1,2\*</sup>

<sup>1</sup>Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Huesca, España.

<sup>2</sup>Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, España.

<sup>3</sup>Instituto Agroalimentario de Aragón-IA2 (Universidad de Zaragoza-CITA), Aragón, España

\* Correspondence: Pilar Catalán (pcatalán@unizar.es)

### Abstract

Taxonomically complex groups are often composed of easily identifiable species but also of undercovered cryptic species. In plants, highly hybridogenous groups with sexually and asexually propagating individuals and several interconnected ploidy levels are a potential source of new cryptic species. In this study we analyzed population samples of the grass *Brachypodium retusum* (45) and *B. phoenicoides* (27) polyploid complexes, which contain allotetraploid and allohexaploid cytotypes of unclear taxonomic adscription, covering their respective native distribution ranges in the western Mediterranean region. 50 and 92 morphoanatomical traits were measured in different subsets of samples. Seven and five morphoanatomical traits significantly discriminate 4x vs 6x cytotypes in *B. retusum* and *B. phoenicoides*, respectively, and most of them reflected expected heterosis patterns. Intraspecific phenotypic analyses detected a large morphoanatomical variation within each complex, though their respective tetraploid and hexaploid groups overlapped in their multivariate PCA morpho-spaces. Up to three and four phenotypic groups were respectively identified within the *B. retusum* and *B. phoenicoides* complexes by multivariate FAMD hierarchical clustering, and some of them correlated with cytotypic, geographic and phyletic lineages. Our data suggest the potential existence of phylogenetic-driven micro-speciation processes for cryptic morpho-groups within *B. phoenicoides* although it is less clear within *B. retusum*.

**Keywords:** allopolyploid complexes, *Brachypodium phoenicoides*, *B. retusum*, cryptic taxa, Mediterranean grasses, morphoanatomical characters, statistics, taxonomy.

## INTRODUCTION

Cryptic species are usually discovered when new genomic and evolutionary approaches detect potential conspecifics through non apparent conspicuous traits (Gustafsson et al., 2014; Vigalondo et al., 2015). The potential crypto-species have shown evidences of cytogenetic disparity, genetic divergence, gene flow (hybridization) or reproductive isolation, or statistical deviations of morphological features from neutral expectations (Harmon et al., 2003; Struck et al., 2018). The increasing number of uncovered cryptic species in plants and animals suggest that cryptic species probably represent a large proportion of the still unveiled biodiversity on earth (Jörger and Schrödl, 2013; Pante et al., 2015; Loxdale et al., 2016). Some estimations points towards to 10-20% of the current single taxonomic species containing two or more species, especially in some megadiverse and still poorly studied organismic groups, but also across all branches of the tree of life (Janzen et al., 2017; Struck et al., 2018). However, the cryptic species concept is also a scientific paradox. Taxonomists have traditionally used morphological characters to determine species boundaries, though speciation and acquisition of diagnostic features are not always directly correlated. The cryptic diversity that can be recognized on a genetic but not on a morphological basis poses a taxonomic challenge (Maxwell et al., 2021), and cryptic species may be also concealed in cases where there is considerable morphological variation but no clear boundaries for species delimitation (Wei et al., 2021).

Cryptic plant species are often associated to highly hybridogenous complexes of sexually reproducing species (Harlan and de Wet, 1963), or to taxa with different chromosomal cytotypes and ploidy levels (Stebbins, 1969, 1971; Stace, 1991; Stebbins & Dremann 1998). The size and taxonomic diversity of these complexes vary enormously depending on the evolutionary history of the group and their capability to interbreed. It is generally assumed, however, that species should be distinguished by stable and prominent phenotypic features, and preferably should comply with the biological species concept (Mayr E, 1942). It has been posited that designation of cryptic species that could not be separated by phenotypic traits and by some amount of genetic isolation would create a threat to nomenclatural stability of taxonomic species (Stevens PF, 1990; Holstein and Luebert, 2017), and consequently those biological entities should be classified as infraspecific ranks or ecotypes (Stace, 1991; Maxwell et al., 2021).

*Brachypodium* (L.) P. Beauv. is a small cool seasonal grass genus which contains about 20 worldwide distributed species (Schippmann, 1991; Catalan et al., 2016). Approximately, half of the recognized *Brachypodium* species are diploids and the other half are confirmed or purported polyploids (Díaz-Pérez et al., 2018; Sancho et al., 2022b; Decena et al., 2023). *Brachypodium* has been resolved as intermediately evolved between the ancestral Pooideae lineages and the recent core-pooid clade (Catalan et al., 2016; Sancho et al., 2018). Its flagship species *B. distachyon* was targeted as a model plant for functional genomics of cereals and biofuel crops (IBI, 2010) and the whole genus has been recently selected as a model system for analysing the evolution and origins of allopolyploidy and perenniality in grasses and monocots (Scholthof et al., 2018; Hasterok et al., 2022). *Brachypodium* shows a remarkable descendant dysploidy trend from ancestral  $x=10$  karyotypes to intermediately-to-recently evolved  $x=9$ ,  $x=8$  and  $x=5$  karyotypes (Lusinska et al., 2019; Sancho et al., 2022; Decena et al., 2023). Phylogenetic, cytogenetic and phenotypic approaches have been used to dissect complex groups of *Brachypodium* cryptic taxa. They were successfully used to disentangle the taxonomic identity and the origins of the three annual species of the *Brachypodium distachyon* complex (Catalán et al., 2012a; López-Alvarez et al., 2012; Catalán et al., 2016a; López-Álvarez et al., 2017a). Considered to be part of a single species for some centuries (*B. distachyon sensu lato*), these studies demonstrated that its three different cytotypes corresponded in fact to three independent species, the diploids *B. stacei* ( $2n=2x=20$ ,  $x=10$ ) and *B. distachyon sensu stricto* ( $2n=2x=10$ ,  $x=5$ ), and their derived allotetraploid *B. hybridum* ( $2n=4x=30$ ,  $x=10+5$ ), which were taxonomically ranked (Catalán et al., 2012). Eight quantitative (Catalán et al., 2012a; López-Álvarez et al., 2017a) and one qualitative (Catalán et al., 2016a) morphological traits were found to discriminate among them, thus unveiling the cryptic diversity of the group.

In contrast to the thoroughly investigated *Brachypodium* annual species, which have reference genomes and known karyotypes (Gordon et al., 2020; Scarlett et al., 2022; Mu et al. 2023), taxonomic uncertainty still persists among the ~17 poorly known *Brachypodium* perennial species, and particularly within some cryptic perennial complex taxa (Catalan et al., 2016; Catalán et al., 2023; Decena et al., 2023). Interestingly, all the studied *Brachypodium* polyploid perennial species have one or more orphan subgenomes, suggesting that their respective diploid progenitor species (genome donors) may be extinct (Sancho et al., 2022; Decena et al., 2023). Recent cytomolecular and



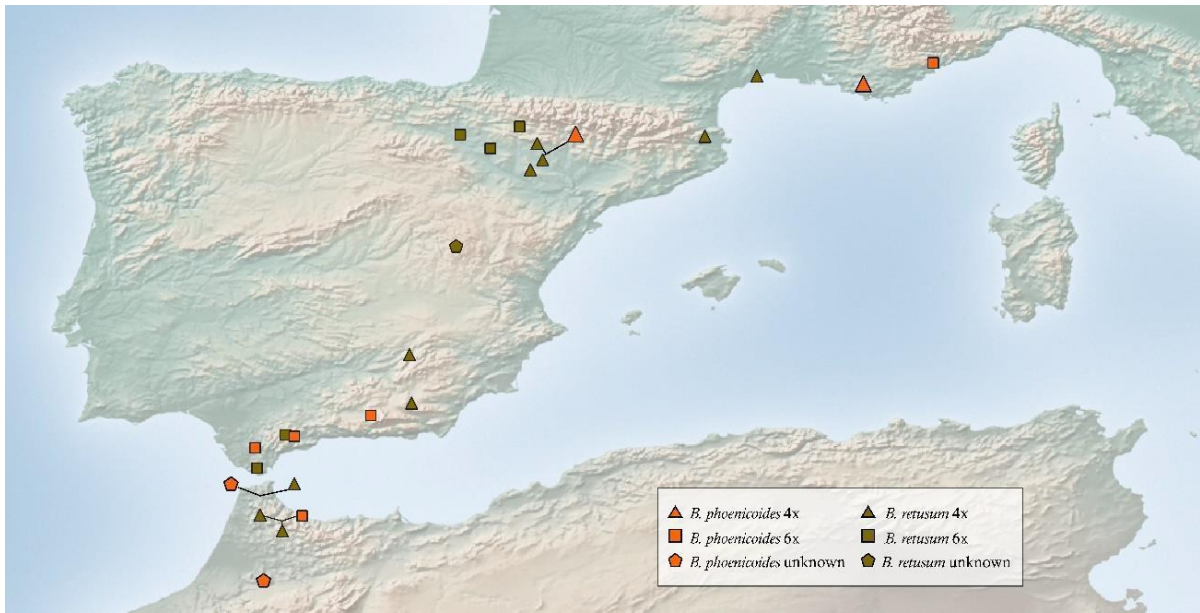
phylogenomic studies of new field collections detected different cytotypes for two Mediterranean perennial species, *B. retusum* and *B. phoenicoides*, and characterized their subgenomes and barcoded karyotypes (Sancho et al., 2022; Decena et al., 2023). *Brachypodium retusum*, a xeric plant distributed along the Mediterranean basin and adapted to dry and arid environments, was resolved as intermediately evolved between the ancestral Miocene outcore *Brachypodium* lineages and recent Plio-Pleistocene core-perennial lineages (Catalan et al., 2016; Sancho et al., 2022). Some of its populations consist of allotetraploid individuals with  $2n=32$  chromosomes ( $n=16$ ) belonging to two different subgenomes, one ancestral ( $x=8$  A2) and the other intermediately evolved ( $x=8$  E1), while other populations contain allohexaploid individuals with  $2n=42$  chromosomes ( $n=21$ ) pertaining to three different subgenomes, A2, E1 and a third more recently evolved subgenome ( $x=5$ , E2) (Sancho et al., 2022; Decena et al., 2023). *Brachypodium phoenicoides*, a plant distributed in the western Mediterranean region and adapted to mesic-dry environments with moist soils, showed a similar intermediately evolved position in the *Brachypodium* phylogeny for some cytotypes but a more recent core-perennial position for others (Catalan et al., 2016; Sancho et al., 2022, Decena et al., 2023). Its populations also consist of allotetraploid or allohexaploid individuals, with the allotetraploids having  $2n=28$  chromosomes ( $n=14$ ) associated to two different subgenomes, one intermediately-to-recently evolved ( $x=5$  E2; shared with the *B. retusum*-6x cytotypes) and the other very recently evolved ( $x=9$  G), and the allohexaploids presenting  $2n=38$  chromosomes ( $n=19$ ) corresponding to three subgenomes, E2, G plus a third extra copy of the E2 subgenome (Sancho et al., 2022; Decena et al., 2023). Despite the striking differences in the subgenomic loads and compositions of *B. retusum* and *B. phoenicoides* allotetraploids and allohexaploids, no independent species have been recognized within any of the two species in modern taxonomic treatments of the genus (Schippmann, 1991) and local floras.

Our study aims to provide an exhaustive preliminary taxonomic framework of the *Brachypodium retusum* and *B. phoenicoides* allopolyploid complexes investigating the morphoanatomical variation of their individuals across their respective geographic distributions and environments, and at the population and cytotypic levels. The specific goals of our search are: i) to identify phenotypic traits that significantly discriminate the different cytotypes (and subgenomic compositions) of each species; ii) to identify other potential infraspecific morphotypic groups within each species and the phenotypic traits



that discriminate them; iii) to establish the potential relationships of infraspecific morphotypic groups to cytotypic, environmental or geographical variables; iv) to uncover cryptic diversity within the *B. retusum* and *B. phoenicoides* polyploid complexes if present.

## MATERIALS AND METHODS



**Figure 1.** Geographical distribution of the 25 studied western Mediterranean populations of *Brachypodium retusum* (olive green) and *B. phoenicoides* (orange). Symbol codes for ploidy levels: tetraploid (triangle), hexaploid (square), unknown (polygon). Colour codes for taxa and symbol codes for cytotypes are indicated in the chart.

### Sampling

A representative sampling of *Brachypodium retusum* and *B. phoenicoides* was performed in the Western Mediterranean region, the native region with the largest diversity of intraspecific tetraploid and hexaploid cytotypes for both species (**Figure 1**; Catalan et al., 2016; Decena et al., 2023). *B. retusum* is characterized by having strongly inrolled distichous leaves and branched woody stems, while *B. phoenicoides* presents partially inrolled non-distichous leaves and non-branched woody stems (Schippmann, 1991; Catalan et al., 2016). The ploidy level of the majority of the studied materials were previously determined through chromosome counting and flow cytometry genome size analysis (Decena et al., 2023). A total of 72 samples of *B. retusum* [45 individuals from 16 populations (10 of tetraploids, 5 of hexaploids, 1 of unknown ploidy)] and *B. phoenicoides* [27 individuals from 9 populations (2 of tetraploids, 5 of hexaploids, 2 of unknown ploidy)] were used in the phenotypic analysis (**Supplementary Table S1**,

**Figure 1).** The samples were collected in the field, studied cytogenetically, and transplanted to our experimental garden at the Escuela Politecnica Superior de Huesca (Universidad de Zaragoza, Spain). Mature individuals from the living collections were used in the phenotypic analyses and their vouchers were deposited at the UZ (University of Zaragoza) and JACA herbaria.

### Phenotypic analysis

A comprehensive phenotypic analysis was performed using 114 characters considered potentially informative to differentiate intraspecific diversity in *Brachypodium* (Schippmann, 1991) and other grasses (Metcalf, 1960; Devesa et al., 1991), or explored taxonomically for the first time. Of these, 56 were macro- and micromorphological traits, 30 anatomical (leaf blade transversal section) traits, and 28 epidermal (leaf blade abaxial epidermis) traits (**Supplementary Table S2**). We employed them to test their potential value to discriminate the tetraploid (4x) and hexaploid (6x) cytotypes of *B. retusum* and *B. phoenicoides* and to detect any other putative intraspecific variation within the studied samples of each species. Due to incompleteness or homogeneity of data for some of the studied variables in the samples (**Appendices S1 and S2**), the number of characters studied were adapted to each statistical analysis. Fifty out of the 114 characters were quantitative [25 morphological (V1, V3, V4, V5, V6, V7, V13, V14, V15, V16, V17, V23, V24, V25, V26, V27, V31, V32, V35, V36, V39, V40, V42, V48, V49), 9 anatomical (V62, V63, V64, V65, V66, V67, V68, V69, V70), and 16 epidermal (V87, V88, V89, V90, V92, V94, V95, V96, V100, V102, V104, V106, V107, V108, V112, V114) traits], and 51 were categorical [24 morphological (V52, V53, V54, V55, V8, V9, V10, V11, V12, V21, V18, V19, V20, V22, V56, V28, V29, V30, V44, V45, V43, V46, V50, V51), 12 anatomical (V81-V86, V71-V76), and 12 epidermal (V91, V93, V97, V98, V99, V101, V103, V105, V109, V110, V111, V113) traits] (**Supplementary Table S2**).

Macromorphological characters were measured on the innovation leaf blade, flag leaf blade, culm, panicle and floral organs with a hard ruler or a calliper under a Nikon dissecting stereomicroscope. The micromorphological anatomical and epidermal characters were measured on transversal sections of the innovation and flag leaf blades, and on leaf blade abaxial epidermis moulds, respectively, under a Zeiss Axiolab microscope. The transversal leaf sections were mounted on a glass slide with glycerol and

observed under the microscope. The abaxial epidermal moulds were obtained from a thin layer of varnish that moulded the epidermis, mounted on a glass slide and observed under the microscope. The two types of preparations were photographed with a Canon digital camera adapted to the microscope at x20 and x100 magnifications. Microscopical measurements were taken with the ImageJ software adjusted to the scale.

### Statistical analyses

Exploratory basic statistic descriptors (Minimum-Maximum range, Median, first quartile, third quartile, and Mean values, and the number of missing values (NA's) were calculated for each quantitative morphological, anatomical and epidermal variable. Normal distribution and homoscedasticity of variables in sampling groups were checked with the Shapiro-Wilk and Levene tests using the *R rstatix* package (Kassambara A, <https://cran.r-project.org/web/packages/rstatix/>). However, as most variables presented a non-normal distribution and failed to adjust to normalization procedures, non-parametric statistic tests were applied to the compared samples. Significant differences between the two ploidy levels (4x, 6x) were estimated for each species at the individual sampling level with the Mann-Whitney test using the *R stats* package (R Core Team, <http://www.R-project.org/>). These analyses included macromorphological traits and microanatomical leaf section traits (**Supplementary Table S2**). In addition, we also tested for significant differences between ploidy levels at the population sampling level for morphological, anatomical, and epidermal traits (**Supplementary Table S2**) using the same approach. Populations with unknown ploidy level were discarded for this analysis. Multiple pairwise comparisons computed with non-parametric Mann-Whitney tests between groups were graphically outputted with the options of the *ggbetween* R package (Patil, 2021).

Multivariate Principal Component Analysis (PCA) of 45 morphoanatomical quantitative variables were performed separately within each species to examine the structure of the samples, and to assess if the observed groupings were consistent with the cytotypic ranks (ploidy levels), geographic distribution or environmental characteristics. The contribution of each character to the coordinate axes that accumulate the highest percentages of variance was calculated by covariance and variance matrices of the samples with respect to the new axes. Furthermore, 96 quantitative and qualitative variables were examined together by means of a Factor analysis of mixed data (FAMD).

This analysis is a principal component method for quantitative variables and a multiple correspondence analysis (MCA) for qualitative variables (Pagès, 2004). All multivariate analyses were first performed using the *missMDA* R package (Josse and Husson, 2016) to account for missing data, and then with the *FactoMineR* R package (Lê et al., 2008) with a previous automatic scaling of the data. For these analyses, all studied samples of each species were included.

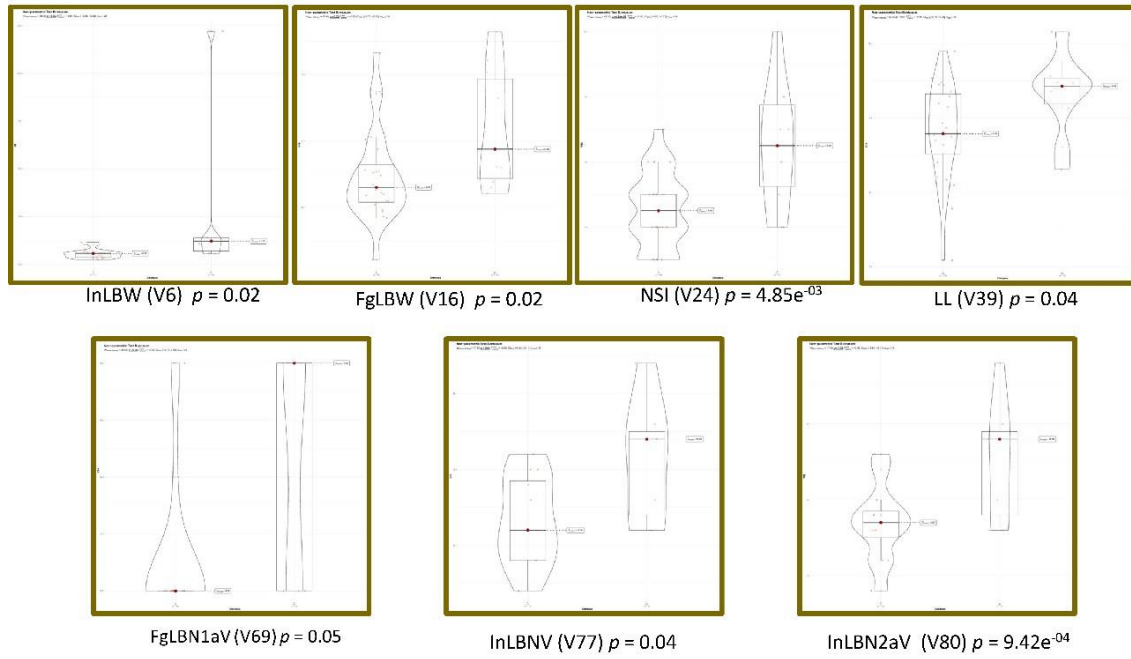
## RESULTS

### Morphological differentiation within *Brachypodium retusum* and *B. phoenicoides* tetraploid and hexaploid cytotypes

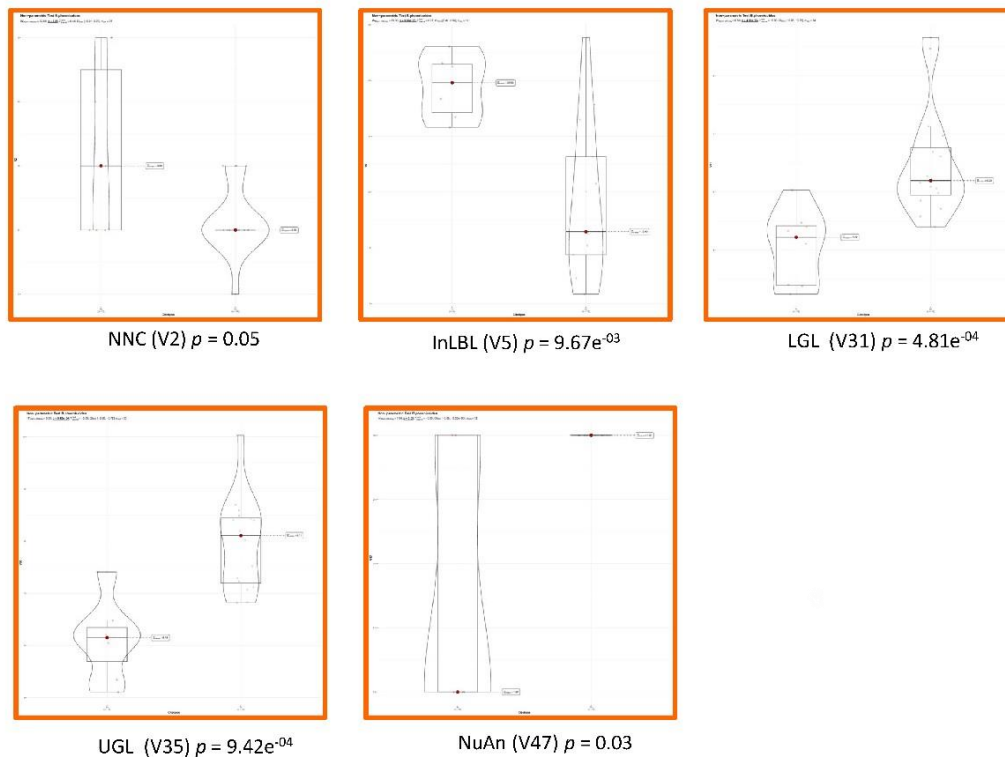
Seven out of the 50 analyzed quantitative morphoanatomical traits were useful in significantly discriminating the *Brachypodium retusum* tetraploid and hexaploid samples and five in significantly separating the *B. phoenicoides* 4x and 6x samples at  $p < 0.05$  (**Table 1, Figure 2**).

The seven traits that discriminated the two *B. retusum* cytotypes (4x vs 6x) were innovation leaf blade width (InLBW, V6), flag leaf blade width (FgLBW, V16), number of spikelets per inflorescence (NSI, V24), lemma length (LL, V39), number of primary veins of flag leaf (FgLBN1aV, V69), number of veins of innovation leaf (InLBNV, V77) and number of secondary veins of innovation leaf (InLBN2aV, V80) (**Table 1A, Figure 2A**). The *B. retusum* tetraploids showed smaller widths for both the innovation (0.57 (0.37-0.62) mm) and the flag leaves (0.54 (0.65-0.82) mm), a lower and constant number of spikelets per inflorescence (3), and smaller lemma lengths 7-9.9 mm (median 8.8 mm) than the hexaploids (1.22(0.70-1.39) mm and 0.93 (0.71-1.46) mm, 3-5 (median 5), 8.3-10.1 mm (median 9.4 mm), respectively) (**Table 1A, Figure 2A, Appendix S1**). The tetraploids also exhibited significantly differentiated anatomical traits values than the hexaploids, such as lower number of total (11 (9-14.25) and secondary veins (8.5 (7.5-9.25) per leaf in the former group than in the second group (17 (12-17.5), and 14 (9-14.5), respectively) (**Table 1A, Figure 2A, Appendix S1**).

**A**



**B**



**Figure 2.** Violin plots and boxplots of the studied morphoanatomical and leaf blade abaxial epidermal variables that showed significant intraspecific differences between the analysed individual samples of *Brachypodium retusum* (A) and *B. phoenicoides* (B). Tetraploid (4x) and hexaploid (6x) samples. Acronyms of variables correspond to those indicated in **Supplementary Table S2**.

The five morphoanatomical traits that significantly differentiated the *B. phoenicoides* cytotypes were the number of culm nodes (NNC, V2), the innovation leaf blade length (InLBL, V5), the lower glume (G1) and upper glume (G2) lengths (LGL, V31 and UGL, V35, respectively) and the number of anthers (NuAn, V47) (**Table 1B, Figure 2B**). The *B. phoenicoides* tetraploids exhibited a higher number of nodes per culm (3 (2-4.5) and longer innovation leaf blades (39.60 (34.26-42.95) cm) than the hexaploids (2 (2-2), and 12.9 (8.76-26.30) cm, respectively) (**Table 2B, Figure 2B, Appendix S2**). By contrast the tetraploids had smaller lengths of lower (5.22 (4.40-5.42) mm) and upper glumes (6.15 (5.70-6.35) mm) and a lower number of anthers (1-3 (mean 1.8) than the hexaploids (6.20 (5.95-6.76) mm, 8.11 (7.20-8.45) mm, and 3, respectively) (**Table 2B, Figure 2B, Appendix S2**).

**Table 1.** Statistics of 50 quantitative morphoanatomical traits (macromorphological and leaf blade section anatomy) from 67 analyzed individuals of tetraploid and hexaploid cytotypes of *Brachypodium retusum* (A) and *B. phoenicoides* (B). In bold, variables showing significant differences in the U-Mann-Whitney test (see **Supplementary Table S2** for abbreviation of variables, and **Supplementary Tables S3** and **S4** for detailed data measurements and U-Mann-Whitney tests, respectively).

(A)	v1	v2	v3	v4	v5	v6	v7	v13	v14	v15	v16	v17	v23	v24	v25	v26	v27	v31	v32	v33	v34	v35	v36	v37	v38
<i>B. retusum</i> 4x																									
Median	44.75	7.50	26.94	0.00	15.68	<b>0.57</b>	1.02	43.10	0.00	11.98	<b>0.65</b>	1.03	7.87	<b>4.50</b>	15.95	26.16	9.00	5.57	0.61	3.00	1.00	6.66	0.75	4.50	1.50
1st Qu.	38.30	5.00	9.46	0.00	7.48	<b>0.37</b>	0.76	12.14	0.00	8.60	<b>0.54</b>	0.83	6.75	<b>4.00</b>	13.79	23.72	7.00	5.03	0.51	3.00	0.00	6.02	0.68	4.00	0.00
3rd Qu.	52.70	9.00	35.60	6.94	53.35	<b>0.62</b>	1.17	59.58	6.47	13.00	<b>0.82</b>	1.50	9.25	<b>5.00</b>	19.06	28.59	11.00	6.02	0.72	4.00	2.00	7.05	0.81	5.00	3.00
Mean	45.68	7.17	25.82	4.13	30.71	<b>0.55</b>	0.97	36.39	6.76	24.92	<b>0.74</b>	1.21	8.19	<b>4.54</b>	20.43	25.99	9.00	5.53	0.62	3.42	1.08	6.52	0.74	4.50	1.54
Min	28.70	4.00	3.72	0.00	1.06	<b>0.25</b>	0.53	6.20	0.00	6.17	<b>0.11</b>	0.61	4.72	<b>3.00</b>	8.30	13.01	3.00	3.55	0.36	2.00	0.00	5.04	0.30	3.00	0.00
Max	64.20	10.00	56.29	19.36	84.63	<b>1.15</b>	1.38	68.45	76.00	156.10	<b>1.66</b>	2.22	13.60	<b>7.00</b>	99.85	39.23	14.00	7.50	0.99	5.00	3.00	8.88	1.19	6.00	4.00
Na's	0.00	0.00	6.00	0.00	6.00	<b>6.00</b>	7.00	6.00	1.00	0.00	<b>0.00</b>	0.00	1.00	<b>0.00</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
<i>B. retusum</i> 6x																									
Median	49.75	6.50	43.13	1.29	38.05	<b>1.22</b>	1.12	50.74	0.00	13.62	<b>0.94</b>	1.20	9.29	<b>6.50</b>	11.60	25.46	7.50	5.80	0.78	3.00	0.00	6.66	0.76	4.00	1.50
1st Qu.	39.65	4.50	16.83	0.20	12.38	<b>0.70</b>	0.93	30.79	0.00	10.46	<b>0.72</b>	0.99	8.29	<b>5.25</b>	10.98	23.46	6.25	5.49	0.66	3.00	0.00	6.03	0.65	3.25	0.25
3rd Qu.	60.02	9.00	46.55	3.66	72.52	<b>1.39</b>	1.29	59.17	2.58	83.11	<b>1.47</b>	1.30	10.17	<b>7.75</b>	15.57	29.99	10.50	6.10	0.80	4.00	1.75	7.19	0.88	5.75	2.75
Mean	48.54	6.60	33.44	2.09	41.94	<b>2.86</b>	1.07	43.72	1.66	40.29	<b>1.09</b>	1.17	8.69	<b>6.50</b>	13.97	25.84	8.30	5.66	0.75	3.70	0.80	6.76	0.76	4.40	1.60
Min	26.40	3.00	7.80	0.00	8.00	<b>0.56</b>	0.61	10.92	0.00	6.77	<b>0.61</b>	0.81	4.01	<b>4.00</b>	10.32	17.68	5.00	4.19	0.45	2.00	0.00	5.31	0.45	3.00	0.00
Max	63.30	9.00	50.81	5.67	79.73	<b>12.20</b>	1.37	62.72	7.40	87.55	<b>1.82</b>	1.53	10.48	<b>10.00</b>	22.53	32.91	13.00	6.77	1.02	6.00	3.00	9.36	1.03	6.00	4.00
Na's	0.00	0.00	4.00	4.00	4.00	<b>4.00</b>	4.00	0.00	0.00	0.00	<b>0.00</b>		2.00	<b>0.00</b>	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00



	v39	v40	v41	v42	v47	v48	v49	v57	v58	v59	v60	v61	v62	v63	v64	v65	v66	v67	v68	v69	v70	v77	v78	v79	v80
<i>B. retusum 4x</i>																									
<b>Median</b>	<b>8.79</b>	0.82	5.00	0.85	2.00	3.19	3.02	148.10	113.50	52.76	27.24	72.40	126.55	98.85	45.75	28.00	71.50	14.00	13.00	<b>3.00</b>	10.00	<b>11.00</b>	10.00	3.00	<b>8.50</b>
<b>1st Qu.</b>	<b>8.52</b>	0.68	4.25	0.49	1.50	2.78	2.60	136.60	105.90	48.60	24.52	55.40	111.55	71.03	43.48	24.68	59.68	13.00	11.00	<b>3.00</b>	10.00	<b>9.00</b>	9.50	3.00	<b>7.50</b>
<b>3rd Qu.</b>	<b>9.33</b>	0.95	5.00	0.99	3.00	3.83	3.44	165.50	138.40	62.50	32.50	85.60	141.18	107.95	62.80	30.35	92.48	15.00	14.00	<b>3.00</b>	12.00	<b>14.25</b>	13.00	3.00	<b>9.25</b>
<b>Mean</b>	<b>8.82</b>	0.86	4.82	0.82	2.18	3.10	3.02	150.10	122.20	55.80	28.59	77.78	120.34	89.88	49.50	27.10	74.67	14.29	13.29	<b>3.29</b>	10.59	<b>11.67</b>	10.67	2.83	<b>8.50</b>
<b>Min</b>	<b>7.09</b>	0.53	3.00	0.13	1.00	0.73	2.18	97.60	80.00	35.70	19.90	42.90	0.11	0.08	0.04	0.02	0.05	10.00	9.00	<b>3.00</b>	3.00	<b>7.00</b>	6.00	1.00	<b>4.00</b>
<b>Max</b>	<b>9.90</b>	1.55	7.00	2.23	3.00	4.43	3.86	207.90	169.30	85.90	42.50	167.90	164.40	141.60	82.40	45.90	153.60	23.00	21.00	<b>5.00</b>	18.00	<b>16.00</b>	15.00	3.00	<b>13.00</b>
<b>Na's</b>	<b>1.00</b>	1.00	2.00	1.00	13.00	14.00	22.00	7.00	7.00	7.00	7.00	7.00	12.00	12.00	12.00	12.00	12.00	7.00	7.00	<b>7.00</b>	7.00	<b>12.00</b>	12.00	12.00	<b>12.00</b>
<i>B. retusum 6x</i>																									
<b>Median</b>	<b>9.43</b>	0.89	5.00	0.80	2.50	3.43	5.69	156.80	120.70	57.60	28.20	69.80	132.10	114.10	59.40	30.70	73.20	16.00	14.00	<b>5.00</b>	11.00	<b>17.00</b>	15.00	3.00	<b>14.00</b>
<b>1st Qu.</b>	<b>9.19</b>	0.76	5.00	0.49	1.75	2.84	5.69	145.60	117.20	51.85	24.40	61.60	112.20	86.75	48.30	29.80	54.30	13.50	13.00	<b>3.00</b>	10.50	<b>12.00</b>	10.00	3.00	<b>9.00</b>
<b>3rd Qu.</b>	<b>9.54</b>	0.96	5.75	0.83	3.00	4.10	5.69	175.20	127.50	62.80	43.85	85.15	169.10	134.05	63.05	37.30	95.25	20.00	18.50	<b>5.00</b>	15.00	<b>17.50</b>	16.50	3.00	<b>14.50</b>
<b>Mean</b>	<b>9.32</b>	0.90	4.70	0.67	2.25	3.51	5.69	316.30	240.20	115.61	63.56	202.93	305.00	232.63	116.73	68.91	158.03	16.86	15.71	<b>4.14</b>	12.86	<b>15.57</b>	13.71	3.00	<b>12.57</b>
<b>Min</b>	<b>8.31</b>	0.49	0.00	0.24	1.00	2.74	5.69	122.40	106.10	42.20	21.00	56.40	94.90	75.70	37.20	22.90	20.90	13.00	11.00	<b>3.00</b>	10.00	<b>11.00</b>	8.00	3.00	<b>8.00</b>
<b>Max</b>	<b>10.16</b>	1.50	7.00	1.00	3.00	4.44	5.69	1293.0	965.00	480.20	259.20	1000.8	1345.0	997.00	497.80	294.60	713.00	22.00	22.00	<b>5.00</b>	18.00	<b>22.00</b>	20.00	3.00	<b>19.00</b>
<b>Na's</b>	<b>0.00</b>	0.00	0.00	1.00	6.00	6.00	9.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	<b>3.00</b>	3.00	<b>3.00</b>	3.00	3.00	<b>3.00</b>

(B)	v1	v2	v3	v4	v5	v6	v7	v13	v14	v15	v16	v17	v23	v24	v25	v26	v27	v31	v32	v33	v34	v35	v36	v37	v38	
B. phoenicoides 4x																										
Mean	73.69	3.29	11.07	0.7	38.91	0.77	1.07	24.78	15.07	30.16	1.22	26.29	14.88	8.75	21.7	29.46	13.12	5.05	0.76	3.25	1.13	6.11	0.84	4.86	1.14	
Min	54.8	2	8	0	31.6	0.54	0.89	10.13	7.1	19.3	1.01	1	11	7	11.08	19.83	8	4.24	0.59	2	0	5.11	0.47	4	0	
Max	95.5	5	14.1	4.2	46.1	1.18	1.26	33.8	20	64.6	1.42	197	18.9	10	30.1	42.62	21	6.03	1.08	6	2	7.41	1.15	6	3	
Na's	1	1	2	2	2	4	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	1	1	1	
1st Qu	59.05	2	8.69	0	34.26	0.62	0.94	20.77	13.15	22.18	1.17	1.75	11.95	8	20.01	22.42	8	4.4	0.66	2.75	0	5.7	0.7	4.5	0	
Median	78.7	3	11.42	0	39.6	0.68	1.05	27.6	16.21	25.4	1.18	1.94	15.4	9	21.46	27.98	12	5.22	0.72	3	1.5	6.15	0.92	5	1	
3rd Qu.	84.37	4.5	13.15	0	42.95	0.82	1.2	29.85	18.2	29.73	1.31	2.24	17.23	9.25	25.23	36.34	18	5.42	0.82	3.25	2	6.35	0.97	5	2	
B. phoenicoides 6x																										
Median	57.6	2	7.6	0	12.9	0.93	1.07	31.2	18.2	27	1.52	1.89	16.6	8	25.54	29.89	11	6.2	0.75	4	0	8.11	0.89	5	0	
1st Qu	51.45	2	5.88	0	8.76	0.76	0.79	21.62	8.78	14.6	1.31	1.34	13.57	7	23.28	26.66	7.75	5.95	0.69	3	0	7.2	0.76	3	0	
3rd Qu.	69.9	2	11.4	0	26.3	1.23	1.26	48.77	29.2	29.65	1.87	1.96	18.65	9	30.42	33.69	12	6.76	0.83	4.25	2	8.45	1.03	6	2	
Mean	60.65	2.14	12.22	0.25	18.24	1.02	1.06	33.68	19.61	23.24	1.56	1.75	16.49	8	26.4	31.31	10.31	6.49	0.77	3.94	1.06	7.95	0.9	4.69	1.13	
Min	31.9	1	3.69	0	1.69	0.62	0.53	17.5	3.7	8.5	0.43	0.83	10	6	15.78	18.3	3	5.4	0.56	2	0	6.82	0.71	3	0	
Max	89.1	3	56.8	3.3	47.7	1.57	1.84	51.5	41.8	35.1	2.31	2.68	26.1	9	39.75	48.96	19	8.66	1.06	7	3	10.03	1.16	7	4	
Na's	1	2	3	3	3	3	3	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	

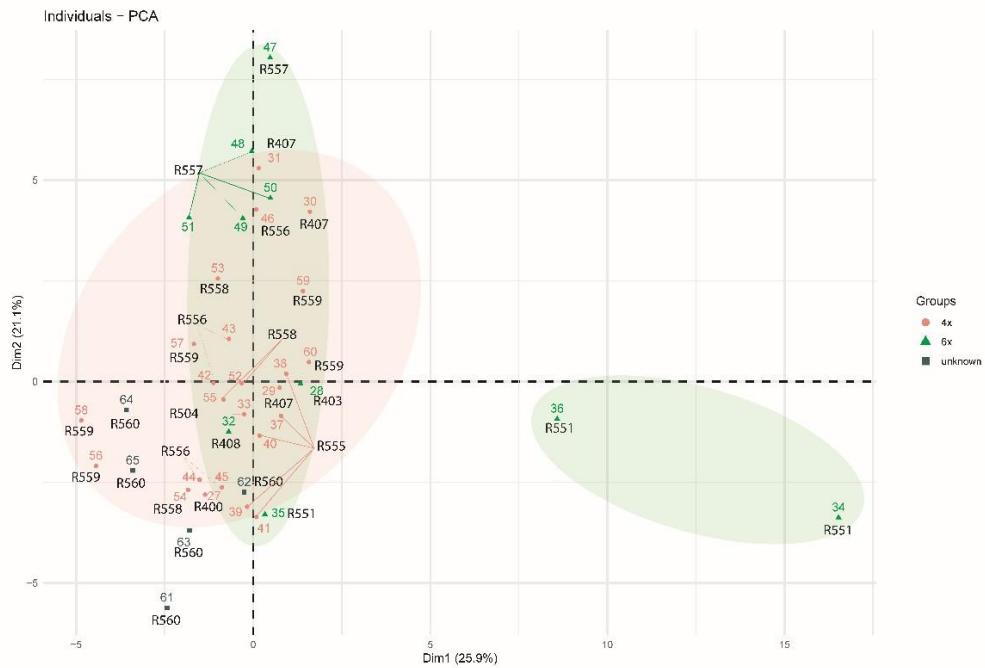
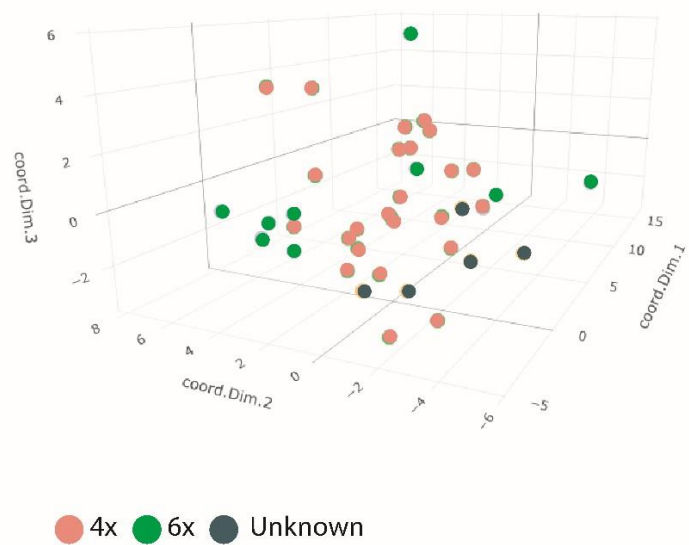
	v39	v40	v41	v42	v47	v48	v49	v57	v58	v59	v60	v61	v62	v63	v64	v65	v66	v67	v68	v69	v70	v77	v78	v79	v80
<i>B. phoenicoides</i> 4x																									
<b>Median</b>	9.7	0.98	6	0.53	1	4.45	4.76	209.5	176.4	75.98	32.42	56.26	206.4	163.9	73.86	31.5	51.26	21	20	5	16	20	19	5	15
<b>1st Qu</b>	9.05	0.84	5	0.39	1	3.9	4.58	185.8	155.8	68.45	29.4	44.82	194.8	154.4	65.9	30.82	36.95	19.5	20	5	12.25	13	12	5	8
<b>3rd Qu.</b>	10.6	1.07	6	1.21	3	4.79	4.93	231.1	199.9	82.5	33.04	63.5	207.3	193.5	83.74	72.88	55.12	25.5	24.5	5.75	19	20	20	5	15
<b>Mean</b>	9.8	0.96	5.63	0.79	1.8	4.25	4.76	482.4	414.5	184.94	82.27	156.44	529.5	428.8	186.33	77.94	146.63	21.83	21.17	5.5	16	17.4	16.6	4.6	12.2
<b>Min</b>	7.79	0.57	5	0	1	3.07	4.4	171.7	149.7	60.68	23.48	39.64	184.1	141.3	64.34	25.5	0	16	15	5	11	13	12	3	7
<b>Max</b>	11.63	1.39	6	1.99	3	5.03	5.11	1887.6	1628	746.4	293	679.6	1854.6	1491	643.8	229	589.8	27	26	7	22	21	20	5	16
<b>Na's</b>	0	0	0	0	3	4	6	2	2	2	3	2	3	3	3	3	3	2	2	2	2	3	3	3	3
<i>B. phoenicoides</i> 6x																									
<b>Median</b>	10.78	1.15	5	0.76	3	4.37	NA	267.3	223.6	100.47	38.76	79.58	209.5	172.9	68.93	31.35	52.43	21	20.5	6	15.5	17.5	16.5	5	12.5
<b>1st Qu</b>	10.04	1.03	5	0.39	3	4.15	NA	224.3	181.7	79.89	36.46	69	186.7	161	61.16	27.27	47.48	20	19	5	14	16	15	5	11
<b>3rd Qu.</b>	11.72	1.34	6	0.84	3	4.6	NA	2003.8	1773.1	826.1	372.6	498.5	229.7	179.8	76.81	35.13	59.7	24	24	6.75	18	18.25	17.25	5	13.25
<b>Mean</b>	10.96	1.16	5.06	0.8	3	4.29	NA	996.8	840.4	398.16	179.46	307.84	206.8	167.9	69.05	31.05	54.75	21.6	21.1	6	15.6	16.75	15.75	5	11.75
<b>Max</b>	14.3	1.63	6	2.37	3	4.67	NA	2405.8	1917.6	1037.6	430.8	869.8	247.4	197.4	79.68	37.26	66.7	26	25	8	18	19	18	5	14
<b>Min</b>	8.8	0.78	3	0.29	3	3.59	NA	0	149.3	62.12	21.6	49.08	160.8	128.3	58.66	24.26	47.44	16	16	5	11	13	12	5	8
<b>Na's</b>	0	1	0	5	9	10	16	6	6	6	7	6	12	12	12	12	12	6	6	6	6	12	12	12	12

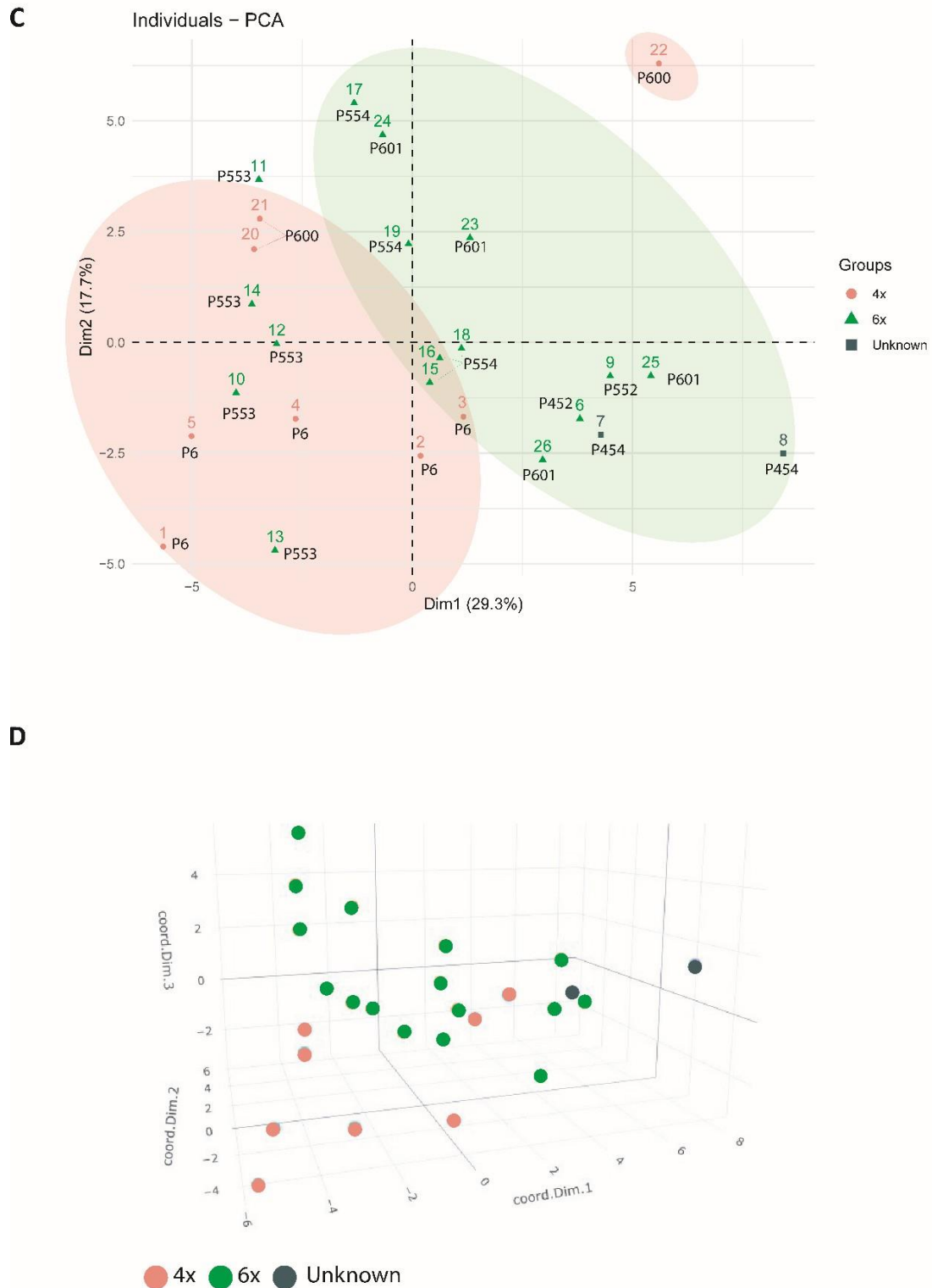
### Multivariate PCA and FAMD analyses of *Brachypodium retusum* and *B. phoenicoides* samples

The PCA of the *Brachypodium retusum* 4x and 6x samples identified three principal components which altogether accounted for 56.59% of the observed morphological variation detected by the 46 analyzed morphoanatomical quantitative characters (PC1: 25.85%, PC2: 21.05%, PC3: 9.68%; **Supplementary Table S3A**). Ten characters were identified as the most important contributors to the negative extreme of the first component (V57-V66), 11 to the positive and negative extremes of the second component (V2, V23, V24, V31, V67-V70, V77, V78, V80), and 10 to both extremes of the third component (V3-V6, V13-V15, V36-V38) **Supplementary Table S3A**). The impact of these components on the hierarchical structure of the *B. retusum* samples was visualized in bidimensional and tridimensional PCA plots (**Figures 3A** and **3B**). Individual samples of the two 4x and 6x cytotypes overlapped in the two-dimensional space created by the first two components (**Figure 3A**), although some hexaploid samples were also spread along the positive extreme of the first component and some tetraploid samples tended to cluster in the lower left corner of the 2D plot. Unknown ploidy level *B. retusum* samples from Northern Spain (R560, Zaragoza) were closer to the tetraploids from the same area (R559 and R400, Huesca) than to the hexaploids in both 2D and 3D plots (**Figures 3A** and **3B**).

The PCA of the *Brachypodium phoenicoides* 4x and 6x samples retrieved three principal components that accounted for 58.43% of the morphological variation detected by the same 46 morphoanatomical traits (PC1: 28.85%, PC2: 18.58%, PC3: 11%; **Supplementary Table S3B**). Ten variables were found to be the most important contributors to the positive extreme of the first component (V57-V66), 9 to the positive and negative extremes of the second component (V4, V6, V67-V69, V77-V80), and 9 to both extremes of the third component (V5, V13, V31, V34, V35, V37, V38, V41, V69) **Supplementary Table S3B**). The 2D and 3D *B. phoenicoides* plots (**Figures 3C** and **3D**) also detected a large overlapping of 4x and 6x samples but with hexaploid samples covering most of the multidimensional phenotypic space and tetraploid samples clustering preferentially in the left half of the bidimensional plot (**Figure 3C**). Unknown ploidy level *B. phoenicoides* samples from Morocco (P454) were closer to the hexaploid than to the tetraploid samples in both 2D and 3D plots and clustered close to samples from a South Spain population (P552) (**Figures 3C** and **3D**).

**A**

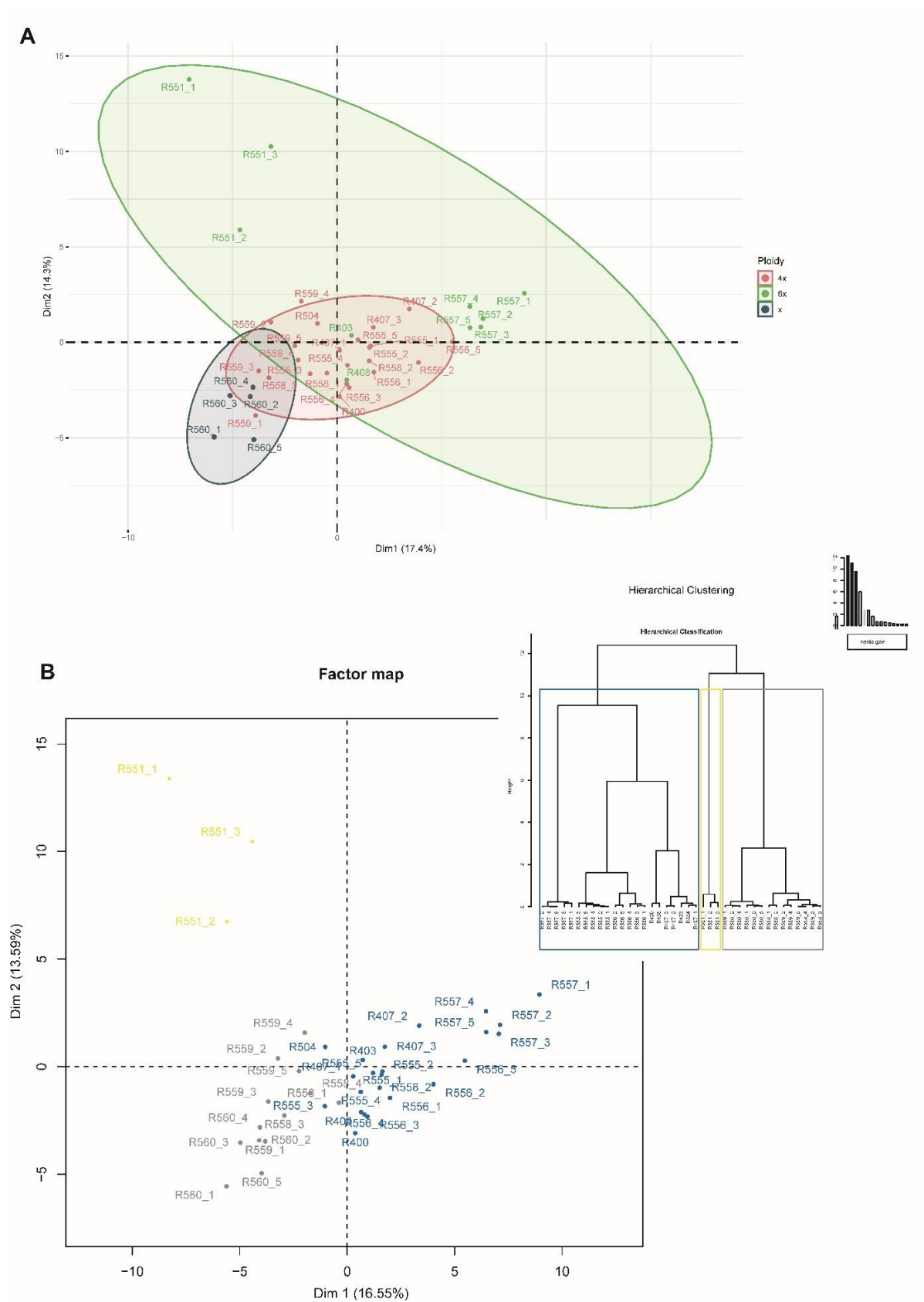
**B**

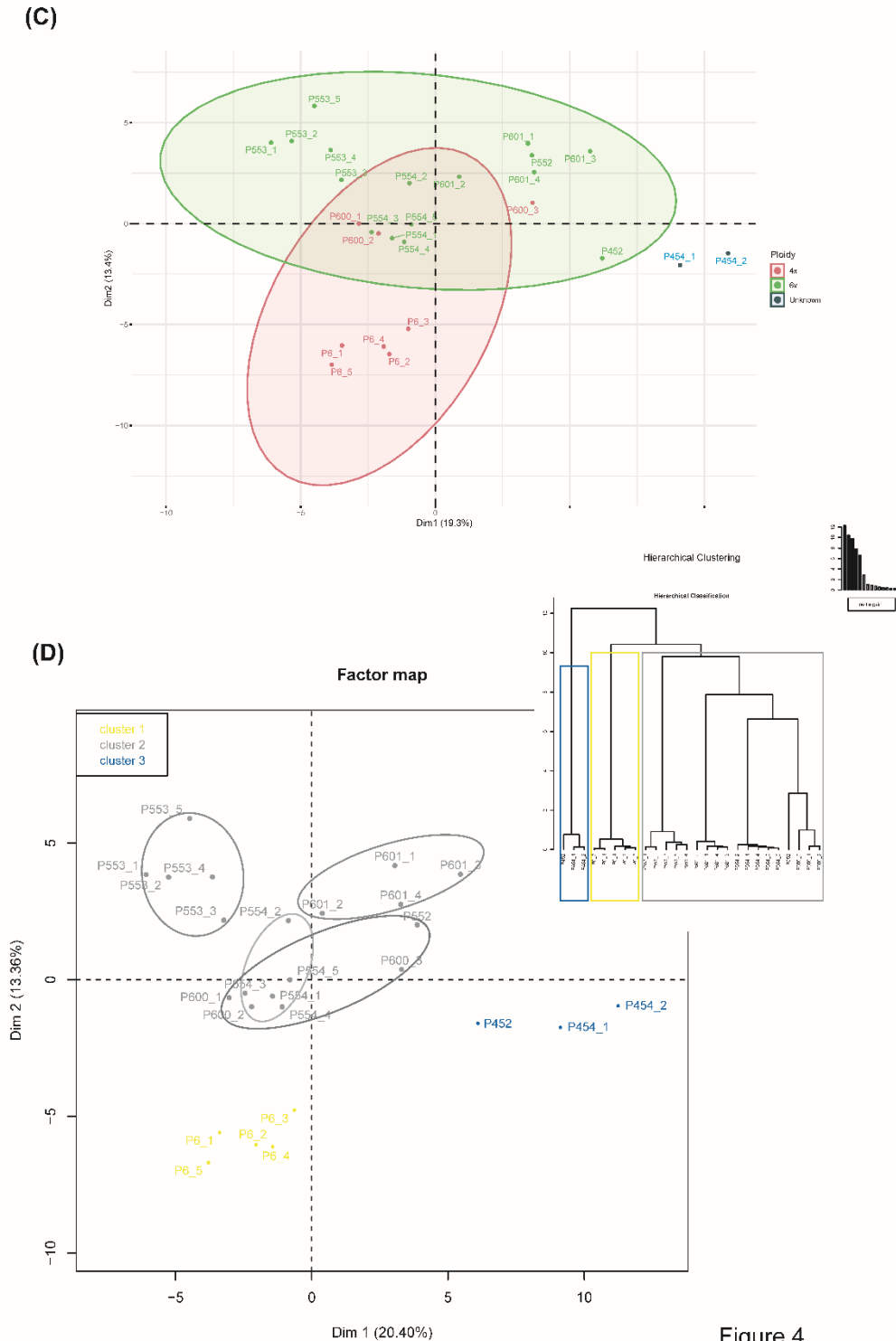


**Figure 3:** Principal Component Analysis (PCA) of 41 samples of *Brachypodium retusum* and 26 samples of *B. phoenicoides* based on 45 quantitative morphoanatomical characters. *B. retusum*: (A) Bidimensional PCA plot with ploidy level as grouping factor; axes 1 and 2 accumulate 25.9% and 21.1% of the variance, respectively. (B) Three-dimensional PCA plot with ploidy level as grouping factor; axis 3 accumulates 9.7% of the variance. *B. phoenicoides*: (C) Bidimensional PCA plot with ploidy level as grouping factor; axes 1 and 2 accumulate 29.3% and 17.1% of the variance, respectively. (D) Three-dimensional PCA plot with ploidy level as grouping factor; axis 3 accumulates 11% of the variance. Tetraploid individuals (red), hexaploid individuals (green), unknown ploidy level individuals (grey).

The morphoanatomical, epidermal and category-group variability of the *B. retusum* samples detected by the FAMD analysis explained 42.04% of the total variance through the first principal components of the PCA (PC1: 17.43%, PC2: 14.30% and PC3:10.30%; **Supplementary Table S4A, Supplementary Figure S1A**). The most important contributing characters to the positive and negative extremes of the PC1 and PC2 axis were quantitative and discrete characters (e. g., V67, V68, V69, V70, V77-V78 to PC1, and V57, V58, V59, V60, V62, V63, V64, V65, V66 to PC2) while those contributing to the extremes of the PC3 axis were mostly qualitative characters (**Supplementary Table S4A, Supplementary Figure S1A**). The 2D PCA plot of this FAMD analysis (according to the ploidy level of samples) showed also an overlap in the distributions of tetraploids and hexaploids although the hexaploids were preferentially distributed towards the positive extremes of the PC1 and PC2 axes and the tetraploids clustered close to each other in the central part of the plot (**Figure 4A**). The individuals with unknown ploidy (R560 population) clustered very closely to tetraploids from the same region (N Spain) (**Figure 4A**). The hierarchical clustering of the FAMD recovered three clusters for the studied *B. retusum* samples (**Supplementary Table S5A; Figure 4B**). Cluster 1 included hexaploid samples from one population in Southern Spain (R551, Malaga), which were characterized by their rounded ligule shape and smooth spikelets, and by their larger veins, valleys, and bulliform cells in the innovation and flag leaves (V57-60, V62-65), longer epidermal macro hairs (V66) and innovation leaf blade length (V15), shorter culms (V1) and upper glumes (V35), and lower number of nodes (V2) than the rest. Clusters 2 contained tetraploid samples from Northern Spain populations (R558, 559, plus the close R560 of unknown ploidy level), characterized by their scabrid glumes and lemma (V44, V45, V46), smooth spikelet rachis (V30), and longer awns (V42) than the others. Cluster 3 reunited hexaploid and tetraploid samples from various zones of South France, North and South Spain and Morocco, which were characterized by their glabrous glumes and lemma (V44, V45, V46), scabrid spikelets (V30), non-ciliated ligule edges of the innovation and flag leaves (V10, V20), glabrous and bifid ligules (V19, V18), and with more nodes per culm (V2), more secondary veins in the upper glume (V38) and smaller innovation leaf sheaths (V3) than the rest (**Supplementary Table S5A; Figure 4B**).







**Figure 4.** Factorial Analysis of Mixed Data (FAMD) of 41 samples of *Brachypodium retusum* and 26 samples of *B. phoenicoides* based on 96 morphoanatomical, epidermal and category-group characters. *B. retusum*: (A) Bidimensional PCA plot with ploidy level as grouping factor; axes 1 and 2 accumulate 15.4% and 13.2% of the variance, respectively. (B) Hierarchical clustering and factor map showing the differentiation of the three main clusters and the distribution of the clusters' samples in a 2D plot, respectively. *B. phoenicoides*: (C) Bidimensional PCA plot with ploidy level as grouping factor; axes 1 and 2 accumulate 19.3% and 13.4% of the variance, respectively. (D) Hierarchical clustering and factor map showing the differentiation of three four main (sub)clusters and the distribution of the clusters' samples in a 2D plot, respectively. (see **Supplementary Table S1**).

The FADM analysis performed with the *B. phoenicoides* samples explained 43.53 % of their global variability by its three main principal components (PC1: 19.27%, PC2: 13.38%, PC3:10.88%; **Supplementary Table S4B, Supplementary Figure S1B**). The most important contributing characters to the positive and negative extremes of the PC1 were 9 out of the ten quantitative variables detected previously (**Supplementary Table S3B**) plus 4 category-group variables (Locality, Population, Geographical region, Country) (**Supplementary Table S4B**), and those contributing most to the extremes of the PC2 and PC3 were category-group variables (Locality, Geographical region, Ploidy) and four morphoanatomical traits (V20, V44, V45, V55), and category-group variables and 3 qualitative (V8, V11, V19) and 6 quantitative (V1, V5, V6, V26, V27, V77, V78, V80) traits, respectively (**Supplementary Table S4B, Supplementary Figure S1B**). The 2D PCA-FAMD plot (ploidy level of samples) showed some overlap in the distributions of tetraploids and hexaploids in the central part of the space although the hexaploids clustered mostly in the positive extreme and the tetraploids in the negative extreme of PC2 (**Figure 4C**). The individuals with unknown ploidy (P454) clustered very closely to hexaploids from the same region (Morocco) (**Figure 4C**). The hierarchical clustering of FAMD recovered tree clearly separated clusters for the studied *B. phoenicoides* samples (**Supplementary Table S5B; Figure 4D**). Cluster 1 contained hexaploid samples from Morocco (P452, P454) and South Spain (P552, Cadiz) which were characterized by absence of prickled epidermal hairs (V75), larger veins, valleys and bulliform cells in the innovation and flag leaves (V57-60, V61-65), higher number of G2 veins (V38) and longer spikelets (V26), and smaller number of culm nodes (V2), and spikelets per inflorescence (V24) than the rest. Cluster 2 included tetraploid samples from one population in Northern Spain (P6, Huesca), which were characterized by their smooth culm (V55), absence of prickled cells (V85), non-ciliate ligule in the flag leaf (V10), longer ligule (V17), innovation leaf blade (V5) and epidermal macro hairs (V66), and smaller glumes (V31, V35), lemma width (V40), shorter distance between the first and the second spikelet (V26), and lower number of culm nodes (V2) than the others (**Supplementary Table S5B; Figure 4D**). Cluster 3 showed the split of two separated subclusters one including hexaploid populations from South Spain (Malaga, Granada) and the other tetraploid and hexaploid populations from South France (**Figure 4D**). Overall, the samples were characterized by their scabrid glumes and culm (V445, V45), glabrous ligule of the innovation leaf (V9), presence of epidermal prickled hairs (V75), longer inflorescence (V23), larger distance between spikelets (V25), more veins in the

lower glume (V33), and shorter length of the leaf sheath and its closing part (V13, V15) than the rest (**Supplementary Table S5B; Figure 4D**).

## DISCUSSION

**Phenotypic variation within the *Brachypodium retusum* and *B. phoenicoides* polyploid complexes is partially concordant with cytotypic, geographical, environmental and phyletic variation.**

Our morphoanatomical study has demonstrated that the *Brachypodium retusum* and *B. phoenicoides* polyploid complexes are aggregates of morphologically distinct phenotypes that correspond to the respective different ploidy-level cytotypes and to geographically or environmentally constrained lineages (**Table 1** and **Supplementary Tables S3 and S4; Figures 2, 3 and 4**). The phenotypic analyses conducted separately at infraspecific level have untapped the large morphological diversity of each complex. The taxonomic characterization of the infraspecific cytotypes could be based on the informative phenetic characters that have shown significant differences between the tetraploid and hexaploid samples and that could be used as diagnostic traits. Within the *B. retusum* polyploid complex five of the seven diagnostic traits corresponded to vegetative characters and two to reproductive characters (**Table 1** and **Supplementary Table S2**). In all cases the allotetraploids showed smaller values than the allohexaploids, exhibiting significantly narrower widths of innovation and flag leaf blades and, consequently, lower number of main, and lateral primary and secondary veins in those leaves, together with lower number of spikelets per inflorescence and shorter lemmas than the allohexaploids (**Table 1, Figure 2**). These expected results would probably be a consequence of the higher genic dosage of the allohexaploids which have an extra subgenome ( $x=5$ , E2) with respect to the two shared subgenomes present in both allotetraploids and allohexaploids ( $x=8$  A2,  $x=8$  E1) (Sancho et al., 2022; Decena et al., 2023). It also complies with the hypothesis of higher plant heterosis for most phenotypic traits in high-polyploids than in diploids or low-polyploids (Lippman and Zamir, 2007; Bansal et al., 2012). Some of these trait differences, like the low number of spikelets per inflorescence and short lemmas in low ploidy individuals were also observed in diploid annuals species (*B. distachyon*) when compared them to allotetraploid *B. hybridum* individuals (Catalán et al., 2012, 2016b). Within the *B. phoenicoides* polyploid complex three of the five diagnostic characters, which correspond to reproductive traits, also had

significantly lower values in allotetraploids than in allohexaploids, with the former presenting shorter lower and upper glumes and a smaller number of anthers than the later (**Table 1, Figure 2**). These differences could be also due to a higher genic dosage of the *B. phoenicoides* allohexaploids, which present an extra duplicated subgenome ( $x=5$  E2) with respect to the two subgenomes present in both cytotypes ( $x=5$  E2,  $x=9$  G) (Sancho et al., 2022; Decena et al., 2023). The upper glume length was also a discriminant character for the separation of the annual diploids *B. distachyon* and *B. stacei* and the allotetraploid *B. hybridum* (Catalán et al., 2012b). Although the number of anthers has been traditionally considered a fixed trait in grasses, which generally exhibit 3 anthers across subfamilies (Kellogg, 2015), analysis of this trait within the *Brachypodium* annual species detected a reduction to 2 anthers in the diploid and cleistogamous *B. distachyon* compared to the 3 anthers of chasmogamous allotetraploid *B. hybridum* (Dinh Thi et al., 2016). Despite *B. phoenicoides* is an allogamous species, the reduction in the number of anthers of some allotetraploids is intriguing (**Table 1, Supplementary Table S3B, Appendix S2**) and should be assessed within a larger sampling of the species. Unexpectedly, two of the diagnostic characters that correspond to vegetative traits, length of the innovation leaf and number of culm nodes, showed significantly higher values in *B. phoenicoides* allotetraploids than in allohexaploids (**Table 1, Figure 2**). It could be a consequence of environmental constraints or other factors influencing the inherent plasticity of the phenotypic traits (Schneider, 2022) rather than genic dosage, or to the negative association between fixed heterosis and classical heterosis (Bansal et al., 2012), and would require a more thoroughly assessment with a larger representative sampling.

In contrast to the results obtained from phenotypic multivariate analyses of the annual *Brachypodium* species where the PCA plots clearly separated the clusters of diploid and allotetraploid samples (Catalán et al., 2012; López-Álvarez et al., 2017), the multivariate PCA explorations of intraspecific phenotypic variation within the *B. retusum* and *B. phoenicoides* polyploid complexes using the 45 quantitative morphological and anatomical traits did not detect a clear cut differentiation of tetraploid and hexaploid groups in any of the two species but rather a large overlapping in their respective distributions in the 2D and 3D morpho-spaces (**Figure 3**). However, these outcomes should not be unexpected, considering that tetraploid and hexaploid individuals from each species share two identical subgenomes (Sancho et al., 2022; Decena et al., 2023) with similar expressions and regulations pathways in the two cytotypes. Moreover, the

*B. phoenicoides* allohexaploids have an extra duplicated copy of the same E2 ( $x=5$ ) subgenome present in the allotetraploids, which theoretically would only increase the fixed heterosis for the encoded traits. Similar overlapping PCA clustering patterns have been found in other angiosperm polyploid complexes (Chansler et al., 2016), supporting the phenotypic closeness of genomically similar polyploids. More explicit structuring was retrieved from the multivariate FAMD searches where qualitative and quantitative morphoanatomical traits, epidermal characters and category-group traits were included in the analysis (**Supplementary Table S4, Figure 4**). Although the PCAs of quantitative data also detected a large overlapping for the *B. retusum* and a less overlap for *B. phoenicoides* tetraploid and hexaploid groups in their respective 2D plots (**Figures 4A, 4C**), the hierarchical clusterings of qualitative data and the 2D factor maps clearly distinguished three clusters of *B. retusum* samples (**Figure 4B**) and three separate clusters (4 subclusters) of *B. phoenicoides* samples (**Figure D**) with groups related not only with ploidy level but also with geographic range and phyletic lineage (Decena et al., 2023). Surprisingly, the segregation of the *B. phoenicoides* clusters, and therefore their corresponding phenotypic groups, was higher than that of the *B. retusum* clusters in all FAMD analyses (Figures 4A, 4B, 4C, 4D), although the subgenomic divergence and disparity is more pronounced between the 4x and 6x *B. retusum* samples (A2, E1, +E2 subgenomes) than between the *B. phoenicoides* samples (G, E2, +E2 subgenomes) (Sancho et al., 2022; Decena et al., 2023). It should be considered, however, that there is no direct association between genetic diversity and phenotypic diversity or heterosis across ploidy; specific combinators at diploid level do not necessarily produce more diverse or larger additive combinators at increasing allopolyploid levels (Bansal et al., 2012). The three FAMD *B. retusum* phenotypic clusters separated only an hexaploid population from S Spain with large leaves and small florets, from tetraploid populations from N Spain with scabrid lemma and glumes and long awns, and the remaining W Mediterranean tetraploid and hexaploid populations with glabrous lemma and glumes and bifid ligule (**Figure 4B**). More exhaustive phenotypic analyses would be required to determine if these features are genetically fixed or caused by ecological adaptation. In contrast, the four FAMD *B. retusum* phenotypic (sub)clusters circumscribed cytotypic and geographic groups and separated hexaploid populations from Morocco and S Spain with large epidermal cells from tetraploid population from N Spain with short floral organs, hexaploid populations from S Spain with large inflorescences, and tetraploid (plus 1 6x) populations from S France with also large panicles (**Figure 4D**). These phenotypic and

cyto-geographic groupings matched the evolutionary relationships recovered for some of these lineages (Decena et al., 2023), thus suggesting the potential existence of phylogenetic-driven microspeciation processes for the cryptic morphogroups within *B. phoenicoides*.

### **Gene flow and taxonomic differentiation of cryptic taxa within the *Brachypodium retusum* and *B. phoenicoides* polyploid complexes**

The potential taxonomic boundaries of the allotetraploid and allohexaploid cytotypic groups of the *B. rupestre* and *B. phoenicoides* complexes could often blur as a consequence of the presumably existence of frequent gene flow between these phylogenetically close lineages (Decena et al., 2023). Khan and Stace (1999) demonstrated that *B. retusum* and *B. phoenicoides* were outbreeders that could hybridize with other species of the core perennial clade, like *B. pinnatum* and *B. sylvaticum*, giving fertile F1 descendants. Noticeably, these authors also showed that some interspecific hybrids were more easily produced and were more fertile than certain intraspecific hybrids involving parents of different chromosome number. Their artificial hybridizations corroborated that bidirectional crosses of tetraploid *B. phoenicoides* and tetraploid *B. pinnatum* were as successful as any of their respective intraspecific crosses, and were even more fertile (>69%) than crosses of diploid and tetraploid individuals of *B. pinnatum*, while those of tetraploid *B. retusum* and tetraploid *B. pinnatum* were also successful but produced lower percentages of fertile descendants (61%) (Khan and Stace, 1999). The higher success of the intercrossing of *B. phoenicoides* with *B. pinnatum* than that of *B. retusum* with the latter species is due to the close evolutionary relationship of *B. phoenicoides* and *B. pinnatum* (*B. pinnatum* complex) and the more divergent relationships of *B. retusum* to them, while the lower hybridization success of both *B. retusum* and *B. phoenicoides* to *B. sylvaticum* (<21%), compared to those with *B. pinnatum*, is also a consequence of the less close relationships of the two species to the isolated *B. sylvaticum* lineage (Decena et al., 2023). Apart from the experimental crosses, other studies have also indicated the existence of spontaneous interspecific *B. retusum* x *B. phoenicoides* and *B. phoenicoides* x *B. pinnatum* hybrids in nature (Schippmann, 1991).

The occurrence of distinct allopolyploidization processes involved in the different origins of the respective allotetraploid and allohexaploid cytotypes of *B. retusum* and



*B. phoenicoides* (Sancho et al., 2022; Decena et al., 2023) would partially explain the statistically significant morphoanatomical differences observed among them in our phenotypic study (**Table 1; Figure 2**). Nonetheless, the genetic outcomes from each of these allopolyploidizations may vary, as hybridization and genome doubling is a recurrent process in *Brachypodium* (e. g., *B. hybridum*, (Gordon et al., 2020; Shiposha et al., 2020; Mu et al. 2023), and the genotypic and phenotypic features of each outcome depend on the constitution of the parents involved in each cross (Dinh Thi et al., 2016). Moreover, the likely existence of extensive gene flow among these *B. retusum* and *B. phoenicoides* cytotypes and other species of the *Brachypodium* perennial core, would counterbalance their genetic isolation, impeding the fixation of phenotypic traits. This would explain the overlapping distribution of 4x and 6x cytotypic samples in the morphoanatomical PCA spaces (**Figures 3 and 4**), and the difficulties to properly classify the cytotypes as separate crypto-taxa based solely on morphoanatomical traits. A more thoroughly study of morphological, anatomical and epidermal characters in all the representative samples would be necessary to elucidate the cryptic diversity in these taxonomically complex groups.

## DATA AVAILABILITY STATEMENT

Input and output data, and Supplementary information are available at Github ([https://github.com/Bioflora/BretusumBphoenicoides\\_Morphology](https://github.com/Bioflora/BretusumBphoenicoides_Morphology)).

## AUTHOR CONTRIBUTIONS

PC designed the study. EP, LAI and PC collected samples. MD and LAI developed the experimental work. MD and PC analyzed the data and interpreted the results. PC and MD prepared the manuscript. All authors revised the manuscript. The authors declare no conflict of interest.

## FUNDING

This study was funded by the Spanish Ministry of Science and Innovation PID2019-108195GB-I00, TED2021-131073B-I00 and PDC2022-133712-I00, and by the Spanish Aragon Government LMP82-21 and Bioflora A01-23R research grants. MD was

supported by a Spanish Ministry of Science predoctoral FPI fellowship. RS was supported by a Bioflora postdoctoral contract.

## ACKNOWLEDGEMENTS

We thank the managers of the Spanish, Moroccan and French National Parks and Autonomous Communities or Regional protected areas for the permissions to collect the studied *Brachypodium retusum* and *B. phoenicoides* samples. The morphoanatomical analyses were performed in the Bioflora laboratory of the Escuela Politécnica Superior de Huesca (EPSHU, Universidad de Zaragoza, Spain).

## SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at [https://github.com/Bioflora/BretusumBphoenicoides\\_Morphology](https://github.com/Bioflora/BretusumBphoenicoides_Morphology)

## References

- Bansal, P., Banga, S., and Banga, S. S. (2012). Heterosis as Investigated in Terms of Polyploidy and Genetic Diversity Using Designed *Brassica juncea* Amphiploid and Its Progenitor Diploid Species. *PLoS One* 7, e29607. doi:10.1371/journal.pone.0029607.
- Catalán, P., Ángeles Decena, M., Sancho, R., Viruel, J., Pérez-Collazos, E., Inda, L. A., et al. (2023). Phylogenetics of the Palearctic model grass *Brachypodium sylvaticum* uncovers two divergent oriental and occidental micro-taxa lineages. *Bot. Pacifica* 12. doi:10.17581/bp.2023.12119.
- Catalán, P., López-Álvarez, D., Bellosta, C., and Villar, L. (2016a). Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei*, and *B. hybridum* (Poaceae). *An. del Jard. Bot. Madrid* 73, 1–14. doi:10.3989/ajbm.2428.
- Catalan, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., and López-Herránz, M. L. (2016). Phylogeny and Evolution of the Genus *Brachypodium*. 9–38. doi:10.1007/7397\_2015\_17.
- Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L. A. J., Langdon, T., et al. (2012). Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann. Bot.* 109, 385–405. doi:10.1093/aob/mcr294.
- Chansler, M. T., Ferguson, C. J., Fehlberg, S. D., and Prather, L. A. (2016). The role of polyploidy in shaping morphological diversity in natural populations of *Phlox amabilis*. *Am. J. Bot.* 103, 1546–1558. doi:10.3732/ajb.1600183.
- Decena, M. A., Sancho, R., Lusinska, J., Hasterok, R., Gorgojo, R., Montes, B., et al. (2023). Cyto-phylogenomics of the grass model genus *Brachypodium* unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends within this largely reticulate diploid-polyploid species-complex. *Front. Plant Sci.*
- Devesa Juan, Ruiz-Tellez Trinidad, Viera-Benitez Carmen, Ortega-Olivencia Ana, Tormo Rafael, C. J. (1991). Las gramíneas de Extremadura.
- Díaz-Pérez, A., López-Álvarez, D., Sancho, R., and Catalán, P. (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. *Mol. Phylogenet. Evol.* 127, 256–271. doi:10.1016/j.ympev.2018.06.003.
- Dinh Thi, V. H., Coriton, O., Le Clainche, I., Arnaud, D., Gordon, S. P., Linc, G., et al. (2016). Recreating Stable *Brachypodium hybridum* Allotetraploids by Uniting the Divergent Genomes of *B. distachyon* and *B. stacei*. *PLoS One* 11, e0167171. doi:10.1371/journal.pone.0167171.
- Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., et al. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.* 11, 1–16. doi:10.1038/s41467-020-17302-5.
- Gustafsson, A. L. S., Skrede, I., Rowe, H. C., Gussarova, G., Borgen, L., Rieseberg, L. H., et al. (2014). Genetics of Cryptic Speciation within an Arctic Mustard, *Draba nivalis*. *PLoS One* 9, e93834. doi:10.1371/journal.pone.0093834.
- Harlan, J. R., and de Wet, J. M. J. (1963). THE COMPILOSPECIES CONCEPT. *Evolution* (N. Y). 17, 497–501. doi:10.1111/j.1558-5646.1963.tb03307.x.
- Harmon, L. J., Schulte, J. A., Larson, A., and Losos, J. B. (2003). Tempo and Mode of Evolutionary Radiation in Iguanian Lizards. *Science* (80-. ). 301, 961–964. doi:10.1126/science.1084786.
- Hasterok, R., Catalan, P., Hazen, S. P., Roulin, A. C., Vogel, J. P., Wang, K., et al. (2022). *Brachypodium*: 20 years as a grass biology model system; the way forward? *Plant Sci.* 2022. doi:10.1016/j.tplants.2022.04.008.
- Holstein, N., and Luebert, F. (2017). Taxonomy: stable taxon boundaries. *Nature* 548, 158–158. doi:10.1038/548158d.
- IBI (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi:10.1038/nature08747.
- Janzen, D. H., Burns, J. M., Cong, Q., Hallwachs, W., Dapkey, T., Manjunath, R., et al. (2017). Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proc. Natl. Acad. Sci.* 114, 8313–8318. doi:10.1073/pnas.1621504114.
- Jörger, K. M., and Schrödl, M. (2013). How to describe a cryptic species? Practical

- challenges of molecular taxonomy. *Front. Zool.* 10, 59. doi:10.1186/1742-9994-10-59.
- Josse, J., and Husson, F. (2016). missMDA : A Package for Handling Missing Values in Multivariate Data Analysis. *J. Stat. Softw.* 70. doi:10.18637/jss.v070.i01.
- Kellogg, E. A. (2015). *The Families and Genera of Vascular Plants. Vol. XIII. Flowering Plants. Monocots. Poaceae.*, ed. K. Kubitzki New York: Springer.
- Khan, M. A., and Stace, C. A. (1999a). Breeding relationships i n the genus *Brachypodium* ( Poaceae :
- Lê, S., Josse, J., and Husson, F. (2008). FactoMineR : An R Package for Multivariate Analysis. *J. Stat. Softw.* 25. doi:10.18637/jss.v025.i01.
- Lippman, Z. B., and Zamir, D. (2007). Heterosis: revisiting the magic. *Trends Genet.* 23, 60–66. doi:10.1016/j.tig.2006.12.006.
- López-Alvarez, D., López-Herranz, M. L., Betekhtin, A., and 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. doi:10.1371/journal.pone.0051058.
- López-Álvarez, D., Zubair, H., Beckmann, M., Draper, J., and Catalán, P. (2017a). Diversity and association of phenotypic and metabolomic traits in the close model grasses *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Ann. Bot.* 119, 545–561. doi:10.1093/aob/mcw239.
- Loxdale, H. D., Davis, B. J., and Davis, R. A. (2016). Known knowns and unknowns in biology. *Biol. J. Linn. Soc.* 117, 386–398. doi:10.1111/bij.12646.
- Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., et al. (2019). Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *Int. J. Mol. Sci.* 20, 1–19. doi:10.3390/ijms20225557.
- Maxwell, S. J., Rymer, T. L., Rowell, M. K., Hernandez Duran, L. C., Berschauer, D. P., Underdown, M., et al. (2021). Defining and Bringing Relevance of Meaning to Species Group-Level Taxa. *Proc. Biol. Soc. Washingt.* 134. doi:10.2988/006-324X-134.1.27-28.

- Mayr E (1942). *Systematic and the origin of species*. New York, NY, USA: Columbia University Press.
- Metcalfe, C. R. (1960). *Anatomy of the Monocotyledons*, Volume 1.
- Mu W, Li K, Yang Y, Breiman A, Yang J, Wu J, Zhu M, Wang S, Catalán P, Nevo E, L. J. (2023). Subgenomic stability of progenitor genomes during repeated allotetraploid origins of the same grass *Brachypodium hybridum*. *bioRxiv*.
- Pagès, J. (2004). Analyse Factorielle de Données Mixtes. *Rev. Stat. Appliquée* 52, 93–111.
- Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., et al. (2015). Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Mol. Ecol.* 24, 525–544. doi:10.1111/mec.13048.
- Patil, I. (2021). Visualizations with statistical details: The “ggstatsplot” approach. *J. Open Source Softw.* 6, 3167. doi:10.21105/joss.03167.
- Sancho, R., Cantalapiedra, C. P., López-Alvarez, D., Gordon, S. P., Vogel, J. P., Catalán, P., et al. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* 218, 1631–1644. doi:10.1111/nph.14926.
- Sancho, R., Inda, L. A., Díaz-Pérez, A., Des Marais, D. L., Gordon, S., Vogel, J. P., et al. (2022a). Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J.* 109, 1535–1558. doi:10.1111/tpj.15650.
- Scarlett, V. T., Lovell, J. T., Shao, M., Phillips, J., Shu, S., Goodstein, D. M., et al. (2022). Multiple origins , one evolutionary trajectory : gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*.
- Schippmann (1991a). Revision der euroäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45.
- Schneider, H. M. (2022). Characterization, costs, cues and future perspectives of phenotypic plasticity. *Ann. Bot.* 130, 131–148. doi:10.1093/aob/mcac087.
- Scholthof, K. B. G., Irigoyen, S., Catalan, P., and Mandadi, K. K. (2018). *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30, 1673–1694.

doi:10.1105/tpc.18.00083.

- Shiposha, V., Marques, I., López-Alvarez, D., Manzaneda, A. J., Hernandez, P., Olonova, M., et al. (2020). Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot. *Ann. Bot.* 125, 625–638. doi:10.1093/aob/mcz169.
- Stace, C. A. (1991). *Plant Taxonomy and Biosystematics*. Cambridge: Cambridge University Press.
- Stebbins, G. (1971). Chromosomal Evolution in Higher Plants. *Edward Arnold Publ. London*. doi:10.1086/407511.
- Stebbins, G. L. (1969). THE SIGNIFICANCE OF HYBRIDIZATION FOR PLANT TAXONOMY AND EVOLUTION. *Taxon* 18, 26–35. doi:10.2307/1218589.
- Stevens PF (1990). “Nomenclatural stability, taxonomic instinct, and flora writing — a recipe for disaster?,” in *The Plant Diversity of Malesia*, ed. G. R. Baas P., Kalkman K. (Dordrecht: Springer).
- Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., et al. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends Ecol. Evol.* 33, 153–163. doi:10.1016/j.tree.2017.11.007.
- Vigalondo, B., Fernández-Mazuecos, M., Vargas, P., and Sáez, L. (2015). Unmasking cryptic species: morphometric and phylogenetic analyses of the Ibero-North African *Linaria incarnata* complex. *Bot. J. Linn. Soc.* 177, 395–417. doi:10.1111/boj.12251.
- Wei, Z., Xia, Z., Shu, J., Shang, H., Maxwell, S. J., Chen, L., et al. (2021). Phylogeny and Taxonomy on Cryptic Species of Forked Ferns of Asia. *Front. Plant Sci.* 12. doi:10.3389/fpls.2021.748562.



## Article

# COMPARATIVE GENOMICS, EVOLUTION, AND DROUGHT-INDUCED EXPRESSION OF DEHYDRIN GENES IN MODEL *BRACHYPODIUM* GRASSES

Maria Angeles Decena<sup>1</sup>, Sergio Gálvez-Rojas<sup>2</sup>, Federico Agostini<sup>2,3</sup>, Rubén Sancho-Cohen<sup>1,4</sup>; Bruno Contreras-Moreira<sup>4,5†</sup>; David L. Des Marais<sup>6</sup>, Pilar Hernández Molina<sup>7‡</sup>, Pilar Catalán<sup>1,4,8\*‡</sup>

<sup>1</sup> E, Dscuela Politécnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain; mdecena@unizar.es (M.A.D.); ruben.sancho.cohen@gmail.com (R.S.)

<sup>2</sup> ETSI Informática, Universidad de Málaga, Blvr Louis Pasteur 35, 29071 Málaga, Spain; galvez@uma.es (S.G.-R.); fagostini@conicet.gov.ar (F.A.)

<sup>3</sup> Instituto de Botánica del Nordeste, UNNE-CONICET, Corrientes W3402, Argentina

<sup>4</sup> Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, 50018 Zaragoza, Spain; bcontreras@eead.csic.es

<sup>5</sup> Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas, Av. Montañana 1005, 50059 Zaragoza, Spain

<sup>6</sup> Civil and Environmental Engineering Department, Faculty of Environmental and Life Science, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139, USA; dldesmar@mit.edu

<sup>7</sup> Instituto de Agricultura Sostenible, IAS-CSIC, Menéndez Pidal Ave, 14004 Córdoba, Spain

<sup>8</sup> Departamento de Ciencias Agrarias y del Medio Natural, Tomsk State University, 36 Lenin Ave, 634050 Tomsk, Russia

\* Correspondence: phernandez@ias.csic.es (P.H.); pcatalan@unizar.es (P.C.)

† Current address: Ensembl Plants, European Bioinformatics Institute, EMBL-EBI, Hinxton CB10 1SD, UK.

‡ Both authors contribute equally.

**Abstract:** Dehydration proteins (dehydrins, DHNs) confer tolerance to water-stress deficit in plants. We performed a comparative genomics and evolutionary study of DHN genes in four model *Brachypodium* grass species. Due to limited knowledge on dehydrin expression under water deprivation stress in *Brachypodium*, we also performed a drought-induced gene expression analysis in 32 ecotypes of the genus' flagship species *B. distachyon* showing different hydric requirements. Genomic sequence analysis detected 10 types of dehydrin genes (*Bdhn*) across the *Brachypodium* species. Domain and conserved motif contents of peptides encoded by *Bdhn* genes revealed eight protein architectures. *Bdhn* genes were spread across several chromosomes. Selection analysis indicated that all the *Bdhn* genes were constrained by purifying selection. Three upstream *cis*-regulatory motifs (BES1, MYB124, ZAT) were detected in several *Bdhn* genes. Gene expression analysis demonstrated that only four *Bdhn1*-*Bdhn2*, *Bdhn3*, and *Bdhn7* genes, orthologs of wheat, barley, rice, sorghum, and maize genes, were expressed in mature leaves of *B. distachyon* and that all of them were more highly expressed in plants under drought conditions. *Brachypodium* dehydrin expression was significantly correlated with drought-response phenotypic traits (plant biomass, leaf carbon and proline contents and water use efficiency increases, and leaf water and nitrogen content decreases) being more pronounced in drought-tolerant ecotypes. Our results indicate that dehydrin type and regulation could be a key factor determining the acquisition of water-stress tolerance in grasses.

**Keywords:** *Bdhn* genes; *cis*-regulatory elements; comparative genomics; dehydrin structure; dehydrin- gene expression; duplicated genes; drought-related traits; drought-tolerant ecotypes; phylogenetics

## INTRODUCTION

Water deprivation is one of the main abiotic stresses that affect plant development and fitness [1,2]. Water deficit stress in plants is mostly caused by low soil water content but also by other abiotic stresses such as salinity and extreme temperature [3,4]—which sometimes interact [2]—affecting the survival of plants in a wide range of environmental conditions [4,5]. Plant species adapted to dry environments have developed mechanisms to protect their cells from water stress deficit. Among the several physiological and genomic regulatory mechanisms triggered by water limitation in plants, there is an almost universal response in the upregulation of dehydrins [1,4,6]. Dehydrins (DHN) belong to group 2 LEA (Late-Embryogenesis-Abundant) proteins [7], and are intrinsically disordered hydrophilic proteins that acquire structure when bound to ligands, such as membranes, acting as chaperones that impede the aggregation or inactivation of other proteins under desiccation to maintain the biological activity of the cell [8,9]. They show a high hydration capacity and can also bind large quantities of cations, retaining water in the drying cells and preventing ionic unbalance and protein denaturation. DHNs are also called RAB proteins because they are usually responsive to abscisic acid [9]. DHNs accumulate in all vegetative tissues under water stress, though with different specificities [4], as well as during seed development [8]. The dehydrin protein family is characterized by a modular sequence domain (Dehydrin) that contains three conserved segments (Y, S, K; [8,10]; Figure 1a). The dehydrin identifier K-segment motif is present in all plant DHNs except for HIRD11 (see Section 2). Perdiguerro et al. demonstrated the induction of dehydrin genes lacking the K-segment under drought stress in *Pinus* [11]. The S-segment is located between the Y and K segments and, when phosphorylated, transfers dehydrins and other nuclear localization signal (NLS) binding proteins from the cytosol to the nucleus [10,12]. The Y-segment motif is located towards the N-terminus of the protein coding region. The Y, S, and K segments are interspersed by other less conserved segments, named interpattern or  $\phi$ -segments that mostly contain small, polar, and charged amino acids [13]. Recent studies have described another conserved F-segment motif in some seed plant DHNs [11,14]. This segment, when present, has a unique copy and is located between the N-terminus of the protein coding region and the S-segment. An additional NLS-segment motif was found in some dehydrin families of maize [12]. Different combinations of conserved motifs have been used to classify the dehydrin domain into major architectures, with five of them ( $K_n$ ,  $SK_n$ ,  $K_nS$ ,  $Y_nK_n$ ,  $Y_nSK_n$ ) being common across angiosperms [10]. The dehydrin domain is occasionally fused with upstream DNAJ-X and DNA-J containing gene domains in some DHN genes [10]. DNA-J is a heat shock protein (Hsp40) that prevents the aggregation of unfolded proteins and functions as a chaperone folding protein when associated with Hsp70 under water-stress [10,15] (Figure 1a). Evolutionary studies of spermatophyte dehydrins identified  $Kn$  or  $SKn$  as the oldest ancestral architectures whereas those containing the Y segment arose only in the angiosperms [10,16]. Two main architectural changes apparently occurred after the ancestral whole genome duplication events of the angiosperms, the Y3SK2-Y3K1 and Y1SK2-K6 dehydrins [10]. Whereas K6 developed a novel function in cold protection (but see [16] for the important role of  $Kn$  dehydrins in drought protection), Y3K1 showed signatures of pseudogenization in some families (e.g., the Brassicaceae) [10]. At the infrageneric level, the dehydrin phylogeny of 11 *Oryza* species showed the splits of four high-to-weakly supported domain clades and a lack of dehydrin-specific subclades [9].

DHNs have been extensively studied in grasses due to their key role as agents conferring water-stress tolerance in cereal and forage crops [7,9,17,18]. Expression of dehydrins under drought stress has been positively associated with plant biomass and grain yield [19].

Several in vitro studies have demonstrated that DHN expression enhances plant stress tolerance [20,21]. Dehydrins maintain the osmotic balance of cells and their chlorophyll contents, bind metals to scavenge ROS, and bind to DNA and phospholipids [20,22]. Despite these advances, some of the biological functions of DHNs have not been fully established yet [10]. Few dehydrins have been characterized biochemically, like those involved in enzyme and membrane bioprotection in *Triticum* [23] and *Zea* [24] and in protection against ROS in *Arabidopsis* and *Brassica* [25,26]. Their protective activities have been experimentally demonstrated in some cases using circular dichroism (e.g., *Zea mays* DHN1 binding to lipid vesicles, [27]; *Hordeum vulgare* Dhn5 protection of lactate deshydrogenase under drought and cold conditions, [28]).

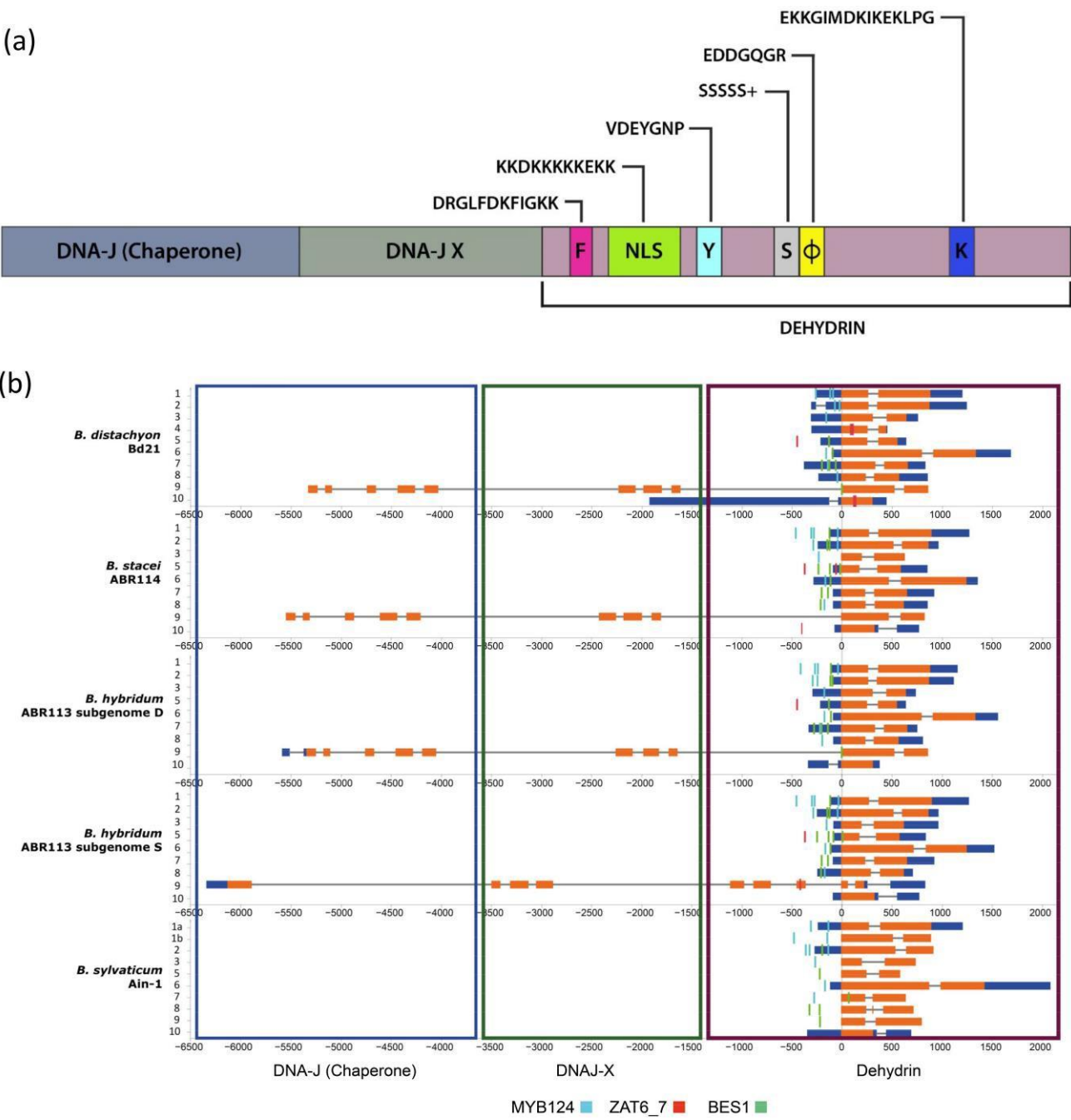
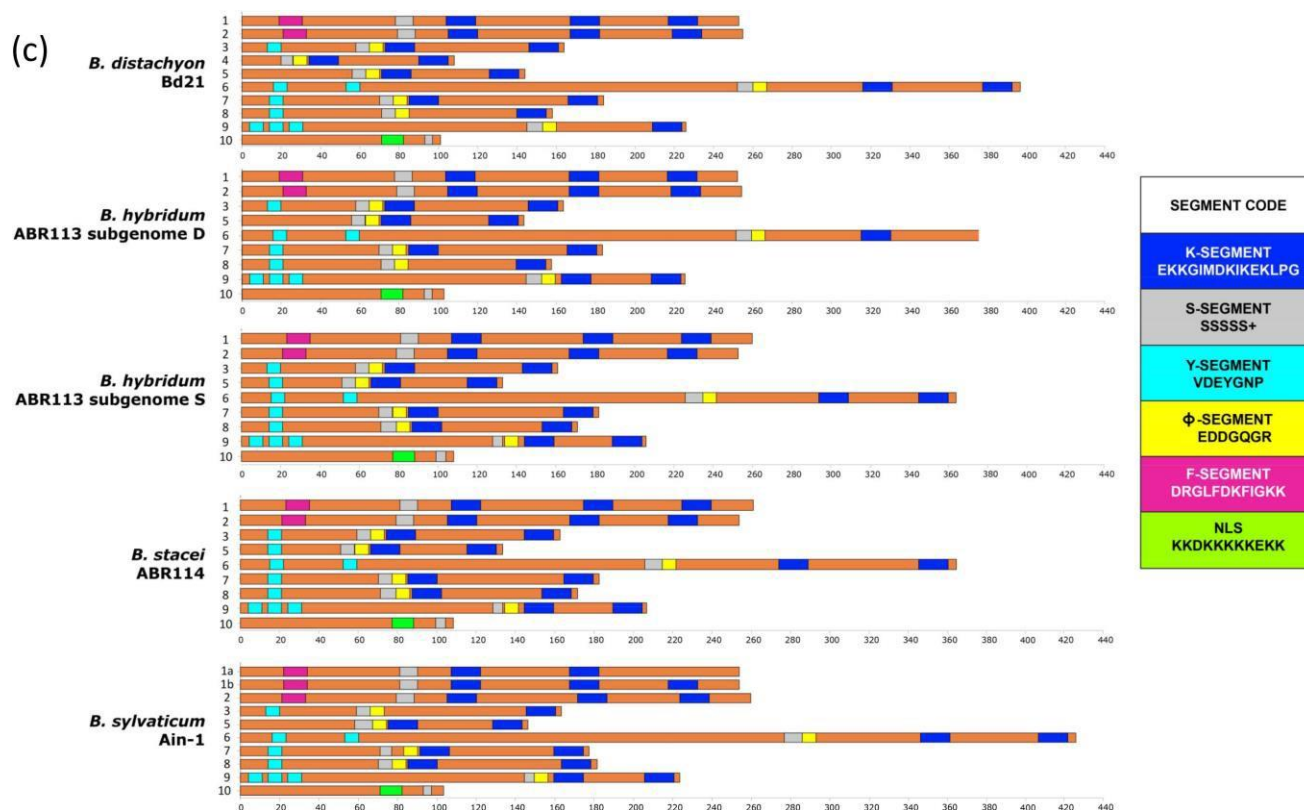


Figure 1. Cont.



**Figure 1.** Schematic structure of a typical grass DHN gene showing the positions of the main domains and coding segments (a). Gene structure of *Brachypodium distachyon* Bd21, *B. stacei* ABR114, *B. hybridum* ABR113 (D and S subgenomes), and *B. sylvaticum* Ain-1 dehydrins showing: (b) the three detected DNA J (purple rectangle), DNA J-X (light green rectangle) and dehydrin (dark green rectangle) domains, and their CDSs (orange bars), introns (grey lines), and 5' and 3' UTRs (blue bars); (c) the conserved motifs of their CDSs with their respective K (dark blue), S (grey), Y (light blue),  $\Phi$  (yellow), F (pink), and NLS (light green) segments. *Brachypodium* dehydrin gene codes (*Bdhn1*–*Bdhn10*) correspond to those indicated in Table 1. *Cis*-regulatory elements BES1, MYB124, and ZAT are mapped in the promoters of each gene (see color codes in the chart; the figure could be also visualized in [http://zeta.uma.es/public/journal/brachy/DHN\\_Brachy\\_4\\_varieties.html](http://zeta.uma.es/public/journal/brachy/DHN_Brachy_4_varieties.html), 29 November 2021).

**Table 1.** Dehydrin genes found in the studied *Brachypodium distachyon*, *B. stacei*, *B. hybridum* (subgenomes D and S), and

*B. sylvaticum* species. *Brachypodium* dehydrin gene codes (*Bdhn1*–*Bdhn10*) given in this study, the protein structure and their corresponding Panther family and subfamilies gene codes. Phytozome dehydrin gene codes correspond to the respective gene numbers in the reference genome of each species deposited in Phytozome.

<i>Bdhn</i>	Structure	Panther	Panther Subfamily	Phytozome Dehydrin Gene Codes				
				<i>B. distachyon</i> (Bd21)	<i>B. stacei</i> (ABR114)	<i>B. hybridum</i> D (ABR113)	<i>B. hybridum</i> S (ABR113)	<i>B. sylvaticum</i> (Ain1)
<i>Bdhn1a</i>	FSK <sub>n</sub>		ERD14	Bradi5g10860	Brast09G089800	Brahy.D05G0138300	Brahy.S09G0091400	Brasy9G143400
<i>Bdhn1b</i>								Brasy9G149400
<i>Bdhn2</i>	FSK <sub>n</sub>	PTH33346	SF23	Bradi3g51200	Brast04G110500	Brahy.D03G0707100	Brahy.S04G0117200	Brasy4G117200
<i>Bdhn3</i>	Y <sub>n</sub> SK <sub>n</sub>			Bradi1g37410	Brast07G152200	Brahy.D01G0507100	Brahy.S07G0169500	Brasy7G144000
<i>Bdhn4</i>	Y <sub>n</sub> SK <sub>n</sub>			Bradi4g22280				

<i>Bdhn5</i>	Y <sub>n</sub> SK <sub>n</sub>			Bradi4g22290	Brast05G049400	Brahy.D04G0319400	Brahy.S05G0054400	Brasy5G271500
<i>Bdhn6</i>	Y <sub>n</sub> SK <sub>n</sub>	SF 19		Bradi4g19525	Brast05G075400	Brahy.D04G0277700	Brahy.S05G0083200	Brasy4G237500
<i>Bdhn7</i>	Y <sub>n</sub> SK <sub>n</sub>			Bradi3g43870	Brast04G194300	Brahy.D03G0604200	Brahy.S04G0208500	Brasy4G220100
<i>Bdhn8</i>	Y <sub>n</sub> SK <sub>n</sub>			Bradi3g43855	Brast04G197200	Brahy.D03G0604000	Brahy.S04G0208900	Brasy4G219800 <i>Bdhn9</i>
	Y <sub>n</sub> SK <sub>n</sub>	XERO I		Bradi2g47575	Brast01G171900	Brahy.D02G0637300	Brahy.S01G0182800	Brasy1G228900 <i>Bdhn10</i>
	K*NLSL-S	PTHR34941	HIRD 11	Bradi1g13330	Brast02G251900	Brahy.D01G017200	Brahy.S02G0268100	Brasy2G277500

*Brachypodium* is a model system for grasses due to its intermediate evolutionary position between the temperate cereals (Triticeae) and the tropical biofuel (*Miscanthus*, *Paspalum*) crops [29]. The three annual species of the genus have been selected as a model complex for polyploidy (diploids *B. distachyon* and *B. stacei* and derived allotetraploid *B. hybridum*; [30]) and one of its perennial species has been selected as a model species for perenniality (diploid *B. sylvaticum*; [31]). Transgenic plants of its flagship species *B. distachyon* have been analyzed to identify candidate genes that enhance drought stress tolerance in plants [32–34] though none of them specifically addressed dehydrins. Inspection of an early version of the *B. distachyon* reference genome Bd21 detected 36 LEA2 encoding genes but did not characterize the dehydrin genes [35]. The dehydrin gene content, structure, evolution, and expression in response to drought among species and accessions of *Brachypodium* has not been investigated yet. Given the significance of DHNs in water stress response of grasses, we analyzed the members of the dehydrin gene family in the four *Brachypodium* model species and in 54 *B. distachyon* ecotypes showing different hydric requirements and drought tolerances using in silico analysis of genome annotations. Due to the lack of studies on dehydrin expression under water deprivation stress in *Brachypodium* we also performed DHN expression analysis in 32 *B. distachyon* ecotypes under different drought and watered conditions. The aims of our study were: (i) to identify and characterize the *Brachypodium* dehydrin genes (*Bdhn*) and the structure and biochemical properties of the encoded proteins, comparing them with those of the closely related cereal crops (*Hordeum*, *Oryza*, *Sorghum*, *Triticum-Aegilops*, *Zea*), and to evaluate the presence of enriched sequence motifs that may have potential regulatory effects (ii) to analyze the syntenic distributions and origins of the *Bdhn* genes, identify gene duplication events, and test the functionality of paralogs, (iii) to investigate their expression profiles under control vs. dry conditions and compare them to those of closely related cereals, and (iv) to correlate the dehydrin gene expressions with the phenotypic and environmental traits of the plants and test their potential phylogenetic signal.

## Results

### Dehydrin Genes of *Brachypodium* Species and Outgroup Grasses

Genome searches in Phytozome and Ensembl Plants retrieved 47 dehydrin gene sequences collected from the reference genomes of the four sequenced *Brachypodium* species [*B. distachyon* Bd21 2n = 2x = 10, x = 5 (10); *B. sylvaticum* Ain-1 2n = 2x = 18, x = 9 (10); *B. stacei* ABR114 2n = 2x = 20, x = 10 (9); *B. hybridum* ABR113 2n = 4x = 30, x = 5 + 10 (18; 9 from its *B. distachyon*-type D subgenome and 9 from its *B. stacei*-type S subgenome)] (Table 1; Figure 1b). A total of 54 orthologous DHN sequences were retrieved from the reference genomes of six outgroup grass species [*Aegilops tauschii* 2n = 2x = 14, x = 7 (9); *Hordeum vulgare* 2n = 2x = 14, x = 7 (8); *Zea mays* 2n = 2x = 20, x = 10 (7); *Oryza sativa* 2n = 2x = 24, x = 12 (6); *Sorghum bicolor* 2n = 2x = 20, x = 10 (5); *Triticum aestivum* 2n = 6x = 42, x = 7 (19); Supplementary Table S1]. A new nomenclature was created for the *Brachypodium* dehydrin genes (*Bdhn1* to *Bdhn10*; Table 1; Figure 1b,c). In several instances we used the same numbers as those of the orthologous *Hordeum vulgare* DHN genes (*Bdhn* 6–7, and *Bdhn* 9–10) (ENA database: AF043086, AF043092; AF043086 and Genbank database: AY681974) and the orthologous *Oryza sativa* DHN genes (*Bdhn1*–2 and *Bdhn8*) (RAP database: Os02g0669100, Os11g0454300). *Bdhn3* was numbered according



to prior annotation in *Brachypodium* (GeneBank: XM\_010229280), whereas the remaining *Bdhn* genes were numbered consecutively as *Bdhn4* and *Bdhn5*. The *Bdhn* genes were also classified according to Panther protein gene families (PTH33346 and PTH34941) and subfamilies (ERD14, XERO1, SF14, SF19, SF23, HIRD11; Table 1). All *Bdhn10* genes found in the studied species of *Brachypodium* lacked the K-segment (Figure 1c). However, those dehydrins showed an extraordinary sequence identity with typical DHNs, including those with a modified K-segment, like DHN-13 from *H. vulgare* [36]. The absence of the K-segment has been also reported in one dehydrin of pine species [11] and four dehydrins of rice species (OnDHN6, OrDHN7, OldHN3, OldHN6; [9]). These *Brachypodium Bdhn10* dehydrins, with architecture K\*(NLS)S, belong to HIRD11 proteins and show orthology with DHN genes from *H. vulgare*, *O. sativa*, *S. bicolor*, and *Z. mays* (Supplementary Table S1).


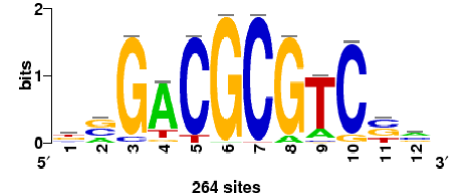
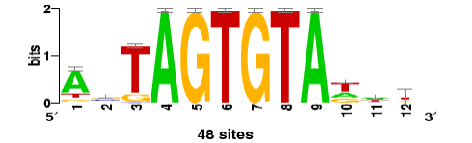
The *Brachypodium* dehydrins showed different lengths, ranging from 86.3 (*Bdhn10*) to 323 amino acid residues (*Bdhn6*), molecular weights from 9762.90 to 30945.54 kDa, and isoelectric points from 4.41 (*Bdhn2*) to 7.6 (*Bdhn7*) (9.4 (*Bdhn4*) in *B. distachyon*) (Supplementary Table S2). All dehydrins presented a negative GRAVY value, indicating that they are hydrophilic proteins. All *Bdhn* genes except *Bdhn5* were structurally conserved across the four *Brachypodium* species (Figure 1b,c). *Bdhn4* was only present in *B. distachyon*, while *Bdhn1* was duplicated in *B. sylvaticum* (*Bdhn1a*, *Bdhn1b*) (Table 1; Figure 1b,c). All *Bdhn* genes but *Bdhn9* showed a single dehydrin domain. *Bdhn9* encoded a protein of 508 aa with DHN–DNAJ–X–DNAJ domains in all species except in *B. sylvaticum*, which showed a gene consisting only of the DHN domain (222 aa) (Supplementary Table S2; Figure 1b). Eight gene architectures were found along the *Bdhn1–Bdhn10* genes (FSK<sub>2</sub>, FSK<sub>3</sub>, S $\phi$ K<sub>2</sub>, YS $\phi$ K<sub>2</sub>, YS $\phi$ K, Y<sub>3</sub>S $\phi$ K, Y<sub>3</sub>S $\phi$ K<sub>2</sub>, NLS-K\*S), with YS $\phi$ K<sub>2</sub> being the most common architecture present in four genes (*Bdhn3*, *Bdhn6*, *Bdhn7*, *Bdhn8*; Table 1, Figure 1c). *Bdhn* 3D modeling indicated that the disordered *Brachypodium* dehydrins lacked tertiary structure whereas their secondary structure consisted of different numbers of  $\alpha$ -helices (1–4) and  $\beta$ -sheets in some inferred proteins (Supplementary Figure S1). Orthology analysis indicates that *Bdhn* genes were also present in the other surveyed grasses. In total, 6 out of 10 *Bdhn* genes were found in *A. tauschii*; *Bdhn1–2* copies were also present in *H. vulgare* and *T. aestivum*, *Bdhn3* in *T. aestivum*, *Bdhn4–5* in *H. vulgare*, *Bdhn7–8* in *T. aestivum*, *Bdhn9* in *O. sativa* and *S. bicolor*, and *Bdhn10* in *H. vulgare*, *O. sativa*, *S. bicolor*, and *Z. mays* (Supplementary Table S1). By contrast, five DHN genes of *O. sativa* (Os01g0702500, Os11g0453900, Os01g0624400, Os01g0225600, Os03g0655400), four of *H. vulgare* (HORVU6Hr1G083960, HORVU6Hr1G011050, HORVU5Hr1G103460, HORVU3Hr1G089300), and three of *Z. mays* (Zm00001d017547, Zm00001d043730, Zm00001d013647) had no orthologous sequences in *Brachypodium*. Pairwise amino acid sequence similarities indicated that *Bdhn4* and *Bdhn5* were the most similar proteins, followed by *Bdhn1* and *Bdhn2*. *Bdhn10* was the most dissimilar dehydrin. Dehydrins with YS $\phi$ K<sub>2</sub> structure were in general similar to each other both in *Brachypodium* and in the other grasses.



### Cis-Regulatory Elements of Bdhc Genes

We performed de novo discovery analysis of cis-regulatory elements (CREs) of Bdhc genes, searching for binding sites of transcription factors (TF) that accumulate around the transcriptional start site and that may control gene expression (Supplementary Figure S2). The analysis consistently identified three clusters, BES1/BZR1, MYB124, and ZAT binding sites (Table 2; Supplementary Figure S2a) that are related with different drought-response signaling pathways. BES1/BZR1 and MYB124 motifs were present in all studied promoters, though MYB124 was predominant in the promoters of Bdhc1 and Bdhc2 and BES1/BZR1 in those of Bdhc7 genes and more abundantly in those of the aridic *B. stacei* and *B. hybridum* species. By contrast, ZAT was only found in the promoters of Bdhc4, Bdhc5, and Bdhc10 genes in annual species (Table 2; Supplementary Figure S2a).

**Table 2.** Upstream putative *cis*-regulatory elements (CREs) found in the promoter region (−500-to+200 bp) of the *Bdhn* genes of *Brachypodium distachyon*, *B. stacei*, *B. hybridum* (subgenomes D and S), and *B. sylvaticum* using Rsat::plants tools and the corresponding reference genome as background. Family identification, motif code (ID), N-cor (normalized correlation), and Sig (significance value) for the highest hit, matches to transcription factor (TF) binding sites, and *Bdhn* genes with the number of sites found within each species and gene. Species and reference genomes: BD, *B. distachyon* Bd21; BS, *B. stacei* ABR114; BHD, *B. hybridum* subgenome D ABR113; BHS, *B. hybridum* subgenome S ABR113; BS, *B. sylvaticum* Ain1. Mapping positions of these *cis*-regulatory motifs are indicated in Figure 1b.

Family ID	Motif ID	Ncor	Sig	TF Binding Site	<i>Bdhn</i> Genes (No. Sites Found)
BRI1-EMS suppressor/brassinazole-resistant	BES1/BZR1	0.719	4.08		<b>BD:</b> <i>Bdhn</i> 5(1) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(3) <i>Bdhn</i> 9(1) <b>BHD:</b> <i>Bdhn</i> 1(1) <i>Bdhn</i> 2(2) <i>Bdhn</i> 5(1) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(3) <i>Bdhn</i> 9(1) <b>BHS:</b> <i>Bdhn</i> 1(1) <i>Bdhn</i> 2(2) <i>Bdhn</i> 5(4) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(2) <i>Bdhn</i> 8(1) <b>BS:</b> <i>Bdhn</i> 1(1) <i>Bdhn</i> 2(2) <i>Bdhn</i> 5(3) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(2) <i>Bdhn</i> 8(1) <b>BSY:</b> <i>Bdhn</i> 2(1) <i>Bdhn</i> 5(1) <i>Bdhn</i> 7(1) <i>Bdhn</i> 8(2) <i>Bdhn</i> 9(1)
Myb	MYB124	0.598	2.18		<b>BD:</b> <i>Bdhn</i> 1(5) <i>Bdhn</i> 2(4) <i>Bdhn</i> 3(2) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(2) <i>Bdhn</i> 8(2) <b>BHD:</b> <i>Bdhn</i> 1(7) <i>Bdhn</i> 2(4) <i>Bdhn</i> 3(2) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(2) <i>Bdhn</i> 8(2) <b>BHS:</b> <i>Bdhn</i> 1(7) <i>Bdhn</i> 2(4) <i>Bdhn</i> 3(2) <i>Bdhn</i> 6(1) <i>Bdhn</i> 8(2) <b>BS:</b> <i>Bdhn</i> 1(8) <i>Bdhn</i> 2(4) <i>Bdhn</i> 3(2) <i>Bdhn</i> 6(1) <i>Bdhn</i> 8(2) <b>BSY:</b> <i>Bdhn</i> 1a(4) <i>Bdhn</i> 1b(4) <i>Bdhn</i> 2(4) <i>Bdhn</i> 3(2) <i>Bdhn</i> 6(1) <i>Bdhn</i> 7(2)
C2H2 zinc finger	ZAT	0.738	4.91		<b>BD:</b> <i>Bdhn</i> 4(3) <i>Bdhn</i> 5(1) <i>Bdhn</i> 10(2) <b>BHD:</b> <i>Bdhn</i> 5(1) <b>BHS:</b> <i>Bdhn</i> 5(1) <i>Bdhn</i> 10(1) <b>BS:</b> <i>Bdhn</i> 5(2) <i>Bdhn</i> 10(1)

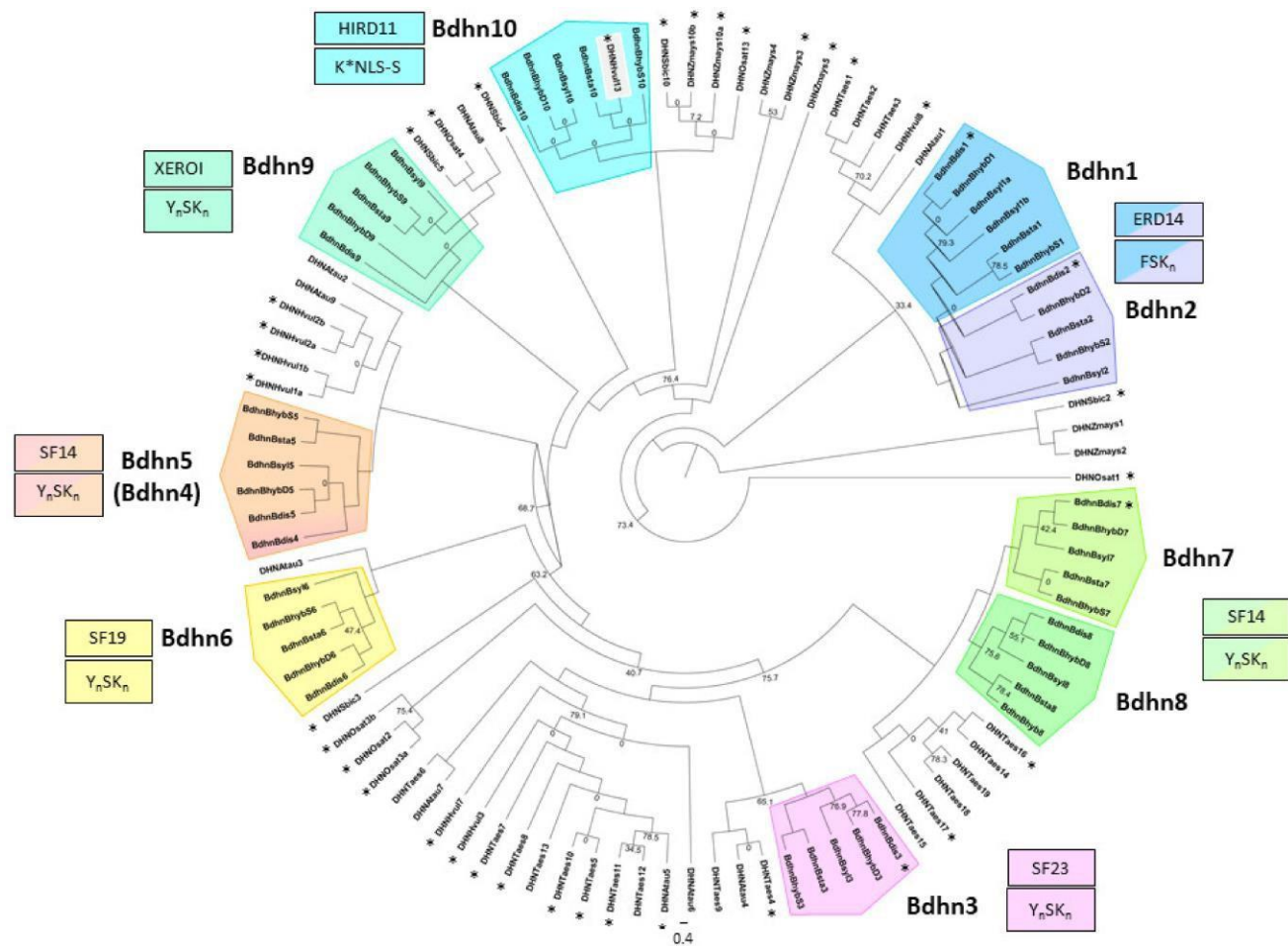
## The *Brachypodium* Dehydrin Tree

To explore the evolutionary relationship among dehydrin genes, we constructed a ML *Brachypodium* dehydrin gene tree from 47 *Bdhn* protein coding regions present in the four studied *Brachypodium* species and in six outgroup grasses (Figure 2). All internal branches but three and all the dehydrin-type clades showed strong to relatively high bootstrap support values ( $BS \geq 70\%$ ). The most divergent split separated the duplicated *Bdhn1-Bdhn2* (ERD14) clade from the rest, followed by the isolated *Bdhn10* (HIRD11) clade. Within the remaining group of  $Y_nSK_n$  dehydrin structural genes, there was a divergence of the *Bdhn9* (XEROI) clade, followed by subsequent divergences of the  $Y_nSK_n$  *Bdhn4-Bdhn5*, *Bdhn6*, *Bdhn3*, and *Bdhn7/Bdhn8* clades (Figure 2). All *Bdhn* clades were monophyletic except the paraphyletic *Bdhn9* and *Bdhn10* clades, which included orthologous sequences from closely related outgroups. Intra-clade branches were overall well-supported except for the poorly informative *Bdhn10* clade.

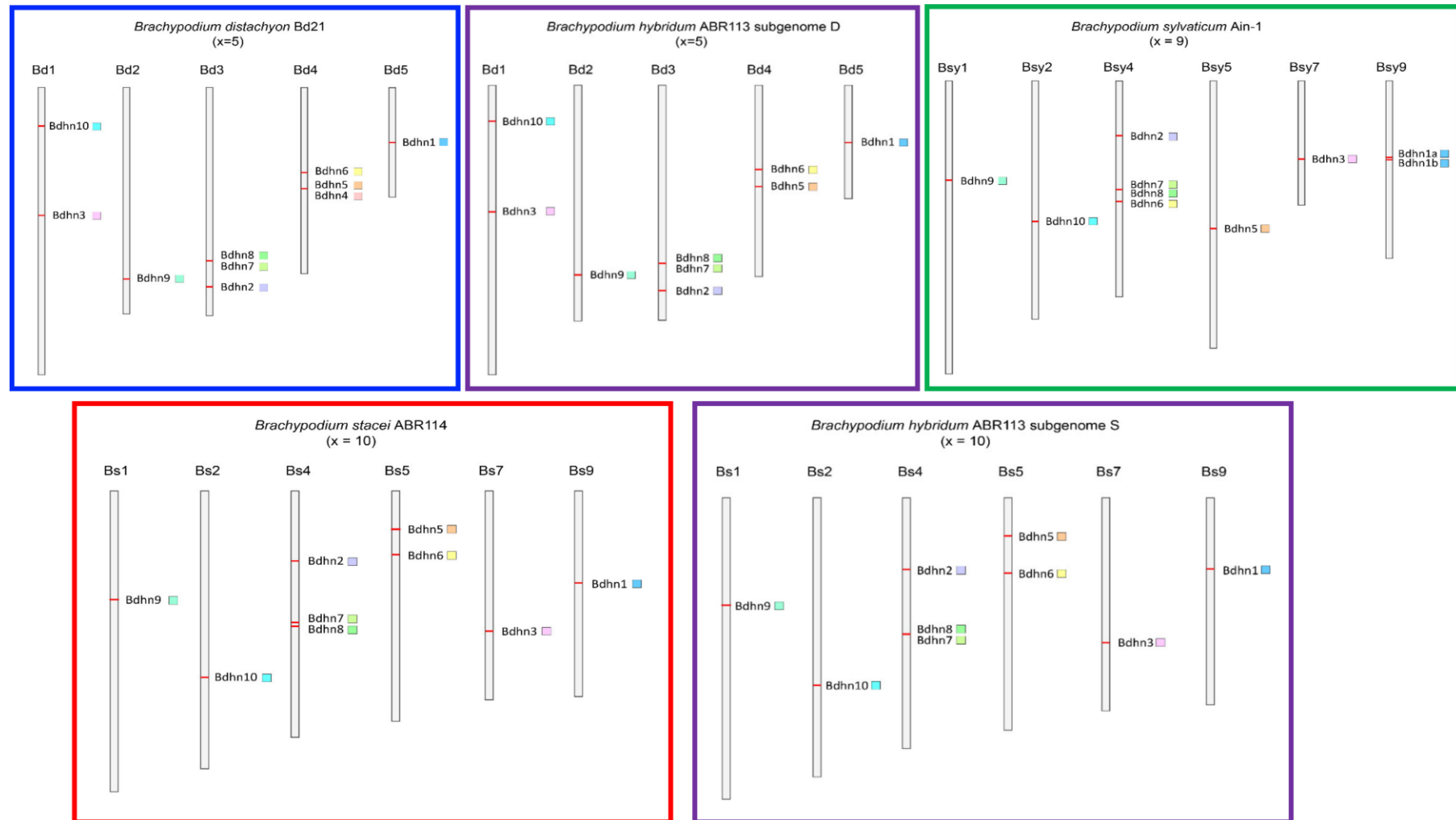
## Chromosome Distributions and Selection Analysis of Duplicated *Bdhn* Genes

We analyzed the physical distributions of *Bdhn* genes on the chromosomes of the studied species to detect the potential occurrence of tandem or segmental duplication. We also performed selection tests on the coding regions of the *Bdhn* genes to explore the potential loss of selective constraints on them. Dehydrin genes were distributed among the five chromosomes of *B. distachyon* Bd21 and the D subgenome of *B. hybridum* ABR113 (Bd), in 6 out of 10 chromosomes of *B. stacei* ABR114 and the S subgenome of *B. hybridum* ABR113 (Bs), and in 6 out of 9 chromosomes of *B. sylvaticum* Ain-1 (Bsy) (Figure 3, Supplementary Table S3). We detected four tandem duplications and two segmental duplications of *Bdhn* genes. In *B. distachyon* and *B. hybridum* D subgenome our analysis detected tandem duplications of *Bdhn7-Bdhn8* in Bd3 and of *Bdhn4-Bdhn5* in Bd4 (only *B. distachyon*), and a segmental duplication of *Bdhn1-Bdhn2* in Bd3 and Bd5 (Figure 3, Supplementary Table S3).

*B. stacei* and *B. hybridum* S subgenome and *B. sylvaticum* showed a tandem duplication of *Bdhn7-Bdhn8* in Bs4 and Bsy4, and a segmental duplication of *Bdhn1-Bdhn2* in Bs4 and Bs9 and in Bsy4 and Bsy9, respectively. In addition, *B. sylvaticum* showed a tandem duplication of *Bdhn1a* and *Bdhn1b* in Bsy9 not found in the other *Brachypodium* species studied (Figure 3, Supplementary Table S3). Overall, all the selection tests performed with branch-sites models (aBSREL (adaptive Branch-Site Random Effects Likelihood) and BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) for internal branches only and for leaf branches only) failed to detect evidence of positive selection in all the *Bdhn* genes and species all ( $p$ -values  $>0.05$ ) except for two significant cases. aBSREL tests for internal branches only detected evidence of positive selection in *Bdhn6* and *Bdhn10* (one branch each) where  $\omega_2$  rate class values were  $>1$  but for very low percentages of tree branches (*Bdhn6*:  $p = 0.04$ ,  $\omega_1 = 0.140$  (97%),  $\omega_2 = 53.9$  (2.6%); *Bdhn10*:  $p = 0.02$ ,  $\omega_1 = 0.0$  (99%),  $\omega_2 = 100,000$  (1.1%); Supplementary Materials S1). Whereas aBSREL tests for positive selection modeling both site-level and branch-level  $\omega$  heterogeneity, BUSTED performs a gene-wide (and not site-specific) test for positive selection. Similarly, selection tests performed at sites [MEME (Mixed Effects Model of Evolution)] did not detect evidence of positive selection at any site across the sequences of the *Bdhn* genes except for two significant or marginally significant positions in genes *Bdhn1-Bdhn2* (site 139,  $p = 0.04$ ; site 163,  $p = 0.02$ ) and two in *Bdhn6* (site 266,  $p = 0.04$ ; site 412,  $p = 0.05$ ) (Supplementary Materials S1). The greater power of MEME indicates



**Figure 2.** Maximum likelihood *Brachypodium Bdhn* tree. Unrooted IQtree cladogram showing the relationships among the dehydrin *Bdhn* gene clades and orthologous grass sequences (Atau: *Aegilops tauchii*; Hvul: *Hordeum vulgare*; Osat: *Oryza sativa*; Sbic: *Sorghum bicolor*; Taes: *Triticum aestivum*; Zmays: *Zea mays*) and among *Brachypodium* species and genomes within each clade. *Bdhn* clades are identified by colors, *Bdhn* codes, gene architecture, and Panther subfamily codes (see Table 1). Duplicated *Bdhn* genes form sister clades or fall within the same clade. Ultrafast bootstrap support (<80%) is shown on branches; the remaining branches are fully supported. Asterisks indicate dehydrin genes differentially expressed under drought vs. control conditions (see text and Supplementary Table S10). Scale bar: number of mutations per site.



**Figure 3.** Chromosomal location of *Bdhn* genes in the studied *Brachypodium* species. *B. distachyon* (blue rectangle), *B. stacei* (red), *B. hybridum* subgenomes S and D (purple), and *B. sylvaticum* (green). *Bdhn* genes are mapped on the chromosome with their respective color flags (see Figure 2).

that selection acting at individual sites is considerably more widespread than constant models would suggest. It also suggests that natural selection is predominantly episodic, with transient periods of adaptive evolution masked by the prevalence of purifying or neutral selection on other branches.

### Dehydrin Gene Clusters, Phylogenetics, and Climate Niche Variation in *B. distachyon*

The genomes of 54 *B. distachyon* ecotypes distributed across the Mediterranean region (Supplementary Table S4) were annotated for the 10 *Bdhn* gene clusters for comparative genomics, evolutionary, and phylogenetic signal of drought-related traits and climate niche analyses. Most ecotypes (74.07%) contained all 10 *Bdhn* genes (Supplementary Figure S3) and were used for downstream analyses. Independent ML trees obtained for each separate *Bdhn* gene, based on exon and intron sequences, showed differently resolved topologies (Supplementary Figure S4). Six of those 10 gene trees (*Bdhn1*, *Bdhn2*, *Bdhn3*, *Bdhn6*, *Bdhn7*, *Bdhn8*) recovered a congruent topology for some *B. distachyon* groups. A ML tree constructed from their concatenated sequences produced a single combined *B. distachyon Bdhn* tree (Figure 4a) showing a resolution similar to those observed previously in the *B. distachyon* nuclear-SNPs [37] (Figure 4b) and plastome [38] (Figure 4c) trees. The *Bdhn* tree revealed a relatively well-supported EDF+ clade (74% BS) and the successive but weakly supported divergences of the S+ and T+ lineages, with a clade of T+ lineages (Bd18-1, Bd21-3, BdTR5i) resolved as sister to the EDF+ clade (Figure 4a). Topological congruence Kishino-Hasekawa (KH), Shimodaira-Hasekawa (SH), and Shimodaira Approximately Unbiased (AU) tests performed between the *B. distachyon Bdhn* tree and the nuclear-SNP and plastome trees indicated that the topology of the *Bdhn* tree did not significantly differ ( $p < 0.001$ ) from the topologies of the two compared trees (Supplementary Table S5), indicating that all three data sets recover congruent evolutionary histories for the divergences of the main *B. distachyon* lineages. However, the *Bdhn* tree visually resembled more the plastome tree than the nuclear tree (Figure 4).

Climate niches were constructed for the studied *B. distachyon* ecotypes to explore if they present differences in climate niche parameters that could be related to distinct responses to drought. The optimal climate niche of each *B. distachyon* ecotype was inferred from PCA of 19 climate variables (Supplementary Table S6 and Figure S5). The main PC1 axis (51.8% of variance) values allowed us to classify the *B. distachyon* ecotypes into cold ( $>2.5$ ; ABR2, ABR3, ABR4, ABR5, ABR6, RON2), warm ( $<2.5$ ; Bd2-3, Bd21, Bd3-1, Bis1, Kah1, Kah5, Koz1, Koz3), and mesic ( $-2.5$  to  $2.5$ , remaining) climate class ecotypes.

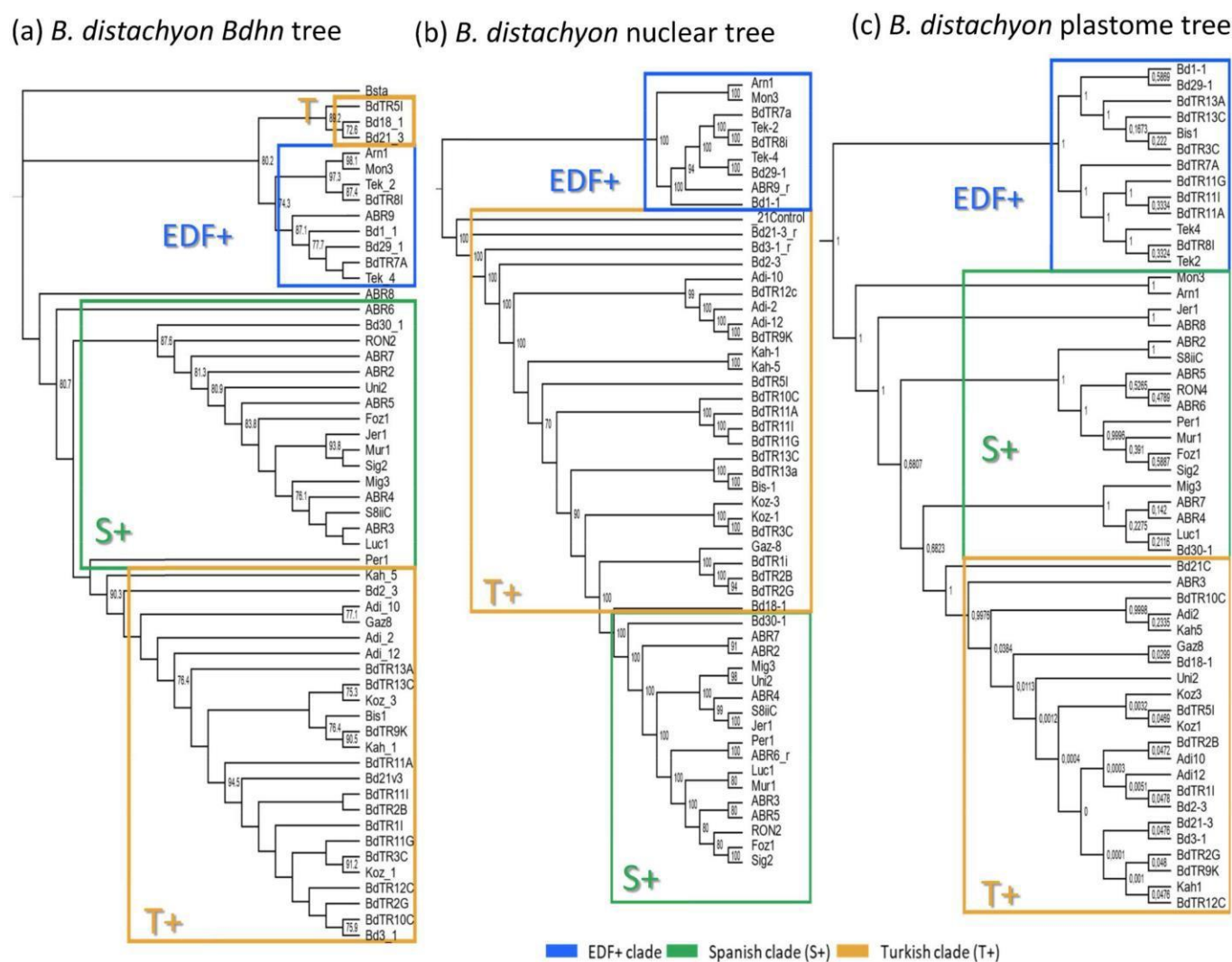
### Differential Expression of *Bdhn* Genes in *Brachypodium distachyon* Ecotypes under Drought and Temperature Stress Conditions

We made use of data from an extensive transcriptome study of 32 *B. distachyon* ecotypes [39], which enabled us to explore intraspecific variation in how *Bdhn* genes are expressed in response to drought in mature plants. Development and tissue specific expression analysis of dehydrin genes was performed in 32 out of the 54 genomically sequenced ecotypes of *B. distachyon* (Supplementary Table S4). Only 4 out of the 10 identified *Bdhn* genes were expressed in mature leaves of all 32 studied *B. distachyon* ecotypes: *Bdhn1a*

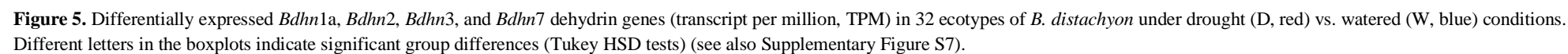
(Bradi5g10860.1), *Bdhn2* (Bradi3g51200.1), *Bdhn3* (Bradi1g37410.1), *Bdhn7* (Bradi3g43870.1); Supplementary Table S7. These annotated dehydrins showed significant differential expression (DE) levels between the watered (W) and dry (D) conditions, independently of temperature conditions in both separate CD-CW-HD-HW and averaged D vs. W comparative tests (Supplementary Table S8 and Figure S6a,b). By contrast, the dehydrin expressions did not show significant differences between cool (C) and hot (H) conditions under drought treatment, and only *Bdhn3* and *Bdhn7* showed significant differences in CW vs. HW conditions, though none of them did in averaged C vs. H comparative tests (Supplementary Figure S6a,c).

The four dehydrins showed significantly increased expression levels in drought conditions in most accessions (Wilcoxon tests, Supplementary Table S8; Tukey tests, Figure 5, Supplementary Figure S7). The DE levels were also significantly different among ecotypes, especially within the dry treatment, being highest in warm ecotypes Koz3 (*Bdhn1a*, *Bdhn3*, *Bdhn7*) and Adi10 (*Bdhn2*) and lowest in cold ecotypes ABR2 (*Bdhn1*) and ABR4 (*Bdhn2*, *Bdhn3*, *Bdhn7*) (Figure 5, Supplementary Figure S7). On average drought increased dehydrin expression by about 5.74% in *Bdhn1a*, 39% in *Bdhn2*, 67.8% in *Bdhn3*, and 97.8% in *Bdhn7* compared to well-watered plants (Supplementary Figure S6b). Overexpression of dehydrins caused by drought stress was significantly correlated between all *Bdhn* gene pairs (Supplementary Table S9 and Figure S8).





**Figure 4.** (a) Maximum likelihood *B. distachyon* Bdhf tree. Consensus tree constructed for 54 *B. distachyon* ecotypes from concatenated exon and intron aligned sequences of six dehydrin genes (*Bdhf1*, *Bdhf2*, *Bdhf3*, *Bdhf6*, *Bdhf7*, and *Bdhf8*). The EDF+ (extremely delayed flowering time, blue) clade, T+ (Turkish and East Mediterranean, orange), and S+ (Spain and West Mediterranean, green) lineages correspond to those indicated in [37,38]. Bootstrap support is indicated on branches. Accession codes correspond to those indicated in Supplementary Table S4. (b) *B. distachyon* nuclear species tree of [37] based on nuclear genome-wide 3.9 million SNPs. (c) *B. distachyon* plastome tree of [38] based on full plastome sequences.

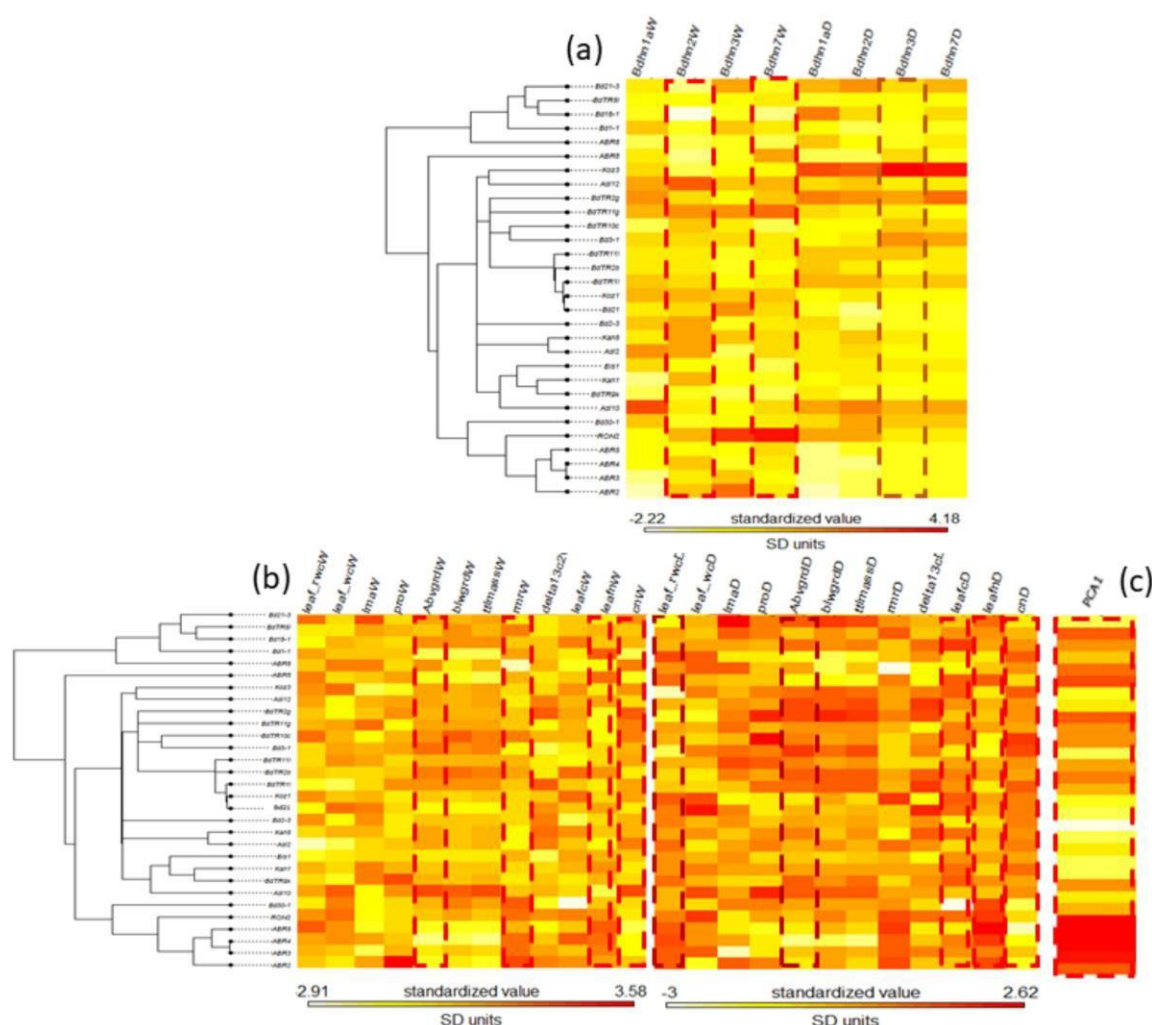


A *B. distachyon*–*T. aestivum* DE comparative analysis found that the drought-induced *Bdhn1*, *Bdhn2*, *Bdhn3*, and *Bdhn7* genes belong to the same ortholog groups as 15 out of the 16 differentially expressed wheat dehydrin (DHN) genes (Supplementary Table S10) under natural field drought stress [40] or greenhouse imposed drought stress [6,41]. In wheat there is a physical clustering of several dehydrin genes in two gene clusters located in the 5L and 6L groups of wheat chromosomes (Supplementary Table S10; Supplementary Figure S9). The 6L cluster contains 25 dehydrins and includes nine DHN3 genes and three DHN4 genes (all orthologs to *Bdhn3*), whereas the 5L cluster has 13 DHN38 genes of which six are orthologs to *Bdhn7*. In addition, the DHN11 genes, located in another portion of 6L chromosomes are orthologs to the *Bdhn1* and *Bdhn2* genes. We observed that orthologs from *B. distachyon* and *T. aestivum* tended to show a similar pattern of expression response to soil drying. Specifically, the duplicated *Bdhn1* and *Bdhn2* genes and the DHN11(A1) genes, the *Bdhn7* gene and the duplicated DHN38 (B1, B2) genes, and the *Bdhn3* gene and the duplicated DHN3 (A1, A6, B6, D1, D4, D6, D8, D9) genes were all upregulated in drought treatments (Supplementary Table S10 and Figure S9; [40]).

### Effects of Drought on Dehydrin Gene Expression and Drought-Response Phenotypic Traits

The potential effect of drought on dehydrin expression levels and on correlated changes of phenotypic and physiological drought-response traits of the plants was evaluated in 32 *B. distachyon* ecotypes. The 12 drought-response phenotypic traits studied (leaf\_rwc: relative water content in leaf; leaf\_wc: water content in leaf; lma: leaf mass per area; pro: leaf proline content; abvgrd: above ground biomass; blwgrd: below ground biomass; ttlmass: total plant mass; rmr: root mass ratio; delta13c: carbon isotope, a proxy for lifetime integrated WUE; leafc: leaf carbon content; leafn: leaf nitrogen content; cn: leaf carbon/nitrogen ratio) also showed significant different values in dry vs. watered conditions across ecotypes (Supplementary Table S11). On average, drought significantly decreased the values of six traits (17.14% in abvgrd, 34.78% in ttlmass, 4% in leaf\_rwc, 36.5% in leaf\_wc, 12.5% in leafn, 2.8% in WUE) and significantly increased those of five traits (33% in pro, 5.5% in rmr, 2.96% in leafc, 5.71% in lma, 21.4% in cn) compared to watered conditions, but did not significantly affect the blwgrd trait (Supplementary Figure S10) [2].

Drought-induced effects caused significant positive and negative correlations between the averaged expressed values of the four *Bdhn* genes and changes in most phenotypic trait values (Supplementary Table S12 and Figure S11). Regression models for independent correlations between the *Bdhn1a*, *Bdhn2*, *Bdhn3*, and *Bdhn7* expressions and the changes in the 12 phenotypic traits showed significant positive correlations for most dehydrin *Bdhn* genes with pro, blwgrd, rmr, WUE, leafc, and cn, negative correlations with leaf\_rwc, leaf\_wc, and leafn, and non-significant correlations with ttlmass and abvgrd (Supplementary Figure S11).



**Figure 6.** Maximum likelihood *B. distachyon* *Bdhn* tree cladogram showing the relationships of 30 ecotypes. Phyloheatmaps of normalized values for different sets of variables: (a) dehydrin (*Bdhn1*, *Bdhn2*, *Bdhn3*, *Bdhn7*) gene expression values under watered (W) and drought (D) conditions; (b) drought-response phenotypic traits (leaf\_rwc: relative water content in leaf; leaf\_wc: water content in leaf; lma: leaf mass per area; pro: leaf proline content; abvgd: above ground biomass; blwgd: below ground biomass; tlmass: total plant mass; rmr: root mass ratio; delta13c: carbon isotope, a proxy for lifetime integrated WUE; leafc: leaf carbon content; leafn: leaf nitrogen content; cn: leaf carbon/nitrogen ratio) values under watered

(W) and drought (D) conditions; (c) climate niche PC1 values. Traits showing significant phylogenetic signal are highlighted with dotted lines (see Supplementary Table S13b). Watered (W): soil irrigated to field capacity every second day; Drought (D): soil water content reduced by ~5% each day (during the 10 days experiment).

## Phylogenetic Signal of Dehydrin Expression, Phenotypic Trait Changes and Climate Variation in the *Brachypodium distachyon* *Bdhn* Tree

The potential phylogenetic signal of dehydrin expression, phenotypic trait changes, and climate variation was evaluated on both the *B. distachyon* nuclear species tree [37] and the *B. distachyon* dehydrin *Bdhn* tree. None of the dehydrin gene expression values under W or D conditions and few phenotypic and climate traits had significant K or lambda values on the *B. distachyon* nuclear species tree (Supplementary Table S13a and Figure S12). By contrast, *Bdhn*2W, *Bdhn*7W, and *Bdhn*3D expression values, phenotypic change of *rmr*W, *leafn*W, *cn*W, *abwgrd*W, *leafc*D, *leafn*D, and *cn*D traits' values and climate niche PCA1 values carried phylogenetic signal (or marginal phylogenetic signal for *leaf\_rwc*D and *abvgrd*D values) when tested on the dehydrin *Bdhn* tree (Supplementary Table S13b, Figure 6).

## Discussion

### The Dehydrin Gene Family in *Brachypodium*

Our comparative genomic analysis of the dehydrin genes in the reference genomes of four *Brachypodium* species and in the genomes of 54 ecotypes of *B. distachyon* has allowed us to identify 47 *Bdhn* genes. This is almost twice the number of LEA2 genes previously found in *B. distachyon* Bd21 [35]. Orthology and evolutionary analysis indicate that most of these proteins were probably present in the ancestor of the grasses (Supplementary Table S1; Figure 2). In contrast to previous non-monophyletic infra-generic phylogenies of dehydrin genes (e.g., *Oryza*; [9]), the *Brachypodium* dehydrin tree showed 8 out of 10 monophyletic and highly supported *Bdhn* clades (Figure 2). Segmental and tandem duplications of ancestral *Bdhn*1-*Bdhn*2 and recent *Bdhn*7-*Bdhn*8 genes have been detected in all studied *Brachypodium* species (Figures 2 and 3; Supplementary Figure S3); the tandemly duplicated *Bdhn*7 and *Bdhn*8 genes probably evolved, in turn, from a duplication of an ancestral *Bdhn*3 gene through insertion/deletions (*Bdhn*7-8) and the loss one K-segment (*Bdhn*8) (Figures 1c and 2). However, *Bdhn*4 is inferred to have originated from a tandem duplication of *Bdhn*5 exclusively in the ancestor of the *B. distachyon* lineage and *Bdhn*1b from a tandem duplication of *Bdhn*1a in the *B. sylvaticum* lineage (this Supplementary Table S3 and Figure S3; Figure 2). The allotetraploid *B. hybridum* exhibits homeologous copies inherited from its diploid progenitor species for the same sets of tandemly and segmentally duplicated genes (Table 1; Figures 2 and 3). Nonetheless, the loss of the *Bdhn*4 gene in its D subgenome probably occurred after the hybridization and whole genome duplication (WGD) event that generated this reference genome [30,42], as this gene is largely present in the *B. distachyon* ecotypes studied (Supplementary Figure S3). Our data support the hypothesis of a highly dynamic evolution of duplications and losses of dehydrin paralogs in *Brachypodium*. This was also evident in other grasses such as *Hordeum vulgare* [43], *Triticum aestivum* [7], and some *Oryza* species [9].

The consensus ML phylogenetic tree of *Brachypodium* species based on 10 *Bdhn* genes (Figure 2) depicted a congruent topology in seven gene clades (*Bdhn*1, *Bdhn*3, *Bdhn*(4)5, *Bdhn*7, *Bdhn*8, *Bdhn*10), resulting in more ancestral *B. stacei*/*B. hybridum*-S sequences followed by the split of more recent *B. distachyon*/*B. hybridum*-D and *B. sylvaticum* copies. This resolution was fully congruent with that of the *Brachypodium* species tree [44]; the branch swaps observed in the three remaining *Bdhn*2, *Bdhn*6, and



*Bdhn9* clades (Figure 2) likely resulted from incomplete lineage-sorting events. Therefore, the evolution of the dehydrin genes was in pace with the organismal evolution of the *Brachypodium* lineages, supporting their species-level evolutionary synchrony. Although the *Bdhn* intraspecific phylogenies of *B. distachyon* ecotypes were more variable or unresolved (Supplementary Figure S4), the consensus topology of the combined *B. distachyon Bdhn* tree and its congruence with the nuclear and plastome trees indicates that the conserved dehydrin genes also track the fast divergences of the recent *B. distachyon* ecotypes (Supplementary Table S5; Figure 4). Selection tests of *Bdhn* genes have consistently failed to detect evidence of positive selection at branch-sites or sites, including the duplicated *Bdhn* genes (Supplementary Materials S1). Our results indicate that all *Brachypodium* dehydrins are likely functional and that the duplicated paralogs are under selective constraint, irrespective of their ancestral or recent origins (Figure 2).

The amino acid composition, structure, and biochemical features inferred for the *Brachypodium* dehydrins (Supplementary Table S2) support their potential roles as regulators of the water-deficit in the cells [8–10]. *Bdhn* dehydrins variation in size, molecular weight, and pI and their GRAVY values fall within those observed for rice dehydrins [9]. The large differences in *Bdhn* pI values [ $\leq 6$  (*Bdhn1*, 2, 9)– $>9$  (*Bdhn3*, 4, 5, 6, 7, 8)] suggest that those proteins may be located in specific compartments of the cell, like the cytoplasm and the nucleus. *Bdhn* dehydrins with high pI values and with phosphorylated S-segments ( $Y_nSK_n$ ; e.g., *Bdhn3*, *Bdhn4*, *Bdhn5*, *Bdhn6*, *Bdhn7*, *Bdhn8*) may bind negatively charged molecules such as DNA and NLS proteins [8]. The *Bdhn10* dehydrins (HIRD11 family), which lack the K-segment, could bind different ions and reduce the formation of reactive oxygen species (ROS), like that observed for the AtHIRD11 ortholog [26]. The three-domain complex architecture of *Bdhn9* (Dna-J–Dna-JX–DHN) has been also observed in other grass dehydrins, like DHN1 of rice [9] and *Setaria italica* [45]; the DnaJ domain may have a chaperone function [9].

### Dehydrin Expression Induction in *Brachypodium distachyon*

Dehydrin expression varies considerably both in plant tissue and developmental stage and under different abiotic stress conditions [7–9]. Most DHN genes have shown high expression profiles in seeds and immature seedling stages in wheat and rice [7,9]; however, their expression decreases in mature or late-development tissues or is even absent in some organs, like mature leaves in rice [9]. Although the expression data was restricted to mature leaf tissue, our analyses demonstrated that four (*Bdhn1a*, *Bdhn2*, *Bdhn3*, *Bdhn7*) out of the ten detected *Brachypodium Bdhn* dehydrin genes were constitutively and inductively expressed in mature leaves of *B. distachyon* plants (Supplementary Tables S7 and S8 and Figures S6 and S7; Figure 5). In silico and RT-PCR expression analysis showed that YSK2- type dehydrins were upregulated in drought-stressed shoots of wheat, whereas Kn-type dehydrins were preferentially expressed in cold-stressed shoots [7]. Other drought-PEG treatments have shown considerable upregulation of eight DHN genes in rice shoots [9]. The effect of drought significantly contributed to transcriptional upregulation of the four genes in dry vs. watered plants in all *B. distachyon* ecotypes (Figure 5), irrespective of the temperature treatment (Supplementary Figures S6 and S7). This analysis demonstrated a 2.5–21.5-fold increase in the expression (TPM) of *Bdhn1a* but a much higher and variable 2–490-fold increase in those of *Bdhn2*, *Bdhn3*, and *Bdhn7* under dry than under water conditions (Supplementary Table S8 and Figure S6; Figure 5). *Bdhn1a* and *Bdhn2* are (FS) $K_n$ - type genes whereas *Bdhn3* and *Bdhn7* are  $Y_nSK_n$ -type genes (Table 1; Figure 2); thus our results partially depart from those found in wheat by [7] suggesting that both types of genes are preferentially upregulated by drought rather than by temperature in mature leaves of *B. distachyon*. By contrast, our comparative structural

analysis between *B. distachyon* differentially expressed *Bdhn* genes in this work, and the previously reported wheat dehydrin genes differentially expressed in flag leaves under natural field [40] or greenhouse-imposed [6,41] drought stress, inferred that 15 out of the 16 DE wheat dehydrins belong to the same *Brachypodium* DE ortholog groups (Supplementary Table S10 and Figure S9). These shared differential gene expression responses to drought by orthologous dehydrin genes' induction in both species reinforce the potential of *B. distachyon* as a model system for cereals such as bread wheat. Our results also highlight the likely importance of these four dehydrins in the protection of *B. distachyon* plants to water stress conditions in mature individuals, the developmental stage when they face the most severe drought conditions of their life cycle [46].

The drought-induced upregulation of *Bdhn* genes was different among the studied *B. distachyon* ecotypes for all four *Bdhn* genes (Supplementary Table S8 and Figure S7; Figure 5). Recent analysis of dehydrin gene expression in *B. distachyon* and its close *B. stacei* and *B. hybridum* species have shown that Bradi1g37410 (*Bdhn3*) is strongly upregulated by drought (>400-fold higher expression in dry plants compared to control plants), although the level of induction depended on genotype [47]. Our data support these results and additionally demonstrate that drought induction also upregulates the expressions of the *Bdhn1a*, *Bdhn2*, and *Bdhn7* genes and that increased expression is most evident in the warm climate ecotypes (Koz3, Adi10) and less so in the cold climate ecotypes (ABR2, ABR4), whereas mesic climate ecotypes show intermediate expression levels (Supplementary Table S8 and Figure S7; Figure 5). Earlier classification of *B. distachyon* ecotypes into drought-tolerant,

-intermediate, and -susceptible, based on phenotypic plant water content and wilting index values [48], roughly correspond to our warm, mesic, and cold ecotype climate classes, though their plants were subjected to uncontrolled severe drought treatments which may have confounded plant size with soil water content. The significant differences found between the drought-induced dehydrin expression levels in our climate class *B. distachyon* ecotypes suggest that ecotypes adapted to warm climates may have developed higher *Bdhn1a*, *Bdhn2*, *Bdhn3*, and *Bdhn7* expression responses as a strategy to protect the mature plants against harsh water deprivation conditions and to ensure the survival and reproduction of the individuals in their habitats. By contrast, in mesic and, especially, in cold climate adapted ecotypes those inductions are much lower possibly due to the absence or mitigated presence of the natural stressor.

The constitutive and induced expression of the stress-responsive dehydrins is upregulated by the presence of specific *cis*-regulatory elements in the promoter region of their genes [49]. Our de novo analysis of *cis*-regulatory elements performed *in silico* consistently found three TF-binding sites, BES1/BZR1, MYB124, and ZAT, across the *Bdhn* promoters of the studied *Brachypodium* species, and more abundantly in the promoter regions of the *Bdhn1a* and *Bdhn2* genes (Table 2; Supplementary Figure S2a). BES1/BZR1 is a brassinosteroid signaling positive regulator (BZR1) family protein involved in the regulation pathway in response to drought [50]. The MYB gene protein MYB124 is related to the abscisic acid (ABA) response, an hormone-regulated pathway implicated in multiple stress response such as drought or cold stress [51,52]. The ZAT C2H2 zinc finger is involved in response to salinity stress [53]. The presence of these *cis*-regulatory elements in the promoters of most *Bdhn* genes suggest that these dehydrins could be highly upregulated in *Brachypodium* plants under different environmental stressors such as drought, cold, and salinity, and that MYB124 and BES1 may play an important role in the induction of the *Bdhn1a* and *Bdhn2* genes in the studied ecotypes, especially in those adapted to warm climates.



## Correlated Dehydrin and Phenotypic Drought Response and Phylogenetic Signal in *Brachypodium distachyon*

Water deficit stress affects the physiology, the phenotypes, and the fitness of plants [1,2]. As shown earlier by [2], drought effect significantly influenced the changes of *B. distachyon* ecotypes' phenotypic traits, reducing the water contents, total plant mass, and leaf nitrogen content of dry plants but increasing their root biomass, leaf carbon content, proline, and WUE (Supplementary Table S11 and Figure S10). Our linear regression models indicated that the expressions of the four *Bdhn* genes were significantly correlated with the changes of most phenotypic traits in the drought treatment (Supplementary Table S12 and Figure S11). The high correlations observed between the expression of *Bdhn1a*, *Bdhn2*, *Bdhn3*, and *Bdhn7* genes and the decrease of leaf water and nitrogen contents and the increase of belowground biomass, root mass, WUE, and leaf carbon and proline content were strongly associated to drought stress. Proline can serve as an osmoprotectant and a signaling molecule triggering adaptive responses to cell water stress. Martinez et al. reported a significant decrease in plant water content but no significant changes in proline content in *B. distachyon* ecotypes under drought stress [46]. By contrast, Fisher et al. found significant differences in both traits across *B. distachyon* ecotypes under mild and severe drought treatments, with drought-tolerant ecotypes showing more prominent water and proline responses than the drought-intermediate and drought-susceptible ecotypes [48]. Our data corroborate the last results and further illustrate that drought-induced proline production is significantly higher in warm-to-mesic climate *B. distachyon* ecotypes (Adi10, Koz3) and lower in cold climate ecotypes (ABR5, ABR4, ABR3) (Supplementary Table S11) and that those differences overlap with the significant differences observed in their dehydrin overexpression profiles (Supplementary Table S8; Figure 5). In cool seasonal plants WUE is expected to increase with aridity [54]. However, Manzaneda et al. found that drought-avoider *B. distachyon* ecotypes showed lower WUE than aridic drought-escape

*B. hybridum* ecotypes, although the former had higher values of WUE plasticity related to climate than the second [55]. Des Marais et al. also found an association of WUE with climate as *B. distachyon* ecotypes from cooler climates were more plastic in their WUE than those from warmer climates [2]. Our data indicate that the overall dehydrin expression is significantly correlated with WUE (Supplementary Figure S11). In addition, WUE shows a great plasticity across ecotypes of any climate class and under both drought and watered conditions (Supplementary Table S12).

Phylogenetic signal measures the statistical dependence among species' trait values due to their phylogenetic relationships [56]. The potential evolutionary signal of dehydrin expression values and phenotypic trait values gave different results when tested on the *B. distachyon* nuclear species tree or the *B. distachyon Bdhn* tree. The absence of phylogenetic signal for the dehydrin expressions and the residual signal for some of their associated drought-response phenotypic traits in the nuclear species tree (Supplementary Table S13a and Figure S12) indicates that these drought-response mechanisms may have evolved independently and at different times along the life history of *B. distachyon*. However, several flowering time traits and their molecular regulators have shown a strongly correlated evolution with the nuclear species tree [37], supporting the important role of flowering time in shaping the divergences of the *B. distachyon* lineages. Conversely, the significant phylogenetic signal of some dehydrin expressions, drought response phenotypic traits changes, and climate niche data variation on the *B. distachyon Bdhn*

tree (Supplementary Table S13b; Figure 6) suggests that the evolution of the *Bdhn* genes is determined by the adaptation of the *B. distachyon* ecotypes to more dry or more wet environmental conditions. It is surprising the high topological similarity found between the *Bdhn* tree and the plastome tree in contrast with its more dissimilar topology with respect to the nuclear species tree (Figure 4) for nuclear dehydrin genes that encode cytoplasmic and nuclear but not chloroplast proteins [8]. The relative congruence detected between the *Bdhn* and plastome trees could be a consequence of incomplete lineage sorting events of the recently evolved *B. distachyon* lineages [38]. However, it could also imply a yet unknown organellar effect on the cellular response mechanism to adaptation to drought, like the role played by the chloroplast in inducing the expression of nuclear heat-response genes during heat stress in plants [57]. Our data open new ways to investigate the potential implication of this organelle in the induction of drought-response nuclear genes, like those encoding for dehydrins, and in their evolutionary history.

## Material and Methods

### Identification of Dehydrin Sequences

Dehydrins of *B. distachyon*, *B. stacei*, *B. hybridum*, *B. sylvaticum*, and other grass outgroups (*Aegilops tauschii*, *Hordeum vulgare*, *Oryza sativa*, *Sorghum bicolor*, *Triticum aestivum*, and *Zea mays*) were identified using three searching approaches. First, the Phytozome v.12.1 database [58]) was searched for DHN gene sequences of *B. distachyon*, *B. stacei*, *B. hybridum*, and *B. sylvaticum* (Table 1). Phytozome dehydrin sequences were retrieved using BioMart to filter sequences having DHN Pfam code (Pfam00257). This search was repeated in the Ensembl [59] and Genbank [60] databases, aiming to retrieve all dehydrin genes present in *Brachypodium*. Redundant sequences and incomplete transcripts were deleted. Second, a consensus K-segment was used as a query sequence to search for complete DHN genes within the retrieved sequences using the BlastP tool [61]. The presence of a K-segment with a maximum threshold of 4 mismatches with respect to the query was used to characterize a protein sequence as a dehydrin. Third, orthologous dehydrin genes from six additional grass species with complete sequenced genomes were searched in Ensembl Plants, Phytozome, Genbank, and Panther (<http://pantherdb.org/data/>, 29 November 2021) databases using BioMart and used as reference outgroups. The Pfam00257 code was used to find DHN orthologous sequences from the six outgroup species, discarding also redundant sequences or sequences without the K-segment.

The theoretical molecular weight (mol. wt), isoelectric point (pI), instability index, and grand hydrophaticity average index (GRAVY) values of the different dehydrins were predicted using the ProtParam tool (<http://web.expasy.org/protparam>, 29 November 2021). In silico dehydrin structures were modeled using the web server version of RaptorX (<http://raptorx.uchicago.edu/BindingSite/>, 29 November 2021; [62,63]) and their structural properties were analyzed using Icn3D 3D structure viewing tool [64].

### Structural Analysis, Conserved Motifs, and Cis-Regulatory Elements (CREs)

The inferred DHN polypeptide sequences were used to analyze the presence of conserved motifs and to characterize the structure of the dehydrins. A custom search tool (Supplementary Data, [https://github.com/Bioflora/Brachypodium\\_dehydrins](https://github.com/Bioflora/Brachypodium_dehydrins) (accessed on 28 October 2021)) was designed to find the conserved K (EKKGIMDKIKEKLP), Y (VDEYGNP), S (SSSS+),  $\Phi$  (EDDGQGR), F (DRGLFDKFIGKK), and NLS (KKDKKKKKEKK)

motifs present in the dehydrin domain. A consensus sequence for each segment was retrieved and used as a query in a BLASTP search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>, 29 November 2021), allowing a maximum threshold of 4 mismatches with respect to the query. The dehydrin architectures were established according to the presence and distribution of their conserved motifs. De novo discovery of CREs was performed on windows of −500-to-+200 bp both sides of the transcriptional start site (TSS) of DHN genes in all the studied *Brachypodium* species and ecotypes. We searched for over-represented motifs using RSAT::Plants [65] tool peak-motifs, as described in [66]. This analysis was run four times, using as genome background model the respective reference genome of each *Brachypodium* species under study (*B. distachyon* Bd21 v3.0.46. JGI, *B. stacei* ABR114 v1.1.JGI, *B. hybridum* ABR113 v1.1.JGI, *B. sylvaticum* Ain1v1.1.JGI). Significant enrichment of the discovered motifs was assessed using as negative control promoters from the same number of randomly picked genes [66]. Candidate motifs were chosen based on their k-mer significance and number of sites and subsequently clustered to avoid redundancies using the matrix-clustering tool [67]. A total of 60 potential *cis*-regulatory motifs were retrieved and subsequently filtered, comparing their k-mer significance and number of sites with the negative controls (Supplementary Figure S2b). Of them, 29 motifs were clustered to avoid redundancies due to different identifications of the same CRE. Selected motifs were finally scanned along each *Bdhn* promoter to locate potential CREs using a matrix scan and a maximum threshold of 9 based on the median length of the 3 motif logos (Table 2).

### Multiple Alignments and Phylogenetic Analysis

Multiple sequence alignment (MSA) of the nucleotide coding sequences of all the *Brachypodium* species and other grass outgroups' dehydrin genes was performed with ClustalW in MEGA v.5 [68] using default settings. The start codon of each dehydrin gene was set using the Phytozome annotations and the sequences were adjusted manually to fit the reading frames. Alignments of dehydrin sequences, including exons and introns were performed with MAFFT v.7.215 [69] in Geneious Prime 2021 (<https://www.geneious.com/prime/> (accessed on 28 October 2021)). These alignments and their respective coding sequences were used for downstream phylogenetic analyses and for the selection tests. There was no missing data in any of the single alignments; the alignment of combined data from the six *Bdhn* genes (*Bdhn1*, *Bdhn2*, *Bdhn3*, *Bdhn6*, *Bdhn7*, *Bdhn8*) used to reconstruct the *B. distachyon* *Bdhn* tree showed an extremely low percentage of missing data (2.7%) caused by the lack of dehydrin copies in some accessions. Maximum likelihood (ML) phylogenetic trees were constructed with IQTREE 1.6.12 [70] imposing the best-fit nucleotide substitution model of each data set according to the Bayesian Information Criterion (BIC). Branch support for the best tree was estimated through 1000 ultrafast bootstrap replicates.

### Chromosomal Location, Gene Duplication, and Selection Analysis

Physical locations of the *Brachypodium* dehydrin genes in the 5 chromosomes of *B. distachyon*, 10 of *B. stacei*, and 9 of *B. sylvaticum* were obtained from Phytozome and Ensembl. They were mapped to their respective chromosomes using gff3 annotation coordinates for each dehydrin gene. Tandemly and segmentally duplicated genes were identified on the chromosomes; tandemly duplicated dehydrin genes were those distributed adjacent to an homologous dehydrin gene on the same chromosome or within a sequence distance of 50 kb [71].

The signature of positive selection ( $\omega > 1$ ) on each *Bdhn* gene for the four studied *Brachypodium* species (five genomes/subgenomes) was tested through both branch-site (aBSREL [72], BUSTED [73]) and site (MEME [74]) tests with Datamonkey2 [75,76]; <https://www.datamonkey.org/>, accessed on 29 November 2021). aBSREL and BUSTED models searched for positive selection at all sites and internal branches only or leaf branches only across the entire phylogeny. MEME (Mixed Effects Model of Evolution) tested for potential diversifying selection at individual sites under a proportion of branches.

### 1.1. Clustering and Phylogeny of Dehydrin Genes in *Brachypodium distachyon* Ecotypes

Annotations of the locations of dehydrin genes in the genomes of 54 *B. distachyon* ecotypes (Supplementary Table S3) were used to map them into the five *B. distachyon* chromosomes using a custom tool (Supplementary Materials, [https://github.com/Bioflora/Brachypodium\\_dehydrins](https://github.com/Bioflora/Brachypodium_dehydrins) (accessed on 28 October 2021)). Protein sequences were obtained from primary transcript files and specific dehydrin genes were extracted from pseudo-molecules using coordinates from gff3 annotation files. The dehydrin protein sequences were aligned using BLOSUM62. Dehydrin genes of all ecotypes were classified into different clusters based on a similarity threshold of 95%. Only clusters containing three or more sequences were selected. Unclassified sequences were iteratively compared to the previous blocks and classified into new clusters following the procedures of the first analysis. The remaining unclassified sequences were identified manually and classified as dehydrins whenever possible. Dehydrin *Bdhn4* and *Bdhn5* were annotated together in certain *B. distachyon* lines. Those sequences were manually curated and their presence in the genomes corroborated using BLASTN (Supplementary Materials S2) at Brachypan database [37]. Maximum-likelihood (ML) phylogenetic analysis was performed with coding and non-coding sequences of the ten *Brachypodium* dehydrin genes across the 54 ecotypes of *B. distachyon* using IQTREE and the procedures indicated above.

## Expression Analysis of Dehydrin Genes in *Brachypodium distachyon*

Expression analysis of dehydrin genes was performed from a transcriptome study of 32 ecotypes of *B. distachyon* [39] (Supplementary Table S4) using replicates of the plant materials employed in the ecophysiological study of [2]. Seeds were stratified at 6 °C for two weeks and then grown in the greenhouse at soil field water capacity for 3 weeks. Light levels were set at 400–1000  $\mu\text{mol m}^{-2}$  photosynthetically active radiation (PAR; mean of 825  $\mu\text{mol m}^{-2}$ ) for 10 h d<sup>-1</sup> (short-day conditions to prevent rapid flowering). Grown plants (21 days from initial pot emergence) were subjected to watered (W) vs. dry (D) conditions for a 10 day experiment, following the experimental design described in [2] (see Supplementary Materials Table S3 for more details). Irrigated plants were watered to soil field capacity every second day whereas soil water content was reduced by ~5% each day in dry plants. The plants under both treatments were simultaneously exposed to cool (C, daytime ~25 °C) or hot (H, daytime ~35 °C) conditions, however the temperature stress conditions did not affect substantially the expression of dehydrin genes (see Section 2). Fully expanded leaves from 31-days old individuals (four replicates) per ecotype and treatment were excised below the lamina, flash-frozen on liquid nitrogen and then stored at –80 °C until RNA extraction. RNA isolation was performed using the Sigma Total Plant RNA kit. RNA-Sequencing of 3' cDNA tag libraries (with fragment of 300–500 bp) was conducted on an Illumina HiSeq2500 platform obtaining 100 bp Single-End (SE) reads. This method yielded only one sequence per expressed transcript in the RNA pool, allowing for

higher sequencing coverage per gene [39]. SE reads were checked for quality with FastQC and adapters and low quality reads were removed and filtered with Trimmomatic-0.32 [77]. Total TPM values were quantified with Kallisto v0.43.1 [78], normalized with Sleuth [79], and annotated for dehydrins using the *B. distachyon* Bd21 v.3.1 reference genome (<http://phytozome.jgi.doe.gov/>, 29 November 2021; [80]). TPM values of annotated dehydrins of plants under the combined WC, WH, DC, and DH treatments were extracted from the large TPM abundance database (Supplementary Table S7). The *B. distachyon* RNAseq data were deposited in the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena> (accessed on 28 October 2021)) under accession codes ERR6133302 to ERR6133575 (project PRJEB45871) and those of *Bdhn* genes in Github ([https://github.com/Bioflora/Brachypodium\\_dehydrins](https://github.com/Bioflora/Brachypodium_dehydrins) (accessed on 28 October 2021)).

Summary statistic (mean, median, SD, range) values and boxplots of differentially expressed (DE) dehydrin (TPM) data were computed for each ecotype and expressed dehydrin gene using the stats package in R. Statistically significant differences between median values of samples under drought (W vs. D) and temperature (C vs. H) stresses, and within each of the W and D treatments of the drought experiment were tested with ggplot and geom\_signif functions in R. Wilcoxon pairwise difference tests for all pairs of compared samples with *p*-values adjusted with the Benjamini–Hochberg procedure to correct for multiple comparisons, Kruskal–Wallis rank tests for the whole group of samples within each group, and posthoc Tukey tests for among ecotypes differences were computed using the base, dplyr, ggplot2, ggpubrr, ggsignif, lm FSA, car, and multcompView packages of R.

### **Drought-Induced Changes in Dehydrin Expressions, Phenotypic, Physiological, and Climatic Niche Traits, and Phylogenetic Signal in *Brachypodium distachyon***

Values of 12 drought-response traits under W and D treatments (leaf\_rwc; leaf\_wc; lma; pro; abvgrd; blwgrd; ttlmass; rmr; delta13c (WUE); leafc; leafn; cn) were measured in the same individual samples (replicates) used in the transcriptomic analyses (Supplementary Materials S4); these phenotypic characters corresponded to those studied by [2]. Summary statistics and significance tests were computed for the 12 traits under W and D treatments following the same procedures mentioned above. The *B. distachyon* dehydrin genes differentially expressed in leaves of plants under W and D conditions in this work were compared to the wheat dehydrin genes differentially expressed under drought conditions in previous transcriptomics analyses. The wheat RNAseq analyses were carried out in flag leaves of individuals under field drought stress [40] and field rain shelter and greenhouse experiments [6,41,81]. The reported DE wheat dehydrins were used to perform the comparisons. Thus, 60 wheat dehydrin gene sequences [82] were retrieved through Blast analysis from the Wheat@URGI portal <https://wheat-urgi.versailles.inrae.fr>, 29 November 2021, [83]. Orthology of the expressed *B. distachyon* and *T. aestivum* dehydrin genes was primarily retrieved from Ensembl Plants using BioMart, and only few cases were retrieved using Blast and the homoeologies previously established in [39] (Supplementary Table S10). Environmental climate data was retrieved for the studied *B. distachyon* ecotypes from worldclim (19 temperature and precipitation variables; Supplementary Table S6). Climatic niche optima were constructed for each ecotype based on occurrence data and the first axis of the ordination of the climatic variables (PCA1) was computed with the dudi.pca function of the ade4 package [84] in R. The climatic niches of the *B. distachyon* ecotypes were classified in climatic classes warm, mesic, or cold according to their PCA1 eigenvalues (see Section 2; Supplementary Table S6).



To address potential correlations between the dehydrin gene expressions and the changes in drought-response phenotypic traits, linear regression model analyses were performed for testing the effect of particular *Bdhn* gene expression on phenotypic changes using the *lm* function of the R stats package.

A consensus ML phylogenetic tree of 30 *B. distachyon* ecotypes based on the expressed dehydrin genes (*Bdhn* tree) was topologically contrasted to that of the *B. distachyon* nuclear species tree based on genome-wide >3.9 million syntenic SNPs [37] using the KH, SH, and AU tests with resampling estimated log-likelihood (RELL) optimization and 1 million bootstrap replicates in PAUP\* [85]. We also tested for topological congruence of the *Bdhn* tree and the *B. distachyon* plastome tree based on full plastome sequences of these ecotypes [38] using the same testing approach.

Dehydrin expression level, drought-response phenotypic change, and climatic niche (PCA1) variation traits were tested for phylogenetic signal using Blomberg's K [86] and Pagel's lambda [87] with the *phylosig* function of the package *phytools* [88] in R. For both tests, values close to 1 indicate that trait values are consistent with the tree topology (phylogenetic signal) and those close to 0 that there is no influence of shared ancestry on trait values (phylogenetic independence). Phylogenetic signal was assessed on both the *B. distachyon* nuclear species tree and the *B. distachyon Bdhn* tree. Phyloheatmaps were generated for the standardized values of these continuous characters with *phytools*.

## Conclusions

We annotated and analyzed the ten *Brachypodium* dehydrin genes (*Bdhn1*–*Bdhn10*) present in the reference genomes of the three annual *B. distachyon*, *B. stacei*, *B. hybridum* and one perennial (*B. sylvaticum*) species of the genus. Most *Bdhn* genes have orthologs in other close grass species. Ancestral segmental and tandem duplications have been, respectively, detected in all species for the *Bdhn1/Bdhn2* and *Bdhn7/Bdhn8* genes, and recent tandem duplications in *B. distachyon* for *Bdhn4/Bdhn5* and in *B. sylvaticum* for *Bdhn1a/Bdhn1b* genes. Structural and biochemical properties of the *Brachypodium* dehydrins indicate that these disordered proteins may be present in the cytoplasmic and nuclear compartments of the cell. The three *cis*-regulatory elements identified in the promoter regions of the *Bdhn* genes suggests that the predominant regulation of the *Bdhn* genes is via ABA- and brassinosteroid-mediated response metabolic pathways. Only four dehydrin genes (*Bdhn1a*, *Bdhn2*, *Bdhn3*, *Bdhn7*) are expressed in mature leaves of *B. distachyon*. Differential expression levels of these dehydrins are mainly induced by drought rather than temperature conditions and are genotype-dependent, being significantly higher in warm than in mesic or cold climate ecotypes. Drought-mediated dehydrin upregulation is significantly correlated with leaf water and nitrogen contents decreases and root biomass and leaf proline increase, which are also genotype-dependent. Reverse genetic experiments of over-expression or silencing of these differentially expressed *Bdhn* genes in *Brachypodium* would be an excellent avenue for future research to confirm their role in response to drought stress.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/plants10122664/s1>, Figure S1: Inferred three dimensional structure of some *Brachypodium* dehydrins. All species' dehydrins showed  $\alpha$ -helices (1 to 4) whereas only some of them presented  $\beta$ -sheets (*Bdhn1* two in all species except *B. distachyon*, *Bdhn2* two to three in *B. sylvaticum*, and *Bdhn6* and *Bdhn9* three in *B. hybridumS* and *B. sylvaticum*): (A) *B. distachyon Bdhn4*; (B) *B. hybridumD Bdhn3*; (C) *B. stacei Bdhn1*; (D) *B. hybridumS Bdhn10*; (E): *B. sylvaticum Bdhn1b*, Figure S2: BES1/BZR1 (Basic leucine zipper; green), MYB124 (Myb gene protein; blue), and ZAT (C2H2 zinc finger; red) *cis*-regulatory elements found in 5'-upstream promoter region (−500-to-+200 bp) of the *Brachypodium Bdhn* genes. (a) Distributions of *cis*-motifs

per species and reference genomes (BD, *B. distachyon* Bd21; BS, *B. stacei* ABR114; BHD, *B. hybridum* subgenome-D ABR113; BHS, *B. hybridum* subgenome-S ABR113; BSY, *B. sylvaticum* Ain1) and per *Bdhn* gene promoter (*Bdhn1* to *Bdhn10*). Values indicate the number of each type of *cis*-motif found in the promoter region of each gene (see color codes in the chart). (b, c) Analysis of *cis*-regulatory element discovery with Rsat:plants tools in the 5-upstream promoter region (−500-to+200 bp) of 47 *Bdhn* genes from four *Brachypodium* species. (b) Maximum k-mer significance values from each analysis are shown for *Bdhn* sequences from each species, (c) Maximum number of sites from each background analysis are shown for *Bdhn* sequences from each species. In each case, another 47 random gene sequences from the respective reference genome were analyzed ten times as controls (grey bars; C1–C10), Figure S3: Chromosomal location of the 10 *Bdhn* genes in 54 *B. distachyon* ecotypes. *Bdhn* genes (clusters) detected in each ecotype were compared against the dehydrins of the reference genome (Bd21 v3) and assigned to the cluster with highest score using global pairwise alignment (Needleman-Wunsch with BLOSUM62). *Bdhn* color codes and the accuracy of the annotations are indicated in the charts, Figure S4: Maximum likelihood *B. distachyon* dehydrin trees obtained from the aligned exon and intron sequences of each independent *Bdhn* gene (*Bdhn1* to *Bdhn10*). Trees were constructed with IQTREE using the *B. stacei* outgroup sequence to root the tree. Bootstrap support is indicated on branches. Accession codes correspond to those indicated in Table S4, Figure S5: Bidimensional PCA plot of 54 *B. distachyon* ecotypes obtained from 19 climate variables (see Table S6). PC1 and PC2 axes comprise 48.5% and 22.4% of the variance, respectively. Ecotype codes correspond to those indicated in Table S4. Ellipses include ecotypes classified within cold (aquamarine), mesic (green), and warm (red) climate classes according to their PC1 values, Figure S6: Boxplots and Wilcoxon pairwise significance tests of differential gene expression values (normalized transcript per million, TPM) of the four expressed dehydrin genes (*Bdhn1a*, *Bdhn2*, *Bdhn3*, *Bdhn7*) under joint and separately analyzed drought and temperature stress conditions in 32 *B. distachyon* ecotypes. Averaged expression values for C (Cool), H (Hot), W (Watered), and D (Drought) treatments and their combinations (see text). (a) Pairwise comparative tests of combined CD-CW-HD-HW treatments; all dehydrins were significantly differentially expressed in all CD vs. CW and HD vs. HW tests, by contrast they were not significantly different in all CD vs. HD tests and in two CW vs. HW tests (*Bdhn1a*, *Bdhn2*); (b) Pairwise comparative tests of D vs. W treatments; all dehydrins were significantly differentially expressed; (c) Pairwise comparative tests of C vs. H treatments; none of the dehydrins were significantly differentially expressed, Figure S7: Differentially expressed *Bdhn1a*, *Bdhn2*, *Bdhn3* and *Bdhn7* dehydrin genes (normalized transcript per million, TPM) across 32 ecotypes of *B. distachyon* under drought (D, red) vs. watered (W, blue) conditions. Different letters in the boxplots indicate significant group differences (Tukey tests), Figure S8: Linear model regression plots of pairwise dehydrin *Bdhn* expression values (normalized transcript per million, TPM) under joint drought and watered conditions, Figure S9: Physical locations of orthologous *Brachypodium distachyon* and *Triticum aestivum* water stress responsive dehydrin genes. Drought-induced wheat dehydrin genes are highlighted in colors. *Brachypodium* chromosomes are drawn at 10x scale with respect to wheat chromosomes, Figure S10: Drought-response phenotypic changes as a function of drought treatment averaged across 32 *B. distachyon* ecotypes [leaf\_rwc (relative water content in leaf); leaf\_wc (water content in leaf); lma (leaf mass per area); pro (leaf proline content); abvrgd (above ground biomass); blwgrd (below ground biomass); tlmass (total mass); rmr (root mass ratio); delta13c (carbon isotope, a proxy for lifetime integrated WUE); leafc (leaf carbon content); leafn (leaf nitrogen content); cn (leaf carbon/nitrogen ratio)]. Asterisks above boxes indicate Wilcoxon pairwise significant difference among drought (D, red) and watered (W, blue) conditions ( $p$ -value < 0.001, \*\*\*), Figure S11: Linear model regression plots of dehydrin *Bdhn1a*, *Bdhn2*, *Bdhn3* and *Bdhn7* expression values (normalized transcript per million, TPM) and drought-response phenotypic traits changes under total dry (D) and watered (W) conditions, Figure S12: Maximum Likelihood *B. distachyon* nuclear species tree cladogram showing the relationships of 30 ecotypes. Phyloheatmaps of normalized values for different sets of variables: (a) dehydrin (*Bdhn1*, *Bdhn2*, *Bdhn3*, *Bdhn7*) gene expression values under watered (W) and drought (D) conditions; (b) drought-response phenotypic traits (leaf\_rwc; leaf\_wc; lma; pro; abvrgd; blwgrd; tlmass; rmr; delta13c; leafc; leafn; cn) values under watered (W) and drought (D) conditions; (c) climate niche PCA1 values. Traits showing significant phylogenetic signal are highlighted with dotted lines (see Table S13a), Table S1: Sampled dehydrin sequences from grass species closely related to *Brachypodium*. The accession code and the protein name correspond to those indicated in Phytozome and Genbank. Asterisks indicate the original names of DHN genes. Outgroup DHN sequences orthologous to the corresponding *Brachypodium* *Bdhn* genes are based on the analyses developed in this study; crosses indicate orthology information retrieved from Ensembl Plants and Phytozome, Table S2: Molecular traits of *Brachypodium* *Bdhn* proteins. No. aa, number of aminoacids; Mwt, molecular weight; pI, isoelectric point; Instability index; GRAVY, grand hydrophaticity average index. Abbreviations of species and reference genomes: BD, *B. distachyon* Bd21; BHD, *B. hybridum* D-subgenome ABR113; BS, *B. stacei* ABR114; BHS, *B. hybridum* S-subgenome ABR113; BSY, *B. sylvaticum* Ain1, Table S3: Chromosomal location of *Bdhn* genes across the four studied *Brachypodium* species and genomes. Chr, chromosome number (*B. distachyon* Bd21: Bd1–Bd5; *B. hybridum* ABR113 subgenome D: Bd1–Bd5; *B. stacei* ABR114: Bs1–Bs10; *B. hybridum* ABR113 subgenome S: Bs1–Bs10; *B. sylvaticum* Ain-1: Bsy1–Bsy9). The highest density of dehydrin genes were found in the syntenic chromosomes Bd3 and Bd4 (*Bdhn2*, *Bdhn4*, *Bdhn5*, *Bdhn6*, *Bdhn7*, *Bdhn8*), Bs4 (*Bdhn2*, *Bdhn6*, *Bdhn7*), the equivalent



*B. hybridum* D and S subgenomic chromosomes (except *Bdhn4*), and Bsy4 (*Bdhn2*, *Bdhn6*, *Bdhn7*, *Bdhn8*). Lengths and positions correspond to the respective reference genomes, Table S4: Sampling origins of the four studied *Brachypodium* species and of 54 ecotypes of *B. distachyon*. All samples were used in the comparative genomic analysis of the dehydrin genes. Asterisks indicate *B. distachyon* accessions additionally used in the dehydrin expression and drought-response phenotypic traits changes analyses (32 ecotypes). Diamonds indicate accessions additionally used in the phylogenetic signal analysis (30 ecotypes), Table S5: Topological congruence tests between (a) the *B. distachyon* nuclear species tree (Gordon et al. 2017) and (b) the *B. distachyon* plastome tree (Sancho et al. 2018) versus the *B. distachyon* dehydrin *Bdhn* tree. Test(s) were performed for significance of likelihood-score differences. KH: Kishino-Hasegawa test using normal approximation, two-tailed test. SH: Shimodaira-Hasegawa test using RELL bootstrap (one-tailed test). AU: Shimodaira Approximately Unbiased test. Values for KH/SH/AU tests are P values for the null hypothesis of no difference between trees. \*the null hypothesis is accepted. Number of bootstrap replicates = 1,000,000, Table S6: *Brachypodium distachyon* climate data. (a) Values of 19 current climate parameters retrieved from worldclim for the sampled localities of the studied *B. distachyon* ecotypes. Climate, climatic class of the *B. distachyon* ecotypes classified according to their PCA1 values (cold: > 2.5; mesic: (−2.5)–(2.5); warm: < −2.5; see Figure S5). (b) PCA1 and PCA2, coordinate values of the first and second PCA axes obtained from the climate PC analysis, Table S7: *Brachypodium distachyon* dehydrin expression data. Filtered and normalized transcripts per million (TPM) values of annotated dehydrins. Only four dehydrin genes (*Bdhn1a*, *Bdhn2*, *Bdhn3*, *Bdhn7*) were expressed in leaves of 31-days grown plants. Plants were subjected to drought (W: watered, D: Drought) and temperature (C: Cold, H: Hot) stress conditions (see text). Code indicates the sampling code used in the RNAseq analysis, Table S8: Summary statistics of dehydrin *Bdhn1a*, *Bdhn2*, *Bdhn3* and *Bdhn7* gene expressions under dry (D) vs. watered (W) conditions and comparative differential expression (DE) tests in *B. distachyon* ecotypes. (a) Kruskal-Wallis rank tests (D vs. W) for each *Bdhn* gene. (b) Wilcoxon pairwise tests of normalized TPM values across ecotypes, *p*-values were adjusted with the Benjamini–Hochberg (BH) procedure, controlling the false discovery rate, to correct for multiple comparisons; n. s., non significant,  $p \leq 0.05$ \* significant values are highlighted in bold, Table S9: Linear model (lm) regression analysis for comparative differential gene expressions of dehydrin *Bdhn* genes in the studied *B. distachyon* ecotypes. W (watered) and D (dry) conditions. Significant *p*-values ( $p \leq 0.001$ \*\*\*), Table S10: Comparative analysis of dehydrin genes showing upregulated expression under drought compared to watered conditions in *Brachypodium distachyon* and *Triticum aestivum*. Orthology between the differentially expressed genes in the two species was retrieved through Ensembl Plants using BioMart and Blast searches using orthologies previously established in Galvez et al., (2019) (§), Table S11: Summary statistics of 12 drought-response phenotypic traits [leaf\_rwc (relative water content in leaf); leaf\_wc (water content in leaf); lma (leaf mass per area); pro (leaf proline content); abvrgd (above ground biomass); blwrgd (below ground biomass); tlmass (total mass); rmr (root mass ratio); delta13c (carbon isotope, a proxy for lifetime integrated WUE); leafc (leaf carbon content); leafn (leaf nitrogen content); cn (leaf carbon/nitrogen ratio)] in dry (D) vs. watered (W) *Brachypodium distachyon* plants. (a) Kruskal-Wallis rank tests (W vs. D) for each phenotypic trait. (b) comparative pairwise Wilcoxon tests in the studied *B. distachyon* ecotypes; *p*-values were adjusted with the Benjamini–Hochberg (BH) procedure, controlling the false discovery rate, to correct for multiple comparisons; n, number of replicates. n. s., non significant, \*  $p \leq 0.05$ \*; significant values are highlighted in bold, Table S12: Linear model (lm) regression analysis for comparative *Brachypodium distachyon* *Bdhn* gene expressions and drought-induced phenotypic trait changes under total watered (W) and dry (D) conditions. Significant *p*-values ( $p \leq 0.05$  \*; 0.01 \*\*; 0.001 \*\*\*), Table S13: Phylogenetic signal of dehydrin gene expressions under watered (W) and dry (D) conditions, drought-induced phenotypic traits changes and climate niche variation assessed in (a) the *B. distachyon* nuclear-SNP tree and (b) the *B. distachyon* *Bdhn* tree using the *phylosig* option of the *phytools* R package. Blomberg's K and Pagel's lambda values close to one indicate phylogenetic signal and values close to zero phylogenetic independence. K, *p*-values based on 1000 randomizations; lambda, *p*-values based on the Likelihood Ratio test. Significant and marginal significant values are highlighted in bold.

**Author Contributions:** P.C., P.H., and S.G.-R. designed the study. M.A.D., S.G.-R., R.S. and D.L.D.M. generated the data. M.A.D., S.G.-R., F.A., R.S., B.C.-M., D.L.D.M. and P.C. analyzed the data. M.A.D., P.H. and P.C. wrote the draft manuscript. All authors contributed to the writing of the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Spanish Ministry of Science and Innovation grant number PID2019-108195GB-I00, European Social Fund/Spanish Aragón Government grant number A01-20R, Spanish Junta de Andalucía grant number P18-RT-992, USDA grant number NIFA-2011-67012- 30663. MD was funded by a Spanish Mineco FPI PhD fellowship. BCM was funded by Spanish Fundación ARAID.

**Data Availability Statement:** The complete protocol, input and output data, and Supplementary information are available at Github ([https://github.com/Bioflora/Brachypodium\\_dehydrins](https://github.com/Bioflora/Brachypodium_dehydrins) (accessed on 28 October 2021)).

**Acknowledgments:** We acknowledge the valuable suggestions of four reviewers which have helped to improve the quality of our manuscript. The bioinformatic and evolutionary analyses were performed at the University of Malaga (Spain) and Escuela Politécnica Superior de Huesca (Universidad de Zaragoza, Spain) laboratories, respectively. Plant growth experiments and RNA Sequencing were performed at M.I.T. The *B. sylvaticum* genome was used with permission under early release conditions of the DOE Joint Genome Institute.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hossain, M.A.; Wani, S.H.; Bhattacharjee, S.; Burritt, D.J.; Tran, L.-S.P. *Drought Stress Tolerance in Plants*; Springer: Cham, Switzerland, 2016; Volume 2.
- Des Marais, D.L.; Lasky, J.R.; Verslues, P.E.; Chang, T.Z.; Juenger, T.E. Interactive effects of water limitation and elevated temperature on the physiology, development and fitness of diverse accessions of *Brachypodium distachyon*. *New Phytol.* **2017**, *214*, 132–144. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mahajan, S.; Tuteja, N. Cold, salinity and drought stresses: An overview. *Arch. Biochem. Biophys.* **2005**, *444*, 139–158. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hanin, M.; Brini, F.; Ebel, C.; Toda, Y.; Takeda, S.; Masmoudi, K. Plant dehydrins and stress tolerance: Versatile proteins for complex mechanisms. *Plant Signal. Behav.* **2011**, *6*, 1503–1509. [\[CrossRef\]](#)
- Tommasini, L.; Svensson, J.T.; Rodriguez, E.M.; Wahid, A.; Malatrasi, M.; Kato, K.; Wanamaker, S.; Resnik, J.; Close, T.J. Dehydrin gene expression provides an indicator of low temperature and drought stress: Transcriptome-based analysis of Barley (*Hordeum vulgare* L.). *Funct. Integr. Genom.* **2008**, *8*, 387–405. [\[CrossRef\]](#)
- Reddy, S.K.; Liu, S.; Rudd, J.C.; Xue, Q.; Payton, P.; Finlayson, S.A.; Mahan, J.; Akhunova, A.; Holalu, S.V.; Lu, N. Physiology and transcriptomics of water-deficit stress responses in wheat cultivars TAM 111 and TAM 112. *J. Plant Physiol.* **2014**, *171*, 1289–1298. [\[CrossRef\]](#)
- Wang, Y.; Xu, H.; Zhu, H.; Tao, Y.; Zhang, G.; Zhang, L.; Zhang, C.; Zhang, Z.; Ma, Z. Classification and expression diversification of wheat dehydrin genes. *Plant Sci.* **2014**, *214*, 113–120. [\[CrossRef\]](#)
- Graether, S.P.; Boddington, K.F. Disorder and function: A review of the dehydrin protein family. *Front. Plant Sci.* **2014**, *5*, 1–12. [\[CrossRef\]](#) [\[PubMed\]](#)
- Verma, G.; Dhar, Y.V.; Srivastava, D.; Kidwai, M.; Chauhan, P.S.; Bag, S.K.; Asif, M.H.; Chakrabarty, D. Genome-wide analysis of rice dehydrin gene family: Its evolutionary conservedness and expression pattern in response to PEG induced dehydration stress. *PLoS ONE* **2017**, *12*, e0176399. [\[CrossRef\]](#)
- Riley, A.C.; Ashlock, D.A.; Graether, S.P. Evolution of the modular, disordered stress proteins known as dehydrins. *PLoS ONE* **2019**, *14*, e0211813. [\[CrossRef\]](#)
- Perdiguer, P.; Collad, C.; Soto, Á. Novel dehydrins lacking complete K-segments in Pinaceae. The exception rather than the rule. *Front. Plant Sci.* **2014**, *5*, 682. [\[CrossRef\]](#)
- Goday, A.; Jensen, A.B.; Culiáñez-Macià, F.A.; Albà, M.M.; Figueras, M.; Serratos, J.; Torrent, M.; Pagès, M. The maize abscisic acid-responsive protein Rab17 is located in the nucleus and interacts with nuclear localization signals. *Plant Cell* **1994**, *6*, 351–360. [\[CrossRef\]](#)
- Rosales, R.; Romero, I.; Escribano, M.I.; Merodio, C.; Sanchez-Ballesta, M.T. The crucial role of Φ- and K-segments in the in vitro functionality of *Vitis vinifera* dehydrin DHN1a. *Phytochemistry* **2014**, *108*, 17–25. [\[CrossRef\]](#) [\[PubMed\]](#)
- Strimbeck, G.R. Hiding in plain sight: The F segment and other conserved features of seed plant SKn dehydrins. *Planta* **2017**, *245*, 1061–1066. [\[CrossRef\]](#)
- Hennessy, F.; Nicoll, W.S.; Zimmermann, R.; Cheetham, M.E.; Blatch, G.L. Not all J domains are created equal: Implications for the specificity of Hsp40-Hsp70 interactions. *Protein Sci.* **2005**, *14*, 1697–1709. [\[CrossRef\]](#) [\[PubMed\]](#)
- Malik, A.A.; Veltri, M.; Boddington, K.F.; Singh, K.K.; Graether, S.P. Genome analysis of conserved dehydrin motifs in vascular plants. *Front. Plant Sci.* **2017**, *8*, 1–18. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kosová, K.; Vítámvás, P.; Prášil, I.T. Wheat and barley dehydrins under cold, drought, and salinity—What can LEA-II proteins tell us about plant stress response? *Front. Plant Sci.* **2014**, *5*, 343. [\[CrossRef\]](#)
- Suprunova, T.; Krugman, T.; Fahima, T.; Chen, G.; Shams, I.; Korol, A.; Nevo, E. Differential expression of dehydrin genes in wild barley, *Hordeum spontaneum*, associated with resistance to water deficit. *Plant Cell Environ.* **2004**, *27*, 1297–1308. [\[CrossRef\]](#)
- Karami, A.; Shahbazi, M.; Niknam, V.; Shobbar, Z.S.; Tafreshi, R.S.; Abedini, R.; Mabood, H.E. Expression analysis of dehydrin multigene family across tolerant and susceptible barley (*Hordeum vulgare* L.) genotypes in response to terminal drought stress. *Acta Physiol. Plant* **2013**, *35*, 2289–2297. [\[CrossRef\]](#)
- Yu, Z.; Wang, X.; Zhang, L. Structural and functional dynamics of dehydrins: A plant protector protein under abiotic stress. *Int. J. Mol. Sci.* **2018**, *19*, 3420. [\[CrossRef\]](#)
- Ly, A.; Su, L.; Liu, X.; Xing, Q.; Huang, B.; An, Y.; Zhou, P. Characterization of Dehydrin protein, CdDHN4-L and CdDHN4-S, and their differential protective roles against abiotic stress in vitro. *BMC Plant Biol.* **2018**, *18*, 299. [\[CrossRef\]](#)
- Liu, Y.; Song, Q.; Li, D.; Yang, X.; Li, D. Multifunctional roles of plant dehydrins in response to environmental stresses. *Front. Plant Sci.* **2017**, *8*, 1018. [\[CrossRef\]](#) [\[PubMed\]](#)
- Drira, M.; Saibi, W.; Brini, F.; Gargouri, A.; Masmoudi, K.; Hanin, M. The K-segments of the wheat dehydrin DHN-5 are essential for the protection of lactate dehydrogenase and β-glucosidase activities in vitro. *Mol. Biotechnol.* **2013**, *54*, 643–650. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Xing, X.; Liu, Y.; Kong, X.; Liu, Y.; Li, D. Overexpression of a maize dehydrin gene, ZmDHN2b, in tobacco enhances tolerance to low temperature. *Plant Growth Regul.* **2011**, *65*, 109–118. [\[CrossRef\]](#)
25. Xu, J.; Zhang, Y.X.; Wei, W.; Han, L.; Guan, Z.Q.; Wang, Z.; Chai, T.Y. BjDHNs confer heavy-metal tolerance in plants. *Mol. Biotechnol.* **2008**, *38*, 91–98. [\[CrossRef\]](#)
26. Hara, M.; Kondo, M.; Kato, T. A KS-type dehydrin and its related domains reduce Cu-promoted radical generation and the histidine residues contribute to the radical-reducing activities. *J. Exp. Bot.* **2013**, *64*, 1615–1624. [\[CrossRef\]](#)
27. Koag, M.C.; Fenton, R.D.; Wilkens, S.; Close, T.J. The binding of maize DHN1 to lipid vesicles. Gain of structure and lipid specificity. *Plant Physiol.* **2003**, *131*, 309–316. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Hughes, S.L.; Scharf, V.; Malcolmson, J.; Hogarth, K.A.; Martynowicz, D.M.; Tralman-Baker, E.; Patel, S.N.; Graether, S.P. The importance of size and disorder in the cryoprotective effects of dehydrins. *Plant Physiol.* **2013**, *163*, 1376–1386. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Scholthof, K.-B.G.; Irigoyen, S.; Catalán, P.; Mandadi, K.K. *Brachypodium*: A monocot grass model system for plant biology. *Plant Cell* **2018**, *30*, 1673–1694. [\[CrossRef\]](#)
30. Gordon, S.P.; Contreras-Moreira, B.; Levy, J.J.; Djamei, A.; Czedik-Eysenberg, A.; Tartaglio, V.S.; Session, A.; Martin, J.; Cartwright, A.; Katz, A.; et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat. Commun.* **2020**, *11*, 3670. [\[CrossRef\]](#)
31. Steinwand, M.A.; Young, H.A.; Bragg, J.N.; Tobias, C.M.; Vogel, J.P. *Brachypodium sylvaticum*, a Model for Perennial Grasses: Transformation and Inbred Line Development. *PLoS ONE* **2013**, *8*, e75180. [\[CrossRef\]](#)
32. Lee, K.; Kang, H. Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. *Mol. Cells* **2016**, *39*, 179–185. [\[CrossRef\]](#)
33. Ryu, H.; Cho, Y.G. Plant hormones in salt stress tolerance. *J. Plant Biol.* **2015**, *58*, 147–155. [\[CrossRef\]](#)
34. Yoon, J.S.; Kim, J.Y.; Lee, M.B.; Seo, Y.W. Over-expression of the *Brachypodium* ASR gene, BdASR4, enhances drought tolerance in *Brachypodium distachyon*. *Plant Cell Rep.* **2019**, *38*, 1109–1125. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Filiz, E.; Ozyigit, I.I.; Tombuloglu, H.; Koc, I. In silico comparative analysis of LEA (Late Embryogenesis Abundant) proteins in *Brachypodium distachyon* L. *Plant Omi. J.* **2013**, *6*, 433–440.
36. Rodríguez, E.M.; Svensson, J.T.; Malatrasi, M.; Choi, D.W.; Close, T.J. Barley Dhn13 encodes a KS-type dehydrin with constitutive and stress responsive expression. *Theor. Appl. Genet.* **2005**, *110*, 852–858. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Gordon, S.P.; Contreras-Moreira, B.; Woods, D.P.; Des Marais, D.L.; Burgess, D.; Shu, S.; Stritt, C.; Roulin, A.; Schackwitz, W.; Tyler, L.; et al. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nat. Commun.* **2017**, *8*, 2184. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Sancho, R.; Cantalapiedra, C.P.; López-Alvarez, D.; Gordon, S.P.; Vogel, J.P.; Catalán, P.; Contreras-Moreira, B. Comparative plastome genomics and phylogenomics of *Brachypodium*: Flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol.* **2018**, *218*, 1631–1644. [\[CrossRef\]](#)
39. Sancho, R.; Catalán, P.; Contreras-Moreira, B.; Juenger, T.E.; Des Marais, D.L. Patterns of gene co-expression under water-deficit treatments and pan-genome occupancy in *Brachypodium distachyon*. *bioRxiv* **2021**. [\[CrossRef\]](#)
40. Gálvez, S.; Mérida-García, R.; Camino, C.; Borrill, P.; Abrouk, M.; Ramírez-González, R.H.; Biyikliglu, S.; Amil-Ruiz, F.; Dorado, G.; Budak, H.; et al. Hotspots in the genomic architecture of field drought responses in wheat as breeding targets. *Funct. Integr. Genom.* **2019**, *19*, 295–309. [\[CrossRef\]](#)
41. Chu, C.; Wang, S.; Paetzold, L.; Wang, Z.; Hui, K.; Rudd, J.C.; Xue, Q.; Ibrahim, A.M.H.; Metz, R.; Johnson, C.D.; et al. RNA-seq analysis reveals different drought tolerance mechanisms in two broadly adapted wheat cultivars ‘TAM 111’ and ‘TAM 112’. *Sci. Rep.* **2021**, *11*, 4301. [\[CrossRef\]](#)
42. Catalán, P.; Müller, J.; Hasterok, R.; Jenkins, G.; Mur, L.A.; Langdon, T.; Betekhtin, A.; Siwinska, D.; Pimentel, M.; López-Alvarez, D. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann. Bot.* **2012**, *109*, 385–405. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Choi, D.W.; Zhu, B.; Close, T.J. The barley (*Hordeum vulgare* L.) dehydrin multigene family: Sequences, allele types, chromosome assignments, and expression characteristics of 11 Dhn genes of cv Dicktoo. *Theor. Appl. Genet.* **1999**, *98*, 1234–1247. [\[CrossRef\]](#)
44. Díaz-Pérez, A.; López-Alvarez, D.; Sancho, R.; Catalán, P. 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. *Mol. Phylogenet. Evol.* **2018**, *127*, 256–271. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Jiménez-Bremont, J.F.; Maruri-López, I.; Ochoa-Alfaro, A.E.; Delgado-Sánchez, P.; Bravo, J.; Rodríguez-Kessler, M. LEA Gene Introns: Is the Intron of Dehydrin Genes a Characteristic of the Serine-Segment? *Plant Mol. Biol. Report.* **2013**, *31*, 128–140. [\[CrossRef\]](#)
46. Martínez, L.M.; Fernández-ocaña, A.; Rey, P.J.; Salido, T.; Amil-ruiz, F.; Manzaneda, A.J. Variation in functional responses to water stress and differentiation between natural allopolyploid populations in the *Brachypodium distachyon* species complex. *Ann. Bot.* **2018**, *121*, 1369–1382. [\[CrossRef\]](#)
47. Martínez, L.M. *Análisis de la Variación Genética, Respuesta Funcional y Expresión Génica Frente a Diferentes Tipos de Estrés en el Complejo de Ploidía Brachypodium Distachyon (Poaceae)*; Universidad de Jaén: Jaén, Spain, 2020.
48. Fisher, L.H.; Han, J.; Corke, F.M.; Akinyemi, A.; Didion, T.; Nielsen, K.K.; Doonan, J.H.; Mur, L.A.; Bosch, M. Linking Dynamic Phenotyping with Metabolite Analysis to Study Natural Variation in Drought Responses of *Brachypodium distachyon*. *Front. Plant Sci.* **2016**, *7*, 1751. [\[CrossRef\]](#)
49. Yamaguchi-Shinozaki, K.; Shinozaki, K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* **2005**, *10*, 88–94. [\[CrossRef\]](#)

50. Cui, X.Y.; Gao, Y.; Guo, J.; Yu, T.F.; Zheng, W.J.; Liu, Y.W.; Chen, J.; Xu, Z.S.; Ma, Y.Z. BES/BZR transcription factor TaBZR2 positively regulates drought responses by activation of TaGST1. *Plant Physiol.* **2019**, *180*, 605–620. [\[CrossRef\]](#)
51. Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: An overview. *Physiol. Mol. Biol. Plants* **2013**, *19*, 307–321. [\[CrossRef\]](#)
52. Khan, S.A.; Li, M.Z.; Wang, S.M.; Yin, H.J. Revisiting the role of plant transcription factors in the battle against abiotic stress. *Int. J. Mol. Sci.* **2018**, *19*, 1634. [\[CrossRef\]](#)
53. Ciftci-Yilmaz, S.; Morsy, M.R.; Song, L.; Coutu, A.; Krizek, B.A.; Lewis, M.W.; Warren, D.; Cushman, J.; Connolly, E.L.; Mittler, R. The EAR-motif of the Cys2/His2-type zinc finger protein Zat7 plays a key role in the defense response of Arabidopsis to salinity stress. *J. Biol. Chem.* **2007**, *282*, 9260–9268. [\[CrossRef\]](#)
54. Cernusak, L.A.; Ubierna, N.; Winter, K.; Holtum, J.A.M.; Marshall, J.D.; Farquhar, G.D. Environmental and physiological determinants of carbon isotope discrimination in terrestrial plants. *New Phytol.* **2013**, *200*, 950–965. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Manzaneda, A.J.; Rey, P.J.; Anderson, J.T.; Raskin, E.; Weiss-Lehman, C.; Mitchell-Olds, T. 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* **2015**, *69*, 2689–2704. [\[CrossRef\]](#)
56. Revell, L.J.; Harmon, L.J.; Collar, D.C. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **2008**, *57*, 591–601. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Hu, S.; Ding, Y.; Zhu, C. Sensitivity and Responses of Chloroplasts to Heat Stress in Plants. *Front. Plant Sci.* **2020**, *11*, 1–11. [\[CrossRef\]](#)
58. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, 1178–1186. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Howe, K.L.; Contreras-Moreira, B.; De Silva, N.; Maslen, G.; Akanni, W.; Allen, J.; Alvarez-Jarreta, J.; Barba, M.; Bolser, D.M.; Cambell, L.; et al. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* **2020**, *48*, D689–D695. [\[CrossRef\]](#)
60. Agarwala, R.; Barrett, T.; Beck, J.; Benson, D.A.; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J.R.; Bryant, S.H.; Canese, K.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [\[CrossRef\]](#)
61. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [\[CrossRef\]](#)
62. Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* **2012**, *7*, 1511–1522. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Xu, J.; McPartlon, M.; Li, J. Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat. Mach. Intell.* **2021**, *3*, 601–609. [\[CrossRef\]](#)
64. Wang, J.; Youkharibache, P.; Zhang, D.; Lanczycki, C.J.; Geer, R.C.; Madej, T.; Phan, L.; Ward, M.; Lu, S.; Marchler, G.H.; et al. ICn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* **2020**, *36*, 131–135. [\[CrossRef\]](#)
65. 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.; et al. RSAT 2018: Regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res.* **2018**, *46*, W209–W214. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Ksouri, N.; Castro-Mondragón, J.A.; Montardit-Tarda, F.; van Helden, J.; Contreras-Moreira, B.; Gogorcena, Y. Tuning promoter boundaries improves regulatory motif discovery in nonmodel plants: The peach example. *Plant Physiol.* **2021**, *185*, 1242–1258. [\[CrossRef\]](#)
67. Castro-Mondragon, J.A.; Jaeger, S.; Thieffry, D.; Thomas-Chollier, M.; Van Helden, J. RSAT matrix-clustering: Dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res.* **2017**, *45*, e119. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [\[CrossRef\]](#)
69. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#)
70. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Rehman, S.; Jørgensen, B.; Aziz, E.; Batool, R.; Naseer, S.; Rasmussen, S.K. Genome Wide Identification and Comparative Analysis of the Serpin Gene Family in *Brachypodium* and Barley. *Plants* **2020**, *9*, 1439. [\[CrossRef\]](#)
72. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [\[CrossRef\]](#)
73. Murrell, B.; Weaver, S.; Smith, M.D.; Wertheim, J.O.; Murrell, S.; Aylward, A.; Eren, K.; Pollner, T.; Martin, D.P.; Smith, D.M.; et al. Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **2015**, *32*, 1365–1371. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **2012**, *8*, e1002764. [\[CrossRef\]](#) [\[PubMed\]](#)
75. Kosakovsky Pond, S.L.; Frost, S.D.W.; Muse, S.V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **2005**, *21*, 676–679. [\[CrossRef\]](#) [\[PubMed\]](#)

76. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777. [[CrossRef](#)]
77. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
78. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [[CrossRef](#)] [[PubMed](#)]
79. Pimentel, H.; Bray, N.L.; Puente, S.; Melsted, P.; Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* **2017**, *14*, 687–690. [[CrossRef](#)]
80. IBI Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **2010**, *463*, 763–768. [[CrossRef](#)]
81. Konstantinov, D.K.; Zubairova, U.S.; Ermakov, A.A.; Doroshkov, A.V. Comparative transcriptome profiling of a resistant vs. susceptible bread wheat (*Triticum aestivum* L.) cultivar in response to water deficit and cold stress. *PeerJ* **2021**, *9*, e11428. [[CrossRef](#)] [[PubMed](#)]
82. Appels, R.; Eversole, K.; Stein, N.; Feuillet, C.; Keller, B.; Rogers, J.; Pozniak, C.J.; Choulet, F.; Distelfeld, A.; Poland, J.; et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, eaar7191. [[CrossRef](#)]
83. Alaux, M.; Rogers, J.; Letellier, T.; Flores, R.; Alfama, F.; Pommier, C.; Mohellibi, N.; Durand, S.; Kimmel, E.; Michotey, C.; et al. Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data. *Genome Biol.* **2018**, *19*, 111. [[CrossRef](#)] [[PubMed](#)]
84. Dray, S.; Dufour, A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **2007**, *22*, 1–20. [[CrossRef](#)]
85. Swofford, D.L. *PAUP\*. Phylogenetic Analysis Using Parsimony(\*and Other Methods), Version 4b10*; Sinauer Associates: Sunderland, MA, USA, 2003; ISBN 0022-2844.
86. Blomberg, S.P.; Garland, T.; Ives, A.R. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* **2003**, *57*, 717–745. [[CrossRef](#)] [[PubMed](#)]
87. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **1999**, *401*, 877–884. [[CrossRef](#)] [[PubMed](#)]
88. Revell, L.J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **2012**, *3*, 217–223. [[CrossRef](#)]



## CONCLUSIONES

1. Los estudios filogenéticos y citogenéticos de representantes de las 21 especies descritas del género *Brachypodium* empleando datos combinados de secuenciación a baja cobertura (genome skimming), tamaños genómicos, recuentos cromosómicos y cariotipado FISH, han permitido detectar tres nuevos citotipos hexaploides para poblaciones de *B. retusum*, *B. phoenicoides* y *B. rupestre*, la obtención de dos familias ribotípicas 5S entre las 20 especies y diferentes subgenomas ancestrales y recientes para los alopoliploides *B. retusum* y *B. phoenicoides*, y una filogenia robusta basada en tres moléculas (plastoma, y genes rDNA 35S y 5S) que muestra la divergencia de linajes y subgenomas ancestrales outcore para *B. stacei*, *B. mexicanum*, *B. boissieri*, *B. retusum* y *B. phoenicoides*, y linajes y subgenomas recientes del clado core perenne para *B. arbuscula*, *B. sylvaticum* oriental y las endémicas afro-asiáticas *B. flexum*, *B. bolusii*, *B. madagascariense* y *B. kawakamii*, *B. sylvaticum* occidental, *B. retusum*, *B. pinnatum*, *B. phoenicoides* y *B. rupestre*.
2. Los análisis filogenéticos revelan la presencia de dos linajes divergentes compuestos de diversos micro-taxa en la especie paleártica diploide *B. sylvaticum* s.l. hasta ahora considerada un grupo monofilético, uno oriental más ancestral, con mayores tasas de diversificación y distribución panHimalaya-Pacífico, y otro más reciente occidental, con una mayor homogenización y distribución Euromediterránea-Siberiana.
3. El análisis del ADN repetitivo nuclear en una representación de todos los linajes evolutivos de *Brachypodium* corroboró la contribución de los elementos repetitivos a la diversificación evolutiva de sus tamaños genómicos. Los elementos repetitivos más abundantes en los genomas de *Brachypodium* fueron los retrotransposones Retand y Tekay. El repeteoma mostró una resolución filogenética en *Brachypodium*, recuperando la divergencia de los principales linajes del género. Los patrones ribotípicos 5S fueron consistentes con el número y la naturaleza de los subgenomas de las especies diploides y poliploides. Las expansiones y contracciones del repeteoma han sido responsables de las tres respuestas diferentes al choque poliploide en *B. mexicanum* (expansión exacerbada por proliferación de transposones), *B. hybridum* (patrón aditivo de repeteomas progenitores por estabilidad genómica postpoliploide) y *B. phoenicoides* - *B. pinnatum* (reducción genómica por pérdida de elementos repetitivos mediante posibles recombinaciones).
4. El análisis morfoanatómico de citotipos alopoliploides de *B. retusum* y *B. phoenicoides* ha detectado una amplia diversidad morfológica entre los dos niveles ploidía descritos para ambos taxones (tetraploide y hexaploide). Siete caracteres vegetativos y reproductores cuantitativos (anchuras de hojas de innovación y bandera y número de venas, número de espiguillas por inflorescencia, y longitud del lema) discriminan a los citotipos 4x y 6x de *B. retusum*, y cinco (longitudes de glumas, número de anteras, longitud de hoja de innovación y número de nodos del culmo) a los de *B. phoenicoides*. Los individuos tetraploides muestran en general valores menores para estos caracteres que los hexaploides, probablemente como resultado de las distintas dosis génicas y la heterosis. Los datos sugieren la potencial existencia de procesos de micro-especiación en los morfotipos crípticos de *B. phoenicoides*, siendo menos claros en *B. retusum*.

5. El estudio genómico comparado y filogenético de genes codificadores de proteínas relacionadas con el estrés hídrico, dehidrinas, en las cuatro especies modelo del género *Brachypodium*, ha identificado 10 genes de dehidrinas en *Brachypodium* (Bdhn) que se encuentran bajo presión selectiva y se caracterizan por presentar motivos cis-reguladores. El análisis de expresión diferencial en 54 ecotipos de *B. distachyon* muestra que 4 de los 10 genes Bdhn se expresan en tejido foliar, y son inducidos en condiciones de estrés hídrico, mostrando sobreexpresión en ecotipos tolerantes a la sequía.

## CONCLUSIONS

1. Phylogenetic and cytogenetic studies on representatives of the 21 species described for the genus *Brachypodium* using combined data from genome skimming sequencing, genomic sizes, chromosome counts, and FISH karyotyping allowed the detection of three new hexaploid cytotypes for *B. retusum*, *B. phoenicoides* and *B. rupestre*, two rDNA 5S ribotypic families among the 20 species, and different ancestral and recent subgenomes for the allopolyploids *B. retusum* and *B. phoenicoides*. In addition, a robust phylogeny based on three molecules (plastome, and rDNA 35S and 5S genes) has been constructed, showing the divergence of outcore ancestral lineages and subgenomes for *B. stacei*, *B. mexicanum*, *B. boissieri*, *B. retusum* and *B. phoenicoides*, and recent lineages and subgenomes of the core perennial clade for *B. arbuscula*, eastern *B. sylvaticum* and the Afro-Asian endemics *B. flexum*, *B. bolusii*, *B. madagascariense* and *B. kawakamii*, western *B. sylvaticum*, *B. retusum*, *B. pinnatum*, *B. phoenicoides* and *B. rupestre*.
2. The phylogenetic analyses revealed the existence of two divergent lineages composed of diverse micro-taxa within the diploid Palearctic species *B. sylvaticum* s.l. considered so far to be a monophyletic group, an eastern ancestral lineage with higher rates of diversification and a pan-Himalayan-Pacific distribution, and a more recent western lineage with greater homogenization and a Euro-Mediterranean-Siberian distribution.
3. Analysis of nuclear repetitive DNA in a representation of all *Brachypodium* evolutionary lineages corroborated the contribution of repetitive elements to evolutionary diversification of their genomic sizes. The most abundant repetitive elements in the *Brachypodium* genomes were the Retand and Tekay retrotransposons. The repeateome showed a phylogenetic resolution in *Brachypodium*, recovering the divergence of the main lineages of the genus. The 5S ribotypic patterns are consistent with the number and nature of the subgenomes of diploid and polyploid species. Repeteome expansions and contractions have caused three different responses to the polyploid shock in *B. mexicanum* (exacerbated expansion by transposon proliferation), *B. hybridum* (additive patterning of parental repeateomes by post-polyploid genomic stability), and *B. phoenicoides* - *B. pinnatum* (genomic reduction by loss of repetitive elements through potential recombinations).
4. The morphoanatomical analysis of the allopolyploid cytotypes of *B. retusum* and *B. phoenicoides* has detected a large morphological diversity between the two ploidy levels of both taxa (tetraploid and hexaploid). Seven diagnostic quantitative vegetative and reproductive characters (innovation and flag leaf widths and number of veins, number of spikelets per inflorescence, and lemma length) discriminate the 4x and 6x cytotypes of *B. retusum*, and five (glume lengths, number of anthers, innovator leaf length, and number of culm nodes) those of *B. phoenicoides*. Tetraploid individuals generally show lower

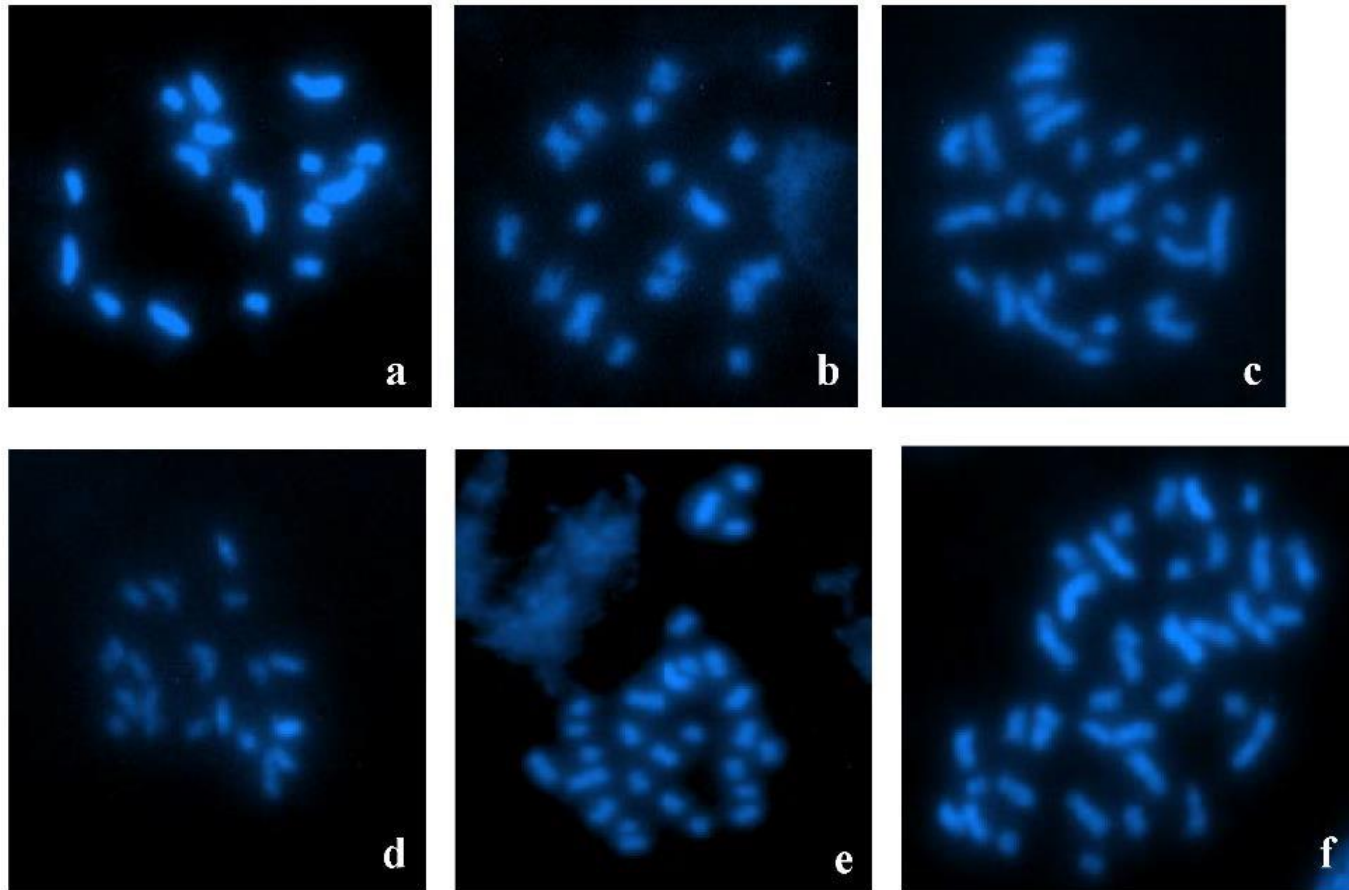


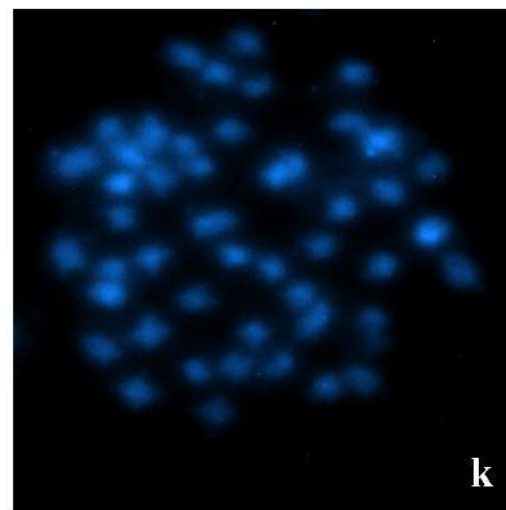
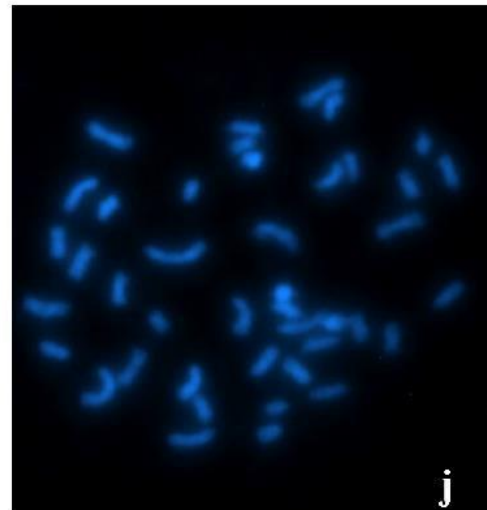
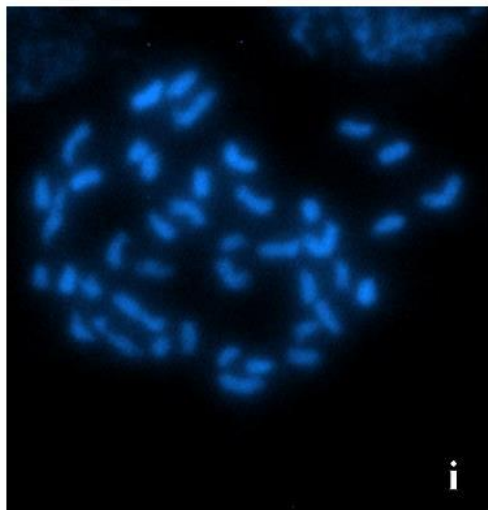
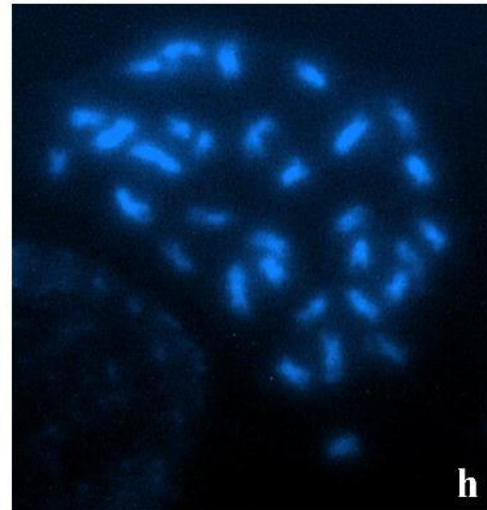
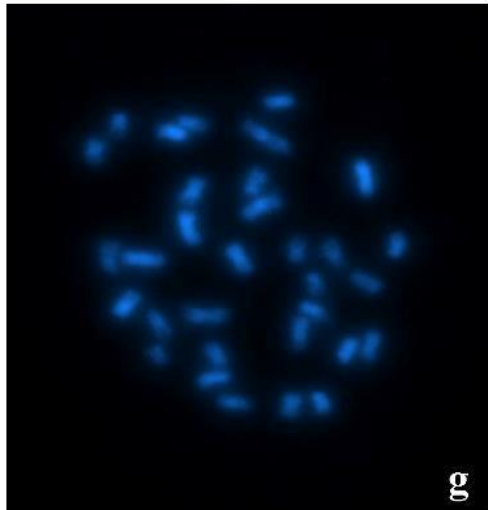
values than hexaploids for these characters, probably as a result of different genetic doses and heterosis. The data suggest the potential existence of micro-speciation processes in the cryptic morphotypes of *B. phoenicoides*, being less clear in *B. retusum*.

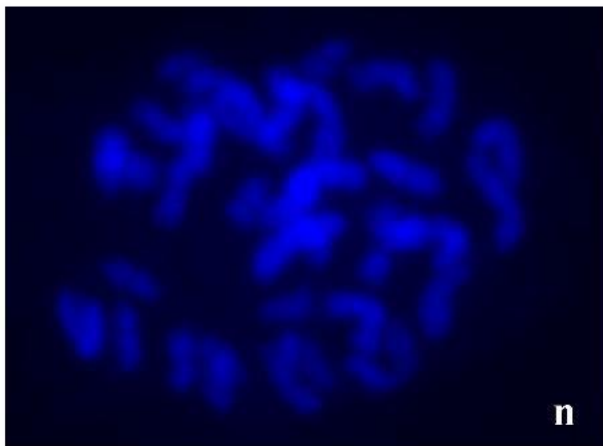
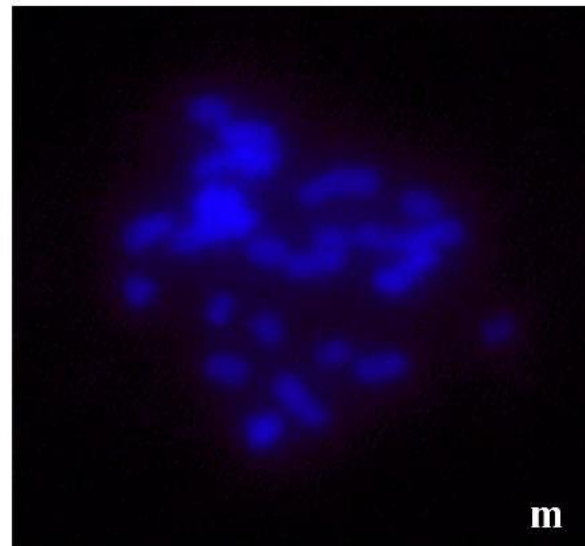
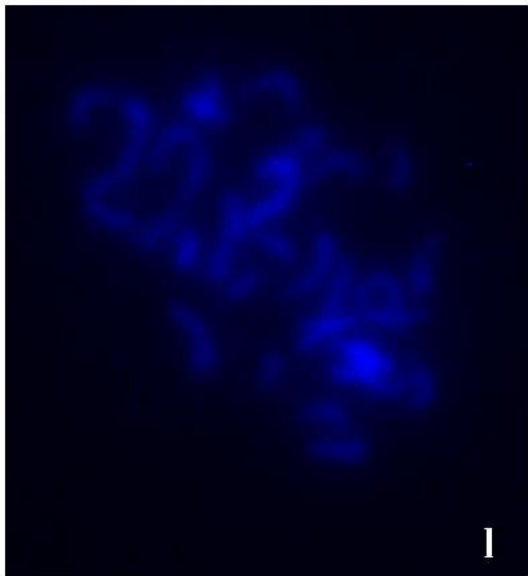
5. The comparative genomic and phylogenetic study of genes encoding proteins related to water stress response, dehydrins, in the four model species of the genus *Brachypodium* have identified 10 dehydrin genes in *Brachypodium* (Bdhn), which are under positive selection and are characterized by presenting cis-regulatory motifs. Analysis of differential gene expression in 54 ecotypes of *B. distachyon* shows that 4 of the 10 Bdhn genes are expressed in leaf tissue, and are induced under water stress conditions, showing overexpression in drought tolerant ecotypes.

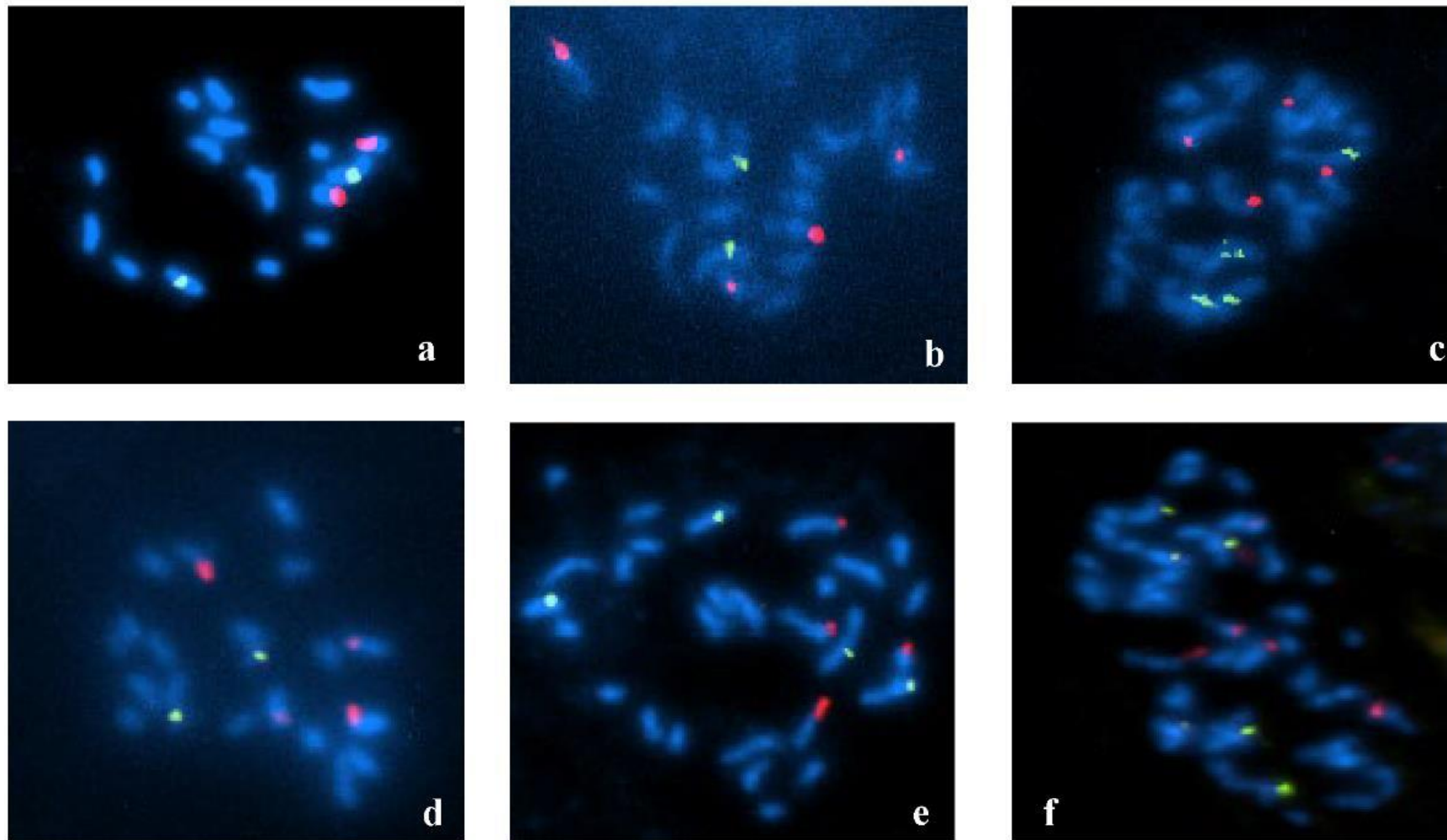
**SUPPLEMENTARY DATA.**

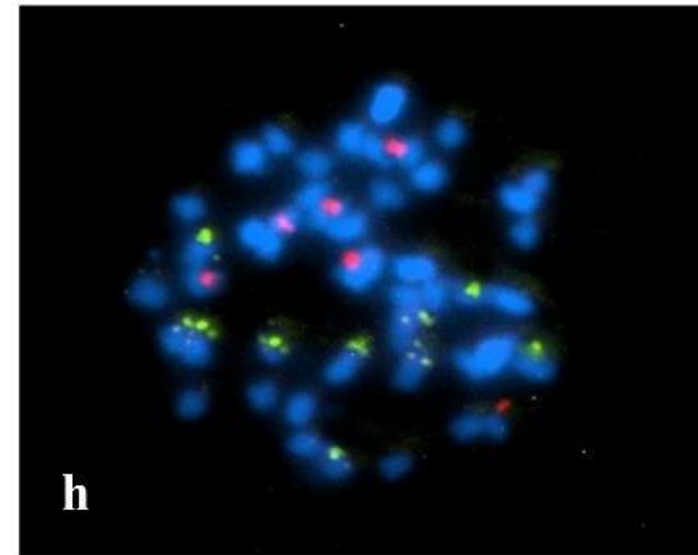
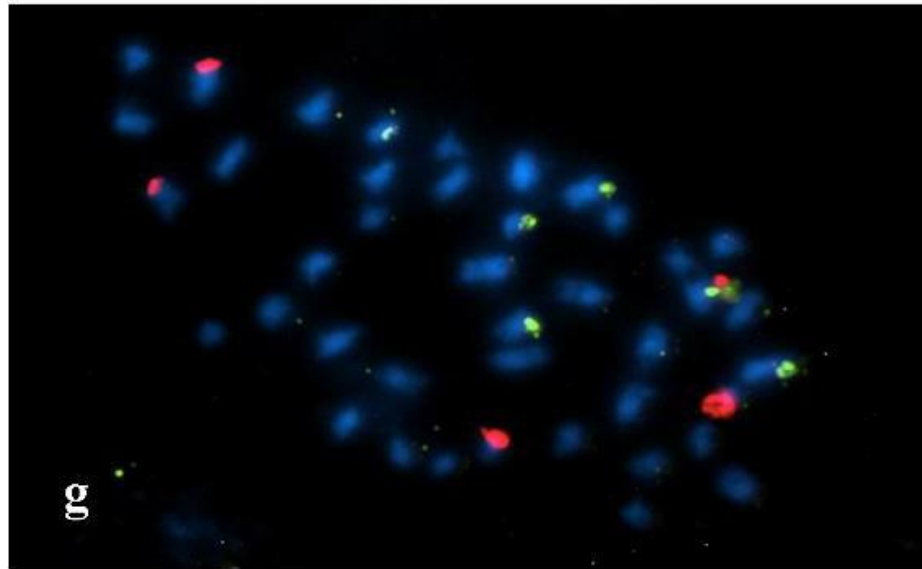
**Supplementary figures and tables Chapter 2**



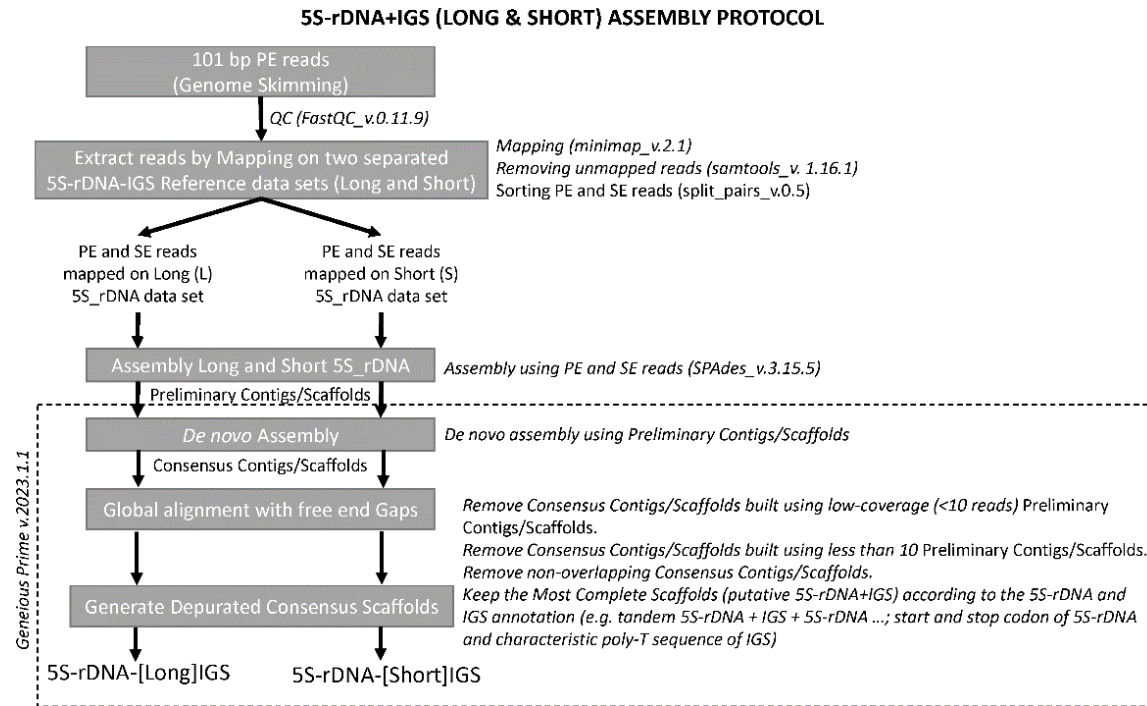








**Supplementary Figure S1.** (A) DAPI-stained somatic metaphase chromosome counts of *Brachypodium* samples **a**: *B. pinnatum* 2n = 16 (Bpinnatum 502). **b**: *B. pinnatum* 2n = 18 (Bpinnatum 515). **c**: *B. pinnatum* 2n = 28 (Bpinnatum 513). **d**: *B. rupestre* 2n = 18 (Brupestre434). **e**: *B. rupestre* 2n = 28 (Brupestre444). **f**: *B. rupestre* 2n = 38 (Brupestre 182). **g**: *B. retusum* 2n = 32 (Bretusum 501). **h**: *B. retusum* 2n = 32 (Bretusum 555). **i**: *B. retusum* 2n = 42 (Bretusum 403). **j**: *B. retusum* 2n = 42 (Bretusum 557). **k**: *B. boissieri* 2n = 48 (Bboissieri 15). **l**: *B. phoenicoides* 2n = 28 (Bphoenicoides 510). **m**: *B. phoenicoides* 2n = 28 (Bphoenicoides 604). **n**: *B. phoenicoides* 2n = 38 (Bphoenicoides 553). **o**: *B. phoenicoides* 2n = 38 (Bphoenicoides 554). (B) FISH of rDNA 25S (red) and 5S (green) probes on somatic metaphase chromosomes of *Brachypodium* samples and number of 25S and 5S loci detected. **a**: *B. pinnatum* 2n = 16 (Bpinnatum 502; 25S:2, 5S:2); **b**: *B. pinnatum* 2n = 18 (Bpinnatum 505; 25S:4, 5S:2); **c**: *B. pinnatum* 2n = 28 (Bpinnatum 514; 25S:4, 5S:4); **d**: *B. rupestre* 2n = 18 (Brupestre 512; 25S:4, 5S:2); **e**: *B. rupestre* 2n = 28 (Brupestre 5; 25S:4, 5S:4S); **f**: *B. rupestre* 2n = 38 (Brupestre 605; 25S:5-6, 5S:6); **g**: *B. retusum* 2n = 42 (Bretusum 561; 25S:5, 5S:5-6); **i**: *B. boissieri* 2n = 48 (Bboissieri 15; 25S:6, 5S:10).



**Supplementary Figure S2.** Workflow of a customized assembly protocol for the generation of the *Brachypodium* rDNA 5S gene Short (5S-S) and Long (5S-L) sequence data sets.



**Supplementary Table S1. (A)** Sampling data for 21 *Brachypodium* taxa under study indicating taxon name and authority, population code, germplasm/garden/herbarium sample source, location (geographic provenance), estimated genome size (2C/pg value), chromosome number (2n), inferred ploidy level, number of rDNA 25S and 5S loci detected through FISH analysis, and data source. Germplasm/garden/herbarium sample source: EPSHU (High Polytechnic School of Huesca, University of Zaragoza, Spain), INIA (Spain), JGI (Joint Genome Institute, USA), RIKEN (Japan), USDA (USA), Herbaria B (Berlin), FI (Firenze), G (Geneva), LD (Lund), M (München), VLA (Vladivostok). Data source: GS (genome size), 2n (chromosome number), 35S and 5S (rDNA 35S and 5S loci). New data from this study are highlighted in bold. References are given for previously published data. Hyphen. unknown data; s. d.. standard deviation. See Supplementary Table S1 (B) for complementary information.

Species	Code	Germplasm/garden /herbarium sample source	Location	Genome size (pg) ± sd	Chrom. No.	Ploidy	25S loci	5S loci	Data source
<i>B. distachyon</i> (L.) P. Beauv.	Bdistachyon_Bd21	JGI	Iraq: Salakudin	0.631±0	10	2x	-	-	GS and 2n: Catalán et al. 2012.
<i>B. stacei</i> Catalán. Joch. Müll.. L.A.J. Mur & T. Langdon	Bstacei_ABR114	JGI	Spain: Balearic islands	0.564±0	20	2x	-	-	GS and 2n: Wolny and Hasterok. 2009; Catalán et al.. 2012; Gordon et al.. 2020.
<i>B. hybridum</i> Catalán. Joch. Müll.. Hasterok & G. Jenkins	Bhybridum_ABR113	JGI	Portugal: Lisbon	1.265±0	30	4x	-	-	GS and 2n: Catalán et al.. 2012.
	Bhybridum_Bhyb26	JGI	Spain: Jaen		30	4x	-	-	GS and 2n: Gordon et al.. 2020.
<i>B. arbuscula</i> Gay ex Knoche	Barbuscula405	EPSHU	Spain: Tenerife	-	-	-	-	-	-
	Barbuscula502	INIA	Spain: La Gomera	0.713±0.004	18	2x	-	-	GS: Sancho et al.. 2022. 2n: Robertson. 1981
<i>B. boissieri</i> Nyman	Bboissieri3	EPSHU	Spain: Granada	3.236±0.072	48	6x	-	-	GS and 2n: Sancho et al.. 2022
	Bboissieri9	EPSHU	Spain: Malaga	<b>3.133±0.074</b>	<b>48</b>	<b>6x</b>	<b>6</b>	<b>8-9-10</b>	-

Supplementary data Chapter 2

	<i>Bboissieri</i> 10	EPSHU	Spain: Granada	<b>3.152±0.04</b>	<b>48</b>	<b>6x</b>	<b>6</b>	<b>8-9-10</b>	-
	<i>Bboissieri</i> 13	EPSHU	Spain: Granada	<b>3.116±0.067</b>	<b>48</b>	<b>6x</b>	<b>6</b>	<b>8-9-10</b>	-
	<i>Bboissieri</i> 14	EPSHU	Spain: Granada	<b>3.111±0.03</b>	<b>48</b>	<b>6x</b>	-	-	-
	<i>Bboissieri</i> 15	EPSHU	Spain: Granada	<b>3.149±0.032</b>	<b>48</b>	<b>6x</b>	<b>6</b>	<b>10</b>	-
<i>B. bolussi</i> Stapf	<i>Bbolussi</i> 046	EPSHU	South Africa: Natal	-	-	-	-	-	-
<i>B. flexum</i> Nees	<i>Bflexum</i> 038	Herbarium LD	South Africa: KwaZulu Natal	-	-	-	-	-	-
	<i>Bflexum</i> 046	Herbarium M	South Africa: KwaZulu Natal	-	-	-	-	-	-
<i>B. genuense</i> (DC.) Roem. & Schult.	<i>Bgenuense</i>	Herbarium FI	Italy: Abruzzo	-	18	2x	-	-	-
<i>B. glaucovirens</i> (Murb.) T. Durand & B.D. Jacks.	<i>Bglaucovirens</i>	EPSHU	Greece: Crete	0.886±0.012	16	2x	2	2	-
<i>B. kawakamii</i> Hayata	<i>Bkawakamii</i>	EPSHU	Taiwan: Kaohsiung	-	-	-	-	-	2n: Lucchese, F.. 1988
<i>B. kurilense</i> (Prob.) Prob.	<i>Bkurilense</i> 9	Herbarium B	Russia: Kuriles. Iturup	-	-	-	-	-	2n: Shi et al.. 1993. GS, 35S and 5S: Wolny & Hasterok. 2009.
<i>B. madagascariense</i> A. Camus	<i>Bmadagascariense</i>	EPSHU	Madagascar: Ankaratra	-	-	-	-	-	-
<i>B. mexicanum</i> (Roem. & Schult.) Link	<i>Bmexicanum</i> 347	EPSHU	Mexico: Hidalgo	3.774±0.033	40	4x	-	-	GS : Sancho et al.. 2022. 2n: Shi et al.. 1993
	<i>Bmexicanum</i> 348H	Herbarium C	Mexico: Puebla	-	-	-	-	-	-

	<i>Bmexicanum</i> 504	Herbarium VLA	Ecuador: Loja	-	-	-	-	-	
<i>B. miserum</i> (Thunb.) Koidz.	<i>Bmiserum</i> 67-1	EPSHU	Japan: Honshu	0.888±0.0232	-	2x	-	-	-
	<i>Bphoenicoides</i> 12	EPSHU	Spain: Teruel	1.499±0.005	28	4x	4	4	-
	<i>Bphoenicoides</i> 6-1R	RIKEN	Spain: Huesca	1.443±0.019	28	4x			-
	<i>Bphoenicoides</i> 10H	Herbarium FI	Italy: Sicily	-	-	-	-	-	GS and 2n: Sancho et al.. 2022.
	<i>Bphoenicoides</i> 11	EPSHU	Spain: Toledo	-	-	-	-	-	-
	<i>Bphoenicoides</i> 452	EPSHU	Morocco: Rift Mts	<b>2.176±0.017</b>	<b>38</b>	<b>6x</b>	-	-	-
	<i>Bphoenicoides</i> 453	USDA	Morocco: Mulay-Idriss	-	-	-	-	-	GS. 2n, 35S and 5S: Wolny & Hasterok. 2009.
<i>B. phoenicoides</i> (L.) P. Beauv. ex Roem. & Schult.	<i>Bphoenicoides</i> 454-6	EPSHU	Morocco: Alcazarseguir	-	-	-	-	-	-
	<i>Bphoenicoides</i> 505	EPSHU	Spain: Huesca	1.476±0.049	28	4x	4	4	-
	<i>Bphoenicoides</i> 509	EPSHU	Spain: Huesca	<b>1.48±0.03</b>	<b>28</b>	<b>4x</b>	-	-	-
	<i>Bphoenicoides</i> 510	USDA	Spain: Huesca	<b>1.435±0.03</b>	<b>28</b>	<b>4x</b>	-	-	GS. 2n, 35S and 5S: Wolny & Hasterok. 2009.
	<i>Bphoenicoides</i> 511	EPSHU	Spain: Huesca	<b>1.474±0.017</b>	<b>28</b>	<b>4x</b>	-	-	-
	<i>Bphoenicoides</i> 512	EPSHU	Spain: Leon	<b>1.457±0.033</b>		<b>4x</b>	-	-	-
	<i>Bphoenicoides</i> 551	EPSHU	Spain: Cadiz	<b>1.446±0.02</b>	<b>28</b>	<b>4x</b>	-	-	-

Supplementary data Chapter 2

<i>Bphoenicoides</i> 552	EPSHU	Spain: Cadiz	<b>2.204±0.039</b>	<b>38</b>	<b>6x</b>	-	-	-
<i>Bphoenicoides</i> 553	EPSHU	Spain: Malaga	<b>2.183±0.013</b>	<b>38</b>	<b>6x</b>	-	-	-
<i>Bphoenicoides</i> 554-1	EPSHU	Spain: Granada	<b>2.155±0.02</b>	<b>38</b>	<b>6x</b>	-	-	-
<i>Bphoenicoides</i> 555	EPSHU	Spain: Granada	<b>1.457±0.037</b>		<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 556	EPSHU	Spain: Jaen	<b>1.444±0.023</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 557	EPSHU	Spain: Huesca	<b>1.494±0.019</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 600	EPSHU	France: Nans le Pins	<b>1.479±0.019</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 601	EPSHU	France: Vence	<b>1.853±0.034</b>	<b>36</b>	<b>6x</b>	-	-	-
<i>Bphoenicoides</i> 602	EPSHU	France: Le Cannet des Maures	<b>1.989±0.022</b>	<b>36</b>	<b>6x</b>	-	-	-
<i>Bphoenicoides</i> 604	EPSHU	France: Bouches du Rhône. Arles	<b>1.453±0.027</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 606	EPSHU	France: Causse de la Selle	<b>1.457±0.046</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 608	EPSHU	France: Vic la Gardiole	<b>1.466±0.034</b>	-	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 609	EPSHU	France: Fleury	<b>1.466±0.027</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bphoenicoides</i> 610	EPSHU	France: Maureillas las Illas	<b>1.467±0.032</b>	-	<b>4x</b>	-	-	-
<i>B. pinnatum</i> (L.) P. Beauv.	<i>Bpinnatum</i> 21	EPSHU	GBR: Yorkshire Dales	-	-	-	-	-
	<i>Bpinnatum</i> 34	EPSHU	GBR: North Wiltshire	<b>1.449±0.018</b>	<b>28</b>	<b>4x</b>	<b>4</b>	<b>4</b>

Supplementary data Chapter 2

<i>Bpinnatum</i> 38H	Herbarium M	Finland: Hämeenlinna	-	-	-	-	-	-
<i>Bpinnatum</i> 39H	Herbarium M	Belgium: Nismes	-	-	-	-	-	-
<i>Bpinnatum</i> 41H	Herbarium G	Croatia: Bunovac	-	-	-	-	-	-
<i>Bpinnatum</i> 42H	Herbarium G	Greece : Evritanias	-	-	-	-	-	-
<i>Bpinnatum</i> 43H	Herbarium M	Russia: Kaluga	-	-	-	-	-	-
<i>Bpinnatum</i> 502	EPSHU	IRAQ	<b>0.968±0.013</b>	<b>16</b>	<b>2x</b>	<b>2</b>	<b>2</b>	-
<i>Bpinnatum</i> 505	EPSHU	Norway	<b>0.822±0.009</b>	<b>18</b>	<b>2x</b>	<b>4</b>	<b>2</b>	-
<i>Bpinnatum</i> 513	USDA	Greece	<b>1.453±0.026</b>	<b>28</b>	4x	4	4	-
<i>Bpinnatum</i> 514	USDA	Turkey	<b>1.537±0.012</b>	<b>28</b>	<b>4x</b>	<b>4</b>	<b>4</b>	-
<i>Bpinnatum</i> 515	USDA	Kazakhstan	<b>0.815±0.004</b>	<b>18</b>	<b>2x</b>	-	-	-
<i>Bpinnatum</i> 519	USDA	Germany: BG. Universität Halle/Saale	<b>1.471±0.044</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bpinnatum</i> 520	USDA	Netherlands: Scherpenzeel	<b>1.499±0.014</b>	<b>28</b>	<b>4x</b>	<b>4</b>	<b>4</b>	-
<i>Bpinnatum</i> 523	EPSHU	Czech Republic	<b>1.435±0.019</b>	<b>28</b>	<b>4x</b>	-	-	-
<i>Bpinnatum</i> 525	EPSHU	Iran	0.882±0.005	18	2x	2-3	2	-

GS: this study. 2n, 35S and  
5S: Wolny & Hasterok.  
2009.

<i>Bretusum</i> 3	EPSHU	Greece : Icaria Island	-	<b>38</b>	-	-	-	-
-------------------	-------	------------------------	---	-----------	---	---	---	---

Supplementary data Chapter 2

<i>B. retusum</i> (Pers.) P. Beauv.	Bretusum400	USDA	Spain: Huesca	1.704±0.024	32	4x	-	-	GS, 35S and 5S: Wolny & Hasterok. 2009.
	Bretusum403	EPSHU	Spain: Huesca	<b>2.373±0.958</b>	<b>42</b>	<b>6x</b>	-	-	-
	Bretusum405	EPSHU	Spain: Huesca	<b>2.386±0.039</b>	<b>42</b>	<b>6x</b>	-	-	GS and 2n: Sancho et al.. 2022.
	Bretusum407	EPSHU	Spain: Huesca	<b>1.715±0.017</b>	<b>32</b>	<b>4x</b>	<b>4</b>	<b>4</b>	-
	Bretusum408	EPSHU	Spain: Navarra	<b>2.431±0.033</b>	<b>42</b>	<b>6x</b>	-	-	-
	Bretusum451	EPSHU	Morocco: Rif Mountains	<b>1.779±0.035</b>	-	<b>4x</b>	-	-	-
	Bretusum452	EPSHU	Morocco: Talembote to Djbel Tazzaot	<b>1.708±0.051</b>	-	<b>4x</b>	-	-	-
	Bretusum453-4	EPSHU	Morocco: Rift Mts.	<b>1.840±0.097</b>	-	<b>4x</b>	-	-	-
	Bretusum454	EPSHU	Morocco: Tazza-Bou Idir	<b>1.862±0.196</b>	-	<b>4x</b>	-	-	-
	Bretusum455	EPSHU	Morocco: Alcazarseguir	-	-	-	-	-	-
	Bretusum500	EPSHU	France: Pourrieres	<b>1.724±0.018</b>	-	<b>4x</b>	-	-	-
	Bretusum501	EPSHU	France: Castries	<b>1.673±0.031</b>	<b>32</b>	<b>4x</b>	-	-	-
	Bretusum504	EPSHU	France: Vic la Gardiole	<b>1.669±0.026</b>	<b>32</b>	<b>4x</b>	-	-	-
	Bretusum506	EPSHU	France: Aude. Fontjoncouse	<b>1.672±0.014</b>	-	<b>4x</b>	-	-	-
	Bretusum507	EPSHU	France: Salses le Château	<b>1.682±0.015</b>	<b>32</b>	<b>4x</b>	-	-	-

Supplementary data Chapter 2

	<i>Bretusum</i> 551	EPSHU	Spain: Malaga	<b>2.109±0.025</b>		<b>6x</b>	-	-	-
	<i>Bretusum</i> 555	EPSHU	Spain: Granada	<b>1.715±0.017</b>	<b>32</b>	<b>4x</b>	-	-	-
	<i>Bretusum</i> 556	EPSHU	Spain: Jaen	<b>1.765±0.014</b>	<b>32</b>	<b>4x</b>	-	-	-
	<i>Bretusum</i> 557	EPSHU	Spain: Cadiz	<b>2.464±0.026</b>	<b>42</b>	<b>6x</b>	-	-	-
	<i>Bretusum</i> 558	EPSHU	Spain: Girona	<b>1.745±0.042</b>	-	<b>4x</b>	-	-	-
	<i>Bretusum</i> 559	EPSHU	Spain: Huesca	<b>1.744±0.018</b>	<b>32</b>	<b>4x</b>	-	-	-
	<i>Bretusum</i> 561	EPSHU	Spain: Zaragoza	<b>2.362±0.046</b>	<b>42</b>	<b>6x</b>	<b>5</b>	<b>5-6</b>	-
	<i>Bretusum</i> 564	EPSHU	Spain: Zaragoza	<b>2.473±0.077</b>	-	<b>6x</b>	-	-	-
<i>B. rupestre</i> (Host) Roem. & Schult.	<i>Brupestre</i> 5	EPSHU	Spain: Huesca	1.469±0.037	28	4x	4	4	-
	<i>Brupestre</i> 7	EPSHU	Russia: Ulitsa Pushkina	<b>1.562±0.016</b>	<b>28</b>	<b>4x</b>	<b>4</b>	<b>4</b>	-
	<i>Brupestre</i> 10	EPSHU	GBR: Cambridgeshire	<b>1.432±0.017</b>	<b>28</b>	<b>4x</b>	-	-	GS and 2n: Sancho et al., 2022.
	<i>Brupestre</i> 11	EPSHU	GBR: NE Yorkshire	-	-	-	-	-	-
	<i>Brupestre</i> 12	EPSHU	GBR	-	-	-	-	-	-
	<i>Brupestre</i> 13H	EPSHU	Greece : Arachowa	-	-	-	-	-	-
	<i>Brupestre</i> 14H	EPSHU	Italy: Iseo Lake	-	-	-	-	-	-
	<i>Brupestre</i> 15H	Herbarium M	Italy: Monte Faravella	-	-	-	-	-	-



Supplementary data Chapter 2

	<i>Brupestre</i> 182	Herbarium M	Croatia: Istria	2.258±0.026	38	6x	5-6	6	-
	<i>Brupestre</i> 439	Herbarium FI	Spain: Huesca	1.55±0.022	28	4x	-	-	-
	<i>Brupestre</i> 441	EPSHU	Spain: Leon	1.483±0.008	28	4x	-	-	-
	<i>Brupestre</i> 442	EPSHU	Spain: Huesca	1.56±0.03	28	4x	-	-	-
	<i>Brupestre</i> 443	EPSHU	Spain: Guipuzcoa	1.498±0.012	28	4x	-	-	-
	<i>Brupestre</i> 444	EPSHU	Spain: Lugo	1.492±0.021	28	4x	-	-	-
	<i>Brupestre</i> 445	EPSHU	France: Indre	1.519±0.039	28	4x	-	-	-
	<i>Brupestre</i> 512	EPSHU	Italy	0.84±0.015	18	2x	4	2	-
	<i>Brupestre</i> 600	EPSHU	France: Nans les Pins	2.216±0.013	38	6x	6	6	-
	<i>Brupestre</i> 605	USDA	France: Pourrieres	2.265±0.013	38	6x	5-6	6	-
	<i>Brupestre</i> 606	EPSHU	France: Causse de la Selle	2.33±0.043	38	6x	-	-	-
<i>B. spryginii</i> Tzvelev	<i>Bspryginii</i> 29	EPSHU	Russia: Krasnodarskii Krai	-	-	-	-	-	-
<i>B. sylvaticum</i> var. <i>breviglume</i> Keng ex Keng f.	<i>Bbreviglume</i> 33H	Herbarium LD	Tibet:	-	-	-	-	-	-
	<i>Bbreviglume</i> 34H	Herbarium M	Pakistan	-	-	-	-	-	-
<i>B.sylvaticum</i> (Huds.) P. Beauv.	<i>Bsylvaticum</i> 8	EPSHU	Russia: Gofitskoye	0.931 ± 0.015	18	2x	2	2	GS, 2n, 35S and 5S: Wolny & Hasterok. 2009.
	<i>Bsylvaticum</i> 29H	Herbarium B	Greece : Evrytania	-	-	-	-	-	GS: Steinwand et al. 2013

Supplementary data Chapter 2

<i>Bsylvaticum</i> 30H	Herbarium B	Denmark: Zealand	-	-	-	-	-	-
<i>Bsylvaticum</i> 31H	Herbarium M	Germany: Bayern	-	-	-	-	-	-
<i>Bsylvaticum</i> 32H	Herbarium M	Russia: Kaluga	-	-	-	-	-	-
<i>Bsylvaticum</i> 51	EPSHU	Morocco: Rif Mountains	<b>0.887±0.016</b>	-	<b>2x</b>			-
<i>Bsylvaticum</i> 53	EPSHU	Morocco: Rif Mountains	<b>0.883±0.004</b>	-	<b>2x</b>	-	-	-
<i>Bsylvaticum</i> 54-1	EPSHU	Morocco: Rif Mountains	<b>0.888±0.008</b>	-	<b>2x</b>	-	-	-
<i>Bsylvaticum</i> 63-4	EPSHU	Morocco: Mulay-Idriss	-	-	-	-	-	-
<i>Bsylvaticum</i> 434	EPSHU	Ukraine: Krym	<b>0.951±0.003</b>	<b>18</b>	<b>2x</b>	<b>4</b>	<b>2</b>	-
<i>Bsylvaticum</i> 445	USDA	Spain: Avila	0.886±0.012	-	2x	2	2	-
<i>Bsylvaticum</i> 446	USDA	Iran: Ardebil	0.844±0.007	-	2x	5	2	-
<i>Bsylvaticum</i> 466-6	EPSHU	<b>Spain: Huesca</b>	<b>0.928±0.013</b>	-	<b>2x</b>	-	-	GS. 2n, 35S and 5S: Wolny & Hasterok. 2009.
<i>Bsylvaticum</i> 467-2	EPSHU	<b>Spain: Huesca</b>	<b>0.934±0.007</b>	<b>18</b>	<b>2x</b>	-	-	GS. 2n, 35S and 5S: Wolny & Hasterok. 2009.
<i>Bsylvaticum</i> 468	EPSHU	<b>Spain: Huesca</b>	<b>0.911±0.017</b>	-	<b>2x</b>	-	-	-
<i>Bsylvaticum</i> 470-8b	EPSHU	<b>Spain: Guipuzcoa</b>	-	-	-	-	-	-
<i>Bsylvaticum</i> 476-9	EPSHU	<b>France: Barbazan</b>	-	-	-	-	-	-
<i>Bsylvaticum</i> 477-1	EPSHU	<b>Spain: Lleida</b>	<b>0.932±0.017</b>	-	<b>2x</b>	-	-	-

*Supplementary data Chapter 2*

<i>Bsylvaticum</i> 500	EPSHU	France:Nans les Pins	0.957±0.012	-	2x	-	-	-
<i>Bsylvaticum</i> 501-6	EPSHU	France: Roquefort les Pins.	0.947±0.01	-	2x	-	-	-
<i>Bsylvaticum</i> 502	EPSHU	France: Le Cannet des Maure	0.939±0.005	18	2x	-	-	-
<i>Bsylvaticum</i> 503	EPSHU	France: Pourrieres	0.953±0.01	-	2x	-	-	-
<i>Bsylvaticum</i> 505	EPSHU	France: Causse de la Selle	0.947±0.025	-	2x	-	-	-
<i>Bsylvaticum</i> 506-3	EPSHU	France: Saint Jean de Fos	-	-	-	-	-	-
<i>Bsylvaticum</i> 507	EPSHU	France: Fleury	0.924±0.022	-	2x	-	-	-
<i>Bsylvaticum</i> 508-5	EPSHU	France:Villeseque des Corbieres	-	-	-	-	-	-
<i>Bsylvaticum</i> 509	EPSHU	France: Maureillas las Illas	0.914±0.019	-	2x	-	-	-
<i>Bsylvaticum</i> 550	EPSHU	Spain: Salamanca	0.926±0.011	-	2x	-	-	-
<i>Bsylvaticum</i> 552-1	EPSHU	Spain: Cadiz	0.917±0.009	18	2x	-	-	-
<i>Bsylvaticum</i> 554-1	EPSHU	Spain: Malaga	0.925±0.01	-	2x	-	-	-
<i>Bsylvaticum</i> 555	EPSHU	Spain: Granada	0.934±0.017	-	2x	-	-	-
<i>Bsylvaticum</i> 556	EPSHU	Spain: Granada	0.923±0.013	-	2x	-	-	-
<i>Bsylvaticum</i> _Ain1	JGI	Tunissia: Ain Draham	0.863±0.029	-	2x	5-6	2	-
<i>Bsylvaticum</i> _Sin1	JGI	Turkey: Sinop	-	-	2x	-	-	-

**Supplementary Table S1 (B)** Supplementary information for sampling data of the 21 *Brachypodium* taxa under study. Taxon's population code, country code, location, geographic coordinates, and Genbank accession codes (plastome. and rDNA 35S and 5S sequences) for each sample. New data from this study are highlighted in bold.

Taxon_pop code	Country code	Location	Geographic coordinates
Bstacei_ABR114	ESP	Spain: Balearic islands	38.678611. 1.40315
Bdistachyon_Bd21	IRQ	Iraq: Salakudin	33.7608833. 44.403075
Bhybridum_ABR113	POR	Portugal: Lisbon	38.7077507. -9.1365919
Bhybridum_Bhyb26	ESP	Spain: Jaen. La Cimbarra	38.3889875. -3.373696
Barbuscula405	ESP	Spain: Canary Island. Tenerife: Teno	28.3427961. -16.87763966
Barbuscula502	ESP	Spain: Canary Island. La Gomera	28.169559. -17.249130
Bboissieri3	ESP	Spain: Granada: Sierra de Huetor. Puerto de La Mora	37.255960. -3.432174
Bboissieri9	ESP	Spain: Málaga. Canillas del Aceituno	36.893000. -4.071000
Bboissieri10	ESP	Spain: Granada. Sierra Nevada	37.115556. -3.443889
Bboissieri13	ESP	Spain: Granada: Sierra Nevada: Dornajo	37.1252778. -3.443611
Bboissieri14	ESP	Spain: Granada. Sierra Nevada: El purche. Monachil	37.1363. -3.485280
Bboissieri15	ESP	Spain: Granada. Sierra Huétor	37.272777. -3.458337
Bbolussi046	RSA	South Africa: Natal. Drakernsbergs.	-29.4546671. 29.47686406
Bbreviglume33H	TBT	Tibet: Gongbogyamda. Alt: 3300m. 20/09/1988. Herb. Lund (LD) Acc. No. 135398	29.192647. 90.704691

Supplementary data Chapter 2

<i>Bbreviglume</i> 34H	PAK	Pakistan: Hazara. kaghan Vy. Paras-Shogran. 2250m 09/07/1997. Botanische Staatssammlung München. M-0175636	29.758506. 70.738451
<i>Bflexum</i> 038	RSA	South Africa: KwaZulu Natal. Weza Forest	-30.591326. 29.751893
<i>Bflexum</i> 046	RSA	South Africa: KwaZulu Natal. Didima	-28.999913. 29.28334238
<i>Bgenuense</i>	ITA	Italy: PN. Della Majella. Brescia	42.120557. 14.026583
<i>Bglaucovirens</i>	GRE	Crete island: Greece. Crete isl. [Berlin Botanical Garden. Dahlem. 3151]	35.256463. 24.827693
<i>Bkawakamii</i>	TWN	Taiwan: Taiwán. Kaohsiung. Taoyuan District. Her. Lugd. Batav 200190	23.27168884. 120.9591758
<i>Bkurilense</i>	RUS	Russia: Kuriles. Iturup Isl.. 2 km SW of Kurilsk town	45.224076. 147.869205
<i>Bmadagascariense</i>	MDG	Madagascar: Ankaratra MTs. Tambunana	-18.349888. 47.3405484
<i>Bmexicanum</i> 347	MEX	Mexico: Chinantla. Puebla. Mexico	18.207800. -98.265178
<i>Bmexicanum</i> 348H	MEX	Mexico: Hidalgo. Sierra de Pachuca	-3.565467. -79.432089
<i>Bmexicanum</i> 504	ECU	Ecuador: Loja.	20.097619. -98.731650
<i>Bmiserum</i> 67-1	JPN	Japan: Honshu. Sendai. Aoba-ku. Aramakiab	38.255903. 140.843407
<i>Bphoenicoides</i> 6	ESP	Spain: Huesca. Panzano	42.211173. -0.187960
<i>Bphoenicoides</i> 10H	ITA	Italy: Palazzo Adriano. Palermo. Sicily. 0.30m. 27/06/1996. Universitatis Florentinae. FI050693	37.65111. 13.30904
<i>Bphoenicoides</i> 11	ESP	Spain: Toledo	39.849748. -4.012598
<i>Bphoenicoides</i> 12	ESP	Spain: Spain: Teruel [USDA PI 253505]	40.339174. -1.052352

## Supplementary data Chapter 2

<i>Bphoenicoides</i> 452	MAR	Morocco: Rift Mts. ca. Djbel Tazzaot	35.25486. -5.153444
<i>Bphoenicoides</i> 453	MAR	Morocco: Mulay-Idriss	34.03331. -5.502972
<i>Bphoenicoides</i> 454	MAR	Morocco: Alcazarseguir	35.8467611 -5.5431361
<i>Bphoenicoides</i> 505	ESP	Spain: Huesca. USDA_PI 89817	42.155281. -0.417135
<i>Bphoenicoides</i> 509	ESP	Spain: Huesca. Jaca. Peña Oroel	42.530556. -0.549167
<i>Bphoenicoides</i> 510	ESP	Spain: Huesca. Sariñena. laguna	41.805. -0.17666667
<i>Bphoenicoides</i> 511	ESP	Spain: Huesca. Alcubierre. Bco. San Caprasio	41.7794444. -0.464166667
<i>Bphoenicoides</i> 512	ESP	Spain: Leon. Ardoncino	42.492011. -5.623341
<i>Bphoenicoides</i> 551	ESP	Spain: Cadiz. Alcala de los Gazules	36.4541667. -5.7555555
<i>Bphoenicoides</i> 552	ESP	Spain: Cadiz. Alcala de los Gazules	36.5319. -5.665000
<i>Bphoenicoides</i> 553	ESP	Spain: Málaga. Canillas del Aceituno	36.74333. -4.924444
<i>Bphoenicoides</i> 554	ESP	Spain: Granada. Sierra Nevada: El purche. Monachil	37.1363. -3.485280
<i>Bphoenicoides</i> 555	ESP	Spain: Granada: Sierra de Baza	37.3325. -2.915555556
<i>Bphoenicoides</i> 556	ESP	Spain: Jaen. Villanueva del arzobispo	38.1786111. -2.9241666
<i>Bphoenicoides</i> 557	ESP	Spain: Huesca: Ainsa-Sobrarbe. Latorre	42.312070. 0.155316
<i>Bphoenicoides</i> 600	FRN	France: Nans les Pins	43.35944. 5.785268
<i>Bphoenicoides</i> 601	FRN	France: Alpes Maritime. Vence.	43.747525. 7.097186
<i>Bphoenicoides</i> 602	FRN	France : Le Cannet des Maures 77 m	43.367684. 6.369455
<i>Bphoenicoides</i> 604	FRN	France : Bouches du Rhône. Arles. 472 m	43.700905. 4.498360
<i>Bphoenicoides</i> 606	FRN	France : Hérault. Causse de la Selle. 95 m	43.802619. 3.643922

Supplementary data Chapter 2

<i>Bphoenicoides</i> 608	FRN	France : Hérault. Vic la Gardiole. 30 m	43.480929. 3.762278
<i>Bphoenicoides</i> 609	FRN	France: Aude. Fleury. 4 m.	43.234664. 3.168557
<i>Bphoenicoides</i> 610	FRN	France: Maureillas las Illas. 100 m.	42.506279. 2.823349
<i>Bpinnatum</i> 21	GBR	Great Britain: Yorkshire Dales	54.196243. -2.163120
<i>Bpinnatum</i> 34	GBR	Great Britain: North Wiltshire. Swillbrook	51.645103. -1.990833
<i>Bpinnatum</i> 38H	FIN	Finland: Hämeenlinna. South Häme. Kitu. 90-95m. Botanische Staatssammlung München. M-0176966	60.968038. 24.582956
<i>Bpinnatum</i> 39H	BEL	Belgium: Nismes. prov. Namur. 16/07/1973. Botanische Staatssammlung München. M-0176263	50.085516. 4.5461986
<i>Bpinnatum</i> 41H	CRO	Croatia: Bunovac. Vagauski. 1300m. 26/07/1983. Genavense (G) G00420389	44.367231. 15.509502
<i>Bpinnatum</i> 42H	GRE	Greece: Evritanias. Eparchia. 1500-1600m. 18/07/1996. Genavense (G) G00420390	39.133333. 21.700000
<i>Bpinnatum</i> 43H	RUS	Yújnov district. Kaluga Province. 15/6/1975. Botanische Staatssammlung München. M-0176963	54.749749. 35.220191
<i>Bpinnatum</i> 502	IRQ	IRAQ: USDA_PI 185135	33.223000. 43.679000
<i>Bpinnatum</i> 505	NOR	Norway: USDA_PI 345964	60.446000. 8.428000
<i>Bpinnatum</i> 513	GRE	Greece : USDA PI 249722	38.142615. 23.673302
<i>Bpinnatum</i> 514	TUR	Turkey: USDA PI 206677	38.963000. 35.243000
<i>Bpinnatum</i> 515	KAZ	Kazakhstan: USDA PI 440176	48.019000. 66.923000
<i>Bpinnatum</i> 519	GER	Germany: BG. Universität Halle/Saale	51.488000. 11.960000
<i>Bpinnatum</i> 520	NED	Netherlands: Scherpenzeel	52.132000. 5.291000
<i>Bpinnatum</i> 523	CZE	Czech Republic: ECN 14G3900002	49.973798. 14.411466



Supplementary data Chapter 2

<i>Bpinnatum</i> 525	IRI	Iran: USDA PI 230114	35.537511. 51.476998
<i>Bretusum</i> 3	GRE	Greece : Icaria Island	37.597989. 26.113647
<i>Bretusum</i> 400	ESP	Spain: Huesca. Angüés	42.126786. -0.161747
<i>Bretusum</i> 403	ESP	Spain: Huesca. Sta. Cruz de la Serós	42.552500. -0.681389
<i>Bretusum</i> 405	ESP	Spain: Huesca: Bernues. Peña Oroel	42.506934. -0.507009
<i>Bretusum</i> 407	ESP	Spain: Huesca. Alcubierre	41.744166. -0.483334
<i>Bretusum</i> 408	ESP	Spain: Navarra. Falces	42.398889. -1.801389
<i>Bretusum</i> 451	MAR	Morocco: Rif Mountains. Talasemtane National Park	35.1781389. -5.147223
<i>Bretusum</i> 452	MAR	Morocco: Talembote to Djbel Tazzaot	35.260055 -5.162416
<i>Bretusum</i> 453	MAR	Morocco: Rift Mts. Djbel Tazzaout	35.260055. -5.162416
<i>Bretusum</i> 454	MAR	Morocco: Tazza-Bou Idir	34.154525. -4.004375
<i>Bretusum</i> 455	MAR	Morocco: Alcazarseguir	35.846761. -5.543136
<i>Bretusum</i> 500	FRN	France: Var. Pourrières	43.516311. 5.740417
<i>Bretusum</i> 501	FRN	France: Hérault. Castries. 85 m	43.668814. 3.968588
<i>Bretusum</i> 504	FRN	France : Hérault. Vic la Gardiole	43.503258. 3.781791
<i>Bretusum</i> 506	FRN	France: Aude. Fontjoncouse. 75 m	43.044149. 2.829018
<i>Bretusum</i> 507	FRN	France : Salses le Château. 75 m	42.787292. 2.874056
<i>Bretusum</i> 551	ESP	Spain: Malaga	36.774722. -5.093889
<i>Bretusum</i> 555	ESP	Spain: Granada. Sierra de Baza	37.372498. -2.718061
<i>Bretusum</i> 556	ESP	Spain: Jaen: Sierra de Segura. Beas del Segura	38.2747222. -2.75805555

*Supplementary data Chapter 2*

<i>Bretusum</i> 557	ESP	Spain: Cádiz. Almovodar. El pedregoso. Tarifa	36.149722. -5.627500
<i>Bretusum</i> 558	ESP	Spain: Girona. Maçanet de Cabrenys	42.3728. 2.803275
<i>Bretusum</i> 559	ESP	Spain: Huesca: Almunia de San Juan	41.9341667. -0.25027777
<i>Bretusum</i> 561	ESP	Spain: Zaragoza. Alhama de Aragón	42.14366. -1.23367
<i>Bretusum</i> 564	ESP	Spain: Zaragoza. Sastago	41.3897222. -0.233055556
<i>Brupestre</i> 5	ESP	Spain: Huesca: Jaca. Castiello de Jaca: Aratorés	42.6544306. -0.56775
<i>Brupestre</i> 7	RUS	Russia: Ulitsa Pushkina. 178. Svetlograd. Stavropolskiy kray. Rusia. 356530	45.330000. 42.863000
<i>Brupestre</i> 10	GBR	Great Britain: Cambridgeshire: New Market. Devil's Ditch	52.227035. 0.367545
<i>Brupestre</i> 11	GBR	Great Britain: England: NE Yorkshire at SE8600465205. chalk pit. Coll. 22.08.2019	54.196243. -2.163120
<i>Brupestre</i> 12	GBR	Great Britain	54.196243. -2.163120
<i>Brupestre</i> 13H	GRE	Greece : Arachowa. Delphi. Botanische Staatssammlung München. M-0176921	38.476653. 22.580611
<i>Brupestre</i> 14H	ITA	Italy: Iseo Lake. between Toline and Vello. Brescia. Botanische Staatssammlung München. M-0176923	45.777391. 10.072254
<i>Brupestre</i> 15H	ITA	Italy: Monte Faravella. Accadia. Foggia. 08/07/2008. dry leaf Universitatis Florentinae. FI050697	41.190057. 15.316094
<i>Brupestre</i> 182	CRO	Croatia: Istria	45.274000. 13.890000
<i>Brupestre</i> 439	ESP	Spain: Huesca. Jaca. Santa cruz de la Seros	42.516111. -0.691389
<i>Brupestre</i> 441	ESP	Spain: León. Campohermoso	
<i>Brupestre</i> 442	ESP	Spain: Huesca. Ribagorza Oriental. Aneto	42.553761. 0.740058
<i>Brupestre</i> 443	ESP	Spain: Guipuzcoa. San Sebastian. Igueldo	43.321367. -2.010319

Supplementary data Chapter 2

<i>Brupestre</i> 444	ESP	Spain: Lugo. Seoane do Caurel	42.614525. -7.096292
<i>Brupestre</i> 445	FRN	France: Indre. Lureuil	46.757217. 1.012292
<i>Brupestre</i> 512	ITA	Italy: USDA	43.160000. 12.240000
<i>Brupestre</i> 600	FRN	France: Nans les Pins	43.341886. 5.767201
<i>Brupestre</i> 605	FRAN	France: Pourrières	43.539017. 5.731643
<i>Brupestre</i> 606	FRN	France : Hérault. Causse de la Selle	43.797892. 3.627851
<i>Bsprygini</i> 29	RUS	Russia: Krasnodarskii Krai. Abinskii Raion	45.111077. 38.818477
<i>Bsylvaticum</i> _Ain1	TUN	Tunissia: Ain Draham	36.771943. 8.692181
<i>Bsylvaticum</i> _Sin1	TUR	Turkey: Sinop	42.030012. 35.158760
<i>Bsylvaticum</i> 8	RUS	Russia: Gofitskoye village [USDA PI 440173]	
<i>Bsylvaticum</i> 29H	GRE	Greece: Evrytania. NO Viniani. 12.06.2007. Quelle. 610m. Mus.Bot. Berol B100281411	38.890649. 21.667174
<i>Bsylvaticum</i> 30H	DEN	Denmark:Zealand. Vrangskov. north of Ringsted. Mus. Bot. Berol B100566906	55.496558. 11.711986
<i>Bsylvaticum</i> 31H	GER	Germany:Bayern. Regierungsbezirk Oberpfalz. Landkreis Regensburg. Vorderer Bayerischer Wald. Forstmühler. 530m. Botanische Staatssammlung München. M-0177011	47.935739. 12.389936
<i>Bsylvaticum</i> 32H	RUS	Russia: Kaluga. distr. Kondrovo. Vallis fluv. Ugra. Botanische Staatssammlung München. M-01769847	54.511058. 36.200533
<i>Bsylvaticum</i> 51	MAR	Morocco: Rif Mountains. Talassemtane National Park	35.12955 -5.1354500001
<i>Bsylvaticum</i> 53	MAR	Morocco: Rif Mountains. Talassemtane National Park	35.1781389 -5.147222223
<i>Bsylvaticum</i> 54-1	MAR	Morocco: Morocco: Rif Mountains. ca. Bab Barre	35.019280. -5.010414
<i>Bsylvaticum</i> 63-4	MAR	Morocco: Mulay-Idriss. Bab Rmila Natural Reserve	34.0356944 -5.4924166

## Supplementary data Chapter 2

Bsylvaticum434	UKR	Ukraine: Krym (USDA PI639821)	45.330000. 42.863000
Bsylvaticum445	ESP	Spain: Avila. Gredos. Candeleda. USDA_PI 318962	40.180845. -5.237279
Bsylvaticum446	IRI	Iran: Ardebil. on east side of grade to Astara. USDA_PI 380758	38.264996. 48.327289
Bsylvaticum466-6	ESP	Spain: Huesca. S. Guara	42.244937. -0.241929
Bsylvaticum467-2	ESP	Spain: Huesca. Bospén	42.0608583. -0.11825001
Bsylvaticum468	ESP	Spain: Huesca. Jaca. Peña Oroel	42.5269444. -0.52083334
Bsylvaticum470-8b	ESP	Spain: Guipuzcoa. San Sebastian. Cristina Enea	43.3156611. -1.974330556
Bsylvaticum476-9	FRN	France: Hautes Pyrenées. Barbazan	43.0325389. -0.62285
Bsylvaticum477-1	ESP	Spain: Lleida: Vall d'Aran: Bausen	42.832489. 0.717812
Bsylvaticum500	FRN	France: Nans les Pins	43.338138. 5.740515
Bsylvaticum501-6	FRN	France : Alpes Maritime. Roquefort les Pins.	43.664392. 7.036268
Bsylvaticum502	FRN	France : Var. Le Cannet des Maure	43.367684. 6.369455
Bsylvaticum503	FRN	France: Pourrières. 250 m	43.484639. 5.720147
Bsylvaticum505	FRN	France : Hérault. Causse de la Selle. 95 m	43.802619. 3.643922
Bsylvaticum506-3	FRN	France : Hérault. Saint Jean de Fos	43.711665. 3.544208
Bsylvaticum507	FRN	France: Aude. Fleury. 4 m	43.234664. 3.168557
Bsylvaticum508-5	FRN	France : Aude. Villesèque des Corbières. 60 m.	43.041336. 2.855819
Bsylvaticum509	FRN	France: Maureillas las Illas. 100 m.	42.506279. 2.823349
Bsylvaticum550	ESP	Spain: Salamanca. San Esteban de la Sierra	40.504151. -6.176246
Bsylvaticum552-1	ESP	Spain: Cádiz. Almóvodar. El pedregoso. Tarifa	36.149722. -5.627500

*Supplementary data Chapter 2*

Bsylvaticum554-1	ESP	Spain: Málaga. Canillas del Aceituno	36.8913889. -4.06361111
Bsylvaticum555	ESP	Spain: Granada: Sierra Nevada: Dornajo	37.1302778. -3.448333334
Bsylvaticum556	ESP	Spain: Granada: Sierra de Baza	37.33111111. -2.90222222

---

**Supplementary Table S2.** Genome skimming sequencing data of the *Brachypodium* samples under study. Taxon's population code, chromosome number, ploidy level, insert size, number of reads (per million) obtained per sample. Mapping strategy for plastome genome assembly.

Taxon-population code	Chrom. Number	Ploidy	Insert size	No. reads (x10 <sup>6</sup> )	Mapping strategy plastome
<i>Barbuscula</i> 405	-	-	288	12090	<i>de novo</i>
<i>Barbuscula</i> 502	18	2x	257.5	22408	<i>de novo</i>
<i>Bboissieri</i> 10	48	6x	248	21676	<i>de novo</i>
<i>Bboissieri</i> 15	48	6x	255	12464	<i>de novo</i>
<i>Bboissieri</i> 3	48	6x	265	12563	<i>de novo</i>
<i>Bboissieri</i> 9	48	6x	256	23691	<i>de novo</i>
<i>Bbolussi</i> 046	-	-	231	29773	read mapping
<i>Bbreviglume</i> 33H	-	-	200	33192	read mapping
<i>Bbreviglume</i> 34H	-	-	200	16629	<i>de novo</i>
<i>Bflexum</i> 038	-	-	181	23532	<i>de novo</i>
<i>Bflexum</i> 046	-	-	198	20126	read mapping
<i>Bgenuense</i>	18	2x	135	26122	<i>de novo</i>
<i>Bglaucovirens</i>	16	2x	261	12508	<i>de novo</i>
<i>Bkawakamii</i>	-	-	150.5	22841	<i>de novo</i>
<i>Bkurilense</i>	-	-	190.5	11611	read mapping
<i>Bmadagascariense</i>	-	-	181.5	22410	read mapping

Supplementary data Chapter 2

Bmexicanum347	40	4x	253	10741	<i>de novo</i>
Bmexicanum348H	-	-	122.5	10090	read mapping
Bmexicanum504	-	-	265	11033	<i>de novo</i>
Bmiserum67-1	-	-	250	26271	read mapping
Bphoenicoides10H	-	-	124.5	25365	<i>de novo</i>
Bphoenicoides11H	-	-	254	10640	<i>de novo</i>
Bphoenicoides452	38	6x	300	20076	<i>de novo</i>
Bphoenicoides453	-	-	297	16162	<i>de novo</i>
Bphoenicoides454	-	-	257	21569	<i>de novo</i>
Bphoenicoides552	38	6x	301	23748	<i>de novo</i>
Bphoenicoides553	38	6x	257	12699	<i>de novo</i>
Bphoenicoides554	38	6x	268	13291	<i>de novo</i>
Bphoenicoides600	28	4x	281	12602	<i>de novo</i>
Bphoenicoides601	36	6x	267	25873	<i>de novo</i>
Bphoenicoides6	28	4x	273	12617	<i>de novo</i>
Bpinnatum21	-	-	260	12649	<i>de novo</i>
Bpinnatum34	28	4x	276	12632	<i>de novo</i>
Bpinnatum38H	-	-	155	20144	<i>de novo</i>
Bpinnatum39H	-	-	135	17276	<i>de novo</i>
Bpinnatum41H	-	-	92.5	20059	<i>de novo</i>



## Supplementary data Chapter 2

Bpinnatum42H	-	-	156	22712	<i>de novo</i>
Bpinnatum43H	-	-	121	17225	<i>de novo</i>
Bpinnatum502	16	2x	284	16224	<i>de novo</i>
Bpinnatum505	18	2x	266	11669	<i>de novo</i>
Bpinnatum514	28	4x	329	28729	<i>de novo</i>
Bpinnatum515	18	2x	260	17966	<i>de novo</i>
Bpinnatum519	28	4x	273	26346	<i>de novo</i>
Bpinnatum520	28	4x	286	11215	<i>de novo</i>
Bretusum3	38	-	227	11525	<i>de novo</i>
Bretusum400	32	4x	261	11182	<i>de novo</i>
Bretusum403	42	6x	273	17596	<i>de novo</i>
Bretusum407	32	4x	286	17981	<i>de novo</i>
Bretusum408	42	6x	316	26345	<i>de novo</i>
Bretusum453	-	4x	252	25608	<i>de novo</i>
Bretusum454	-	4x	275	22616	<i>de novo</i>
Bretusum455	-	-	271	20902	<i>de novo</i>
Bretusum504	32	4x	319	27892	<i>de novo</i>
Bretusum551	-	6x	306	29839	<i>de novo</i>
Bretusum555	32	4x	291	30216	<i>de novo</i>
Bretusum557	42	6x	297	24823	<i>de novo</i>

## Supplementary data Chapter 2

Bretusum561	42	6x	247	14162	<i>de novo</i>
Brupestre11	-	-	277	11618	<i>de novo</i>
Brupestre12	-	-	260	13204	<i>de novo</i>
Brupestre13H	-	-	120	18859	<i>de novo</i>
Brupestre14H	-	-	109	17732	read mapping
Brupestre15H	-	-	149	21871	read mapping
Brupestre182	38	6x	277	14523	<i>de novo</i>
Brupestre439	28	4x	286	19514	<i>de novo</i>
Brupestre441	28	4x	322	15849	<i>de novo</i>
Brupestre442	28	4x	322	18421	<i>de novo</i>
Brupestre443	28	4x	293	14913	<i>de novo</i>
Brupestre444	28	4x	269	10590	<i>de novo</i>
Brupestre512	18	2x	295	13316	<i>de novo</i>
Brupestre600	38	6x	272	12141	<i>de novo</i>
Brupestre605	38	6x	250	11457	<i>de novo</i>
Brupestre7	28	4x	290	15196	<i>de novo</i>
<hr/>					
B spryginii 29	-	-	217	10836	read mapping
<hr/>					
Bsylvaticum29H	-	-	200 (default)	38710164	read mapping
Bsylvaticum30H	-	-	200 (default)	43089680	read mapping

## Supplementary data Chapter 2

Bsylvaticum31H	-	-	200 (default)	40115538	read mapping
Bsylvaticum32H	-	-	200 (default)	41607486	read mapping
Bsylvaticum434	18	2x	258	11418	<i>de novo</i>
Bsylvaticum466-6	-	2x	13802	278	<i>de novo</i>
Bsylvaticum467-2	18	2x	274	19704	<i>de novo</i>
Bsylvaticum470-8b	-	-	294	18296	<i>de novo</i>
Bsylvaticum476-9	-	-	271	18554	<i>de novo</i>
Bsylvaticum501-6	-	2x	294	10677	<i>de novo</i>
Bsylvaticum506-3	-	-	282	15311	<i>de novo</i>
Bsylvaticum508-5	-	-	249	21518	<i>de novo</i>
Bsylvaticum54-1	-	2x	285	12984	<i>de novo</i>
Bsylvaticum552-1	18	2x	271	12039	<i>de novo</i>
Bsylvaticum554-1	-	2x	308	19288	<i>de novo</i>
Bsylvaticum63-4	-	-	305	18300	<i>de novo</i>
Bsylvaticum Ain1	18	2x	JGI; Genome Portal	JGI; Genome Portal	<i>de novo</i>
Bsylvaticum Sin1	18	2x	JGI; Genome Portal	JGI; Genome Portal	<i>de novo</i>
Bhybridum 26	30	4x	JGI; Genome Portal	JGI; Genome Portal	<i>de novo</i>

**Supplementary Table S3.** Statistics of the *Brachypodium* plastome parsimony haplotypic network for the main plastomic groups (I-XII) indicating number of haplotypes, number of variable sites, haplotype diversity index (Hd), nucleotide diversity (per site) ( $\pi$ ), and Tajima's D for each group. NA (not applicable; four or more sequences are need to compute Tajima's D statistic). Significant values are highlighted in bold.

Plastomic Groups	No. of haplotypes (H) [Accessions]	No. variable sites	Hd	$\pi$	Tajima's D [Statistical significance]
(I) <i>B.stacei</i> + <i>B. hybridum</i> -S	2 [BstaceiABR114_2x; BhybridumABR113_4x]	3	1	0.00074	NA
(II) <i>B. mexicanum</i>	3 [Bmexicanum347_4x; Bmexicanum348H; Bmexicanum504]	116	1	0.01900	NA
(III) <i>B. distachyon</i> + <i>B. hybridum</i> -D	2 [BdistachyonBd21-3_2x; Bhybridum26_4x]	452	1	0.11106	NA
(IV) <i>B. boissieri</i>	4 [Bboissieri3_6x; Bboissieri9_6x; Bboissieri10_6x; Bboissieri15_6x]	56	1	0.00762	0.15506 [P > 0.10]
(V) <i>B. arbuscula</i>	2 [Barbuscula502_2x; Barbuscula405]	25	1	0.00614	NA
(VI) <i>B. phoenicoides</i>	8 [Bphoenicoides11; Bphoenicoides452_6x; Bphoenicoides453; Bphoenicoides454; Bphoenicoides552_6x; Bphoenicoides553_6x; Bphoenicoides554_6x; Bretusum551_6x]	227	1	0.01548	-1.53596 [0.10 > P > 0.05]
(VIII) <i>B. sylvaticum</i> s.l.	21 [Bglaucovirens_2x; Bpinnatum519_4x; Bspryginii_2x; Bsylvaticum29H_2x; Bsylvaticum32H; Bsylvaticum54_2x; Bsylvaticum63_2x; Bsylvaticum434_2x; Bsylvaticum467_2x; Bsylvaticum470_2x; Bsylvaticum476_2x; Bsylvaticum501_2x; BsylvaticumAin1_2x; BsylvaticumSin1_2x; [haplotype <u>H1</u> : Bsylvaticum466_2x Bsylvaticum506_2x Bsylvaticum508_2x Bsylvaticum552_2x Bsylvaticum554_2x]; [haplotype <u>H2</u> : Bsylvaticum30H_2x Bsylvaticum31H]]	174	0.9476	0.00514	-2.33073 [P < 0.05]
	(excluding <i>B. spryginii</i> : 92)	(0.9421 excluding <i>B. spryginii</i> )	(0.00334 excluding <i>B. spryginii</i> )	(-1.95264 [P < 0.05] excluding <i>B. spryginii</i> )	

Supplementary data Chapter 2

(IX) <i>B. retusum</i>	7 [Bretusum400_4x; Bretusum407_4x; Bretusum453_4x; Bretusum454_4x; Bretusum455; Bretusum504_4x; Bretusum555_6x]	87	1	0.00676	-1.31153 [P > 0.10]
(X) <i>B. pinnatum</i>	10 [Bpinnatum21; Bpinnatum38H; Bpinnatum41H; Bpinnatum42H; Bpinnatum43H; Bpinnatum502_2x; Bpinnatum514_4x; Bpinnatum515_2x; Brupestre7_4x; Brupestre13H]	108	1	0.00645	-1.55016 [P > 0.10]
(XI) MISCELLANEOUS ( <i>B. pinnatum</i> + <i>B. phoenicoides</i> + <i>B. rupestre</i> + <i>B.</i> <i>retusum</i> )	11 [Bphoenicoides10H; Bpinnatum39H; Bpinnatum505_2x; Bphoenicoides601_6x; Bretusum557_6x; Brupestre15H; Brupestre605_6x; [haplotype <u>H4</u> : Bpinnatum34_4x Bretusum408_6x]; [haplotype <u>H5</u> : Bphoenicoides600_4x Brupestre600_6x]]	98	0.9636	0.00537	-1.66154 [0.10 > P > 0.05]
(XII) <i>B. rupestre</i> 4x	13 [Brupestre11; Brupestre439_4x; Brupestre441_4x; Brupestre442_4x; Brupestre443_4x; Brupestre444_4x; Bretusum561_6x; [haplotype <u>H3</u> : Brupestre12 ; Bpinnatum520_4x]; Brupestre182_6x ; Brupestre512_2x ; Bgenuense_2x] ; Brupestre14H]]	149	0.9872	0.00922	-1.00129 [P > 0.10]

**Supplementary Table S4. (A)** Nuclear rDNA 5S reference sequences and accession used to build the *Brachypodium* 5S gene Short (5S-S) and 5S gene Long (5S-L) data sets.

Species	Accession	Source
Ae. tauschii	EU925124.1	Baum et al. 2009
A tauschii	EU924977.1	Baum et al. 2009
H. vulgare var morex (L)	H. vulgare Morex v3	Baum et al. 1994 ( U07376). Accession used to blast against the updated genome in JGI portal
Osativa indica	ENSRNA046494644	
O. sativa japonica	ENSRNA049475447	
S. bicolor	AF242643	Pan et al. 2000
T. turgidum	KY439150.1	Genbank database
T. turgidum	KY439200.1	Genbank database
T. urartu	AY049865.1	Baum et al 2004
T. urartu	AY049864.1	Baum et al 2004
Z. mays	AF242646	Pan et al. 2000
T. aestivum	JGI; Genome Portal	Lagudah et al 1989. Accessions used to blast against the updated genome in JGI portal
T. aestivum	JGI; Genome Portal	Lagudah et al 1989. Accessions used to blast against the updated genome in JGI portal

<i>B. distachyon</i>	JN831367.1	Qiwei et al 2010
<i>B. stacei</i>	<i>B.stacei</i> ABR114 316 v2.1	retrieved from reference genome Blast against <i>Bdistachyon</i> reference
<i>B. mexicanum</i>	<i>B. mexicanum</i> v1.1.	retrieved from reference genome Blast against <i>Bdistachyon</i> and <i>Bstacei</i> ABR114 reference
<i>B. hybridum</i> ABR113	<i>B. hybridum</i> var ABR113	retrieved from reference genome Blast against <i>Bdistachyon</i> and <i>Bstacei</i> ABR114 reference
<i>B. hybridum</i> Bhyb26	<i>B. hybridum</i> Bhyb26_693_v2.0	retrieved from reference genome Blast against <i>Bdistachyon</i> and <i>Bstacei</i> ABR114 reference
<i>B. arbuscula</i>	BARB1.v3.1	retrieved from reference genome Blast against <i>Bdistachyon</i> and <i>Bstacei</i> ABR114 reference

---



**Supplementary Table S4. (B)** Length in base pair for each 5S sequence (short and long) recovered.

Accession	Length (bp)
<b>5S-rDNA-SHORT (S)</b>	
<i>Bstacei</i> ABR114_2x_S	270
<i>Bhybridum</i> ABR113_4x_S	272
<i>Bboissieri</i> 10_6x_S	271
<i>Bboissieri</i> 15_6x_S	271
<i>Bboissieri</i> 3_6x_S	271
<i>Bboissieri</i> 9_6x_S	271
<i>Bmexicanum</i> 347_4x_S	273
<i>Bmexicanum</i> 348H_S	273
<i>Bphoenicoides</i> 11_S	272
<i>Bphoenicoides</i> 452_6x_S	272
<i>Bphoenicoides</i> 453_S	272
<i>Bphoenicoides</i> 454_S	272
<i>Bphoenicoides</i> 553_6x_S	272
<i>Bretusum</i> 400_4x_S	271
<i>Bretusum</i> 403_6x_S	271
<i>Bretusum</i> 453_4x_S	271

<i>Bretusum</i> 454_4x_S	271
<i>Bretusum</i> 504_4x_S	271
<i>Bretusum</i> 555_4x_S	271
<i>Bretusum</i> 557_6x_S	271
<i>Bretusum</i> 561_6x_S	271
<hr/>	
5S-rDNA-LONG (L)	
<hr/>	
<i>Bdistachyon</i> Bd21_2x_L	371
<i>Bhybridum</i> ABR113_4x_L	383
<i>Barbuscula</i> 405_L	406
<i>Barbuscula</i> 502_2x_L	406
<i>Bbolussi</i> _L	395
<i>Bbreviglume</i> 33H_L	400
<i>Bflexum</i> 038_L	463
<i>Bflexum</i> 046_L	463
<i>Bgenuense</i> _2x_L	373
<i>Bglaucovirens</i> _2x_L	392
<i>Bkawakamii</i> _L	400
<i>Bkurilense</i> _L	400

<i>Bmadagascariense_L</i>	400
<i>Bmiserum67_L</i>	400
<i>Bphoenicoides10H_L</i>	358
<i>Bphoenicoides11_L</i>	373
<i>Bphoenicoides453_L</i>	374
<i>Bphoenicoides454_L</i>	373
<i>Bphoenicoides552_6x_L</i>	373
<i>Bphoenicoides553_6x_L</i>	395
<i>Bphoenicoides554_6x_L</i>	376
<i>Bphoenicoides6_4x_L</i>	375
<i>Bphoenicoides600_4x_L</i>	374
<i>Bpinnatum21_L</i>	373
<i>Bpinnatum34_4x_L</i>	373
<i>Bpinnatum38H_L</i>	373
<i>Bpinnatum39H_L</i>	373
<i>Bpinnatum41H_L</i>	373
<i>Bpinnatum42H_L</i>	373
<i>Bpinnatum43H_L</i>	373

<i>Bpinnatum</i> 502_2x_L	373
<i>Bpinnatum</i> 505_2x_L	373
<i>Bpinnatum</i> 514_4x_L	373
<i>Bpinnatum</i> 515_2x_L	373
<i>Bpinnatum</i> 519_4x_L	391
<i>Bretusum</i> 400_4x_L	392
<i>Bretusum</i> 403_6x_L	373
<i>Bretusum</i> 407_4x_L	391
<i>Bretusum</i> 408_6x_L	373
<i>Bretusum</i> 453_6x_L	375
<i>Bretusum</i> 454_4x_L1	409
<i>Bretusum</i> 454_4x_L2	375
<i>Bretusum</i> 504_4x_L	375
<i>Bretusum</i> 551_6x_L	391
<i>Bretusum</i> 555_4x_L1	408
<i>Bretusum</i> 555_4x_L2	375
<i>Bretusum</i> 561_6x_L	373
<i>Brupestre</i> 11_L	390

<i>Brupestre</i> 12_L	373
<i>Brupestre</i> 13H_L	373
<i>Brupestre</i> 15H_L	373
<i>Brupestre</i> 182_6x_L	390
<i>Brupestre</i> 439_4x_L	373
<i>Brupestre</i> 441_4x_L	388
<i>Brupestre</i> 442_4x_L	395
<i>Brupestre</i> 443_4x_L	373
<i>Brupestre</i> 444_4x_L	373
<i>Brupestre</i> 512_2x_L	373
<i>Brupestre</i> 600_6x_L	390
<i>Brupestre</i> 605_6x_L1	395
<i>Brupestre</i> 605_6x_L2	389
<i>Bsprygini</i> _L	390
<i>Bsylvaticum</i> 29H_L	392
<i>Bsylvaticum</i> 30H_L	392
<i>Bsylvaticum</i> 31H_L	392
<i>Bsylvaticum</i> 32H_L	391

Bsylvaticum434_2x_L	391
Bsylvaticum466_2x_L	392
Bsylvaticum467_2x_L	392
Bsylvaticum476_2x_L	392
Bsylvaticum501_2x_L	392
Bsylvaticum506_2x_L	392
Bsylvaticum508_2x_L	392
Bsylvaticum54_2x_L	392
Bsylvaticum552_2x_L	392
Bsylvaticum554_2x_L	392
Bsylvaticum63_2x_L	392

**Supplementary Table S5. (A)** Discordance metrics for the ASTRAL species tree phylogeny (Fig. 7) of the *Brachypodium* supertree (Plastome+35S+5S gene trees). Quartet support values (proportion of quartets in the gene trees that agrees with the species tree) for the main topology (q1), the first alternative (q2), and the second alternative topology (q3). Total number of nodes of the Astral topology: 118 (see tree node numbers in Fig. 7). NA (not applied) indicates that the effective number of gene trees for that node is zero, meaning that any gene tree matches the ASTRAL tree for the node.

No. Nodes	Q1	Q2	Q3
1	NA	NA	NA
2	NA	NA	NA
3	100.000	0.000	0.000
4	100.000	0.000	0.000

5	100.000	0.000	0.000
6	100.000	0.000	0.000
7	100.000	0.000	0.000
8	100.000	0.000	0.000
9	97.802	0.000	2.198
10	100.000	0.000	0.000
11	50.000	50.000	0.000
12	50.000	50.000	0.000
13	100.000	0.000	0.000
14	100.000	0.000	0.000
15	NA	NA	NA
16	NA	NA	NA
17	NA	NA	NA
18	NA	NA	NA
19	NA	NA	NA
20	100.000	0.000	0.000
21	NA	NA	NA
22	100.000	0.000	0.000

23	100.000	0.000	0.000
24	100.000	0.000	0.000
25	100.000	0.000	0.000
26	NA	NA	NA
27	100.000	0.000	0.000
28	NA	NA	NA
29	100.000	0.000	0.000
30	100.000	0.000	0.000
31	100.000	0.000	0.000
32	100.000	0.000	0.000
33	81.844	11.028	7.129
34	71.909	1.942	26.149
35	100.000	0.000	0.000
36	55.564	17.909	26.527
37	100.000	0.000	0.000
38	49.577	23.433	26.989
39	82.379	5.874	11.747
40	52.939	45.031	2.031



41	100.000	0.000	0.000
42	52.094	1.042	46.864
43	56.883	3.034	40.082
44	50.351	37.674	11.975
45	43.116	50.352	6.532
46	100.000	0.000	0.000
47	100.000	0.000	0.000
48	64.320	35.395	0.285
49	100.000	0.000	0.000
50	51.240	0.826	47.934
51	98.361	1.639	0.000
52	100.000	0.000	0.000
53	63.640	15.569	20.791
54	75.366	18.608	6.026
55	77.038	6.489	16.473
56	68.700	8.552	22.748
57	50.564	39.320	10.117
58	61.041	37.290	1.669

59	50.868	26.496	22.636
60	39.901	14.094	46.006
61	73.861	13.310	12.830
62	60.692	28.139	11.169
63	58.452	0.495	41.053
64	72.611	5.136	22.254
65	49.931	47.188	2.881
66	67.251	30.994	1.754
67	100.000	0.000	0.000
68	100.000	0.000	0.000
69	NA	NA	NA
70	NA	NA	NA
71	NA	NA	NA
72	NA	NA	NA
73	NA	NA	NA
74	68.795	11.282	19.922
75	54.123	24.887	20.990
76	68.044	9.243	22.713

*Supplementary data Chapter 2*

77	49.327	34.511	16.162
78	69.484	18.497	12.019
79	52.662	1.683	45.655
80	3.704	0.000	96.296
81	100.000	0.000	0.000
82	75.112	8.975	15.913
83	63.487	33.085	3.428
84	56.851	36.443	6.706
85	100.000	0.000	0.000
86	100.000	0.000	0.000
87	56.096	23.317	20.588
88	61.986	13.204	24.809
89	61.957	19.565	18.478
90	69.231	30.769	0.000
91	47.189	36.722	16.090
92	78.626	9.972	11.402
93	53.092	8.421	38.487
94	45.336	32.197	22.467

95	56.745	18.035	25.220
96	67.626	5.755	26.619
97	100.000	0.000	0.000
98	58.192	31.638	10.169
99	75.489	4.477	20.034
100	53.771	2.002	44.227
101	72.932	11.123	15.945
102	44.177	21.543	34.280
103	48.346	28.499	23.155
104	51.337	48.663	0.000
105	55.314	14.478	30.208
106	0.000	100.000	0.000
107	NA	NA	NA
108	100.000	0.000	0.000
109	85.933	6.584	7.484
110	57.526	14.342	28.132
111	34.466	33.131	32.403
112	36.863	17.843	45.294

113	100.000	0.000	0.000
114	61.948	14.362	23.689
115	60.841	0.000	39.159
116	68.016	15.992	15.992
117	50.000	36.979	13.021
118	100.000	0.000	0.000

**Supplementary Table S5. (B)** Range of overall quartets' support values for the *Brachypodium* ASTRAL species tree (Fig. 7) and number of nodes with different quartets' support percentages (100%, >50%, <50%, 0%) and NA.

Quartets	Support	No. nodes with percentage of support and NA				
	Overall	100%	>50%	<50%	0%	NA
Q1	3.7-100	35	89	13	1	16
Q2	0 - 100	1	4	98	38	16
Q3	0-96.3	0	1	101	41	16

## Supplementary figures and tables Chapter 3

**Supplementary Table S1.** Supplementary information on samples included in the repeatome analysis of *Brachypodium*. Taxon authorship, sample's code, chromosome number (2n), inferred ploidy level (nx), genome size (2C/pg), detailed localities and vouchers, and sources of cytogenetic and genomic data [Chromosome number (2n), Genome size (2C), Repeat content (RC)]. Ploidy levels were inferred from chromosome counting, genome sizes or repeat content. Asterisks indicate inferred 2n values according to those of the same taxonomic cytotypes (this study and previous records).

Taxon	Code	Chromosome number (2n)	Ploidy (nx)	Genome Size (2C/pg)	Locality	Data source
<i>B.distachyon</i> (L.) P. Beauv.	Bdis_Bd21-3	10	2x	0.631±0	Iraq: Salakudin (SRR3944701)	2n, 2C: Wolny and Hasterok, 2009; Catalán et al., 2012; Gordon et al., 2020. RC: this study
<i>B.stacei</i> Catalán, Joch. Müll., L.A.J. Mur & T. Langdon	Bsta_ABR114	20	2x	0.564±0	Spain: Balearic isles. Formentera (SRR423681)	2n, 2C: Catalán et al., 2012; Gordon et al. 2020; Wolny and Hasterok. 2009. RC: this study
<i>B. hybridum</i> Catalán, Joch. Müll., Hasterok & G. Jenkins	Bhyb_ABR113	30	2x	1.265±0	Portugal: Lisbon (SRR3945056; SRR3945058)	2n, 2C: Catalán et al., 2012. RC: this study
<i>B.arbuscula</i> Gay ex Knoche	Barb502	18	4x	0.713±0.004	Spain: La Gomera	2n, 2C : Sancho et al., 2022. RC: this study
<i>B.boissieri</i> Nyman	Bboi3	48	2x	3.236±0.072	Spain: Granada	2n, 2C : Sancho et al. 2022. RC: this study

<i>B.mexicanum</i> (Roem. & Schult.) Link	Bboi10	48	6x	3.152±0.04	Spain: Granada	2n, 2C: Decena et al., 2023 <sup>a</sup> . RC: this study
	Bboi15	48	6x	3.149±0.032	Spain: Granada	2n, 2C: Decena et al., 2023 <sup>a</sup> RC: this study
	Bmex347-2	40	6x	0.944	Mexico: Hidalgo	2C : Sancho et al. 2022. 2n b: Shi et al., 1993. RC: this study
	Bmex348H*	40*	6x*	0.944*	Mexico: Puebla	2n, 2C, RC: this study. From Herbarium B
	Bmex504*	40*	6x*	0.944*	Ecuador: Loja	2n, 2C RC: this study
<i>B.phoenicoides</i> (L.) P. Beauv. ex Roem. & Schult.	Bpho6-1R	28	4x	1.469±0.012	Spain: Huesca	2n, 2C: Decena et al., 2023 <sup>a</sup> . RC: this study
	Bpho422	28*	4x	1.443±0.019	Slovakia: Ružomberok	2C: Sancho et al., 2022. 2n, RC: this study
	Bpho452	38	4x	2.176±0.017	Morocco: Rift Mts	2n, 2C, RC: this study
	Bpho552	36	6x	2.204±0.039	Spain: Cadiz	2n, 2C: Decena et al., 2023a. RC: this study
	Bpho553	38	6x	2.183±0.013	Spain: Malaga	2n, 2C: Decena et al., 2023a. RC: this study
	Bpho554-1	38	6x	2.155±0.02	Spain: Granada	2n, 2C: Decena et al., 2023a
<i>B.pinnatum</i> (L.) P. Beauv.	Bpin505	18	6x	0.822±0.009	Norway: Oslo. USDA_PI 345964	2n, 2C: Decena et al., 2023a. RC: this study

<i>B. retusum</i> (Pers.) P. Beauv.	Bpin34	28	2x	1.449±0.018	Great Britain: North Wiltshire	2n, 2C: Decena et al., 2023a. RC: this study
	Bpin514	28	4x	1.537±0.012	Turkey: Samsun. USDA PI 206677	2n, 2C: Decena et al., 2023a. RC: this study
	Bpin520	28	4x	1.499±0.014	Netherlands: Scherpenzeel	2n, 2C: Decena et al., 2023a. RC: this study
	Bret400	32	4x	1704±0.024	Spain: Huesca	2n: Decena et al., 2023a. 2C: Wolny & Hasterok. 2009. RC: this study
	Bret407	32	4x	1.715±0.017	Spain: Huesca	2n, 2C: Decena et al., 2023a. RC: this study
	Bret453-4	32*	4x	1.855±0.1	Morocco: Rift Mts.	2n, 2C: Decena et al., 2023a. RC: this study
	Bret454	32*	4x	1.836±0.115	Morocco: Tazza-Bou Idir	2n, 2C: Decena et al., 2023a. RC: this study
	Bret504	32	4x	1.669±0.026	France: Vic la Gardiole	2n, 2C: Decena et al., 2023a. RC: this study
	Bret555	32	4x	1.715±0.017	Spain: Granada	2n, 2C: Decena et al., 2023a. RC: this study
	Bret403	42	4x	2.373±0.958	Spain: Huesca	2n, 2C: Decena et al., 2023a. RC: this study
	Bret408	42	6x	2.431±0.033	Spain: Navarra	2n, 2C: Decena et al., 2023a. RC: this study



<i>B. rupestre</i> (Host) Roem. & Schult.	Bret551	42*	6x	2.109±0.025	Spain: Malaga	2C: Decena et al., 2023a. 2n, RC: this study
	Bret557	42	6x	2.464±0.026	Spain: Cadiz	2n, 2C: Decena et al., 2023a. RC: this study
	Bret561	42	6x	2.362±0.046	Spain: Zaragoza	2n, 2C: Decena et al., 2023a. RC: this study
	Brup7	28	6x	1.562±0.016	Russia: Ulitsa Pushkina	2n, 2C: Decena et al., 2023a. RC: this study
	Brup439-1	28	4x	1.55±0.022	Spain: Huesca	2n, 2C: Decena et al., 2023a. RC: this study
	Brup441	28	4x	1.483±0.008	Spain: Leon	2n, 2C: Decena et al., 2023a. RC: this study
	Brup442	28	4x	1.56±0.03	Spain: Huesca	2n, 2C: Decena et al., 2023a. RC: this study
	Brup443	28	4x	1.498±0.012	Spain: Guipuzcoa	2n, 2C: Decena et al., 2023a. RC: this study
	Brup444	28	4x	1.492±0.021	Spain: Lugo	2n, 2C: Decena et al., 2023a. RC: this study
	Brup182	38	4x	2.258±0.026	Croatia: Istria	2n, 2C: Decena et al., 2023a. RC: this study
	Brup600	38	6x	2.216±0.013	France: Nans les Pins	2n, 2C: Decena et al., 2023a. RC: this study

	Brup605	38	6x	2.265±0.013	France: Pourrieres	2n, 2C: Decena et al., 2023a. RC: this study
	Bsyl154-1	18*	2x	0.888±0.008	Morocco: Rif Mountains	2n, 2C: Decena et al., 2023a. RC: this study
<i>B.sylvaticum</i> (Huds.) P. Beauv.	Bsyl466-6	18*	2x	0.928±0.013	Spain: Huesca	2n, 2C: Decena et al., 2023a. RC: this study
	Bsyl477-1	18*	2x	0.932±0.017	Spain: Lleida	2n, 2C: Decena et al., 2023a. RC: this study
	Bsyl501-6	18*	2x	0.947±0.01	France: Roquefort les Pins.	2n, 2C: Decena et al., 2023a. RC: this study

**Supplementary Table S2.** *Brachypodium* samples used in the repetitive DNA analysis. Genome skimming paired-end (PE) reads per sample and PE reads selected by Repeat Explorer2 per sample in the comparative analyses. Plastid reads were excluded from the analysis. *B. distachyon*, *B. stacei* and *B. hybridum* reads retrieved from NCBI accessions (see Table 1 and Supplementary Table S1); input reads: *B. distachyon* (Bd21-3): Fw (190.762.798) and Rv (190.762.798); *B. stacei* (ABR114): Fw (53.997.876) and Rv (53.997.876); *B. hybridum* (ABR113) Fw (1.203.802) and Rv (1.203.802) plus Fw (1.064.326) and Rv (1.064.326).

Taxon	Code	Ploidy	Genome Skimming (million reads )	RE2 input Reads (101 bp) 0.5x	RE2 input analysed Reads (101 bp)	Proportion of reads in top clusters (%) and 1Cx (Mbp)							Nreads
						(from RE2 summary)							
						Total		Plastid		Nuclear			
						%	1Cx (Mbp)	%	1Cx (Mbp)	%	1Cx (Mbp)		
<i>B.distachyon</i> (L.) P. Beauv.	Bdis_Bd21-3	2x	SRR4236817	1527520	1527520	26	80.23	2.15	6.63	23.85	73.59	19787	
<i>B.stacei</i> Catalán. Joch. Müll.. L.A.J. Mur & T. Langdon	Bsta_ABR114	2x	SRR3944701	1365328	1365328	22	60.68	0.85	2.34	21.15	58.33	37471	
<i>B.hybridum</i> Catalán. Joch. Müll.. Hasterok & G. Jenkins	Bhyb_ABR113	4x	SRR3945056; SRR3945058	3062302	3062302	23	71.14	0.66	2.04	22.34	69.10	0	
<i>B.arbuscula</i> Gay ex Knoche	Barb502	2x	12.090	1726024	1726024	29	101.11	5.99	20.88	23.01	80.23	0	
<i>B.boissieri</i> Nyman	Bboi3	6x	12.563	7833684	1127035	33	174.06	3.31	17.46	29.69	156.61	54655	
	Bboi10	6x	21.676	7630336	1296259	35	179.82	3.95	20.29	31.05	159.53	84870	
	Bboi15	6x	12.464	7623074	1049297	34	174.52	1.39	7.13	32.61	167.38	56906	

	Bmex347-2	4x	11.833	9136070	2988496	68	627.47	0.17	1.57	67.83	625.90	0
<b><i>B.mexicanum</i> (Roem. &amp; Schult.) Link</b>	Bmex504	4x	11.033	9136070	3783791	60	553.65	1.89	17.44	58.11	536.21	14447
	Bmex348H	4x	10.741	9136070	3820787	57	525.96	0.31	2.86	56.69	523.10	44018
<b><i>B.phoenicoides</i> (L.) P. Beauv. ex Roem. &amp; Schult.</b>	Bpho6-1R	4x	12.617	3493202	2211485	27	95.26	0.80	2.82	26.20	92.44	16660
	Bpho422	4x	13.586	3556144	1721391	28	100.57	0.05	0.18	27.95	100.39	570
	Bpho452	6x	20.076	5267644	2021816	26	92.22	1.65	5.85	24.35	86.37	152902
	Bpho552	6x	23.748	5335426	1946085	24	86.22	1.44	5.17	22.56	81.05	0
	Bpho553	6x	12.699	5284590	2315228	29	103.19	5.82	20.71	23.18	82.48	0
	Bpho554-1	6x	13.291	5216806	2109699	24	84.30	1.33	4.67	22.67	79.63	119355
<b><i>B.pinnatum</i> (L.) P. Beauv.</b>	Bpin505	2x	11.669	1989892	1989892	32	128.63	5.59	22.47	26.41	106.16	89568
	Bpin34	4x	12.632	3507728	1756795	29	102.74	4.71	16.69	24.29	86.05	0
	Bpin514	4x	28.729	3720758	1390935	27	101.47	2.22	8.34	24.78	93.12	0
	Bpin520	4x	11.215	3628768	1980512	25	91.63	2.69	9.86	22.31	81.77	81146
<b><i>B.retusum</i> (Pers.) P. Beauv.</b>	Bret400	4x	11.182	4125030	1835553	29	120.82	2.83	11.79	26.17	109.03	88389
	Bret407	4x	17.981	4151658	1914254	29	121.60	1.57	6.58	27.43	115.02	64751

	Bret453-4	4x	25608	4454258	1572982	28	125.9664	0.41	1.844508	27.59	124.121892	0
	Bret454	4x	22616	4507514	1763978	30	136.5777	2.36	10.7441124	27.64	125.8335876	83733
	Bret504	4x	27.892	4040302	1674933	31	126.50	4.07	16.61	26.93	109.89	62596
	Bret555	4x	30.216	4151658	1662812	29	121.60	1.26	5.28	27.74	116.32	79391
	Bret403	6x	17.596	5744540	2211465	30	116.04	3.75	14.50	26.25	101.53	84206
	Bret408	6x	26.345	5884946	1643744	26	103.03	1.56	6.18	24.44	96.84	0
	Bret551	6x	29.839	5105450	1787407	31	106.57	5.33	18.32	25.67	88.24	61278
	Bret557	6x	24.823	5964832	1862813	28	112.46	1.59	6.39	26.41	106.07	83175
	Bret561	6x	14.162	5717910	1734687	29	111.65	1.33	5.12	27.67	106.53	20965
<b><i>B. rupestre</i> (Host) Roem. &amp; Schult.</b>	Brup7	4x	15.196	3781278	1695272	28	106.93	2.22	8.48	25.78	98.46	0
	Brup439-1	4x	19.514	3752228	1505166	27	102.32	1.59	6.03	25.41	96.30	0
	Brup441	4x	15.849	3590034	1832351	25	90.65	1.44	5.22	23.56	85.43	88449
	Brup442	4x	18.421	3776436	1602794	26	99.17	2.27	8.66	23.73	90.51	89171
	Brup443	4x	14.913	3626346	1851291	27	98.89	2.11	7.73	24.89	91.16	0
	Brup444	4x	10.590	3611822	1788089	26	94.85	1.95	7.11	24.05	87.73	0

	Brup182	6x	14.523	5466148	1857595	29	106.74	3.74	13.77	25.26	92.97	0
	Brup600	6x	12.141	5364476	1402913	27	97.53	3.08	11.13	23.92	86.40	0
	Brup605	6x	11.457	5483094	1525977	26	95.99	1.36	5.02	24.64	90.97	0
<b><i>B.sylvaticum</i> (Huds.) P. Beauv.</b>	Bsyl54-1	2x	12.984	2149664	2149664	35	151.98	2.07	8.99	32.93	142.99	92343
	Bsyl466-6	2x	13.802	2246496	2246496	37	167.90	1.11	5.04	35.89	162.87	0
	Bsyl477-1	2x	14.281	2256178	2256178	37	168.63	2.77	12.62	34.23	156.00	0
	Bsyl501-6	2x	10.677	2292490	2292490	35	162.08	1.76	8.15	33.24	153.93	46623

**Supplementary Table S3.** Estimated sizes (Mbp) of repeats elements according to their proportions in the monoploid genome (1Cx). Kruskal-Wallis tests for significant differences in repeat proportions for each repetitive element across the studied samples. Significant values are highlighted in bold.

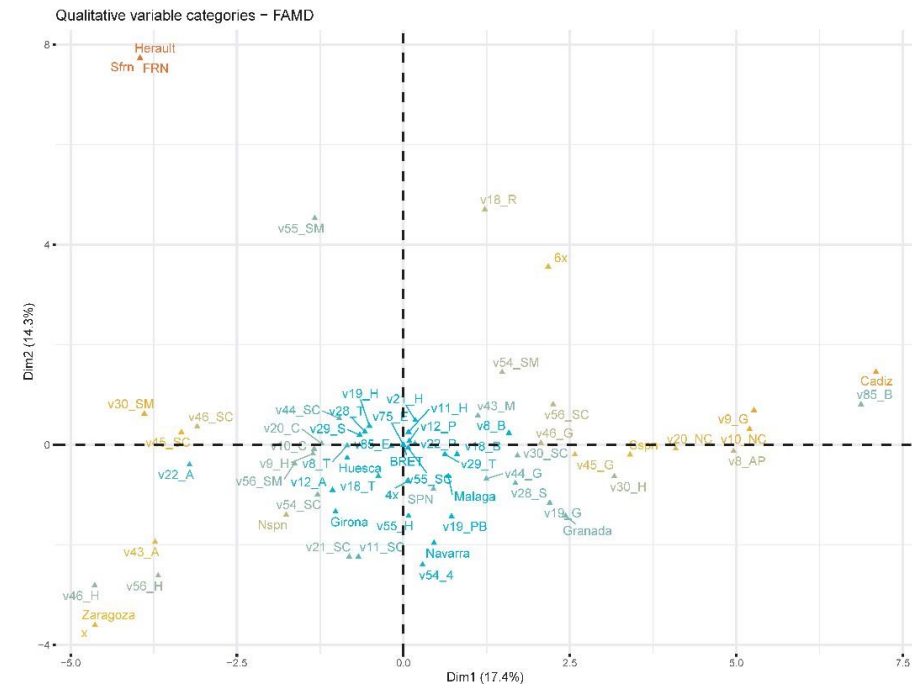
Code	Class I	Class LTR	Class I/Ty1 Copia										Class I/Ty3 gypsy										Class I			Class II/TIR			Class II		mobile element	satellite	4S rDNA	18S rDNA	25S rDNA	5S rDNA	Unclassified repeat	Unclassified	TOTAL NUCLEAR	unclustered (single-copy) nuclear
			Ty1 copia	Alu	Alu2	Angela	Blanca	Ikeros	Ivans	SIRE	TAR	Tork	Ty3 gypsy	Athila	Ogre	Retard	CRN	Teka	Reina	parametrovirus	LINE	EnSpm CACTA	hAT	MuDR/Mutator	PIF Harbinger	Helitron														
Bdis_Bd21-3	0.00	0.77	0.00	0.00	0.28	2.84	1.05	0.34	0.34	3.61	1.23	0.06	1.39	0.00	0.00	21.07	0.03	8.08	0.06	0.34	0.03	2.35	0.00	1.11	0.06	0.00	0.00	0.03	6.63	0.00	0.00	0.28	4.01	14.19	70.20	238.3				
Bsta_ABR114	0.00	2.68	0.00	0.08	0.00	4.22	0.19	0.41	0.11	2.92	0.80	0.03	0.00	4.66	0.00	3.64	1.77	9.05	0.00	0.00	0.19	0.88	0.00	0.41	0.00	0.00	0.11	0.33	1.35	0.00	0.00	0.19	7.56	16.16	57.76	218.0				
Bhyb_ABR113	0.00	6.03	0.00	0.34	0.00	1.82	1.42	0.49	0.00	2.94	0.59	0.62	0.00	2.07	0.00	0.19	3.68	3.71	0.00	0.87	0.09	0.62	0.06	0.15	0.06	0.00	0.00	0.49	0.00	0.12	0.03	0.03	0.00	41.75	68.20	241.0				
Barb502	0.00	0.14	0.00	0.00	0.00	0.70	0.87	0.94	0.10	6.55	0.77	0.87	0.00	0.00	0.00	14.02	14.71	3.84	0.00	0.63	0.59	1.19	0.00	5.19	0.59	0.00	0.00	0.21	3.38	0.00	0.00	0.98	0.00	22.31	78.59	270.0				
Bboi3	0.00	5.59	0.00	0.00	0.00	0.95	1.00	1.00	0.00	7.65	2.64	1.37	0.00	0.00	0.00	55.17	3.01	7.49	0.00	0.79	0.84	2.90	0.00	2.11	1.11	0.00	0.00	6.07	3.48	0.00	0.00	0.53	25.58	25.53	154.8	372.6				
Bboi10	0.00	5.04	0.00	0.00	0.00	0.87	0.51	0.82	0.00	6.27	2.41	1.03	0.00	0.00	0.00	60.47	2.31	7.55	0.00	0.36	0.72	3.19	0.00	2.21	1.08	0.00	0.00	1.13	8.73	0.00	0.00	0.82	33.65	21.22	160.4	353.3				
Bboi15	0.00	5.75	0.00	0.00	0.00	0.87	0.62	1.08	0.05	6.31	2.41	1.28	0.00	0.00	0.00	67.70	2.36	6.57	0.00	0.21	0.82	3.23	0.00	2.16	1.08	0.00	0.00	7.80	4.05	0.00	0.00	0.62	27.82	22.43	165.2	348.0				
Bmex347-2	0.00	12.92	0.00	1.85	0.00	6.64	1.85	4.89	1.85	20.48	8.03	0.00	0.00	2.58	17.81	191.0	7.47	249.4	0.00	0.00	0.18	19.47	0.00	31.28	0.28	0.09	0.00	15.41	7.01	0.00	0.00	1.66	0.00	25.01	627.1	295.5				
Bmex504	0.00	2.77	0.00	1.85	0.00	0.00	5.44	6.09	2.58	14.49	7.84	0.00	0.00	13.38	10.52	155.3	20.58	173.0	0.18	1.11	0.46	21.22	3.05	24.27	3.23	3.23	8.58	12.27	4.15	0.00	0.00	2.40	3.51	32.02	533.5	389.2				
Bmex348H	0.00	9.04	0.00	1.48	0.00	7.01	2.03	4.71	1.48	9.41	7.01	3.14	0.00	0.83	17.81	153.5	1.85	201.5	0.00	0.65	0.74	8.95	1.66	21.13	1.11	0.37	0.37	13.20	6.09	0.00	0.00	1.57	10.61	35.34	522.6	400.1				
Bpho6-1R	0.00	5.29	0.00	0.11	0.00	0.18	0.35	0.60	0.11	5.96	2.43	1.09	0.00	0.00	0.00	16.69	15.56	6.88	0.00	0.07	0.64	2.96	0.00	3.56	1.23	0.00	0.00	1.62	0.00	0.00	1.55	2.65	21.94	91.48	261.3					
Bpho422	0.00	6.72	0.00	0.00	0.00	0.25	0.00	0.65	0.00	5.85	2.59	1.26	0.00	0.00	0.00	19.65	3.09	7.43	0.00	0.00	0.32	4.85	0.00	3.70	1.19	0.00	0.00	17.67	5.42	0.00	0.00	0.57	0.11	18.82	100.1	259.0				
Bpho452	0.00	5.78	0.00	0.04	0.00	0.21	0.43	0.46	0.11	6.77	1.63	0.64	0.00	0.00	0.00	15.57	0.11	4.04	0.00	0.60	0.25	2.52	0.07	3.16	0.99	0.00	0.00	0.60	0.00	0.00	0.35	26.81	14.90	86.05	268.6					
Bpho552	0.00	1.44	0.00	0.07	0.00	0.22	0.43	0.50	0.14	6.50	1.58	0.65	0.00	0.40	0.00	15.05	2.19	7.08	0.00	0.47	0.50	2.66	0.00	2.84	0.97	0.00	0.00	18.93	2.69	0.00	0.00	0.61	0.00	15.88	81.80	277.4				
Bpho553	0.00	6.90	0.00	0.11	0.00	0.21	0.39	0.53	0.18	5.59	1.57	0.00	0.00	0.04	0.00	14.84	2.17	2.21	0.00	0.89	0.78	2.17	0.00	2.99	1.28	0.00	0.00	15.73	5.27	0.00	0.00	0.53	0.00	18.33	82.69	273.1				
Bpho554-1	0.00	7.24	0.00	0.07	0.00	0.28	0.42	0.49	0.00	6.29	1.62	0.74	0.00	0.56	0.00	11.91	0.11	1.97	0.00	0.39	0.63	2.63	0.00	3.06	0.70	0.00	0.00	1.65	0.00	0.00	0.39	19.88	18.76	79.77	271.4					

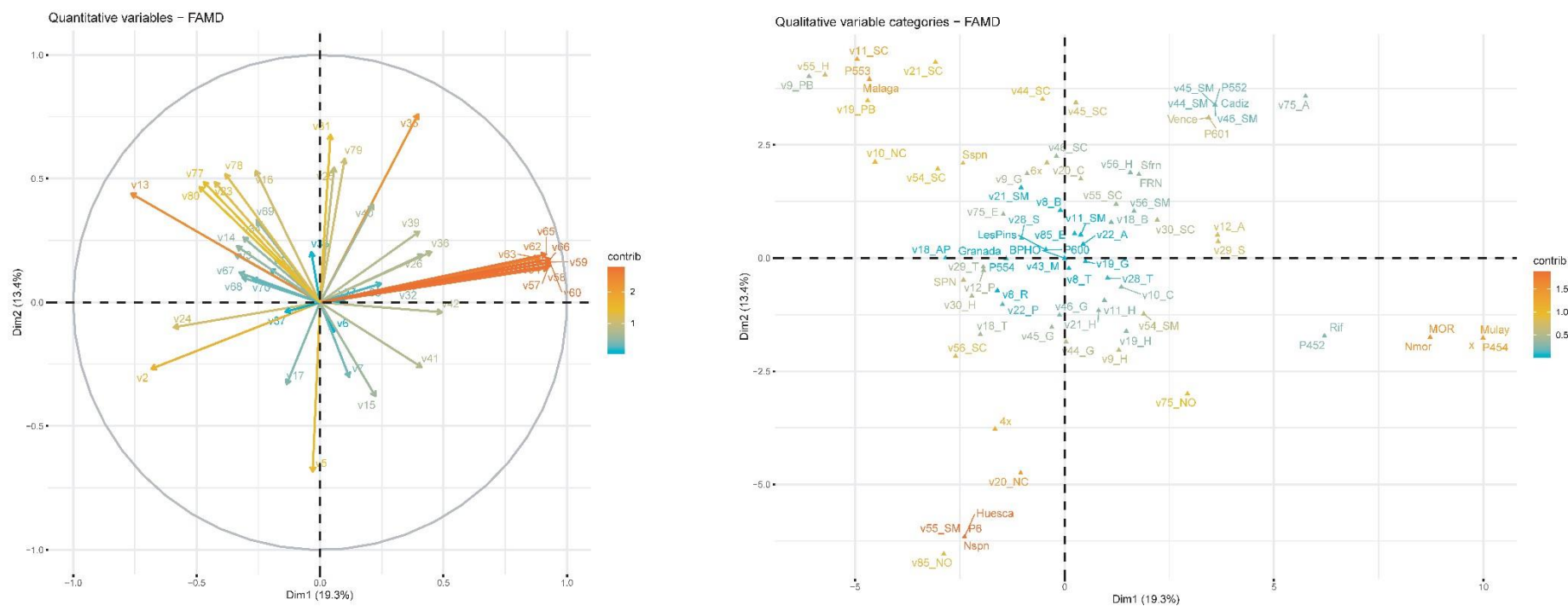
Bpin505	0.00	4.58	0.00	0.04	0.00	0.04	1.00	0.72	0.04	7.80	2.33	1.09	0.00	0.04	0.00	22.67	0.12	9.93	0.00	0.00	0.76	4.86	0.00	4.78	1.00	1.81	0.00	1.17	5.63	0.00	0.00	0.32	18.09	18.41	107.2	294.7
Bpin34	0.00	8.36	0.00	0.00	0.00	0.00	0.46	0.71	0.00	5.63	1.70	0.92	0.00	0.00	0.00	18.78	2.16	3.33	0.00	0.00	0.25	3.12	0.00	3.01	1.10	0.00	0.00	17.43	4.04	0.00	0.00	0.21	0.00	14.95	86.16	268.1
Bpin514	0.00	6.76	0.00	0.00	0.00	0.49	0.56	0.79	0.00	6.43	2.14	0.64	0.00	0.00	0.00	20.59	2.18	4.55	0.00	0.00	0.00	3.91	0.00	3.34	1.13	0.00	0.00	22.74	2.71	0.00	0.00	0.38	0.00	14.28	93.61	282.1
Bpin520	0.00	9.46	0.00	0.07	0.00	0.37	0.66	0.70	0.00	6.93	1.72	1.14	0.00	0.62	0.00	14.04	0.00	3.74	0.00	0.00	0.48	3.67	0.00	3.30	0.99	0.00	0.00	0.95	2.31	0.00	0.00	0.26	15.03	15.94	82.35	284.1
Bret400	0.00	5.58	0.00	0.00	0.00	0.71	0.79	0.87	0.25	6.83	1.83	1.00	0.00	0.00	0.00	27.46	2.79	8.33	0.00	0.62	0.87	2.62	0.00	2.92	0.92	0.00	0.00	1.21	0.00	0.00	0.00	0.33	20.08	21.87	107.9	308.7
Bret407	0.00	4.07	0.00	0.00	0.00	0.50	0.63	0.80	0.34	6.67	1.84	0.88	0.00	0.00	0.00	32.62	2.10	10.27	0.00	0.55	0.29	3.19	0.04	2.89	1.59	0.00	0.00	4.65	4.95	0.00	0.00	0.50	14.17	21.26	114.8	304.5
Bret453-4	0.00	9.31	0.00	0.04	0.00	0.63	0.72	0.90	0.36	7.92	2.16	1.03	0.00	0.00	0.00	34.87	2.43	7.96	0.00	0.81	0.63	4.45	0.00	2.88	1.12	0.00	0.00	22.99	4.00	0.00	0.00	0.31	0.00	20.15	125.7	324.1
Bret454	0.00	6.83	0.00	0.05	0.00	0.68	0.82	0.82	0.41	7.38	1.82	0.91	0.00	0.00	0.00	33.37	2.73	8.88	0.00	0.64	0.18	2.82	0.00	3.23	1.37	0.00	0.00	2.37	4.32	0.00	0.00	0.77	21.62	21.72	123.7	331.5
Bret504	0.00	5.22	0.00	0.00	0.00	0.16	0.65	0.82	0.41	5.75	1.75	0.90	0.00	0.00	0.00	32.44	2.20	7.75	0.00	0.65	0.16	4.53	0.00	2.69	1.35	0.00	0.00	4.82	4.94	0.00	0.00	0.53	15.26	17.02	110.0	298.0
Bret555	0.00	1.55	0.00	0.00	0.00	0.67	0.55	0.80	0.38	6.79	1.89	0.80	0.00	0.00	0.00	34.26	2.01	12.62	0.00	0.63	0.13	3.14	0.00	3.06	1.43	0.00	0.00	1.89	3.94	0.00	0.00	0.55	20.00	19.20	116.2	303.0
Bret403	0.00	5.26	0.00	0.19	0.00	0.50	0.58	0.73	0.19	6.61	1.86	1.24	0.00	0.00	0.00	27.54	2.36	8.51	0.00	0.35	0.70	2.36	0.00	2.94	1.43	0.00	0.00	5.22	0.00	0.00	0.00	0.46	14.74	18.30	102.0	284.7
Bret408	0.00	5.51	0.00	0.00	0.00	0.00	0.63	0.83	0.00	7.41	1.98	1.03	0.00	0.16	0.00	20.88	2.50	8.12	0.00	0.00	0.40	2.77	0.00	3.69	1.07	0.00	0.00	20.96	4.36	0.00	0.00	0.24	0.00	16.40	98.94	297.3
Bret551	0.00	3.30	0.00	0.00	0.00	0.24	0.52	0.58	0.31	5.40	1.48	0.00	0.00	0.00	0.00	24.34	2.06	6.88	0.00	0.48	0.10	3.23	0.00	2.34	1.03	0.00	0.00	4.78	4.92	0.00	0.00	0.31	11.79	14.58	88.66	255.1
Bret557	0.00	5.78	0.00	0.00	0.00	0.40	0.56	0.80	0.28	8.27	1.77	0.68	0.00	0.08	0.00	27.83	2.01	5.26	0.00	0.48	0.84	2.73	0.00	3.01	1.33	0.00	0.00	2.13	4.34	0.00	0.00	0.56	17.95	17.83	104.9	296.6
Bret561	0.00	5.66	0.00	0.12	0.00	0.27	0.50	0.77	0.19	6.31	1.89	1.23	0.00	0.00	0.00	29.38	2.43	7.05	0.00	0.46	0.15	2.54	0.00	2.73	1.12	0.00	0.00	18.52	0.00	0.00	0.00	0.39	4.66	18.48	104.8	280.1
Brup7	0.00	5.84	0.00	0.19	0.00	0.15	0.50	0.80	0.00	6.45	2.18	0.80	0.00	0.00	0.00	19.32	2.18	6.30	0.00	0.00	0.15	2.83	0.00	3.40	1.34	0.00	0.00	20.81	5.84	0.00	0.00	0.42	0.00	17.53	97.04	284.8
Brup439	0.00	5.84	0.00	0.00	0.00	0.00	0.53	0.87	0.00	6.67	1.78	0.99	0.00	0.11	0.00	19.29	2.20	7.35	0.00	0.00	0.08	2.73	0.00	3.26	0.87	0.00	0.00	22.17	5.68	0.00	0.00	0.45	0.00	15.05	95.92	283.0
Brup441	0.00	4.90	0.00	0.00	0.00	0.25	0.47	0.54	0.00	5.58	1.92	1.02	0.00	0.00	0.00	15.70	2.28	5.87	0.00	0.73	0.33	1.99	0.00	2.72	1.23	0.00	0.00	1.23	3.26	0.00	0.00	0.47	17.51	16.90	84.92	277.6
Brup442	0.00	9.27	0.00	0.00	0.00	0.19	0.57	0.69	0.00	6.60	1.75	0.92	0.00	0.04	0.00	18.46	1.98	3.74	0.00	0.00	0.00	3.28	0.00	3.01	1.14	0.00	0.00	1.30	3.39	0.00	0.00	0.57	21.21	14.23	92.34	289.0
Brup443	0.00	1.21	0.00	0.04	0.00	0.00	0.55	0.70	0.04	6.45	1.72	0.92	0.00	0.04	0.00	19.59	2.12	13.41	0.00	0.00	0.44	3.74	0.00	3.59	1.43	0.00	0.00	18.97	2.23	0.00	0.00	0.37	0.00	14.80	92.33	273.93
Brup444	0.00	4.63	0.00	0.00	0.00	0.04	0.58	0.62	0.00	6.24	1.68	0.00	0.00	0.04	0.00	18.53	2.12	7.44	0.00	0.00	0.15	2.19	0.00	3.06	0.95	0.00	0.00	17.04	4.41	0.00	0.00	0.55	0.00	17.84	88.10	276.7
Brup182	0.00	4.60	0.00	0.00	0.00	0.26	0.55	0.81	0.00	7.25	2.21	0.88	0.00	0.04	0.00	21.72	2.02	6.55	0.00	0.00	0.04	2.91	0.00	3.50	1.29	0.00	0.00	16.38	3.42	0.00	0.00	0.44	0.00	16.89	91.76	276.3
Brup600	0.00	1.52	0.00	0.00	0.00	0.00	0.47	0.61	0.00	6.18	1.95	0.72	0.00	0.00	0.00	19.14	1.95	8.34	0.00	0.00	0.07	4.05	0.00	2.82	0.76	0.00	0.00	19.94	3.54	0.00	0.00	0.33	0.00	15.17	87.56	273.6



Brup605	0.00	8.09	0.00	0.00	0.00	0.26	0.52	0.70	0.00	6.57	2.22	0.78	0.00	0.00	0.00	19.64	2.14	3.03	0.00	0.04	0.04	2.66	0.00	3.25	1.26	0.00	0.00	19.27	3.47	0.00	0.00	0.26	0.00	17.13	91.30	277.8
Bsy154-1	0.00	0.13	3.13	0.00	0.00	1.61	0.83	1.65	0.48	7.90	0.00	0.00	0.00	1.26	0.00	22.75	2.61	45.81	0.00	0.26	0.26	4.99	0.00	6.99	1.52	0.00	0.00	1.65	0.00	0.00	0.00	0.96	18.67	20.32	143.7	290.4
Bsy1466-6	0.00	6.13	3.27	0.00	0.00	1.59	0.91	1.72	0.36	8.71	0.00	0.00	0.00	1.18	0.00	27.09	21.87	42.43	0.00	0.05	0.73	4.40	0.36	8.21	1.32	0.14	0.00	3.90	7.44	0.00	0.00	0.32	0.00	21.60	163.7	290.0
Bsy1477-1	0.46	3.14	3.01	0.00	0.00	1.78	0.73	1.50	0.14	7.66	0.00	0.00	0.00	0.50	0.00	28.39	9.30	46.76	0.00	0.09	0.09	6.29	0.00	7.66	1.28	0.09	0.00	11.21	11.62	0.00	0.00	0.64	0.00	15.91	158.6	297.0
Bsy1501-6	0.00	6.16	3.06	0.00	0.00	1.81	0.69	1.39	0.14	7.78	0.00	0.00	0.00	1.16	0.00	24.59	8.80	40.24	0.00	0.23	0.28	4.03	0.05	6.76	1.39	0.09	0.00	2.45	13.48	0.00	0.00	0.42	9.40	18.15	152.5	310.5
mean	0.01	5.29	0.28	0.15	0.01	0.93	0.78	1.10	0.26	7.04	2.06	0.76	0.03	0.68	1.05	33.89	3.92	23.87	0.01	0.33	0.37	4.04	0.12	4.87	1.11	0.13	0.21	9.12	4.01	0.00	0.00	0.58	9.14	19.47	135.6	292.0
sd	0.07	2.70	0.91	0.44	0.04	1.54	0.81	1.18	0.51	2.71	1.67	0.58	0.21	2.14	4.03	38.67	5.01	52.24	0.03	0.32	0.28	3.86	0.52	5.98	0.50	0.55	1.29	8.26	2.91	0.02	0.00	0.45	10.06	5.55	120.1	36.45
Kruskall wallis test	4.50	<u>7.14</u>	<u>19.30</u>	1.91	4.50	<u>7.77</u>	<u>8.11</u>	2.30	2.98	1.09	<u>11.97</u>	<u>6.09</u>	4.50	4.82	3.44	0.25	1.91	<u>8.39</u>	1.76	0.89	1.42	3.78	1.52	<u>6.32</u>	0.91	<u>9.06</u>	1.67	<u>7.48</u>	2.94	1.10	1.10	0.78	0.93	0.71	0.66	0.78
p-value	0.11	<u>0.03</u>	<u>0.00</u>	0.39	0.11	<u>0.02</u>	<u>0.02</u>	0.32	0.23	0.58	<u>0.00</u>	<u>0.05</u>	0.11	0.09	0.18	0.88	0.38	<u>0.02</u>	0.41	0.64	0.49	0.15	0.47	<u>0.04</u>	0.64	<u>0.01</u>	0.43	<u>0.02</u>	0.23	0.58	0.58	0.68	0.63	0.70	0.72	0.68

**A** *B. retusum*



**B** *B. phoenicoides*

**Supplementary Figure S2.** Principal Component Analysis (PCA) of 26 samples of *Brachypodium phoenicoides*. Contribution of quantitative variables to the two first dimensions. Contribution is represented by a colour gradient with higher levels in orange and lowest in blue.

## Supplementary data Chapter 4

**Supplementary Table S1:** Samples included in the morphoanatomical and epidermal study of *Brachypodium retusum* and *B. phoenicoides*. Species names and authorities, population code, sampling size (number of individuals studied), chromosome number (2n), ploidy level (nx), locality of origin, geographical coordinates (latitude, longitude), and geographical region (MOR: Morocco; NSPA: North Spain; SSPA: South Spain; SFRN: South France) are indicated for each sample. Asterisks and diamond indicate populations included only in the epidermal or in the morphoanatomical analysis, respectively.

Specie	Population Code	Sampling size	2n	Ploidy	Locality	Geographical coordinates	Geographical region
<i>B. phoenicoides</i> (L.) P. Beauv. ex Roem. & Schult.	Bphoenicoides6	5	28	4x	Spain: Huesca	42.21173, -0.187960	NESP
	Bphoenicoides452	1	38	6x	Morocco: Rift Mts	35.25486, -5.153444	NMAR
	Bphoenicoides453*	1	-	-	Morocco: Mulay-Idriss	34.03331, -5.502972	NMAR
	Bphoenicoides454	2	-	-	Morocco: Alcazarseguir	35.8467611 -5.5431361	NMAR
	Bphoenicoides552	1	38	6x	Spain: Cadiz	36.5319, -5.665000	SESP
	Bphoenicoides553	5	38	6x	Spain: Malaga	36.74333, -4.924444	SESP
	Bphoenicoides554	5	38	6x	Spain: Granada	37.1363, -3.485280	SESP
	Bphoenicoides600	3	28	4x	France: Nans le Pins	43.35944, 5.785268	SFRN
	Bphoenicoides601	4	36	6x	France: Vence	43.747525, 7.097186	SFRN
<i>B. retusum</i> (Pers.) P. Beauv.	Bretusum400	1	32	4x	Spain: Huesca	42.126786, -0.161747	NESP
	Bretusum403	1	42	6x	Spain: Huesca	42.552500, -0.681389	NESP
	Bretusum407	3	32	4x	Spain: Huesca	41.744166, -0.483334	NESP
	Bretusum408	1	42	6x	Spain: Navarra	42.398889, -1.801389	NESP
	Bretusum451*	1	32	4x	Morocco: Rif Mountains	35.1781389, -5.147223	NMAR
	Bretusum452*	1	32	4x	Morocco: Talembote to Djbel Tazzaot	35.260055 -5.162416	NMAR
	Bretusum455*	1	-	4x	Morocco: Alcazarseguir	35.846761, -5.543136	NMAR
	Bretusum504	1	32	4x	France: Vic la Gardiole	43.503258, 3.781791	SFRN
	Bretusum551	5	42	6x	Spain: Malaga	36.774722, -5.093889	SESP
	Bretusum555	5	32	4x	Spain: Granada	37.372498, -2.718061	SESP
	Bretusum556	5	32	4x	Spain: Jaen	38.2747222, -2.75805555	SESP
	Bretusum557	5	42	6x	Spain: Cadiz	36.149722, -5.627500	SESP
	Bretusum558	4	32	4x	Spain: Girona	42.3728, 2.803275	NESP
	Bretusum559	5	32	4x	Spain: Huesca	41.9341667, -0.25027777	NESP
	Bretusum560	5	-	-	Spain: Zaragoza	40.3040667, -1.877830557	NESP
	Bretusum561*	1	42	6x	Spain: Zaragoza	42.14366, -1.23367	NESP

**Supplementary Table S2.** Morphological, anatomical and epidermal trait variables used in the phenotypic analysis of the studied *Brachypodium retusum* and *B. phoenicoides* samples and their respective measurement scales (quantitative variables) or character states (qualitative variables) and their acronyms.

Variable code	Variable description		
	plant organ	trait	acronym
<b>Morphological traits:</b>			
V52	Rhizomes	Rhizome presence (present, P; absent, A)	RP
V53	Rhizomes	Rhizome type (strong, S; weak, W)	RT
V1	Culm	Culm length (cm)	CL
V2	Culm	Number of nodes of culm	NNC
V54	Culm	Culm hairiness (hairy, H; scabrid, SC; smooth, SM)	CH
V55	Culm	Culm node hairiness (hairy, H; scabrid, SC; smooth, SM)	CNH
V3	Innovation leaf	Leaf sheath length (cm)	InLSL
V4	Innovation leaf	Leaf sheath closing length (cm)	InLSCL
V5	Innovation leaf	Leaf blade length (cm)	InLBL
V6	Innovation leaf	Leaf blade width (mm)	InLBW
V11	Innovation leaf	Leaf blade hairiness (hairy, H; scabrid, SC; smooth, SM)	InLBH
V7	Innovation leaf	Ligule length (mm)	InLiL
V8	Innovation leaf	Ligule apex shape (Truncated, T; rounded, R; Apiculate, AP; bifid, B)	InLiAS
V9	Innovation leaf	Ligule pubescence (Hairy, H; Pubescent, PB; Glabrous, G)	InLiP
V10	Innovation leaf	Ligule edge (Ciliate, C; Non-Ciliate: NC)	InLiE
V12	Innovation leaf	Ligule auricles (present, P; absent, A)	InLiAU
V13	Flag leaf	Leaf sheath length (cm)	FgLSL

## Supplementary data Chapter 4

V14	Flag leaf	Leaf sheath closing length (cm)	FgISCL
V15	Flag leaf	Leaf blade lenght (cm)	FgLBL
V16	Flag leaf	Leaf blade width (mm)	FgLBW
V21	Flag leaf	Leaf blade hairiness (hairy, H; scabrid, SC; smooth, SM)	FgLBH
V17	Flag leaf	Ligule length (mm)	FgLiL
V18	Flag leaf	Ligule apex shape (Truncated, T; rounded, R; Apiculate, AP; bifidus, B)	FgLiAS
V19	Flag leaf	Ligule pubescence (Hairy, H; Pubescent, PB; Glabrous, G)	FgLiP
V20	Flag leaf	Ligule edge (Ciliate, C; Non-Ciliate: NC)	FgLiE
V22	Flag leaf	Ligule auricles (present, P; absent, A)	FgLiAU
V23	Inflorescence	Inflorescence length (cm)	IL
V24	Inflorescence	Number of spikelets/inflorescence	NSI
V25	Inflorescence	Distance between 1st and 2nd spikelet (mm)	DFSSp
V56	Inflorescence	Inflorescence axis hairiness (hairy, H; scabrid, SC; smooth, SM)	IAH
V26	Inflorescence	Spikelet length (mm)	SL
V27	Inflorescence	Number of florets/spikelet	NFSp
V28	Inflorescence	Spikelet apex (Straight, S; Twisted, T)	SpA
V29	Inflorescence	Spikelet rachis shape (Straight, S; Twisted, T)	SpR
V30	Inflorescence	Spikelet hairiness (hairy, H; scabrid, SC; smooth, SM)	SpH
V31	Inflorescence	Lower glume (G1) length (mm)	LGL
V32	Inflorescence	Lower glume (G1 ) width (mm)	LGW
V33	Inflorescence	Lower glume (G1) number of veins	LGNV
V34	Inflorescence	Lower glume (G1) number of secondary veins	LGSV
V44	Inflorescence	Lower glume (G1) hairiness (Hairy, H; Scabrid, SC; Glabrous, G; Smooth, SM)	LGH
V35	Inflorescence	Upper glume (G2) length (mm)	UGL

## Supplementary data Chapter 4

V36	Inflorescence	Upper glume (G2) width (mm)	UGW
V37	Inflorescence	Upper glume (G2) number of veins	UGNV
V38	Inflorescence	Upper glume (G2) number of secondary veins	UGSC?
V45	Inflorescence	Upper glume (G2) hairiness (Hairy, H; Scabrid, SC; Glabrous, G; Smooth, SM)	UGH
V43	Inflorescence	Lower and upper glume (G1, G2) apex (Mucronate, M; Awned, A)	LGA?
V39	Inflorescence	Lemma length (mm)	LL
V40	Inflorescence	Lemma width (mm)	LW
V41	Inflorescence	Lemma number of veins	LNV
V46	Inflorescence	Lemma hairiness (Hairy, H; Scabrid, SC; Glabrous, G; Smooth, SM)	LH
V42	Inflorescence	Awn length (mm)	AwL
V47	Inflorescence	Number of anthers	NuAn
V48	Inflorescence	Anther length (mm)	AnL
V49	Inflorescence	Caryopsis length (mm)	CL
V50	Inflorescence	Caryopsis color (Brown, B; Yellow, Y)	CC
V51	Inflorescence	Ovary tip (hairy, H; scabrid, SC; smooth, SM)	OT
<b>Anatomical traits:</b>			
V77	Innovation leaf	Leaf blade number of veins	InLBNV
V78	Innovation leaf	Leaf blade number of valleys	InLBNVa
V79	Innovation leaf	Leaf blade number of 1ary veins	InLBN1aV
V80	Innovation leaf	Leaf blade number of 2ary veins	InLBN2aV
V62	Innovation leaf	Leaf blade 1ary veins height (µm)	InLB1aVH
V63	Innovation leaf	Leaf blade 2ary veins height (µm)	InLB2aVH
V64	Innovation leaf	Leaf blade valleys' height (µm)	InLBVaH
V65	Innovation leaf	Leaf blade bulliform cells' height (µm)	InLBBuCH

## Supplementary data Chapter 4

V66	Innovation leaf	Leaf blade hairs' length (µm)	InLBHaL
V81	Innovation leaf	Leaf blade main vein adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	InLBScBr
V82	Innovation leaf	Leaf blade 1ary lateral veins adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	InLBMVScBr
V83	Innovation leaf	Leaf blade 2ary lateral veins adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	InLB1aLVScBr
V84	Innovation leaf	Leaf blade bulliform cells (Presence, P; Absence, A)	InLBBuC
V85	Innovation	Leaf blade hook-type trichomes (absence, NO; abaxial, E ; adaxial, H)	InLBHoT
V86	Innovation	Leaf blade long-type trichomes (absence, NO; abaxial, E ; adaxial, H)	InLBLoT
V67	Flag leaf	Leaf blade number of veins	FgLBNV
V68	Flag leaf	Leaf blade number of valleys	FgLBNVa
V69	Flag leaf	Leaf blade number of 1ary veins	FgLBN1aV
V70	Flag leaf	Leaf blade number of 2ary veins	FgLBN2aV
V57	Flag leaf	Leaf blade 1ary veins height (µm)	FgLB1aVH
V58	Flag leaf	Leaf blade 2ary veins height (µm)	FgLB2aVH
V59	Flag leaf	Leaf blade valleys' height (µm)	FgLBVaH
V60	Flag leaf	Leaf blade bulliform cells' height (µm)	FgLBBuCH
V61	Flag leaf	Leaf blade hairs' length (µm)	FgLBHaL
V71	Flag leaf	Leaf blade adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	FgLBSBr
V72	Flag leaf	Leaf blade 1ary lateral veins adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	FgLB1aLVScBr
V73	Flag leaf	Leaf blade 2ary lateral veins adaxial and abaxial sclerenchyma bridges (Bridge, B; Absence, A)	FgLB2aLVScBr
V74	Flag leaf	Leaf blade bulliform cells (Presence, P; Absence, A)	FgLBBuC
V75	Flag leaf	Leaf blade hook-type trichomes (absence, NO; abaxial, E ; adaxial, H)	FgLBHoT



V76	Flag leaf	Leaf blade long-type trichomes (absence, NO; abaxial, E ; adaxial, H)	FgLBLot
<b>Epidermal traits (abaxial epidermis):</b>			
V87	Abaxial leaf blade	Number of stomata/field x40	AbEpStpF
V88	Abaxial leaf blade	Number of Epidermal cells/field 40x	AbEpNuEpCpF
V89	Abaxial leaf blade	Stomatal index	AbEpStIn
V90	Abaxial leaf blade	Stomata Guardian cells length (µm)	AbEpSGCL
V91	Abaxial leaf ribs	Long cells type (suberous, SZ; siliceous rounded, So; siliceous no-rounded, Z)	AbEpRLCT
V92	Abaxial leaf ribs	Long cell length (µm)	AbEpRLCL
V93	Abaxial leaf ribs	Short cells type (suberous, SZ; siliceous rounded, So; siliceous no-rounded, Z)	AbEpRSCT
V94	Abaxial leaf ribs	Short cells length (µm) [suberous: SZ]	AbEpRSCSZL
V95	Abaxial leaf ribs	Short cells length (µm) [siliceous rounded: So]	AbEpRSCSoL
V96	Abaxial leaf ribs	Short cells length (µm) [siliceous no-rounded:Z]	AbEpRSCZL
V97	Abaxial leaf ribs	Suberous cells (absence,NA; presence, SZ)	AbEpRSCSZ
V98	Abaxial leaf ribs	Siliceous cells (absence, NA; presence, Z)	AbEpRSCSo
V99	Abaxial leaf ribs	Hook trichomes (absence, No; presence, Y)	AbEpRHtr
V100	Abaxial leaf ribs	Hook trichomes (length. µm)	AbEpRHtrL
V101	Abaxial leaf ribs	Long trichomes (absence, No; presence, Y)	AbEpRLtr
V102	Abaxial leaf ribs	Long trichomes (length. µm)	AbEpRLtrL
V103	Abaxial leaf valleys	Long cells type (suberous, SZ; siliceous rounded, So; siliceous no-rounded, Z)	AbEpVLCt
V104	Abaxial leaf valleys	Long cell length (µm)	AbEpVLCI
V105	Abaxial leaf valleys	Short cells type (suberous, SZ; siliceous rounded, So; siliceous no-rounded, Z)	AbEpVSCt
V106	Abaxial leaf valley	Short cells length (µm) [suberous: SZ]	AbEpVSCSZI
V107	Abaxial leaf valley	Short cells length (µm [siliceous rounded: So]	AbEpVSCSol
V108	Abaxial leaf valley	Short cells length (µm [siliceous:Z]	AbEpVSCZI

## Supplementary data Chapter 4

V109	Abaxial leaf valley	Suberous cells (absence,NA; presence, SZ)	AbEpVSCSZ
V110	Abaxial leaf valley	Siliceous cells (absence, NA; presence, Z)	AbEpVSCZ
V111	Abaxial leaf valley	Hook trichomes (absence, No; presence, Y)	AbEpVHtr
V112	Abaxial leaf valley	Hook trichomes length (μm)	AbEpVHtrL
V113	Abaxial leaf valley	Long trichomes (absence, No; presence, Y)	AbEpVLtr
V114	Abaxial leaf valley	Long trichomes length (μm)	AbEpVLtrL

**Supplementary Table S3.** Loadings of phenotypic traits onto the first Factor Analysis of Mixed Data (FAMD) components of the intraspecific multivariate analyses of *Brachypodium retusum* (A) and *B. phoenicoides* (B). Highest contributions of characters to each component are highlighted in bold. Variable codes correspond to those indicated in Supplementary Table S2.

(A)

Components	<i>B. retusum</i>		
	comp 1	comp 2	comp 3
eigenvalue	11.63441	9.472808	4.3576
percentage of variance	25.85425	21.05068	9.683556
cumulative percentage of variance	25.85425	46.90493	56.58849

Variable contribution			
Variables	Dim.1	Dim.2	Dim.3
v59	8.2108	0.0713	0.4717
v63	8.1813	0.1512	0.3490
v57	8.1228	0.1316	0.5657
v62	8.1219	0.1973	0.4969
v65	8.1017	0.2110	0.6647
v58	8.0960	0.1784	0.5046
v64	8.0301	0.2104	0.6280
v66	7.9457	0.4941	0.1853
v61	7.8835	0.3984	0.4766
v60	7.8233	0.1955	0.8551
v5	2.2563	0.4490	9.8334
v7	1.8621	0.5540	0.7920
v35	1.6432	0.3743	0.3085
v38	1.5739	0.0658	8.8736
v15	1.4399	0.0567	2.8740
v37	1.2640	0.1329	11.8655
v39	1.2472	1.4134	0.0001
v23	1.1281	3.7997	0.2880
v42	1.0676	0.1232	0.5775
v79	1.0370	2.3244	0.3601
v40	0.9991	1.1147	1.9421
v2	0.8226	2.7506	0.9236
v33	0.6911	0.3526	2.0843
v26	0.4585	0.1240	2.2890
v3	0.3851	0.0003	11.5339
v6	0.3488	0.1076	7.5027

### Supplementary data Chapter 4

<b>v31</b>	0.3110	<b>3.0076</b>	0.5237
<b>v70</b>	0.2153	<b>7.8297</b>	1.3148
v16	0.1510	<b>6.3393</b>	0.5735
<b>v36</b>	0.1365	0.0030	<b>3.6512</b>
<b>v67</b>	0.0839	<b>9.1139</b>	0.0389
<b>v77</b>	0.0780	<b>9.5871</b>	0.0139
<b>v4</b>	0.0705	2.3067	<b>3.4158</b>
<b>v78</b>	0.0615	<b>9.0928</b>	0.0320
<b>v24</b>	0.0218	<b>6.2022</b>	0.7491
<b>v68</b>	0.0201	<b>8.9269</b>	0.0598
<b>v80</b>	0.0190	<b>8.6859</b>	0.0066
v32	0.0186	1.1801	2.4242
v34	0.0172	1.5229	0.0899
<b>v14</b>	0.0169	0.1281	<b>3.5688</b>
<b>v69</b>	0.0139	<b>7.8588</b>	0.2628
<b>v13</b>	0.0102	0.7457	<b>12.4435</b>
v41	0.0058	0.5144	1.4344
v25	0.0052	0.5054	0.0012
v27	0.0023	0.4669	2.1498

(B)

Components	<i>B. phoenicoides</i>		
	comp 1	comp 2	comp 3
<b>eigenvalue</b>	12.6934562	8.17692068	4.83919614
<b>percentage of variance</b>	28.8487641	18.5839106	10.9981731
<b>cumulative percentage of variance</b>	28.84876	47.43267	58.43085

# Supplementary data Chapter 4

Variable contribution			
Variables	Dim.1	Dim.2	Dim.3
<b>v66</b>	<b>7.4123</b>	0.3373	0.0106
<b>v63</b>	<b>7.3847</b>	0.2640	0.1087
<b>v64</b>	<b>7.3847</b>	0.2609	0.1239
<b>v62</b>	<b>7.3376</b>	0.3239	0.1029
<b>v65</b>	<b>7.1981</b>	0.4381	0.2017
<b>v59</b>	<b>7.1915</b>	0.6955	0.0001
<b>v58</b>	<b>7.1834</b>	0.6752	0.0004
<b>v60</b>	<b>7.1201</b>	0.7040	0.0019
<b>v57</b>	<b>6.9921</b>	0.7615	0.0017
<b>v61</b>	<b>6.8078</b>	0.3618	0.0346
v2	4.4504	0.0357	1.9195
v42	2.6726	0.4419	0.0061
<b>v13</b>	2.2546	2.1473	<b>5.3441</b>
v36	2.0939	0.0402	1.8922
v14	2.0826	0.5888	1.0299
v24	1.9667	2.0443	0.5351
v39	1.7421	0.0447	2.0829
v26	1.5083	0.9087	0.2211
<b>v35</b>	1.3712	0.2584	<b>10.0501</b>
<b>v41</b>	1.0225	0.0435	<b>5.7534</b>
v23	0.7801	3.6015	3.5858
v40	0.7468	0.3886	3.6849
<b>v38</b>	0.7437	0.0873	<b>5.1301</b>
v3	0.6092	2.9214	1.7179
v15	0.5347	0.0097	0.5261
<b>v6</b>	0.4970	<b>6.2636</b>	0.0236
<b>v4</b>	0.4394	<b>5.7728</b>	0.9773
<b>v67</b>	0.3615	<b>5.6223</b>	0.6791
v25	0.3399	3.6062	2.6681
<b>v68</b>	0.3367	<b>5.8648</b>	0.8441
v32	0.2423	2.2615	0.4890
<b>v34</b>	0.2400	0.0326	<b>6.1197</b>
v27	0.2264	2.3307	1.1872
v70	0.2119	3.2529	0.2227
<b>v69</b>	0.1886	<b>5.9290</b>	<b>4.0520</b>
<b>v78</b>	0.1544	<b>9.9757</b>	0.4688
v16	0.0644	4.1971	3.6335

#### Supplementary data Chapter 4

v33	0.0367	0.0101	1.4676
<b>v37</b>	0.0358	1.1097	<b>5.2329</b>
v7	0.0128	3.9024	2.3177
<b>v80</b>	0.0073	<b>9.7762</b>	0.7506
<b>v77</b>	0.0070	<b>9.9944</b>	0.5431
<b>v31</b>	0.0033	0.2683	<b>11.6586</b>
<b>v5</b>	0.0028	1.4454	<b>12.5988</b>

**Supplementary Table S4.** Hierarchical clustering analysis results of multivariate Factor Analysis of Mixed Data (FAMD) of the studied *Brachypodium retusum* (A) and *B. phoenicodes* (B) samples. Descriptions of each cluster by categories and quantitative variables, respectively. Highest contributions of categories and characters to each cluster are highlighted in bold. Categories correspond to those indicated in Supplementary Table S1 and variables to those indicated in **Supplementary Table S2**.

**(A) *B. retusum***

Clusters	Description of each cluster by the categories						Description of each cluster by quantitative variables						
	Vars	Cla/Mod	Mod/Cla	Global	p.value	v.test	Vars	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Cluster 1	Geo=Sfrn	75.00	100.00	10.26	0.00	3.52	<b>v59</b>	5.24	0.73	-0.30	0.41	0.35	0.00
	Loc=Herault	75.00	100.00	10.26	0.00	3.52	<b>v58</b>	5.23	0.68	-0.32	0.39	0.34	0.00
	Country=FRN	75.00	100.00	10.26	0.00	3.52	<b>v60</b>	5.19	0.90	-0.28	0.58	0.40	0.00
	Ploidy=6x	30.00	100.00	25.64	0.01	2.48	<b>v57</b>	5.18	0.73	-0.31	0.48	0.36	0.00
	v18=v18_R	50.00	66.67	10.26	0.02	2.26	v63	5.13	1.64	-0.07	0.80	0.59	0.00
	v30=v30_SM	21.43	100.00	35.90	0.04	2.06	v62	5.09	1.72	-0.05	0.87	0.62	0.00
	Ploidy=4x	0.00	0.00	61.54	0.05	-1.96	v64	5.04	1.80	-0.02	0.92	0.64	0.00
	Country=SPN	0.00	0.00	89.74	0.00	-3.52	v65	4.83	2.20	0.06	1.36	0.79	0.00
							v66	4.76	2.19	0.14	1.28	0.77	0.00
							v15	2.66	1.87	0.09	0.12	1.19	0.01
Cluster 2	<b>v45=v45_SC</b>	76.47	100.00	43.59	0.00	5.13	v42	2.97	<b>0.53</b>	-0.04	1.02	0.83	0.00
	v43=v43_A	100.00	69.23	23.08	0.00	4.65	v40	-2.09	-0.77	-0.38	0.39	0.83	0.04
	v54=v54_SC	65.00	100.00	51.28	0.00	4.43	v16	-2.24	-0.83	-0.44	0.34	0.75	0.03
	Geo=Nspn	65.00	100.00	51.28	0.00	4.43	v78	-2.26	-0.80	-0.44	0.33	0.68	0.02
	<b>v30=v30_SM</b>	78.57	84.62	35.90	0.00	4.34	v7	-2.31	-0.66	-0.25	0.50	0.78	0.02
	v44=v44_SC	59.09	100.00	56.41	0.00	4.01	v77	-2.33	-0.82	-0.43	0.34	0.71	0.02

# Supplementary data Chapter 4

	<b>v46=v46_SC</b>	76.92	76.92	33.33	0.00	3.90	v35	-2.43	-0.78	-0.33	0.55	0.80	0.01
	Ploidy=x	100.00	38.46	12.82	0.00	3.06	v70	-2.51	-0.95	-0.53	0.44	0.74	0.01
	Loc=Zaragoza	100.00	38.46	12.82	0.00	3.06	v69	-2.51	-0.95	-0.64	0.20	0.55	0.01
	v18=v18_T	46.43	100.00	71.79	0.00	2.83	v38	-2.65	-0.56	0.05	0.62	0.99	0.01
	v56=v56_SM	55.56	76.92	46.15	0.01	2.62	v67	-2.80	-1.03	-0.60	0.37	0.67	0.01
	v20=v20_C	43.33	100.00	76.92	0.01	2.44	v39	-2.97	-0.82	-0.45	0.69	0.55	0.00
	v19=v19_H	43.33	100.00	76.92	0.01	2.44	v68	-3.08	-1.11	-0.63	0.39	0.68	0.00
	v29=v29_S	52.63	76.92	48.72	0.02	2.39	v17	-3.21	-0.15	-0.14	0.01	0.02	0.00
	v10=v10_C	41.94	100.00	79.49	0.03	2.24	v79	-3.24	-0.99	-0.61	0.48	0.50	0.00
	v8=v8_T	41.94	100.00	79.49	0.03	2.24							
	v56=v56_H	100.00	23.08	7.69	0.03	2.15							
	v21=v21_SC	71.43	38.46	17.95	0.03	2.13							
	v18=v18_B	0.00	0.00	17.95	0.04	-2.03							
	v21=v21_H	25.00	61.54	82.05	0.03	-2.13							
	v10=v10_NC	0.00	0.00	20.51	0.03	-2.24							
	v29=v29_T	15.00	23.08	51.28	0.02	-2.39							
	v20=v20_NC	0.00	0.00	23.08	0.01	-2.44							
	Ploidy=6x	0.00	0.00	25.64	0.01	-2.64							
	v30=v30_SC	5.88	7.69	43.59	0.00	-3.18							
	Geo=Sspn	0.00	0.00	38.46	0.00	-3.61							
	v54=v54_SM	0.00	0.00	43.59	0.00	-4.01							
	v44=v44_G	0.00	0.00	43.59	0.00	-4.01							
	v56=v56_SC	0.00	0.00	46.15	0.00	-4.21							
	v43=v43_M	13.33	30.77	76.92	0.00	-4.65							
	v46=v46_G	4.17	7.69	61.54	0.00	-4.83							
	v45=v45_G	0.00	0.00	56.41	0.00	-5.13							
<b>Cluster 3</b>	<b>v45=v45_G</b>	<b>100.00</b>	95.65	56.41	0.00	6.24	<b>v2</b>	<b>3.04</b>	<b>0.88</b>	<b>0.57</b>	0.64	0.75	0.00



## Supplementary data Chapter 4

<b>v46=v46_G</b>	<b>91.67</b>	95.65	61.54	0.00	5.31	v38	2.08	<b>0.32</b>	<b>0.05</b>	1.00	0.99	0.04
<b>v44=v44_G</b>	<b>100.00</b>	73.91	43.59	0.00	4.76	v3	-2.04	<b>0.25</b>	<b>0.48</b>	0.92	0.83	0.04
<b>Geo=Sspn</b>	<b>100.00</b>	65.22	38.46	0.00	4.27	v17	2.87	-0.13	-0.14	0.02	0.02	0.00
v43=v43_M	76.67	100.00	76.92	0.00	4.04	v16	2.35	-0.20	-0.44	0.86	0.75	0.02
<b>v30=v30_SC</b>	<b>94.12</b>	69.57	43.59	0.00	3.94	v66	-3.40	-0.21	0.14	0.22	0.77	0.00
v56=v56_SC	88.89	69.57	46.15	0.00	3.47	v77	2.27	-0.22	-0.43	0.81	0.71	0.02
<b>v20=v20_NC</b>	<b>100.00</b>	39.13	23.08	0.00	2.89	v78	2.10	-0.25	-0.44	0.78	0.68	0.04
<b>v10=v10_NC</b>	<b>100.00</b>	34.78	20.51	0.01	2.65	v70	2.52	-0.27	-0.53	0.81	0.74	0.01
v29=v29_T	80.00	69.57	51.28	0.01	2.64	v65	-3.22	-0.28	0.06	0.16	0.79	0.00
v54=v54_SM	82.35	60.87	43.59	0.01	2.53	v39	2.08	-0.29	-0.45	0.32	0.55	0.04
<b>v18=v18_B</b>	<b>100.00</b>	30.43	17.95	0.02	2.41	v64	-3.25	-0.30	-0.02	0.15	0.64	0.00
<b>v19=v19_G</b>	<b>100.00</b>	26.09	15.38	0.03	2.16	v62	-3.14	-0.31	-0.05	0.13	0.62	0.00
v56=v56_SM	38.89	30.43	46.15	0.02	-2.26	v63	-3.07	-0.32	-0.07	0.14	0.59	0.00
v29=v29_S	36.84	30.43	48.72	0.01	-2.64	v68	3.30	-0.32	-0.63	0.68	0.68	0.00
v10=v10_C	48.39	65.22	79.49	0.01	-2.65	v24	2.21	-0.33	-0.55	0.82	0.75	0.03
v8=v8_T	48.39	65.22	79.49	0.01	-2.65	v67	2.91	-0.34	-0.60	0.70	0.67	0.00
Ploidy=x	0.00	0.00	12.82	0.01	-2.67	v42	-2.73	-0.34	-0.04	0.50	0.83	0.01
Loc=Zaragoza	0.00	0.00	12.82	0.01	-2.67	v60	-2.15	-0.39	-0.28	0.09	0.40	0.03
v20=v20_C	46.67	60.87	76.92	0.00	-2.89	v79	2.59	-0.43	-0.61	0.41	0.50	0.01
v19=v19_H	46.67	60.87	76.92	0.00	-2.89	v69	2.69	-0.44	-0.64	0.61	0.55	0.01
v54=v54_SC	35.00	30.43	51.28	0.00	-3.04							
Geo=Nspn	35.00	30.43	51.28	0.00	-3.04							
v43=v43_A	0.00	0.00	23.08	0.00	-4.04							
v46=v46_SC	7.69	4.35	33.33	0.00	-4.55							
v44=v44_SC	27.27	26.09	56.41	0.00	-4.76							
v30=v30_SM	0.00	0.00	35.90	0.00	-5.77							
v45=v45_SC	5.88	4.35	43.59	0.00	-6.24							

(B) *B. phoenicoides*

Clusters	Description of each cluster by the categories						Description of each cluster by quantitative variables						
	Vars	Cla/Mod	Mod/Cla	Global	p.value	v.test	Vars	v.test	Mean in category	Overall mean	sd in category	Overall sd	p.value
Cluster 1	<b>v55=v55_SM</b>	100.00	100.00	19.23	0.00	4.33	v17	2.02	1.47	0.21	3.17	1.52	0.04
	<b>Geo=Nspn</b>	100.00	100.00	19.23	0.00	4.33	v5	2.28	0.30	-0.25	0.21	0.59	0.02
	Loc=Huesca	100.00	100.00	19.23	0.00	4.33	v79	-2.57	0.29	0.94	0.63	0.62	0.01
	<b>v85=v85_NO</b>	100.00	40.00	7.69	0.03	2.16	v25	-2.53	-0.20	0.31	0.33	0.49	0.01
	<b>v20=v20_NC</b>	71.43	100.00	26.92	0.00	3.60	v40	-2.30	-0.26	0.64	0.87	0.95	0.02
	Ploidy=4x	62.50	100.00	30.77	0.00	3.34	v26	-2.16	-0.51	0.44	0.60	1.07	0.03
	v56=v56_SC	50.00	100.00	38.46	0.00	2.89	v2	2.27	-0.56	-1.02	0.56	0.49	0.02
	v18=v18_T	50.00	80.00	30.77	0.02	2.31	v31	-2.72	-0.83	0.36	0.72	1.07	0.01
	v85=v85_E	12.50	60.00	92.31	0.03	-2.16	v35	-3.38	-0.95	0.49	0.44	1.03	0.00
	v18=v18_B	5.88	20.00	65.38	0.04	-2.09							
	Geo=Sspn	0.00	0.00	42.31	0.05	-2.00							
	Ploidy=6x	0.00	0.00	61.54	0.00	-2.89							
	v55=v55_SC	0.00	0.00	73.08	0.00	-3.60							
	v20=v20_C	0.00	0.00	73.08	0.00	-3.60							
Cluster2	<b>v54=v54_SC</b>	100.00	58.82	38.46	0.00	2.91	v23	2.52	1.20	0.91	0.76	0.80	0.01
	<b>v44=v44_SC</b>	100.00	47.06	30.77	0.02	2.42	v16	2.27	1.04	0.76	0.84	0.83	0.02
	<b>v45=v45_SC</b>	100.00	41.18	26.92	0.03	2.18	v78	3.57	0.86	0.48	0.49	0.73	0.00
	<b>Geo=Sfrn</b>	100.00	41.18	26.92	0.03	2.18	v77	3.57	0.79	0.40	0.47	0.75	0.00
	<b>Country=FRN</b>	100.00	41.18	26.92	0.03	2.18	v80	3.70	0.59	0.12	0.50	0.88	0.00
	<b>v9=v9_G</b>	91.67	64.71	46.15	0.01	2.49	v25	2.11	0.46	0.31	0.46	0.49	0.03
	<b>Geo=Sspn</b>	90.91	58.82	42.31	0.03	2.23	v33	2.18	0.30	-0.01	0.84	0.99	0.03
	<b>v75=v75_E</b>	88.89	94.12	69.23	0.00	3.54	<b>v13</b>	<b>3.31</b>	<b>-0.06</b>	<b>-0.48</b>	0.59	0.87	0.00

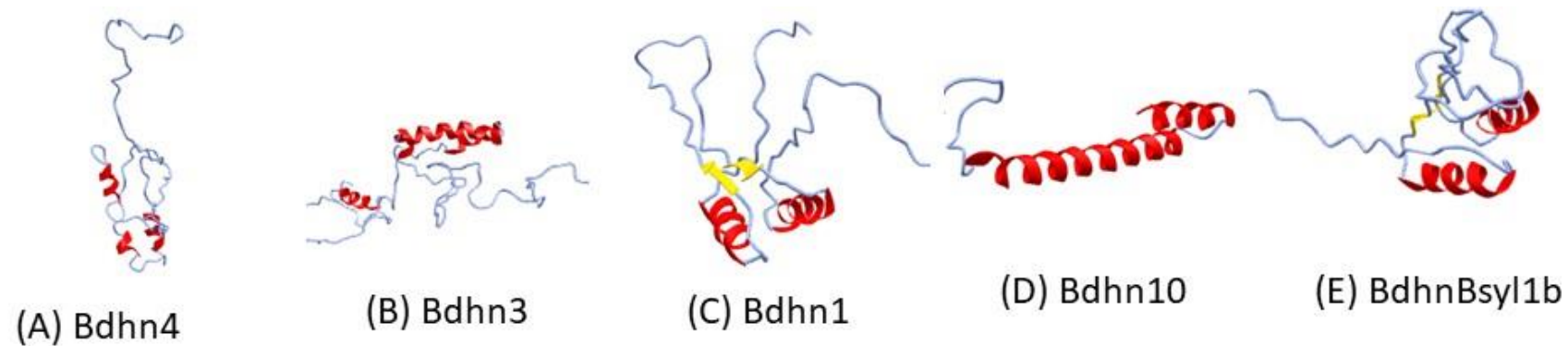
# Supplementary data Chapter 4

	Ploidy=6x	87.50	82.35	61.54	0.01	2.80	v15	-2.60	-0.27	-0.12	0.27	0.40	0.01
	v20=v20_C	84.21	94.12	73.08	0.00	3.05							
	v55=v55_SC	78.95	88.24	73.08	0.03	2.16							
	v54=v54_SM	43.75	41.18	61.54	0.00	-2.91							
	v9=v9_H	38.46	29.41	50.00	0.01	-2.76							
	v20=v20_NC	14.29	5.88	26.92	0.00	-3.05							
	Geo=Nmor	0.00	0.00	11.54	0.03	-2.14							
	Country=MOR	0.00	0.00	11.54	0.03	-2.14							
	v55=v55_SM	0.00	0.00	19.23	0.00	-3.10							
	Geo=Nspn	0.00	0.00	19.23	0.00	-3.10							
	Loc=Huesca	0.00	0.00	19.23	0.00	-3.10							
	v75=v75_NO	0.00	0.00	26.92	0.00	-4.03							
Cluster 3	<b>Geo=Nmor</b>	100.00	75.00	11.54	0.00	3.17	v57	3.80	<b>3.20</b>	<b>0.81</b>	0.28	1.34	0.00
	<b>Country=MOR</b>	100.00	75.00	11.54	0.00	3.17	v58	3.67	<b>3.13</b>	<b>0.83</b>	0.33	1.34	0.00
	<b>Ploidy=x</b>	100.00	50.00	7.69	0.02	2.36	v59	3.46	<b>3.00</b>	<b>0.82</b>	0.41	1.34	0.00
	<b>Loc=Mulay</b>	100.00	50.00	7.69	0.02	2.36	v63	2.67	<b>2.98</b>	<b>1.22</b>	0.73	1.41	0.01
	<b>v75=v75_NO</b>	57.14	100.00	26.92	0.00	3.04	v60	3.61	<b>2.97</b>	<b>0.76</b>	0.36	1.30	0.00
	v12=v12_A	44.44	100.00	34.62	0.01	2.63	v62	2.67	<b>2.97</b>	<b>1.21</b>	0.75	1.41	0.01
	v56=v56_SM	40.00	100.00	38.46	0.01	2.46	v64	2.66	<b>2.96</b>	<b>1.19</b>	0.81	1.42	0.01
	v30=v30_SC	33.33	100.00	46.15	0.03	2.13	v66	2.91	<b>2.86</b>	<b>0.97</b>	0.90	1.39	0.00
	v9=v9_H	30.77	100.00	50.00	0.05	1.98	v65	2.68	<b>2.64</b>	<b>1.03</b>	0.79	1.28	0.01
	v30=v30_H	0.00	0.00	53.85	0.03	-2.13	v26	2.69	<b>1.79</b>	<b>0.44</b>	1.02	1.07	0.01
	v12=v12_P	0.00	0.00	65.38	0.01	-2.63	v38	2.86	<b>1.25</b>	<b>-0.06</b>	0.90	0.98	0.00
	v75=v75_E	0.00	0.00	69.23	0.00	-2.83	v24	-2.51	0.03	0.83	0.90	0.68	0.01
							v78	-2.85	-0.50	0.48	0.71	0.73	0.00
							v77	-3.12	-0.70	0.40	0.75	0.75	0.00
							v33	-1.96	-0.93	-0.01	0.41	0.99	0.05

### Supplementary data Chapter 4

			v37	-2.39	-1.17	0.05	0.38	1.09	0.02
			v80	-3.40	-1.28	0.12	0.75	0.88	0.00
			v2	-2.38	-1.57	-1.02	0.41	0.49	0.02
			v13	-3.55	-1.92	-0.48	0.59	0.87	0.00

## Supplementary figures and tables Chapter 5



Complete sets of 3D structures for all *Brachypodium* species and proteins are available in the following links:

**Brachypodium distachyon:** <http://zeta.uma.es/public/journal/brachy/raptorx/distachyon.html>

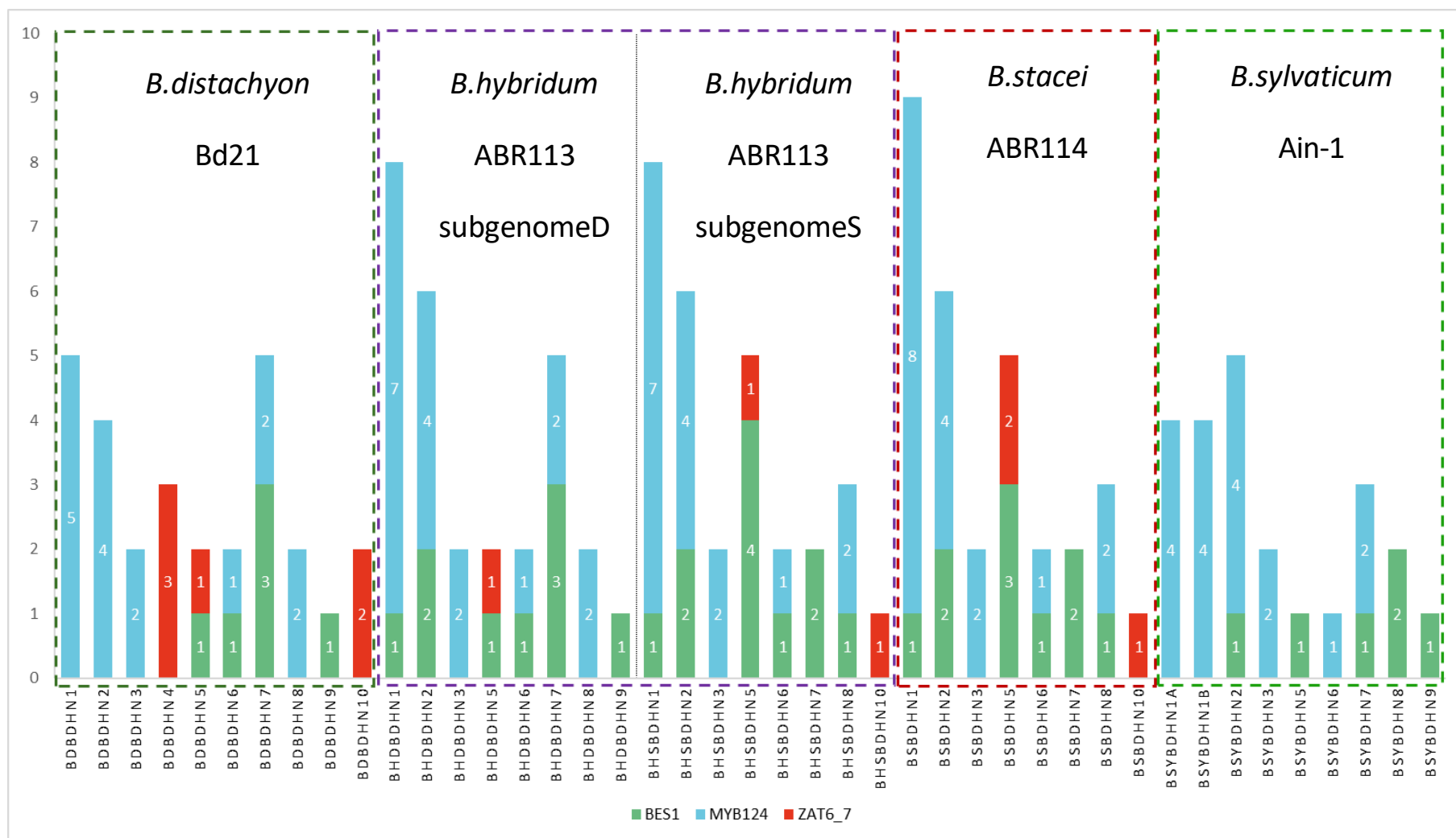
**Brachypodium hybridum D:** <http://zeta.uma.es/public/journal/brachy/raptorx/hybridumD.html>

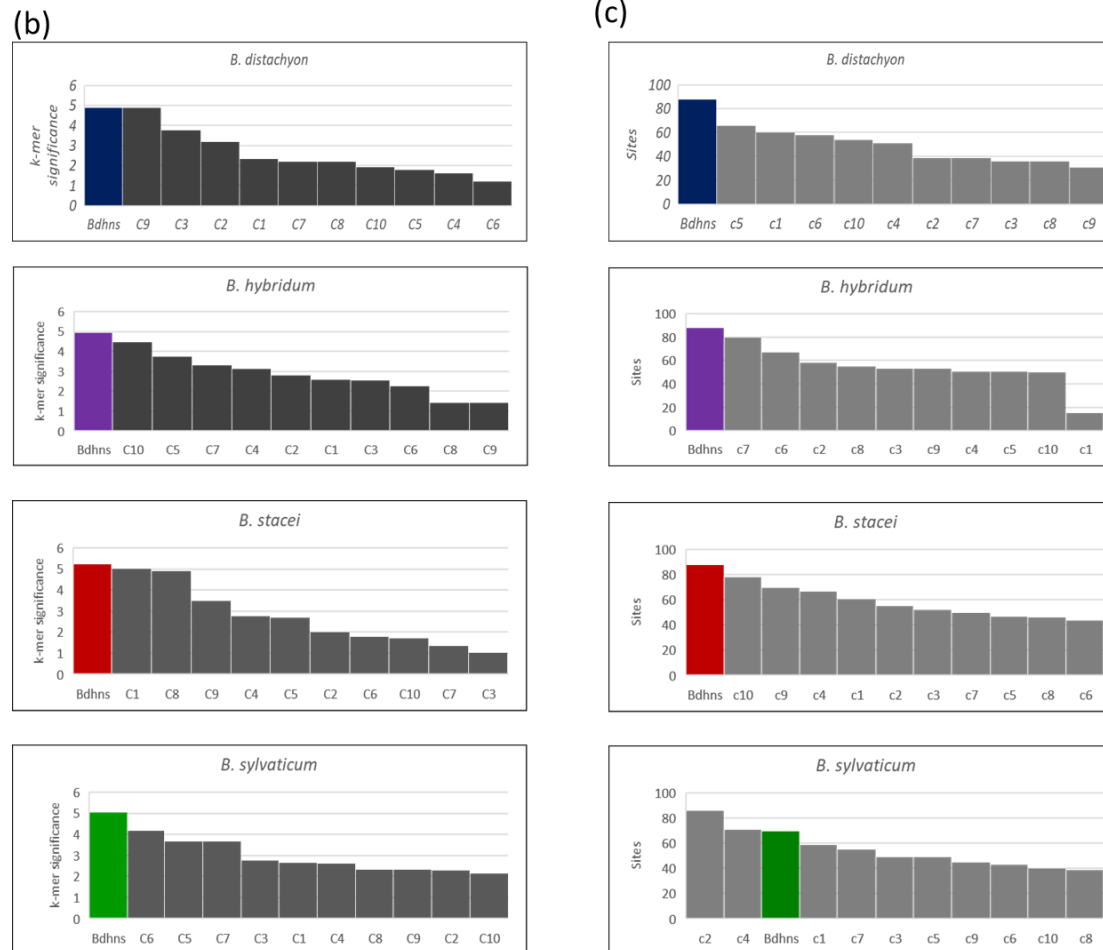
**Brachypodium stacei:** <http://zeta.uma.es/public/journal/brachy/raptorx/stacei.html>

**Brachypodium hybridum S:** <http://zeta.uma.es/public/journal/brachy/raptorx/hybridumS.html>

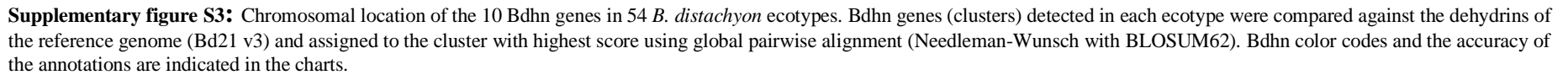
**Brachypodium sylvaticum:** <http://zeta.uma.es/public/journal/brachy/raptorx/sylvaticum.html>

**Supplementary figure S1:** Figure S1: Inferred three dimensional structure of some *Brachypodium* dehydrins. All species' dehydrins showed  $\alpha$ -helices (1 to 4) whereas only some of them presented  $\beta$ -sheets (Bdhn1 two in all species except *B. distachyon*, Bdhn2 two to three in *B. sylvaticum*, and Bdhn6 and Bdhn9 three in *B. hybridumS* and *B. sylvaticum*): (A) *B. distachyon* Bdhn4; (B) *B. hybridumD* Bdhn3; (C) *B. stacei* Bdhn1; (D) *B. hybridumS* Bdhn10; (E): *B. sylvaticum* Bdhn1b,



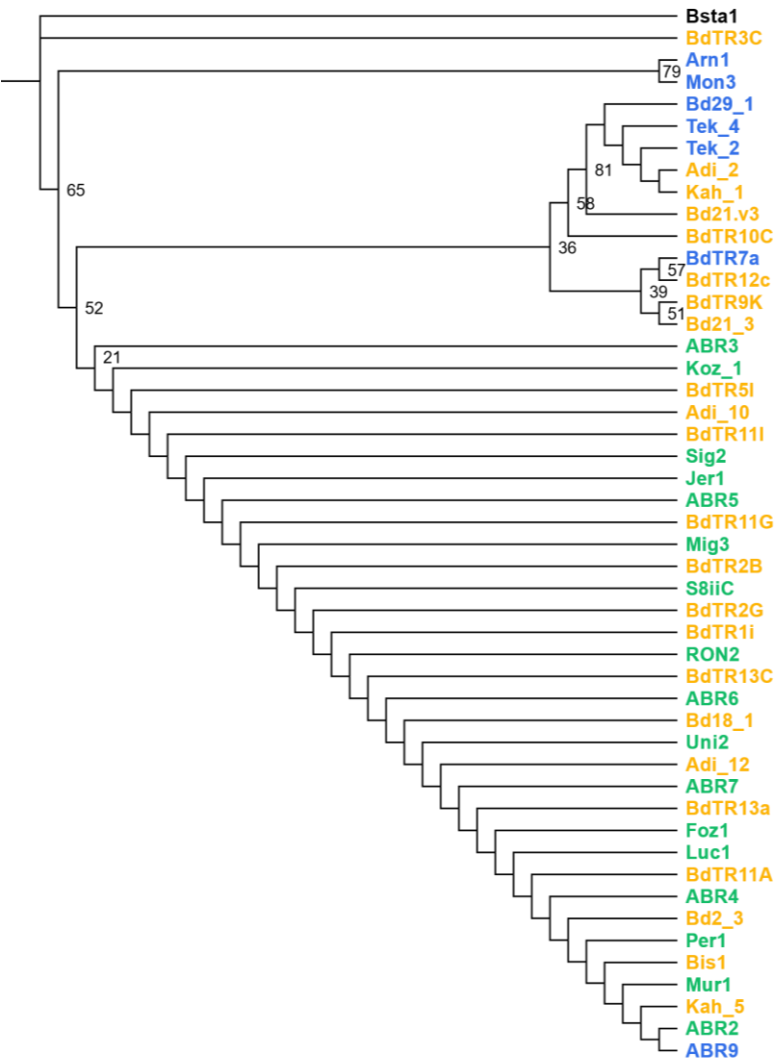


**Supplementary figure S2:** BES1/BZR1 (Basic leucine zipper; green), MYB124 (Myb gene protein; blue), and ZAT (C2H2 zinc finger; red) cis-regulatory elements found in 5'-upstream promoter region (−500-to-+200 bp) of the *Brachypodium* Bdhn genes. (a) Distributions of cis-motifs per species and reference genomes (BD, *B. distachyon* Bd21; BS, *B. stacei* ABR114; BHD, *B. hybridum* subgenome-D ABR113; BHS, *B. hybridum* subgenome-S ABR113; BSY, *B. sylvaticum* Ain1) and per Bdhn gene promoter (Bdhn1 to Bdhn10). Values indicate the number of each type of cis-motif found in the promoter region of each gene (see color codes in the chart). (b, c) Analysis of cis-regulatory element discovery with Rsat::plants tools in the 5-upstream promoter region (−500-to-+200 bp) of 47 Bdhn genes from four *Brachypodium* species. (b) Maximum k-mer significance values from each analysis are shown for Bdhn sequences from each species, (c) Maximum number of sites from each background analysis are shown for Bdhn sequences from each species. In each case, another 47 random gene sequences from the respective reference genome were analyzed ten times as controls (grey bars; C1-C10)

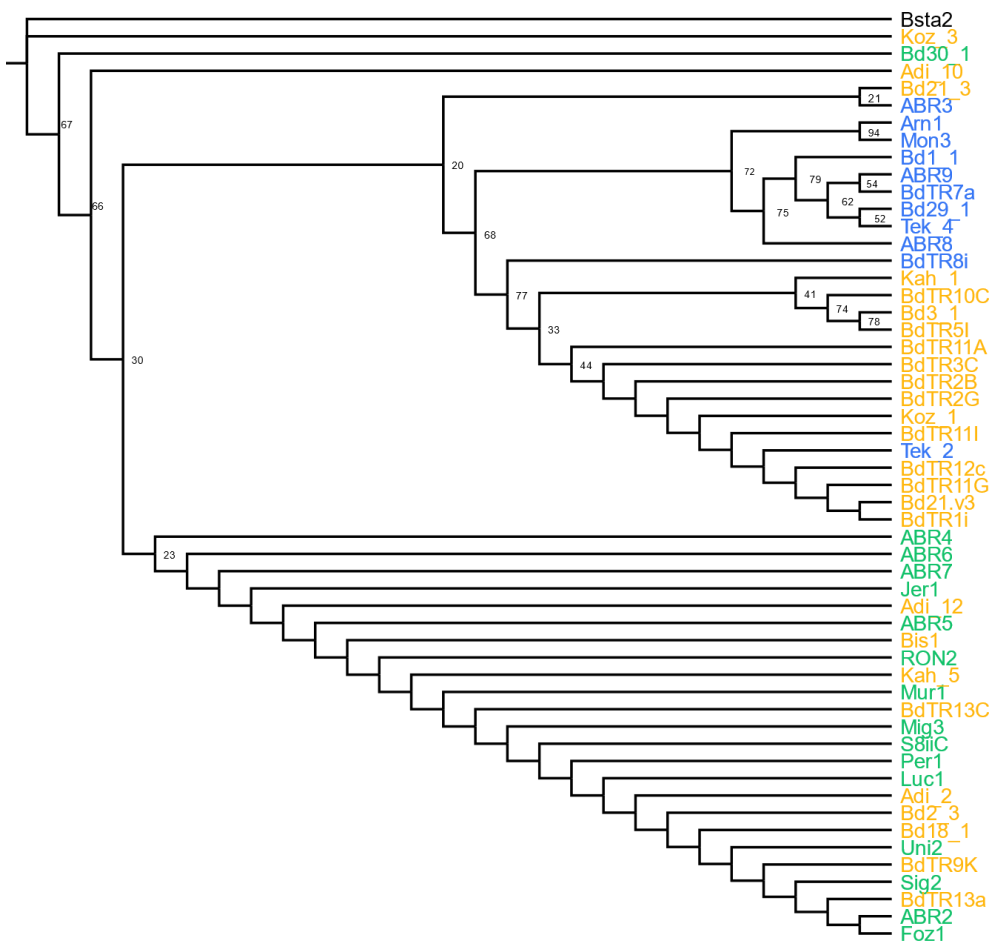




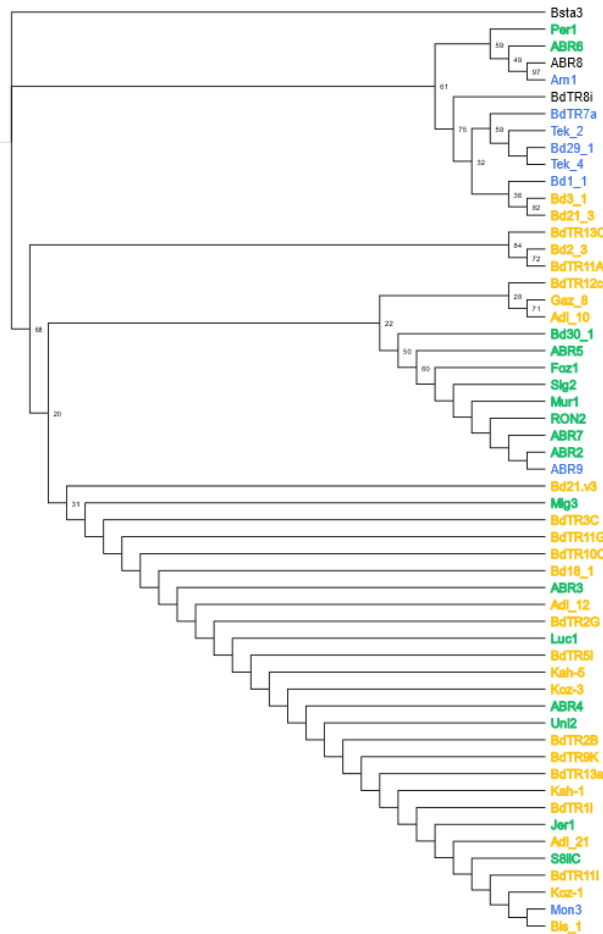
Bdhn1



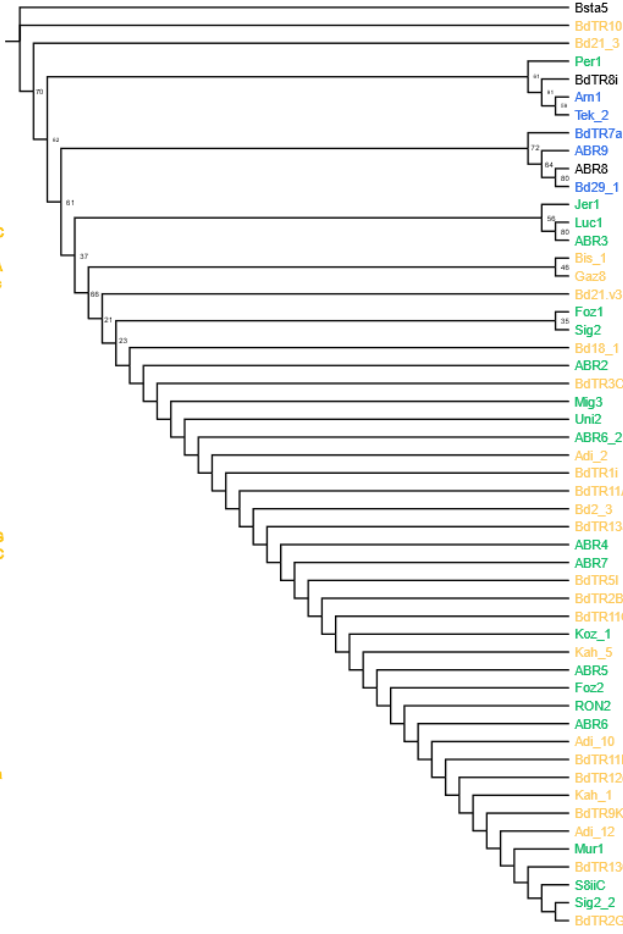
Bdhn2



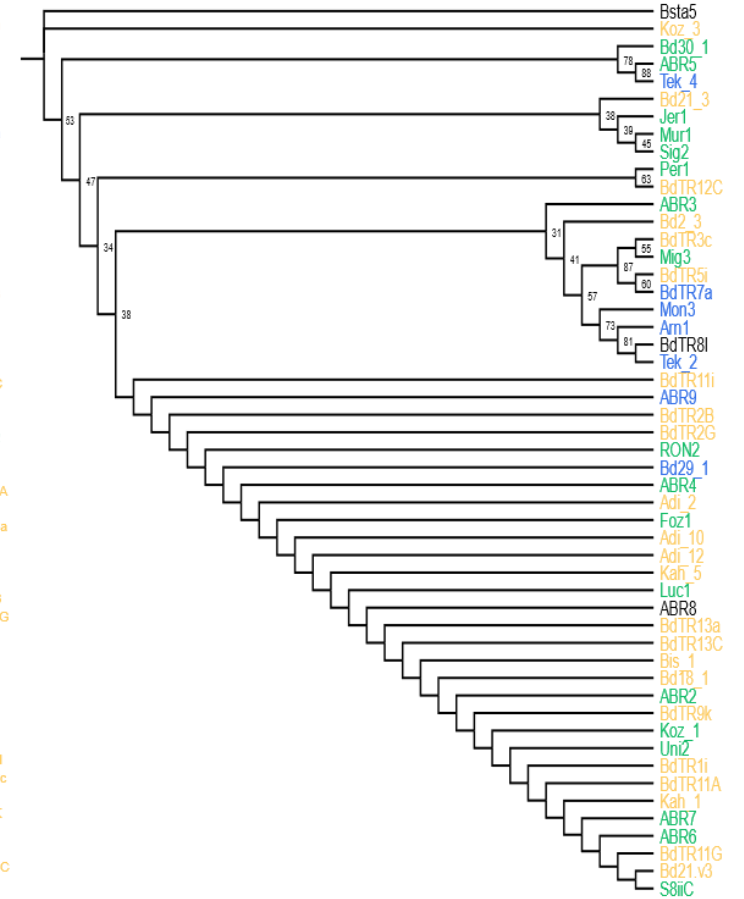
Bdhn3



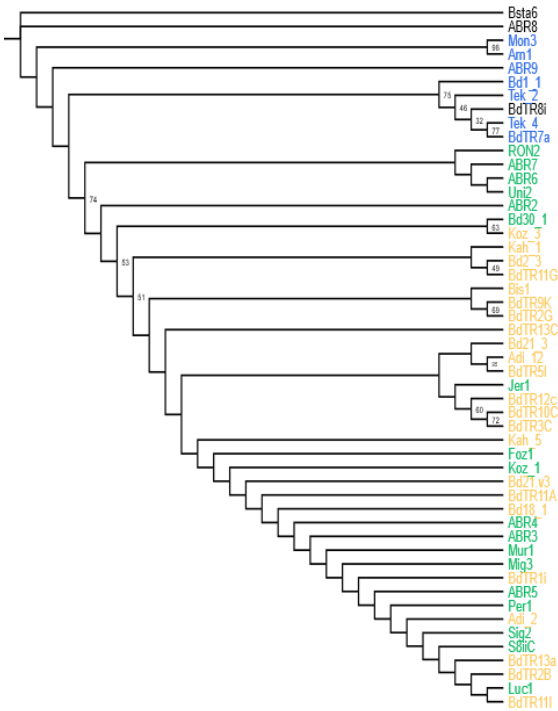
Bdhn4



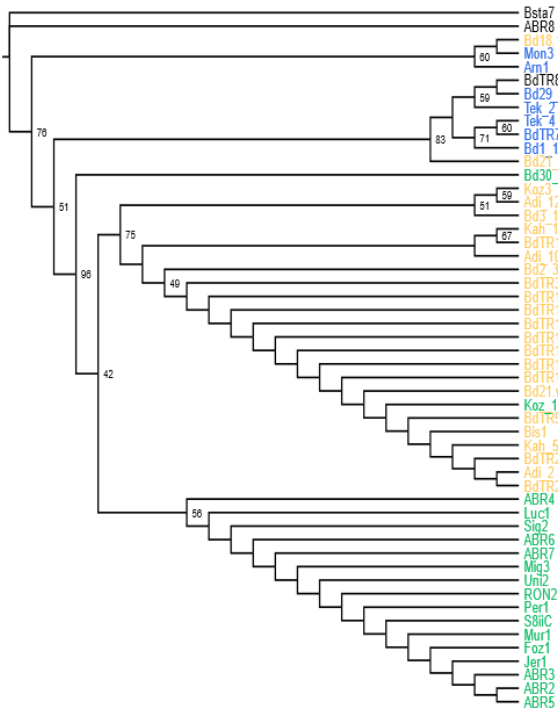
Bdhn5



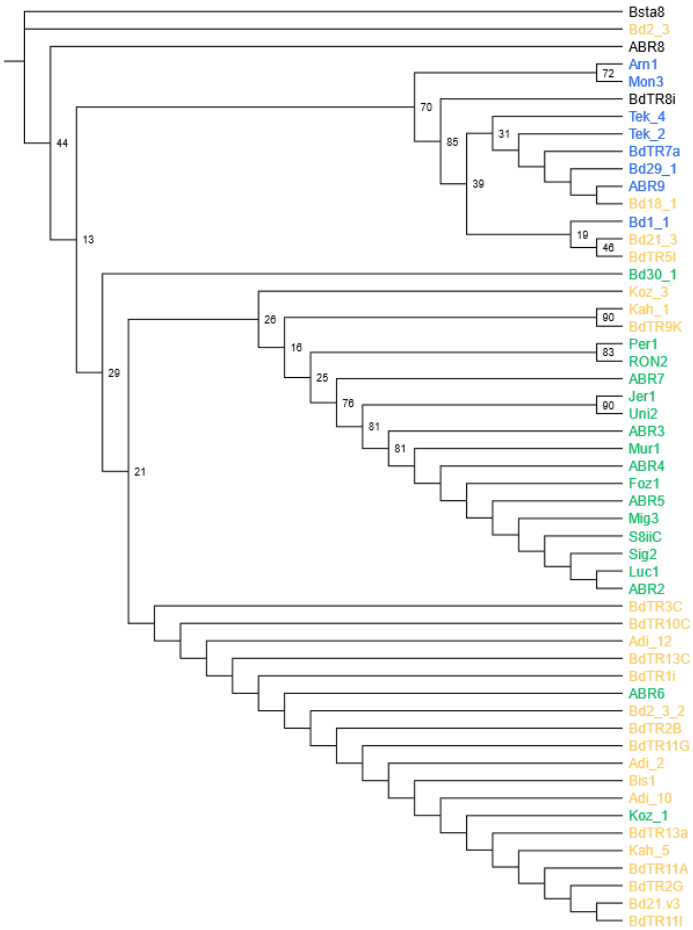
Bdhn6

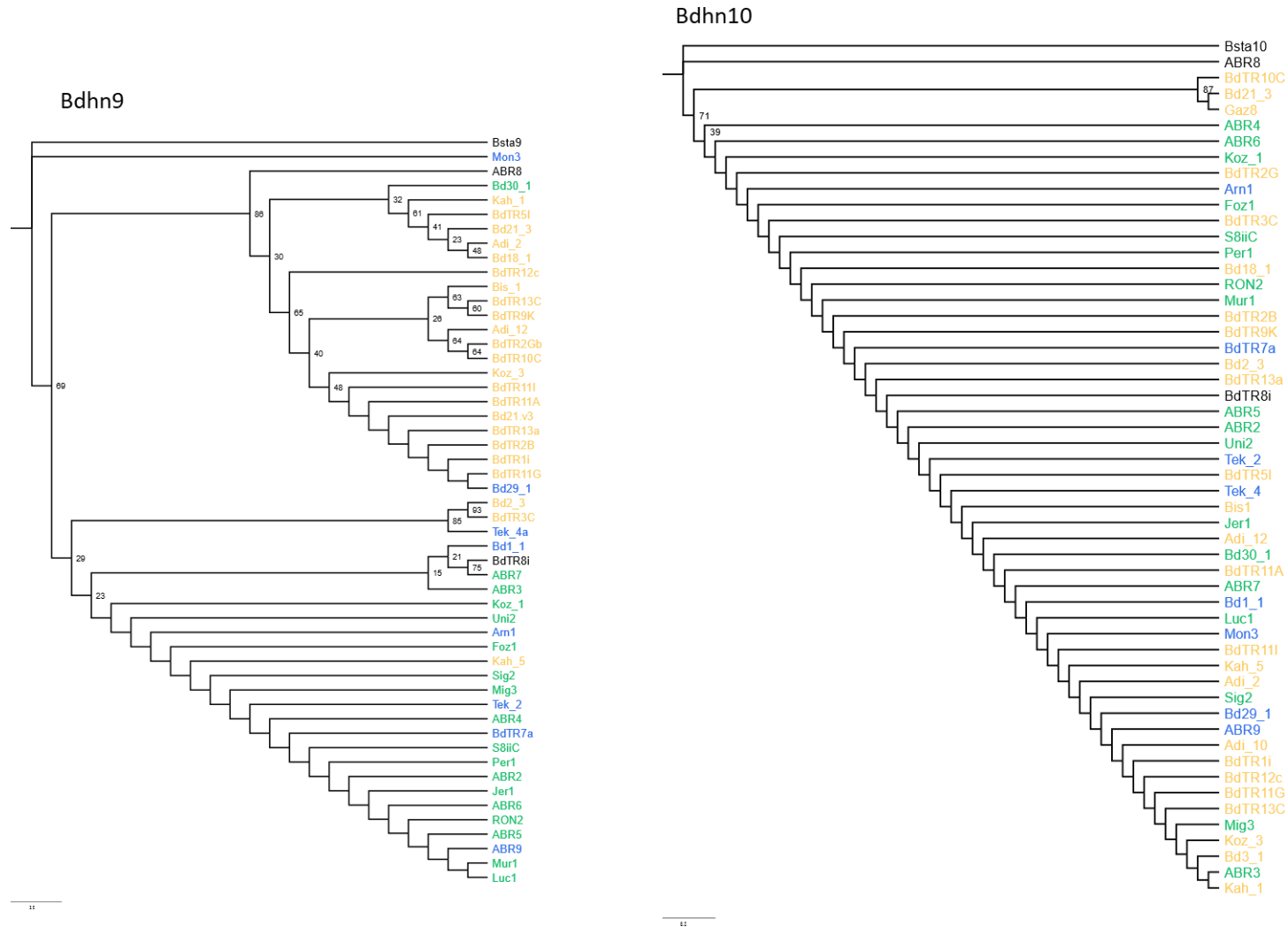


Bdhn7

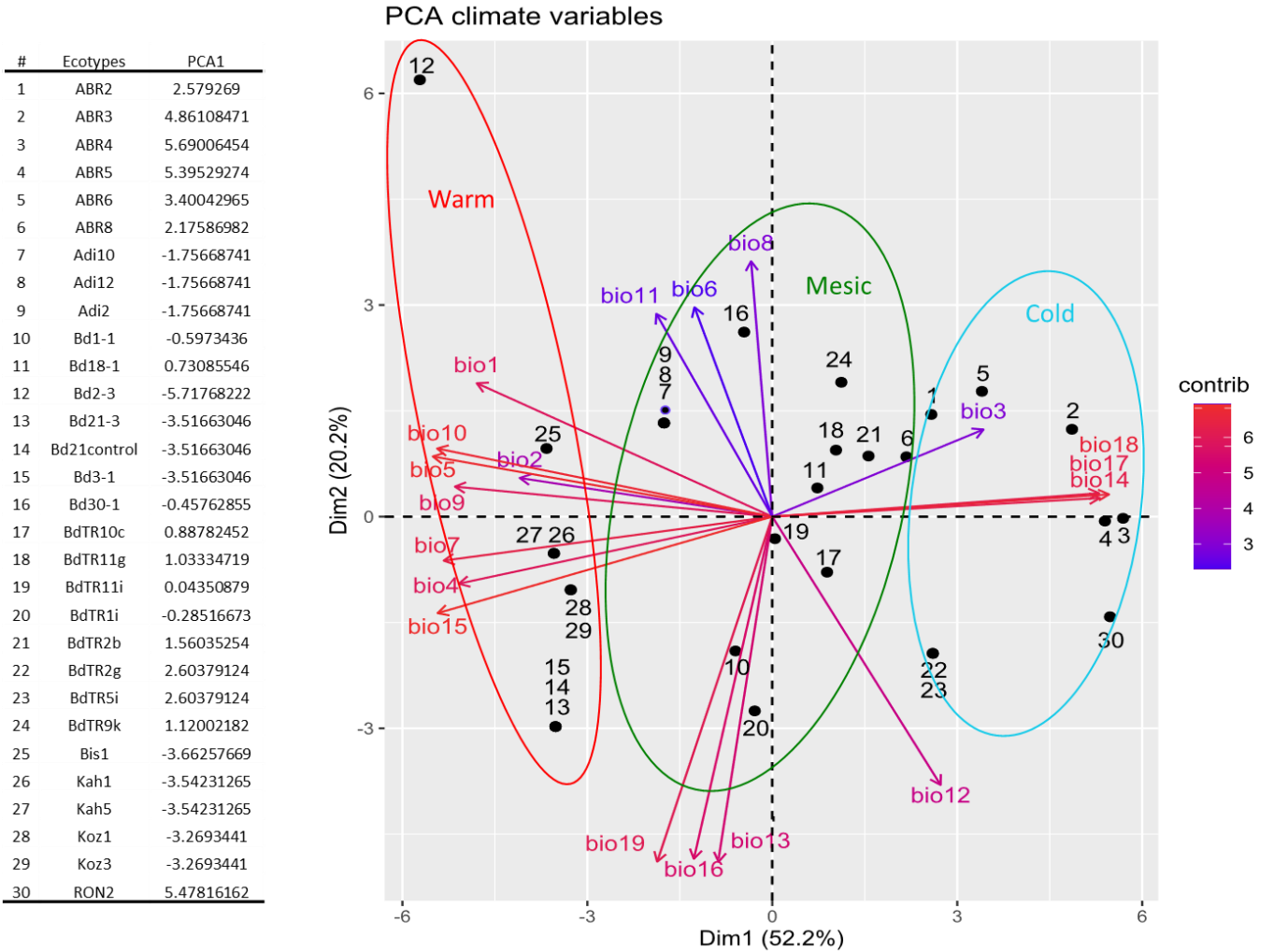


Bdhn8

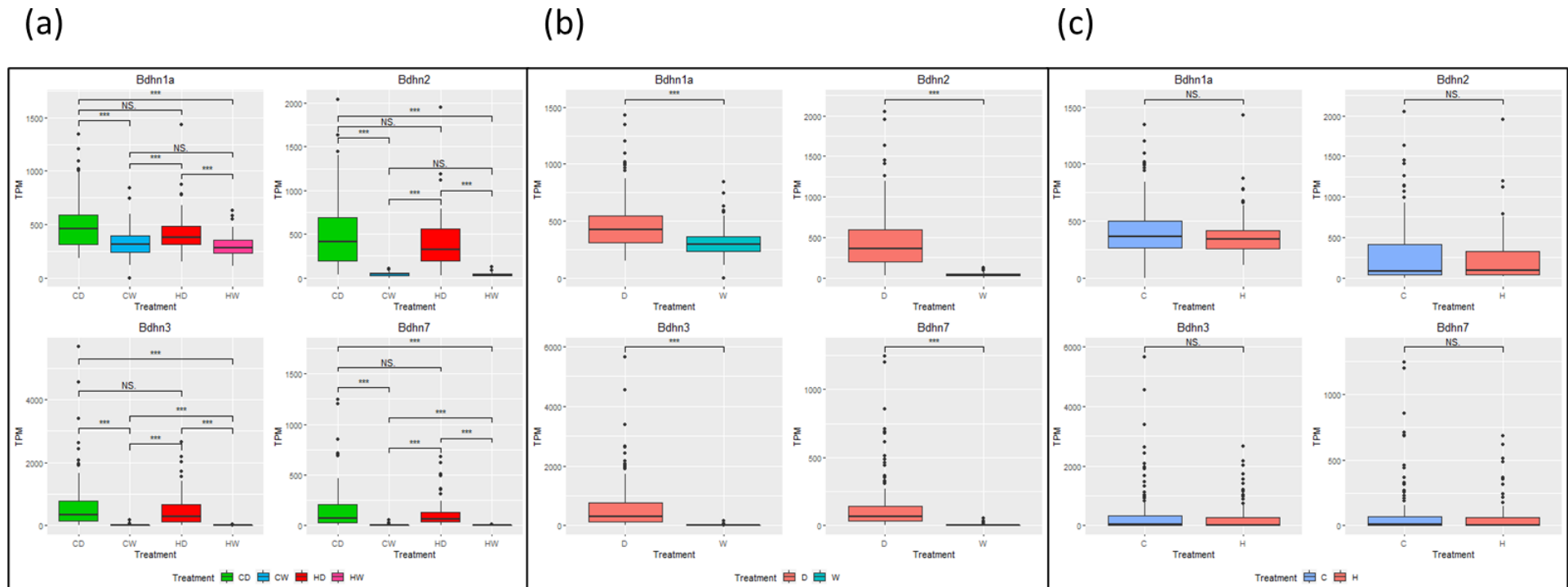




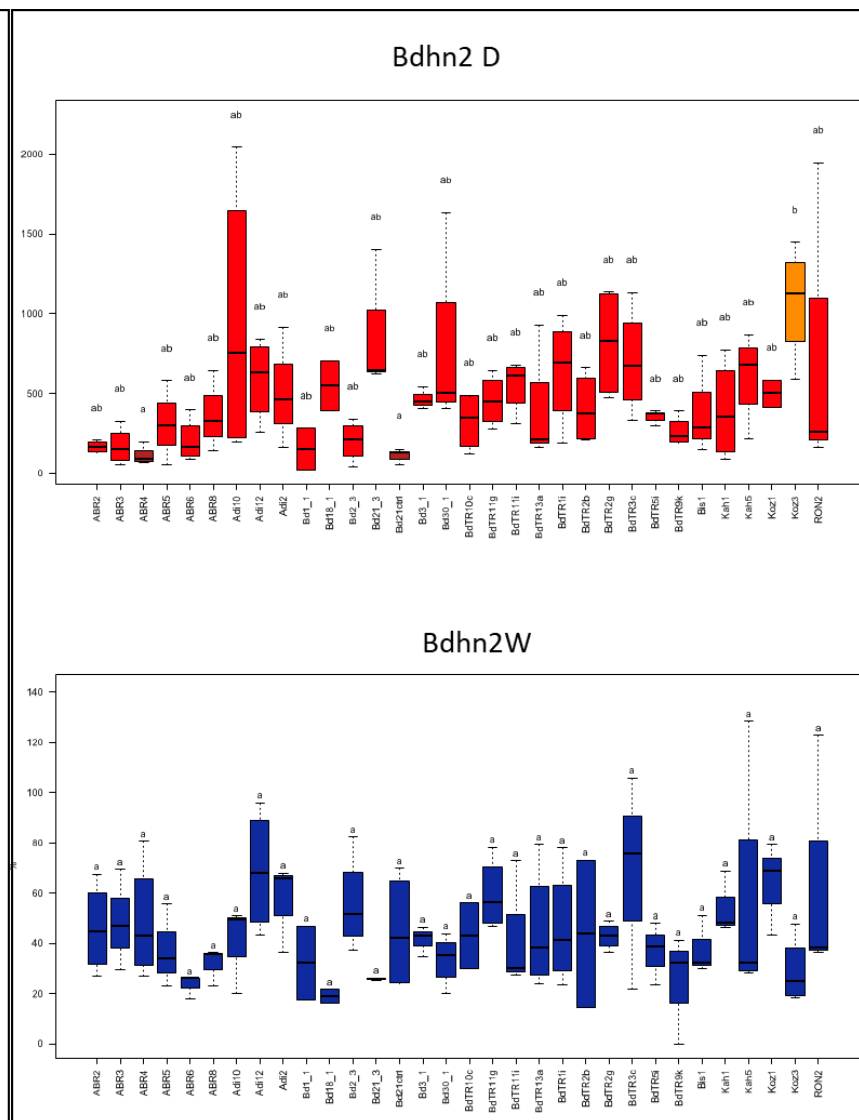
**Supplementary figure S4:** Maximum likelihood *B. distachyon* dehydrin trees obtained from the aligned exon and intron sequences of each independent *Bdhn* gene (*Bdhn1* to *Bdhn10*). Trees were constructed with IQTREE using the *B. stacei* outgroup sequence to root the tree. Bootstrap support is indicated on branches. Accession codes correspond to those indicated in Table S4.

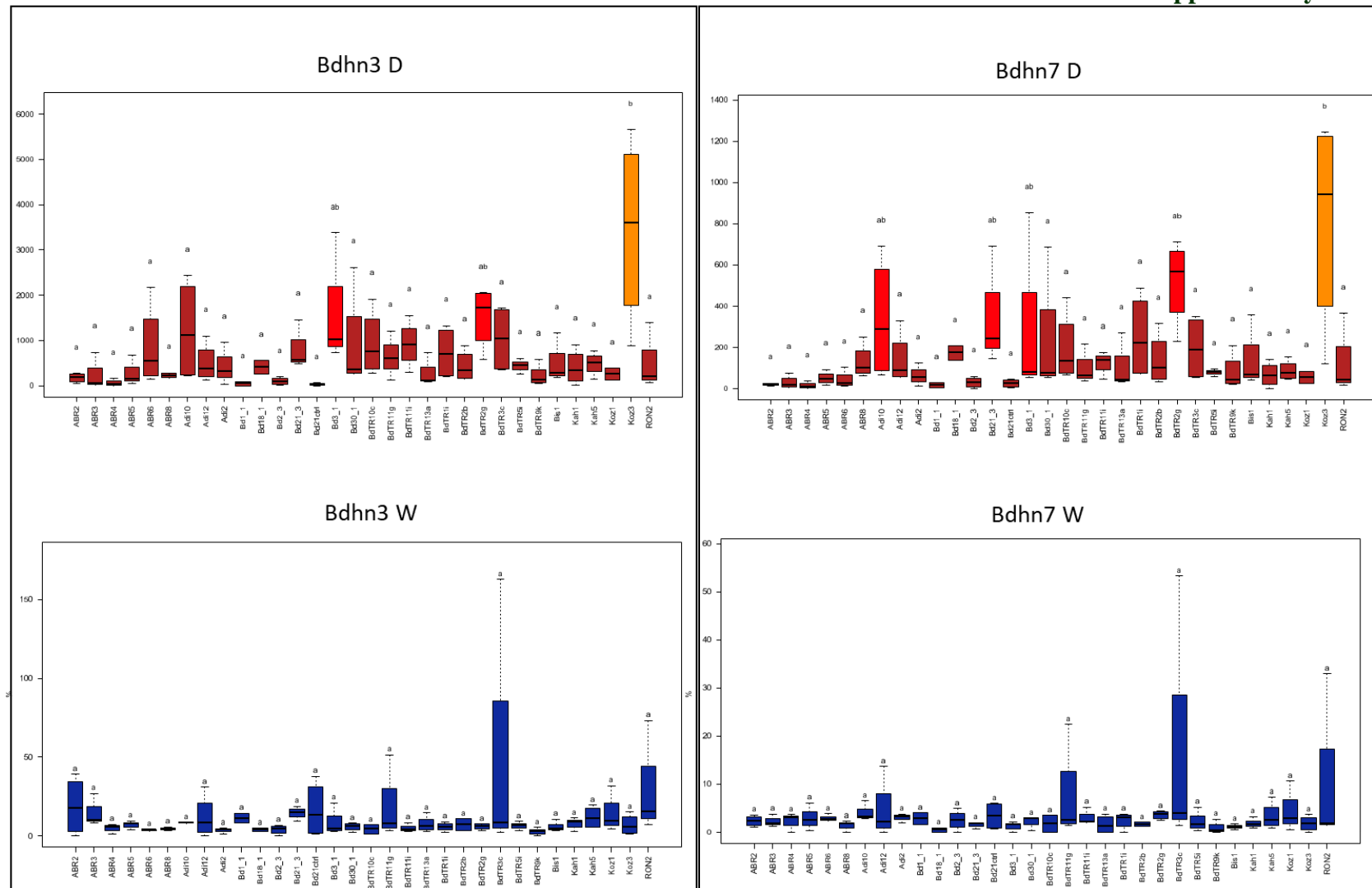


**Supplementary figure S5:** Bidimensional PCA plot of 54 *B. distachyon* ecotypes obtained from 19 climate variables (see Table S6). PC1 and PC2 axes comprise 48.5% and 22.4% of the variance, respectively. Ecotype codes correspond to those indicated in Table S4. Ellipses include ecotypes classified within cold (aquamarine), mesic (green), and warm (red) climate classes according to their PC1 values



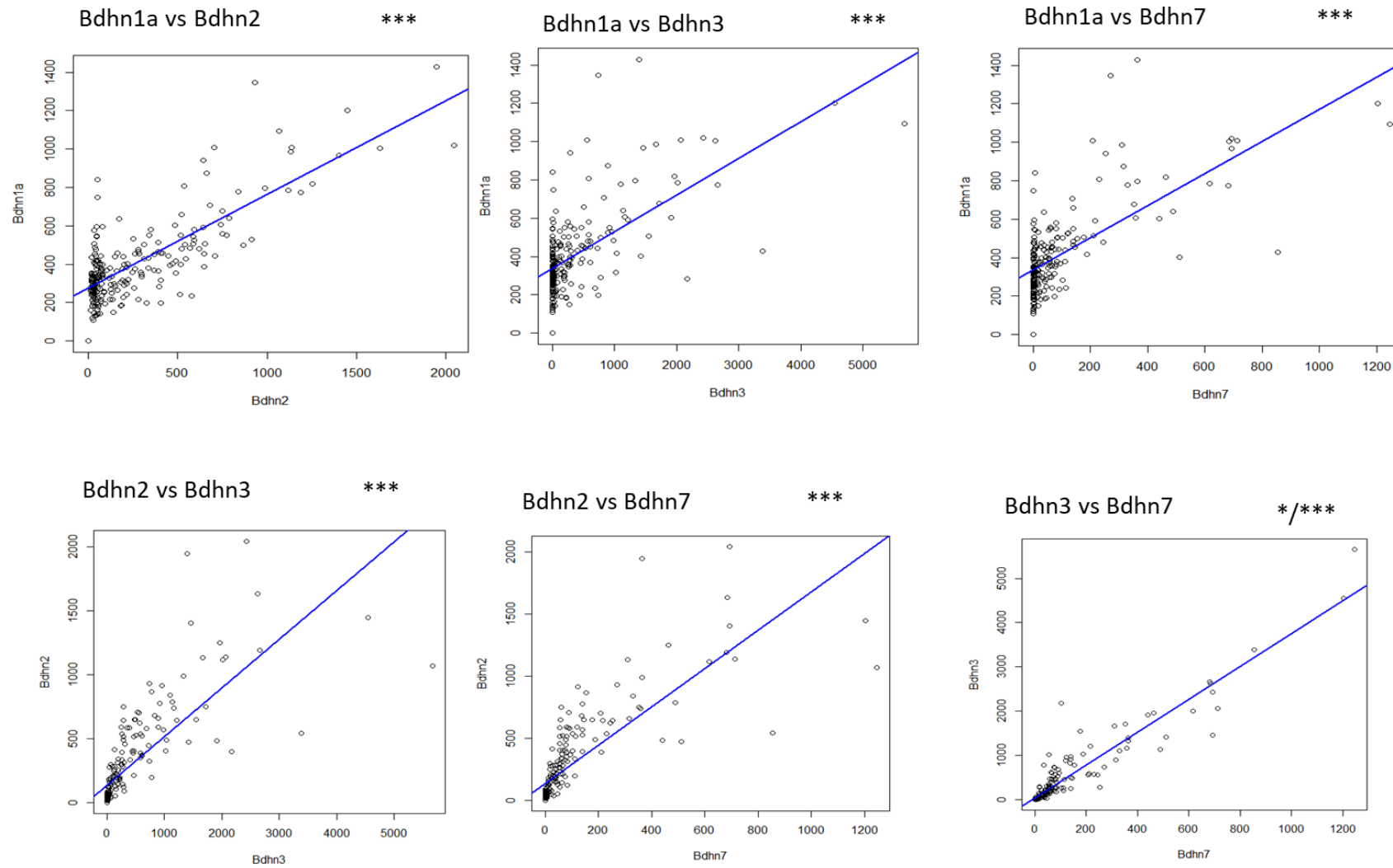
**Supplementary figure S6:** Boxplots and Wilcoxon pairwise significance tests of differential gene expression values (normalized transcript per million, TPM) of the four expressed dehydrin genes (Bdhn1a, Bdhn2, Bdhn3, Bdhn7) under joint and separately analyzed drought and temperature stress conditions in 32 *B. distachyon* ecotypes. Averaged expression values for C (Cool), H (Hot), W (Watered), and D (Drought) treatments and their combinations (see text). (a) Pairwise comparative tests of combined CD-CW-HD-HW treatments; all dehydrins were significantly differentially expressed in all CD vs. CW and HD vs. HW tests, by contrast they were not significantly different in all CD vs. HD tests and in two CW vs. HW tests (Bdhn1a, Bdhn2); (b) Pairwise comparative tests of D vs. W treatments; all dehydrins were significantly differentially expressed; (c) Pairwise comparative tests of C vs. H treatments; none of the dehydrins were significantly differentially expressed.



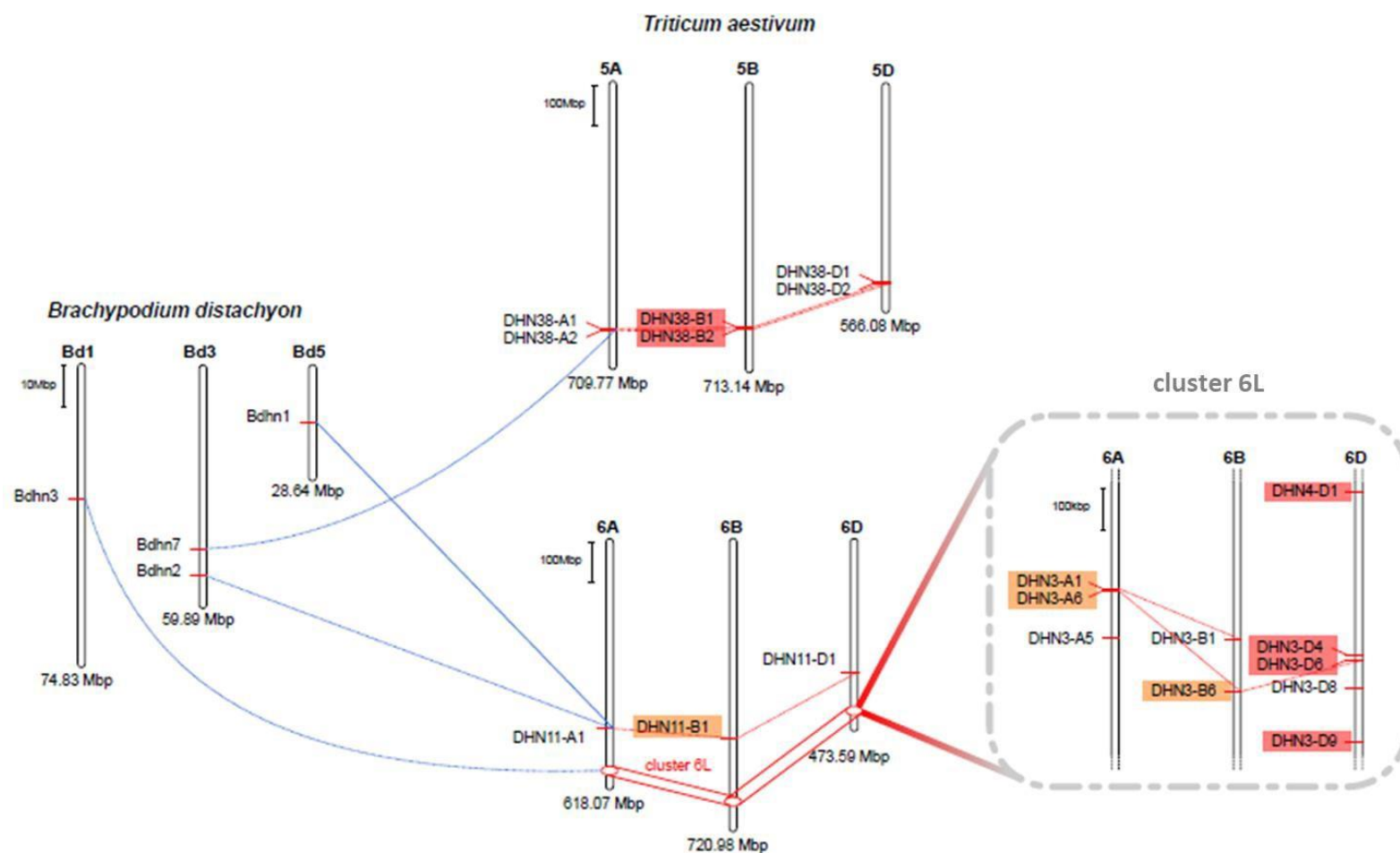


**Supplementary figure S7:** Differentially expressed Bdhn1a, Bdhn2, Bdhn3 and Bdhn7 dehydrin genes (normalized transcript per million, TPM) across 32 ecotypes of *B. distachyon* under drought (D, red) vs. watered (W, blue) conditions. Different letters in the boxplots indicate significant group differences (Tukey tests)

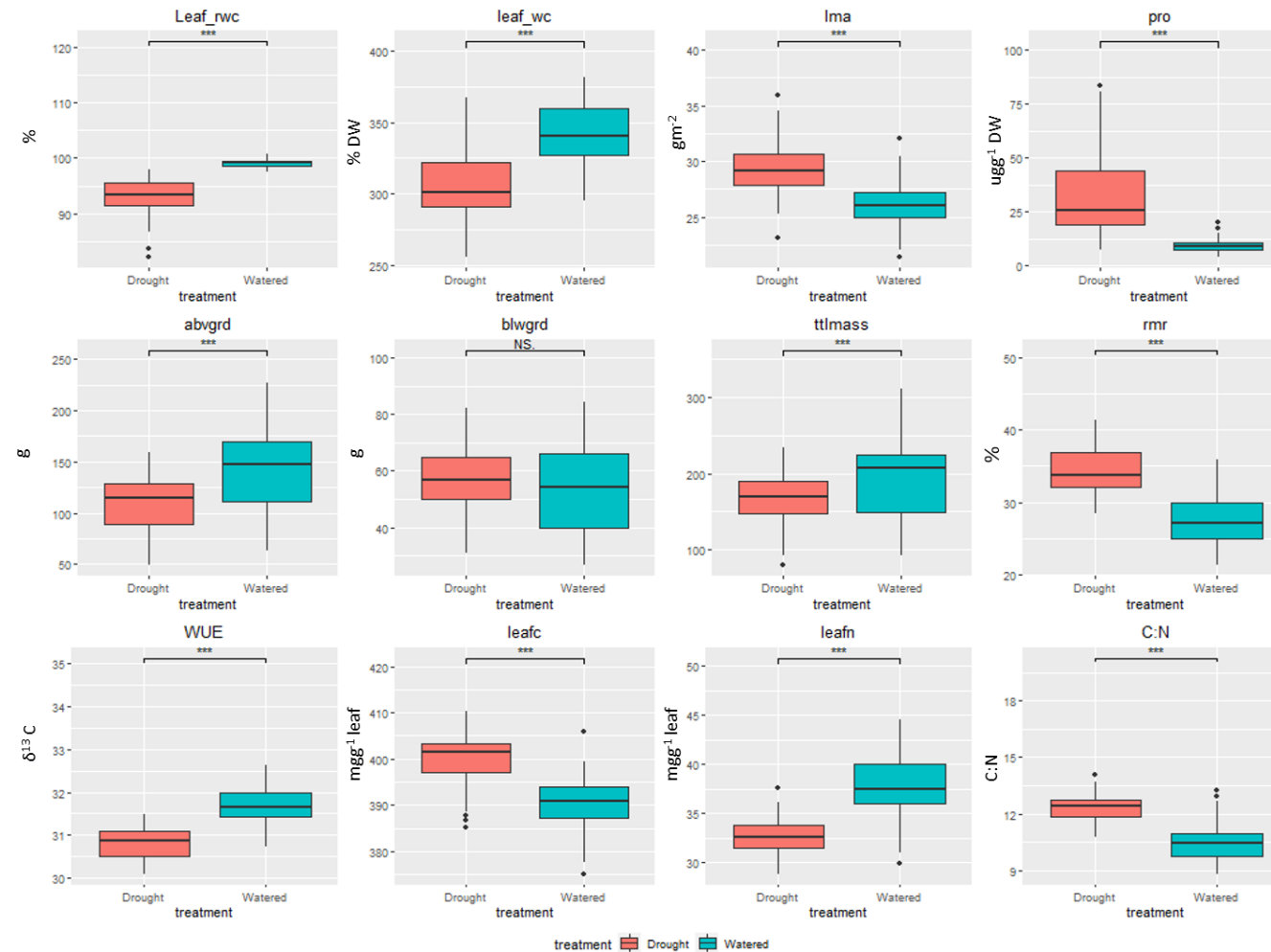




**Supplementary figure S8:** Linear model regression plots of pairwise dehydrine Bdhn expression values (normalized transcript per million, TPM) under joint drought and watered conditions.

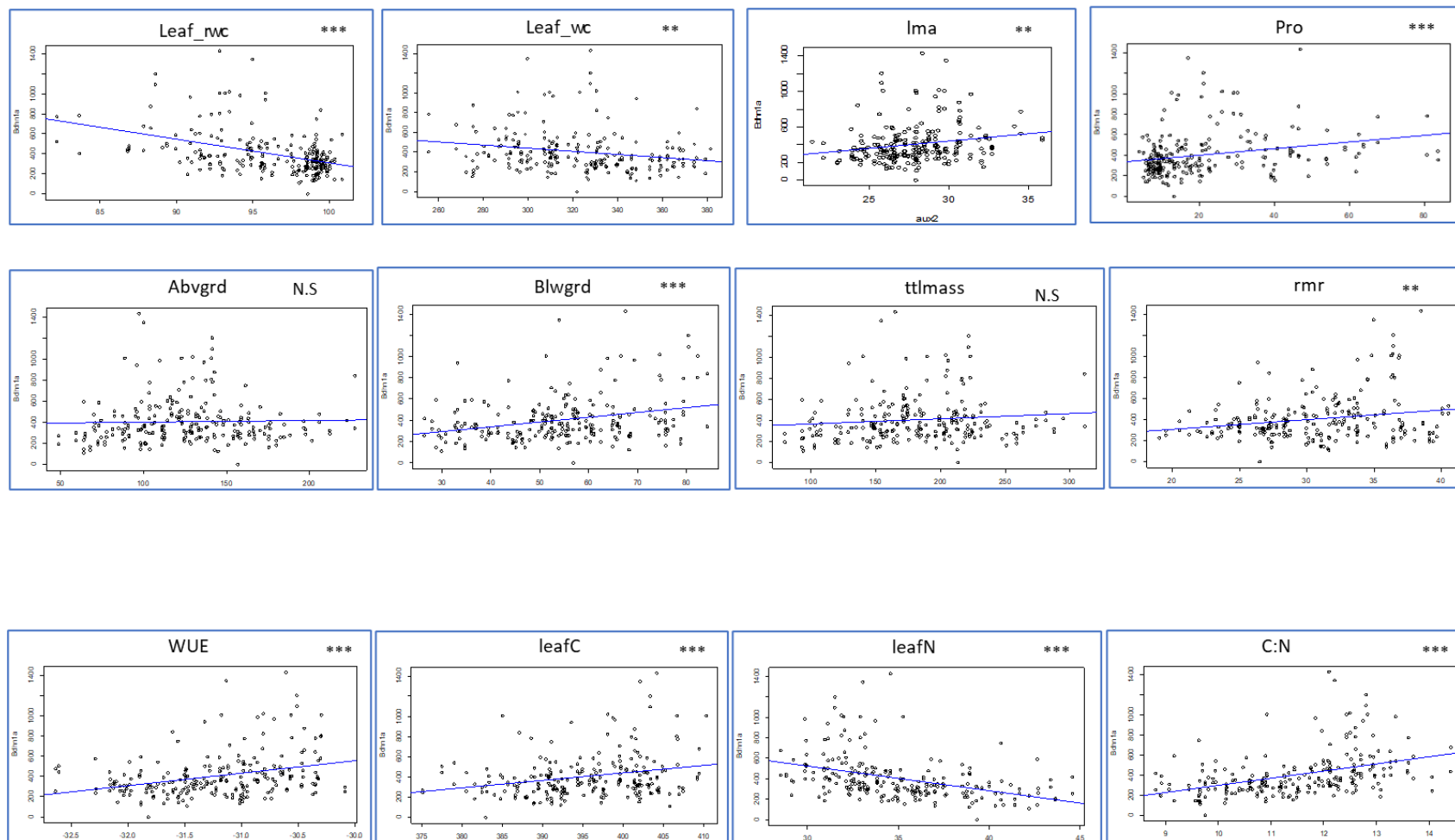


**Supplementary figure S9:** Physical locations of orthologous *Brachypodium distachyon* and *Triticum aestivum* water stress responsive dehydrin genes. Drought induced wheat dehydrin genes are highlighted in colors. *Brachypodium* chromosomes are drawn at 10x scale with respect to wheat chromosomes.

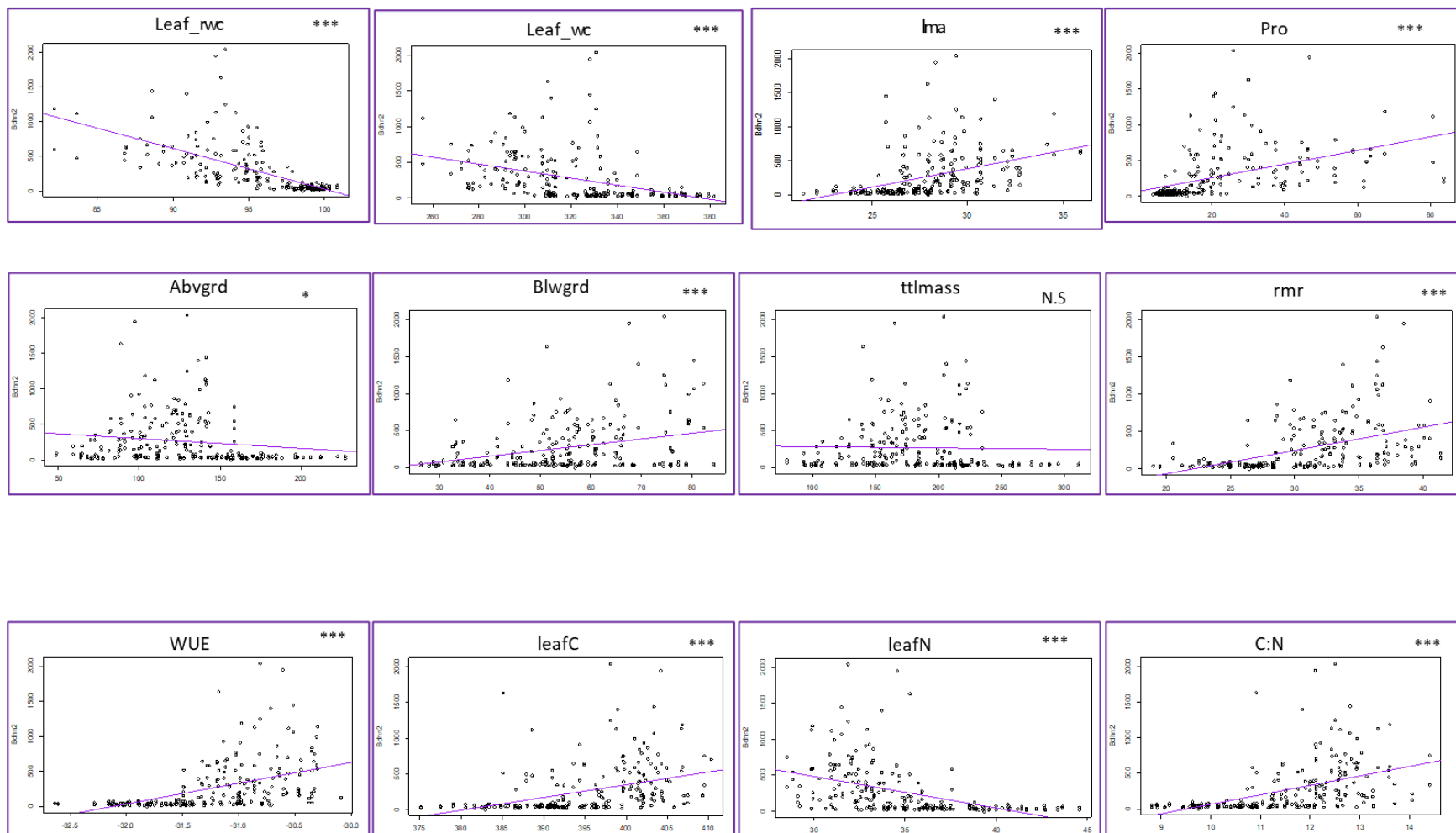


**Supplementary figure S10:** Drought-response phenotypic changes as a function of drought treatment averaged across 32 *B. distachyon* ecotypes [leaf\_rwc (relative water content in leaf); leaf\_wc (water content in leaf); lma (leaf mass per area); pro (leaf proline content); abvgrd (above ground biomass); blwgrd (below ground biomass); ttlmass (total mass); rmr (root mass ratio); delta13c (carbon isotope, a proxy for lifetime integrated WUE); leafc (leaf carbon content); leafn (leaf nitrogen content); cn (leaf carbon/nitrogen ratio)]. Asterisks above boxes indicate Wilcoxon pairwise significant difference among drought (D, red) and watered (W, blue) conditions (p-value < 0.001, \*\*\*)

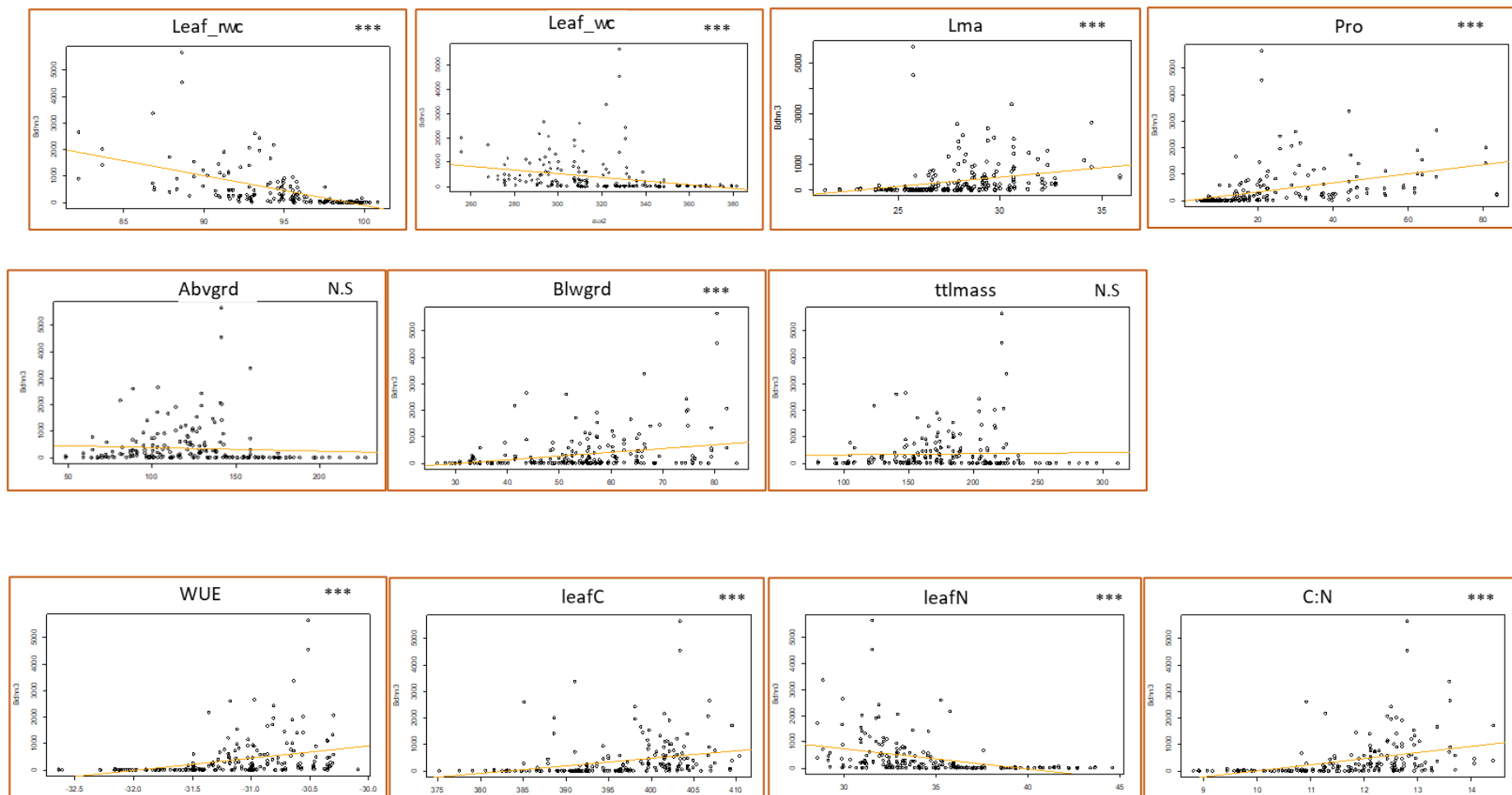
Bdhn1a vs phenotypic traits



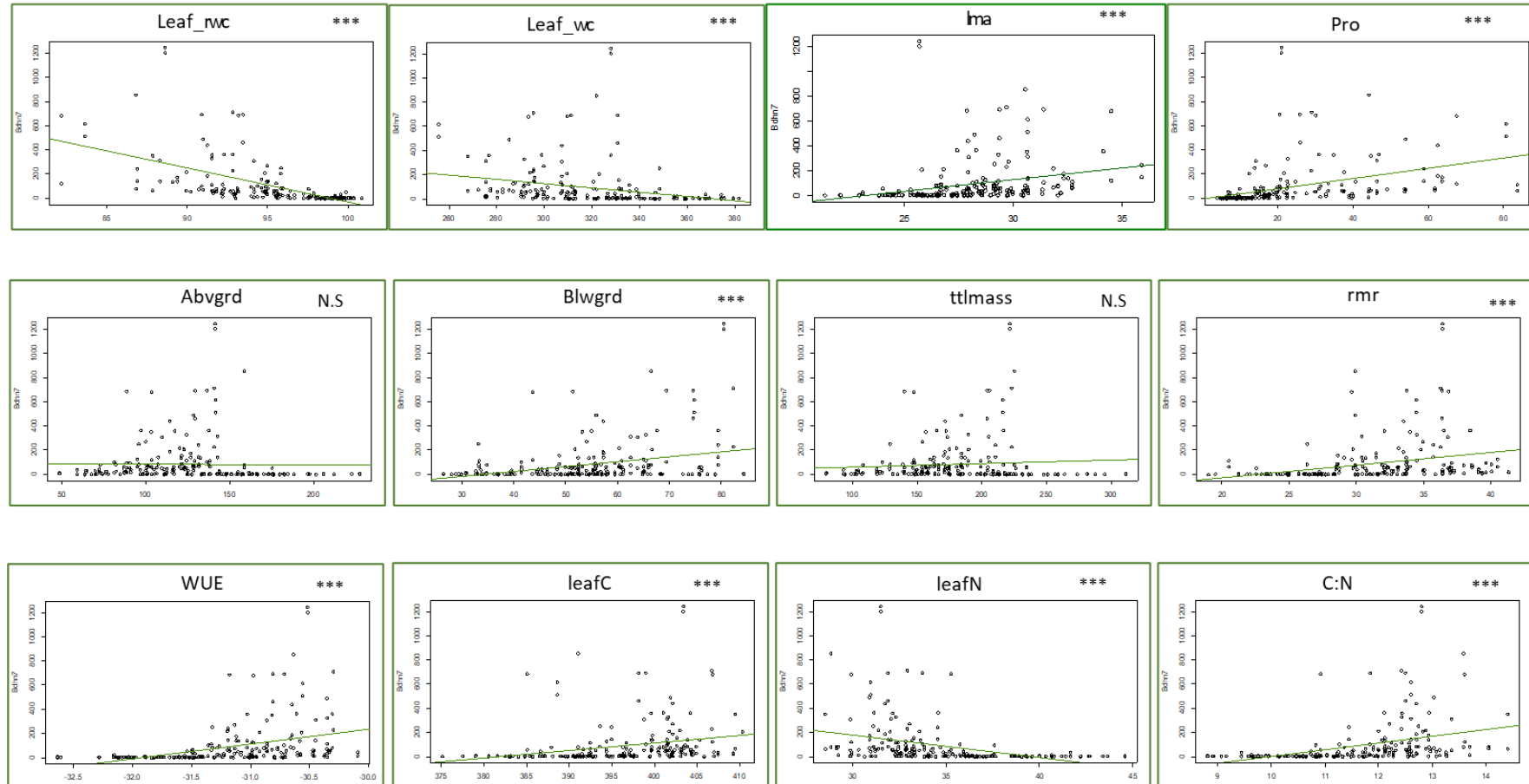
Bdhn2 vs phenotypic traits



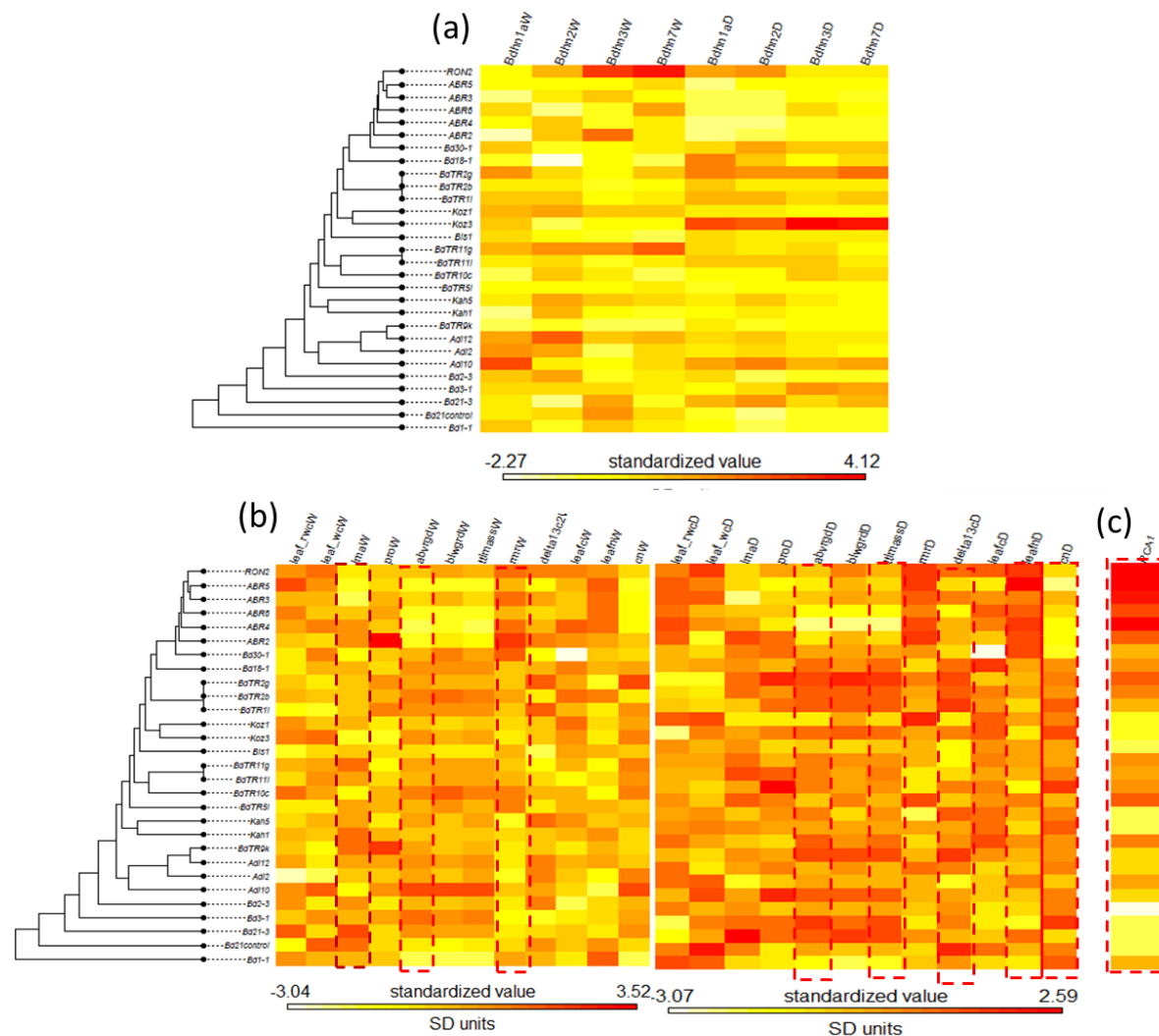
Bdhn3vs phenotypic traits



Bdhn7vs phenotypic traits



**Supplementary figure S11:** Figure S11: Linear model regression plots of dehydrins Bdhn1a, Bdhn2, Bdhn3 and Bdhn7 expression values (normalized transcript per million, TPM) and drought response phenotypic traits changes under total dry (D) and watered (W) conditions.



**Supplementary figure S12:** Maximum Likelihood *B. distachyon* nuclear species tree cladogram showing the relationships of 30 ecotypes. Phyloheatmaps of normalized values for different sets of variables: (a) dehydrin (Bdhn1, Bdhn2, Bdhn3, Bdhn7) gene expression values under watered (W) and drought (D) conditions; (b) drought-response phenotypic traits (leaf\_rwc; leaf\_wc; lma; pro; abvrgd; blwgrd; tlmass; rnr; delta13c; leafc; leafn; cn) values under watered (W) and drought (D) conditions; (c) climate niche PCA1 values. Traits showing significant phylogenetic signal are highlighted with dotted lines (see Table S13a) appendix



### APPENDICES

Supplementary and additional data (Appendices) for each of the PhD chapters can be accessed in the following links:

- <https://github.com/Bioflora/BsylvaticumBotPacifica> (Chapter 1)
- [https://github.com/Bioflora/Brachypodium\\_phylogenomics](https://github.com/Bioflora/Brachypodium_phylogenomics) (Chapter 2)
- [https://github.com/Bioflora/Brachypodium\\_Repeteome](https://github.com/Bioflora/Brachypodium_Repeteome) (Chapter 3)
- [https://github.com/Bioflora/BretusumBphoenicoides\\_Morphology](https://github.com/Bioflora/BretusumBphoenicoides_Morphology) (Chapter 4)
- [https://github.com/Bioflora/Brachypodium\\_dehydrins](https://github.com/Bioflora/Brachypodium_dehydrins) (Chapter 5)

## PUBLICATIONS OF THE PhD THESIS

**Chapter 1:** Article published in *Botanica Pacifica*. A journal of plant science and conservation.

**Reference:** Phylogenetics of the palearctic model grass *Brachypodium sylvaticum* uncovers two divergent oriental and occidental micro-taxa lineages. 2022. *Botanica Pacifica* 12(1): 21-28. (No. ART-2023-134080). DOI: 10.17581/bp.2023.12119

**Authors:** Catalán, P., Probatova, N. S., **Decena, M. Á.**, Sancho, R., Pérez-Collazos, E., Inda, L. A., & Viruel, J. (2023).

**Chapter 2:** Article under final revisions by the authors

**Reference:** Cyto-phylogenomics of the *Brachypodium* grass species-complex unveils highly divergent cryptic diversity and different pre- and post-polyploidization descendant dysploidy trends.

**Authors:** **María Ángeles Decena**, Rubén Sancho, Joanna Lusinska, Robert Hasterok, Rubén Gorgojo, Beatriz Montes, Ernesto Pérez-Collazos, Luís. A. Inda, Pilar Catalán.

**Chapter 3:** Article under final revisions by the authors

**Reference:** Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in grass *Brachypodium* polyploids.

**Authors:** **María Ángeles Decena**, Rubén Sancho, Luis A. Inda, Ernesto Pérez-Collazos, Pilar Catalán.

**Chapter 4:** Article under revision by the authors

**Reference:** Morphoanatomical analysis of cryptic and evolutionary divergent polyploid cytotypes of model Mediterranean *Brachypodium retusum* and *B. phoenicoides* grasses

**Authors:** **María Ángeles Decena**, Luis A. Inda, Ernesto Pérez-Collazos, Pilar Catalán.

## Publications of the PhD thesis

**Chapter 5:** Article published in *Plants*, a journal ranked in the first quartile (Q1) of the Plant Sciences area (Web Of Science)

**Reference:** Comparative Genomics, Evolution, and Drought-Induced Expression of Dehydrin Genes in Model *Brachypodium* Grasses. 2021, *Plants* 10, 2664. <https://doi.org/10.3390/plants10122664>

**Authors:** Decena, M.A.; Gálvez-Rojas, S.; Agostini, F.; Sancho, R.; Contreras-Moreira, B.; Des Marais, D.L.; Hernandez, P.; Catalán, P.