TESIS DE LA UNIVERSIDAD DE ZARAGOZA

2024 274

Valeriia Shiposha

Phylogeograghy, morphological variation and ecological niche modeling of Eurasian and Mediterranean ***Brachypodium*** P. Beauv. (Poaceae)

Director/es
Marqués, Isabel
Olonova Vladimirovna, Marina
Catalán Rodríguez, Pilar

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

PHYLOGEOGRAGHY, MORPHOLOGICAL VARIATION AND ECOLOGICAL NICHE MODELING OF EURASIAN AND MEDITERRANEAN ***BRACHYPODIUM*** P. BEAUV. (POACEAE)

Autor

Valeriia Shiposha

Director/es

Marqués, Isabel
Olonova Vladimirovna, Marina
Catalán Rodríguez, Pilar

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

2019

PhD THESIS

JOINT THESIS SUPERVISION

UNIVERSITY OF ZARAGOZA – TOMSK STATE UNIVERSITY

Phylogeography, morphological variation and ecological niche modeling of Eurasian and Mediterranean *Brachypodium* P. Beauv. (Poaceae) species

Author:

Valeriia Shiposha

Supervisors:

Dr. Pilar Catalán Rodríguez

Dr. Marina Olonova

Dr. Isabel Marques

Huesca-Tomsk, 12 December 2019

High Polytechnic School

Agricultural and Environmental Sciences

University of Zaragoza

Institute of Biology

Department Botany Department

Tomsk State University

High Polytechnic School

Agricultural and Environmental Sciences

University of Zaragoza

Institute of Biology

Department Botany Department

Tomsk State University

Phylogeography, morphological variation and ecological niche modeling of Eurasian and Mediterranean *Brachypodium* P. Beauv. (Poaceae) species

PhD thesis presented by

Ms Sc Valeriia Shiposha

To attain the Philosophy Doctor degree jointly issued by the

University of Zaragoza and Tomsk State University

Supervisors of the PhD thesis:

Dr. Pilar Catalán, Dr. Marina Olonova, Dr. Isabel Marques

Dr. Pilar Catalán, Professor in Botany at the University of Zaragoza, Dr. Marina Olonova, Professor in Botany at Tomsk State University and Dr. Isabel Marques, Assistant Researcher at the University of Lisbon certify that the work presented in the current PhD thesis has been developed under their co-supervision and authorize its submission and defense.

Huesca-Tomsk, 12 December 2019

The current PhD thesis has been developed under a specific collaboration agreement between the University of Zaragoza and the University of Tomsk for joint thesis supervision. The PhD thesis work was developed in the laboratories of the research group Bioflora, at the University of Zaragoza (Unizar, High Polytechnic School of Huesca, Agricultural and Environmental Sciences Department), under the supervisions of Drs. Pilar Catalan and Isabel Marques, and the research group in Plant Biology and Ecology, at Tomsk State University (TSU, Department of Botany, Institute of Biology Herbarium), under the supervision of Dr. Marina Olonova.

The molecular laboratory work was conducted at the High Polytechnic School of Huesca (Unizar), and the ecological niche modeling work at the Institute of Biology Herbarium (TSU). The systematic work was developed in the two institutions and during two research stays at the Komarov Institute (LE) and the Moscow University (MW) herbaria. The bioinformatics resources used were provided by the Unizar High Polytechnic School of Huesca cluster and the TSU Computing center.

The PhD thesis was funded by the research grant project "Evolution of biological traits and speciation processes in the model genus *Brachypodium* (Poaceae) based on comparative and functional genomic analysis" subsidized by the Spanish Ministry of Economy and Competitiveness (Grant No. CGL2016-79790-P). The PhD student benefited from a pre-doctoral fellowship from TSU, and one of its research stays at Unizar (6 months) was funded by the Russian Ministry of Education and Science.

To my supervisors

ACKNOWLEDGEMENTS

I would like to thank all the people who accompanied and supported me in this demanding and important research activity.

My supervisors, Dr. Pilar Catalan, Dr. Marina Olonova, Dr. Isabel Marques for giving me the opportunity to carry on a thesis under their leading, for the tremendous efforts spent on teaching me the classical and modern method of researching plants, for making research plans, for spending time on analysis, revision and improvement of all manuscripts that make up this thesis, for huge support, attention and advice.

Alexander Kuznetsov and Tomsk State University for financial support, for the opportunity to go on scientific internships, conferences, seminars, for support and friendly advice.

To the leadership of the University of Zaragoza and the High Polytechnic School of Huesca for the warm greeting and the opportunity to work in these institutions.

Herbarium curators AAU, B, C, FI, G, JACA, LD, LE, MW, NS, NSK, SEV, TC, UGR, Unizar-EPS for providing research material.

Employees of the Bioflora Research Group, Laboratory of Molecular and Structural Analysis, Herbarium of P.N. Krylov (TSU), Department of Botany (TSU) for suitable conditions for scientific work, for help and advice.

Dr. Luis Ángel Inda for his contribution to the morphological studies that are part of this thesis.

Dr. Diana López for her contribution to the amplification and sequencing of the chloroplast genome, as well as for her scientific support in this dissertation.

Marisa López for her help and teaching various laboratory methods in the early stages of the thesis.

Dmitry Feoktistov for his help and training in programs for statistical analysis and the construction of ecological niches.

Alexandra Selezneva for assistance in anatomical studies of stomata guard cells.

Roman Romanets for his help in researching pollen grains on a scanning microscope.

Researchers, professors and employees of the High Polytechnic School of Huesca, who made me feel them like friends every time of my stay in Spain, showed picturesque places and provided support: Dr. Ernesto Pérez-Collazos, Dr. Ruben Sancho, Dr. Ester Sales Clemente, Dr. Rocío López Flores, Dr. José Antonio Cuchí, Dr. Luis Villar Pérez.

Researchers, friends of the laboratory of molecular and structural analysis (TSU) Evgeny, Natalia, Denis, Polina, as well as the head Dr. Irina Gureeva of the Herbarium of TSU.

PhD students with whom I had the pleasure to coincide during my thesis in Tomsk State University Alena, Tamara, Lydia, Tatiana, Eugene.

My family and close friends for their unconditional support and for the fact that I was able to write this work.

Thanks to everyone and everyone who I met on this long and difficult path!

INDEX

SUMMARY

The genus *Brachypodium* represents a singular, exquisite model system genus in the post-genomic era of plant biology that has emerged as the right tool for much needed advances in the biology and evolution of grasses. It has been greatly accelerated by next-generation sequencing technologies and rapidly subsumed into fundamental research on evolutionary biology, systems biology, ecology research, and for the development of new tools and concepts towards improving other temperate C_3 grasses. Additionally, *Brachypodium* species have maintained their wildness—providing a treasure-trove of resources for plant systematics and ecology researchers to study the plants *in situ*.

Plant perenniality is an outstanding issue in biological research, aiming to identify the mechanisms causing the allocation of resources and the subsequent life strategy. The perennial-annual transition (or its converse) has occurred several times across the evolution of the *Poaceae*. Two evolutionary switches from more ancestral annual lineages to more recently evolved perennial lineages (*B. stacei* → *B. mexicanum*, *B. distachyon* → core perennial clade), have occurred along the *Brachypodium* phylogeny, framing an exceptional evolutionary scenario for the ongoing investigation of annuality/perenniality examples in the model genus. Allopolyploidy is recognized as a main evolutionary event in angiosperms, which are all considered descendants of palaeopolyploid ancestors, and with some subsequently diploidized lineages accumulating more recent mergings (hybridization) and genome doublings in meso and neopolyploid species. It is especially remarkable in the grass family, where allopolyploids account for 70-80% of its species. *Brachypodium* again serves as an optimal model genus, in this case to explore the evolution and the functional outcomes of allopolyploidy. Within the genus, of ~20 species half are diploids and half are allopolyploids with the latter showing differently inherited (more ancestral to more recently evolved) compact subgenomes of unknown origin. The three annual *Brachypodium* species were selected as a tractable grass polyploid system to investigate the effects of subgenome dominance and origin in the phenotypic, physiological and adaptive responses of the

allotetraploid hybrids. However, the putative multiple origins of the hybrids have not yet been disclosed.

Despite the outstanding advances acquired on the genomic, evolutionary and biological features of the model *Brachypodium* species, the genus is still in lack of deep systematic, evolutionary and ecological studies for some of its species and at infraspecific population level. In this PhD thesis we have uncovered the existing gap of knowledge on the systematic, ecological and population-genetic characteristic of several annual and perennial Eurasian and Mediterranean *Brachypodium* species.

Our phylogeographic and population genetic studies of the annual *B. stacei*, *B. distachyon* and *B. hybridum* have contributed to discover new germplasm sources and to disentangle the genetic diversity and structure of their populations, the spatio-temporal scenarios of their origins and dispersals, the relationships of genetic groups with environmental variables, and the multiple founder events of the hybrid allopolyploid from distinct progenitor sources. Our research has also comprised the investigation of the perennial Eurasian and Mediterranean species of the *B. pinnatum* complex and of the Eurasian *B. sylvaticum* species. Our environmental niche modeling analyses of the perennial *B. sylvaticum*, *B. pinnatum* s. l. and *B. phoenicoides* species have shown that the Atlantic *B. sylvaticum* and the continental *B. pinnatum* species likely had western and northern glacial refugia in Eurasia, whereas *B. phoenicoides* probably evolved and survived *in situ* in the western Mediterranean region, and that *B. sylvaticum* occupies a stable niche in relict black taiga and broad-leaved forests in Siberia that experienced a considerable niche reduction since the last glaciation, apparently caused by global change. Our studies on the phenotypic variation of the diploid and tetraploid cytotypes of *B. pinnatum*, *B. rupestre* and *B. phoenicoides* have detected significant differences in discriminant traits that separate the species and a large infraspecific morphoanatomical variation within potential hybrid swarms and diverse cytotypes of this "compilospecies" group.

The compilation of studies developed in this thesis has contributed to increase the current knowledge on the phylogeography, environmental niche features and phenotypic variation of species of the genus *Brachypodium*

PhD THESIS STRUCTURE

The PhD thesis is structured in three general chapters (Introduction, Objectives, Conclusions) and five specific chapters related to the research conducted during the PhD work (Chapters 1 to 5). A further section of Appendices includes supplementary information from the research chapters. The last section of the thesis lists the Publications obtained from the PhD research.

The order of the chapters and sections is as follows:

- **INTRODUCTION.**
- **OBJECTIVES**: Main and specific objectives of the PhD work.
- Chapters of studies conducted in the thesis (each chapters contains the Summary, Introduction, Material and Methods, Results and Discussion):
    - **Chapter 1.** Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands.
    - **Chapter 2.** Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*
    - **Chapter 3.** Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot.
    - **Chapter 4**. Glacial refugia and postglacial range shifts of the widespread Eurasian *Brachypodium sylvaticum* and *B. pinnatum* and the western Mediterranean *B. phoenicoides* grasses: insights from environmental niche modeling analysis and conservation status of relict Siberian populations of *B. sylvaticum*.
    - **Chapter 5.** Morphoanatomical study of the *Brachypodium pinnatum* complex species (*B. pinnatum, B. rupestre, B. phoenicoides*) (Poaceae) and their diploid and polyploid cytotypes.
- **CONCLUSIONS:** General conclusions of the thesis.
- **APPENDICES:** This section includes supporting information from each research chapter (supplementary methods, results, tables and figures):

- **Appendix I:** Supporting Information of Chapter 1
- **Appendix II:** Supporting Information of Chapter 2
- **Appendix III:** Supporting Information of Chapter 3
- **Appendix IV:** Supporting Information of Chapter 4
- **Appendix V:** Supporting Information of Chapter 5

- **PUBLICATIONS OF THE PhD THESIS:** This section lists the publications obtained from and contributed to by the PhD thesis.

INTRODUCTION

Phylogenetic position of the temperate grasses within the Poaceae evolutionary framework

The Poaceae include approximately 11,000-11,506 species (Kellogg 2015; Soreng et al. 2017) classified into 698 (Kellogg, 2015) to 768 (Soreng et al. 2017) genera. They are arranged into 12 subfamilies and 51-52 (Soreng et al. 2015, 2017) to 30 (Kellogg 2015) tribes. Evolutionary studies of grass representatives indicate a diverging grade of ancient Anomochlooideae, Pharoideae, and Puelioideae subfamilial lineages that preceded the split of the main BOP (Bambusoideae, Oryzoideae and Pooideae) and PACMAD (Panicoideae, Aristoideae, Chloridoideae, Micrairoideae, Arundinoideae, Danthonioideae) clades (Kellogg, 2015; Soreng *et al.*, 2017). Comparative genomics studies indicate that all the Poaceae derive from a grass ancestor that likely experienced a whole genome duplication (WGD) event between 90 to 70 Ma (Paterson *et al.*, 2004; Salse *et al.*, 2008; Murat et al. 2010). Evidence suggests that the ancient grass paleopolyploidization was followed by subsequent diploidizations, involving differential losses of many duplicated heterologous copies in the subgenomes (Paterson *et al.*, 2004) or by profound distinct genomic rearrangements, including successive centromeric chromosome fusions along the divergent grass lineages (Murat et al. 2010). By contrast, new polyploidization events apparently led to the rising of mesopolyploids, originated some million years ago, and of neopolyploids, considered to have arisen during or after the Quaternary glaciations (Stebbins, 1985; Marcussen *et al.*, 2015). Allopolyploids account for 70% of the current grass species (Stebbins, 1949; Kellogg, 2015).

The largest subfamily Pooideae contains ca. 3968 cold and temperate grass species ranked into 202 genera (Soreng et al. 2017). The pooids include some of the most prominent crops such as wheat, rye, oats and barley. Recent molecular phylogenies support the monophyly of the Pooideae within the Poaceae, and recover it as sister to the Bambusoideae in the BOP clade (Saarela *et al.*, 2015; Sancho et al. 2018). The tribal arrangement of the

Pooideae has varied widely over the last century. In the most recent classification twelve subtribes (plus the incertae sedis Avenula – Homalotrichon) belong to the Poeae-type plastid DNA clade and seven tribes to the Aveneae-type plastid DNA clade (Soreng et al., 2015). The systematic positions of the different tribes and subtribes within the Pooideae are currently under discussion, and their evolutionary relationships are not totally resolved (Kellogg, 2015; Soreng et al., 2015, 2017). The pooids show a karyotype evolutionary trend of increasing chromosome sizes and decreasing chromosome base numbers (Catalán *et al.*, 1997) ranging from basal tribes with small chromosomes and high chromosome base numbers (Brachyelytreae=11; Lygeae=10; Nardeae=13; Phaenospermatae=12; Meliceae=10, 9, 8; Stipeae=12, 11, 10; Diarrheneae=10), through the intermediate ones of Brachypodieae (10, 9, 8, but also 5) (Catalán & Olmstead, 2000), to the large chromosomes and almost constant chromosome base number of x=7 present in the more recently evolved Triticodae + Poodae, although x=6, 5, 4, 2 occasionally occur in Aveneae (Catalán *et al.*, 2016). The increase in the rate of diversification detected in the temperate Pooideae grasses, (Pimentel *et al.*, 2017) was correlated with the drop in global temperatures that took place in the Middle to Late Eocene and the Oligocene (Beerling & Royer, 2011). 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).

Systematics and evolution of *Brachypodium*

Brachypodium (L.) P. Beauv. was separated from *Bromus* L. in 1812 by Palisot de Beauvois. Decades of systematic and phylogenetic studies were necessary, however, to frame its evolutionary position within the grasses. *Brachypodium* is considered today the single representative genus of the monotypic tribe Brachypodieae, which constitutes one of the intermediate diverging lineages of the temperate Pooideae grasses (Catalán et al. 2016). Its controversial position was caused by its shared or similar morphological and anatomical traits with distinct pooid groups. Consequently, it was classified in different tribes, based on the

possession of embryo with mesocotyl (Poeae), hairy terminal ovary appendage and long narrow caryopsis and hilum (Bromeae), or spicate to racemose inflorescence and hairy lodicles (Triticeae), until its definitive adscription to its own tribe Brachypodieae (Jacques-Félix 1962; Schippmann 1991; Watson & Dallwitz 1992). Subsequently, its separate tribal treatment was confirmed by a number of private biological (embryo development), biochemical (exclusive seed storage proteins, seed globulins, seed storage polysaccharides and stem and leaf fructosans) (Schippmann 1991), karyotype (large disploidy) (Robertson 1981; Khan 1984) and genetic landmark (Catalán et al. 2016) characters.

The most recent phylogenetic works have consistently resolved Brachypodieae as the sister lineage of the recently evolved core pooid clade of temperate cereals and forages [Triticodae (Triticeae + Bromeae) /Poodae (Poeae + Aveneae)]. Its intermediate placement between the basal (Brachyelytreae, Lygeae-Nardeae, Phaenospermatae, Meliceae, Stipeae) and the recently evolved (Triticodae/Poodae) Pooideae lineages has been recovered from both plastid and nuclear based topologies (Catalán et al. 1997; Schneider et al. 2011; Catalán et al. 2016) and from combined analysis of molecular and morphological data (GPWG 2001). An intermediate position in the Pooideae tree is also reconstructed for the isolated *Diarrhena* (Diarrheneae) lineage, which apparently split earlier than *Brachypodium* (Catalán et al. 1997; Schneider et al. 2011; Sancho et al. 2018). A recent phylogenomic study based on plastome analysis showed, however, that *Brachypodium* and *Diarrhena* could be closer to the basal pooids than to the recently evolved core pooid clade (Sancho et al. 2018). The two independent and small monogeneric Brachypodieae and Diarrheneae tribes present remarkable embryo features (bambusoid-like in *Diarrhena*, first lateral stem developing from coleoptile in *Brachypodium*), with *Brachypodium* also showing intermediate chromosome base numbers when mapped into the pooid tree (Catalán et al. 2016).

The genus *Brachypodium* is widely distributed throughout the world, but probably the center of its diversity and species formation is the Mediterranean region, where most of the current species are concentrated and the highest intraspecific diversity is observed

(Schippmann, 1991, Catalán et al., 2016). According to various estimates, *Brachypodium* comprises from about 18 species (Tsvelev, 1976; Schippmann, 1991; Catalán et al., 2016; Diaz-Perez et al. 2018) to several proposed names (TROPICOS, https://www.tropicos.org/NameSearch.aspx?name=Brachypodium&commonname=). The most recent taxonomic updatings indicate that three of the 18 recognized taxa are annual species and 15 are perennial taxa (Catalán et al. 2016; Diaz-Perez et al. 2018). The three annual species have a large distribution in their native circum-Mediterranean region (*B. distachyon*, *B. stacei*, *B. hybridum*) (Catalán et al. 2012; López-Alvarez et al. 2012, 2015). 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. bolusii*), tropical and South Africa (*B. flexum*), Madagascar (*B. madagascariense*), Taiwan (*B. kawakamii*), SE Asia – New Guinea (*B. sylvaticum* var. *pseudodistachyon*), and America (*B. mexicanum*)] (Schippmann 1991; Catalán et al. 2016; Diaz-Pérez et al. 2018). Two *Brachypodium* species, the annual *B. hybridum* and the perennial *B. sylvaticum*, are invasive plants. *B. sylvaticum* has been introduced and is spread in western N America and in Australia, and *B. hybridum* has successfully colonized C Europe, western N America (California), S America (Uruguay, Argentina), South Africa and Oceania (Australia, New Zealand) (Catalán et al. 2016).

The annual species are characterized by their short life-cycle, ephemeral habit and self-fertility (Catalán & Olmstead 2000; Catalán et al. 2012). Recent analysis of cryptic phenotypic, cytogenetic and molecular traits allowed researchers to separate the three species (*B. distachyon*, *B. stacei* and *B. hybridum*; Catalán et al. 2012). By contrast, most of the perennial taxa show long-rhizomes and self-incompatibility (Catalán et al. 1995; Khan & Stace 1999), except *B. mexicanum* and *B. sylvaticum* that are self-compatible and slender-rhizomatose plants (Khan & Stace 1999; Steinwand et al. 2013). *B. sylvaticum i*s characterized by its nodding panicle, densely hairy habit and long-awned lemma. The long-rhizomatose *B.*

pinnatum, *B. rupestre* and *B. phoenicoides* show erect panicles. *B. phoenicoides*, adapted to dry places, is glabrous and presents partially inrrolled leaves, semi-patent twisted spikelets and awnless lemmas, whereas the mesic *B. pinnatum* and *B. rupestre* have short awns and bright green colored leaves. *B. rupestre*, considered until recently a subspecies of *B. pinnatum*, differs from it in its glabrous leaves and spikelets and in leaf epidermal traits (Schippmann 1991; Catalán et al. 2016). Taxonomic uncertainty still persists among some poorly known extra-European *Brachypodium* taxa and within some Eurasian cryptic complex taxa (Schippmann 1991; Catalán & Olmstead 2008; Catalan et al. 2016). Among the less known extra-European taxa, up to 5 different species have been described in America, 11 in Africa and 15 in Asia. Regarding the Eurasian cryptic taxa, they correspond to ploidy complexes of putative diploid parents and their derived allopolyploids, involving different cytotypes of *B. pinnatum* (2x, 4x) and *B. rupestre* (2x, 4x) (Khan & Stace 1999; Wolny & Hasterok 2009; Betekhtin et al. 2014). The intraspecific cytotypes could hardly be differentiated based on morphological traits; however, cytogenetic studies suggest that the allotetrapolyploids derive from interspecific crosses of distinct diploid progenitors. Tetraploid species with 2n=28 chromosomes have been found to be hybrid allopolyploids, potentially derived from diploid 2n=18 (x=9) and 2n=10 (x=5) progenitors (Khan & Stace 1999; Betekhtin et al. 2014). The genus shows a remarkable disploidy, with chromosome base numbers of diploids ranging from the presumably more ancestral x=10 (*B. stacei*), through x= 9 (*B. arbuscula*, *B. sylvaticum*, *B. pinnatum*, *B. rupestre*) and x=8 (*B. glaucovirens*), to x=5 (*B. distachyon*) (Robertson 1981; Betekhtin et al. 2014). Betekhtin et al. (2014) proposed two alternative hypotheses for karyotype evolution in *Brachypodium*, continuous descendant disploidy (x= 10 to x= 9, 8 to x= 5) vs. descendant + ascendant disploidy (x=10 to x=5 to x=9, 8), with allotetraploid 2n=28 species originating always in a later stage.

A recent phylogenetic study provided new insights into the evolutionary history of *Brachypodium* (Diaz-Pérez et al., 2018). All the analyzed plastid and cloned nuclear loci agreed with previous studies (Catalán et al. 2012, 2014) in the more ancestral divergences of the annual diploid *B. stacei*, and in the sister relationship of the annual diploid *B. distachyon* to

the recenmost core perennial clade, which successively split into the *B. arbuscula*, *B. genuense*, *B. sylvaticum*, *B. glaucovirens*, and *B. pinnatum* 2x (2n=18)/*B. rupestre* 2x (2n=18) diploid lineages. The grafting of *Brachypodium* polyploid alleles, inferred from a minimum evolution approach, along the branches of the diploid skeleton species tree, suggested there were six homeologous genomes that could have participated in allopolyploidization events within *Brachypodium*, spanning several levels of phylogenetic depth. The results indicated that the alleles of *B. mexicanum* originated from ancestral and *B. stacei*-type genomes, whereas the *B. hybridum* alleles were strongly associated with *B. stacei*-type and *B. distachyon*-type ancestors. The perennial species *B. boissieri* had alleles strongly related to ancestral and *B. stacei*-type genomes and to the recent core *B. sylvaticum*-type genome. Grafting allelic copies of the remaining polyploid or unknown-ploidy *Brachypodium* species was restricted to the recent stem branch and internal branches of the core perennial clade. The *B. arbuscula*-type, *B. sylvaticum*-type and *B. pinnatum*-type genomes were potentially involved in the origins of seven allopolyploid core perennial species: *B. phoenicoides*, *B. kawakamii*, *B. madagascariense*, *B. retusum*, *B. flexum* and *B. bolusii*. Different allotetraploid *B. pinnatum* and *B. rupestre* cytotypes showed the overall participation of *B. sylvaticum*-type and *B. arbuscula*-type diploid genomes in most of them, though additional sources of genome ancestry associated to *B. glaucovirens*-type and *B. pinnatum*-type diploid genomes were also found in some of them (Diaz-Pérez et al., 2018). New phylogenomic analysis of *Brachypodium* based on transcriptome data have corroborated most of these findings but have also detected additional alternative genomic sources for allohexaploid *B. retusum* (ancestral genome) and for the allotetraploids *B. rupestre* and *B. phoenicoides* (unkown non-core x=5 and core x=9 genomes) (Sancho 2018).

A molecular dating analysis of the *Brachypodium* tree indicated that the *Brachypodium* lineage branched off from its stem node in the Late Eocene (38.8 Ma) and the split of the crown node occurred in the Mid-Miocene (12.6 Ma) (Diaz-Pérez et al. 2018). The long time elapsed between the stem and crown splits would explain the evolutionary and genomic isolation of *Brachypodium* from its closest pooid relatives (Catalan et al. 2016). The dating

analysis also showed successive Late-Miocene and Early-Pliocene divergences for the basalmost currently extant *Brachypodium* genome lineages (*B. stacei*, 6.8 Ma; *B. distachyon*, 5.1 Ma). This was followed by a rapid radiation of the core perennial genome lineages from the end of the Pliocene (2.4 Ma) through the Quaternary, showing the sequential divergence of *B. arbuscula* (1.5 Ma), *B. genuense* (0.7 Ma), *B. sylvaticum* (0.6 Ma), *B. glaucovirens* (0.5 Ma), and *B. rupestre*/*B. pinnatum* lineages (0.3 Ma). According to a coalescence-based Isolation Migration model, the American *B. mexicanum* originated by the hybridization of its two genomes approximately 3.3 Ma and the Mediterranean *B. hybridum* originated from its progenitor genomes in the Quaternary (0.04 Ma). The Mediterranean *B. retusum* and *B. boissieri*, the African *B. flexum* and the eastern-Asian *B. kawakamii* species were inferred to have resulted from the merging of three distinct genomes between 0.03 and 0.07 Ma, whereas the sister eastern Asian *B. sylvaticum* and *B. sylvaticum* var*. pseudodistachyon* diverged from the Eurasian *B. sylvaticum* lineage in the late Quaternary (0.2 Ma) (Diaz-Perez et al. 2018).

A surge of scientific interest in the model grass genus *Brachypodium*

Recently, a group of grasses of the cool season genus *Brachypodium* have emerged as model systems for crop and biofuel grasses (Vogel, 2016). *B. distachyon* (L.) P.Beauv. and its close congeners have been proposed as suitable models for grasses and monocots based on its optimal biological and genomic features and its close phylogenetic relatedness to the temperate cereals (IBI, 2010; Catalán et al., 2014, 2016; Gordon et al., 2016; Scholthof et al., 2018).

The impact of the new model plant *Brachypodium distachyon* on grass genomic research gathered pace since the publication of the full genome sequence of the diploid genotype Bd21 by the International Brachypodium Initiative (IBI 2010). Over the last years, a large number of researchers have investigated the biology, genomics, evolution and ecology of *B. distachyon* and its congeners (Vogel 2016; Scholthof et al. 2018). *Brachypodium*

represents an excellent resource for comparative evolutionary genomic studies, and recent work has identified the important role played by hybridization in the history and ecology of its species (Betekhtin et al. 2014; Diaz-Perez et al., 2018). The small genome sizes, compact genomes (e. g. low levels of repetitive DNA), diverse ecological tolerances, ready propagation under controlled growth conditions, and considerable existing molecular and genomic resources make this genus an excellent candidate for addressing fundamental questions in ecological and comparative genomics. The new taxonomic and phylogenetic findings, and the advent of inexpensive next generation sequencing technology, has set the stage for high definition investigation of the unusual genomic diversity and evolutionary relationships in *Brachypodium* (Catalan et al. 2016).

The demonstration that the model plant was not one but three species (Catalán et al. 2012) opened the way to a thoroughly comparative genomic study of this diploid-polyploid complex. The nuclear and organellar genomes of *B. stacei* and *B. hybridum* have been sequenced and will serve as a model for the origins and consequences of the speciation and polyploidization events that might parallel those of economically important cereals (e. g., wheats; Marcussen et al. 2014). Phylogenomic analysis of the nuclear intraspecific diversity in *B. distachyon* showed a main split of intraspecific lineages characterized by their flowering-time features, whereas comparative genomics of nuclear and plastome data further identified introgressions and chloroplast capture events between the groups, suggesting that flowering time variation is a main factor driving rapid intraspecific divergence in *B. distachyon*, although it is counterbalanced by repeated introgression between previously isolated lineages (Gordon et al. 2017; Sancho et al. 2018). Genomic resources have been also developed for the perennial species *B. sylvaticum*, including its reference genome (*B. sylvaticum* Ain1) and a second resequenced line (*B. sylvaticum* Sin1) (Scholthof et al. 2018), and the sequencing of other perennial genomes is on the way. Comparative genomics of annual vs. perennial species of *Brachypodium* aim to identify the genome donors and the hybridization processes involved in the origin of allopolyploid species (approximately half of the studied taxa) and the switches from perenniality to annuality. Since the various *Brachypodium* species and *B. distachyon*

accessions are native to a wide geographic region with varied climates, these new sequence resources will facilitate genome-wide association mapping of genes controlling tolerance to drought and other abiotic stresses and to phenotypic and biological traits that might have triggered different speciation processes (Catalan et al. 2016).

A need of new studies of Eurasian and Mediterranean *Brachypodium* species

Despite the outstanding advances acquired on the genomic, evolutionary and biological features of the model *Brachypodium* species (Scholthof et al. 2018), the genus is still in lack of deep systematic, evolutionary and ecological studies for some of its species and at infraspecific population level. Taxonomically, its European species are the best studied taxa after the comprehensive works of Schippmann (1991), Catalan et al. (2012) and Lopez-Alvarez et al. (2017). However, few studies have been conducted on the Asian species, except for those connected with some regional Floras, like Siberia (Krylov 1928; Reverdatto 1964; Peshkova 1990) and Malesia (Veldkamp & van Scheindelen, 1989). Even the largely distributed Eurasian species *B. sylvaticum* and *B. pinnatum* have only been extensively studied in the European part of their Holarctic range (Schippmann 1991; Paszko 2007), whereas there is an absence of systematic analyses for their infraspecific taxa and populations distributed in the central, northern and eastern Asian region. 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 *B. distachyon*, *B. stacei* and *B. hybridum* (Catalan et al. 2012, 2014). Cryptic taxa also exist among the perennial *Brachypodium* species, namely among the diploid and tetraploid cytotypes (2n=2x=18; 2n=4x=28) of *B. pinnatum* and *B. rupestre* distributed across Eurasia (Betekhtin et al. 2014; Catalan et al. 2016). However, the cytogenetic hypothesis about the potential origin of these Eurasian tetraploids and of the close western Mediterranean tetraploid *B. phoenicoides* (2n=4x=28) from putative crosses of x=9 perennial diploids and an unknown ancestral x=5 diploid has not yet been confirmed. Noticeably, no taxonomic analysis has been performed yet in this interesting hybrid complex.

Environmental niche modeling analysis has been extensively used to predict the potential distribution ranges of a large number of species and to test the suitability of their environmental niches under different ecological and evolutionary hypotheses. However, detailed ecological studies have been conducted only on the three annual circum-Mediterranean species of *Brachypodium* to date (Lopez-Alvarez et al. 2015; Manzaneda et al. 2012, 2015; Rey et al. 2017; Martinez et al. 2018). An environmental niche modeling analysis of the *B. distachyon* complex indicated that, overall, *B. distachyon* grows in higher, cooler and wetter places, *B. stacei* in lower, warmer and drier places, and *B. hybridum* in places with intermediate ecological features and across latitudinal boundaries but also overlapping with those of its parents (Lopez-Alvarez et al. 2015). Furthermore, paleoenvironmental modeling data supported the Mediterranean basin and adjacent areas as long-term refugia for *B. stacei* and *B. distachyon*, and some of them as potential hybrid zones which could have favored the recurrent origins of *B. hybridum* since the late Pleistocene (López-Álvarez et al. 2015). No other environmental niche modeling studies have been performed with the perennial *Brachypodium* species. It is known that in the Russian territory *B. pinnatum* s. l. is represented by two chromosomal races, but it is not known whether they differ morphologically, ecologically and geographically. Also, the widespread *B. sylvaticum* is considered a rare and threatened species in central Asia but its potential niche breath, extension and temporal dynamic across its large geographic range are totally unknown. Classical morphological, ecological and geographical studies would serve to clarify the systematic positions of these perennial *Brachypodium* taxa and to characterize their niche features and ranges since the last glaciation.

Although the thee annual circum-Mediterranean *Brachypodium* species have been thoroughly studied from many different systematics, evolutionary, functional genomics, ecophysiological and translational aspects (Vogel 2016; Scholthof et al. 2018), very few population-level studies have been conducted on them so far. Deep genomic level analyses have been performed on genotyped (Tyler et al, 2016) or resequenced (Gordon et al. 2017;

Sancho et al. 2018) *B. distachyon* lines, but none of them have explored the genetic and environmental isolation of populations. Only two population-genetic analyses have been performed on the allotetraploid *B. hybridum*, though within very restricted native (Neji et al. 2015) and invaded (Bakker et al. 2009) areas and with different aims. Noticeably, there is a complete lack of population genetic studies on the poorly know *B. stacei* species.

The aim of this PhD thesis is to uncover the existing gap of knowledge on the systematic, ecological and population-genetic characteristic of several annual and perennial Eurasian and Mediterranean *Brachypodium* species. Within this framework, we have investigated the phylogeography and population genetics of the annual *B. stacei*, *B. distachyon* and *B. hybridum* species along one of their most diverse distribution ranges, the Iberian Peninsula hotspot. Our results would contribute to discover new germplasm sources and to disentangle the genetic diversity and structure of their populations, the spatio-temporal scenarios of their origins and dispersals, the relationships of genetic groups with environmental variables, and the potential multiple founder events of the hybrid allopolyploid from distinct progenitor sources.

Our research also comprises the investigation of the perennial Eurasian and Mediterranean species of the *B. pinnatum* complex and of the Eurasian *B. sylvaticum* species, with special emphasis in the less explored Asian territory. We will analyse the phenotypic variation of the diploid and tetraploid cytotypes of *B. pinnatum*, *B. rupestre* and *B. phoenicoides*, studying a large set of macromorphological and microanatomical vegetative and reproductive characters. Our data would allow us to detect significant differences in the phenotypic traits that separate the species and other potential infraspecific groups and cytotypes. A large infraspecific morphoanatomical variation has been observed among populations of each species, suggesting the potential existence of hybrid swarms or new, as yet unknown, cytotypes in this taxonomically complex “compilospecies” group. We have performed environmental niche modeling (ENM) analyses of the perennial *B. sylvaticum*, *B. pinnatum* s. l. and *B. phoenicoides* species. We aim to reconstruct current and

paleoenvironmental models that would indicate if the Atlantic *B. sylvaticum* and the continental *B. pinnatum* species could have had western and northern glacial refugia in Eurasia, and if *B. phoenicoides* could have evolved and survived *in situ* in the western Mediterranean region. We will examine the trends of potential range contractions or expansions of the species' ENMs since the last glaciation to the present. We also aim to determine if *B. sylvaticum* has occupied a stable niche in relict black taiga and broad-leaved forests in Siberia during the last glacial-interglacial phases, and that if its range shifts are related to global change.

References

Bakker EG, Montgomery B, Nguyen T, Eide K, Chang J, Mockler TC, Liston A, Seabloom EW, Borer ET. 2009. Strong population structure characterizes weediness gene evolution in the invasive grass species *Brachypodium distachyon*. *Molecular Ecology* 18: 2588-2601.

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

Betekhtin A, Jenkins G, Hasterok R. 2014. Reconstructing the evolution of *Brachypodium* genomes using Comparative Chromosome Painting. *PLoS ONE* 9(12): e115108.

Bouchenak-Khelladi, Y., Verboom, G.A., Hodkinson, T.R., Salamin, N., Francois§, O., Ní Chonghaile, G., & Savolainen, V. 2009. The origins and diversification of C4 grasses and savanna-adapted ungulates. *Global Change Biology*, 15, 2397–2417.

Catalán P, Shi Y, Amstrong L, Draper J, Stace CA. 1995. Molecular phylogeny of the grass genus *Brachypodium* P. Beauv. based on RFLP and RAPD analysis. *Botanical Journal of the Linnean Society* 177: 263-280.
Catalán P, Kellogg EA, Olmstead RG. 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast *ndh*F gene sequencing. *Molecular Phylogenetics and Evolution* 8: 1-18.

Catalán P, Olmstead RG. 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.

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

Catalán P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LAJ, Pecchioni N, Rasmussen SK, Vogel JP, Voxeur A. 2014. Update on the genomics and basic biology of Brachypodium. *Trends in Plant Science* 19: 414-418.

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

Díaz-Pérez A, López-Álvarez D, Sancho R, Catalan P. 2018. Reconstructing 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.

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

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

Gordon SP, Contreras-Moreira B, Daniel Woods D, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin A, Schackwitz W, Tyler L, Martin J, Lipzen A, Dochy N, Phillips J, Barry K, Geuten K, Juenger TE, Amasino R, Caicedo AL, Goodstein D, Davidson P, Mur L, Figueroa M, Freeling M, Catalan P, Vogel JP. 2017. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with phenotypic variation. *Nature Communications* 19; 8(1): 2184

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

International Brachypodium Initiative (IBI). 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463: 763-768.

Jacques-Félix H. 1962. *Les Graminées d'Afrique Tropicale*. Paris. Institut des Recherches Agronomiques Tropicales et des Cultures Vivrières, 168-178.

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

Khan MA. 1984. Biosystematic studies in *Brachypodium* (Poaceae). PhD thesis. University of Leicester.

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

Krylov, P. N. 1928. Poaceae, pp. 137–385. In Flora of Western Siberia, Vol. 2. Tomsk University, Tomsk, Russia. [In Russian.]

López-Alvarez D, López-Herranz ML, Betekhtin A, 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). PLoS ONE 7(12): e51058.

López-Alvarez D, Manzaneda A J, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, Mur LAJ, Caicedo AL, Hazen SP, 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: 1-16.

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

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. 2012. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytologist* 193: 797-805.

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

Marcussen T., Sandve S.R., Heier L., Spannagl M., Pfeifer M., IWGSC, Jakobsen K.S.,Wulff B.B.H., Steuernagel B., Mayer K.F.X., Olsen O.A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092.

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

Martínez, L.M., Fernández-ocaña, A., Rey, P.J., Salido, T., Amil-ruiz, F., & Manzaneda, A.J. 2018. Variation in functional responses to water stress and differentiation between natural allopolyploid populations in the *Brachypodium distachyon* species complex. *Annals of Botany* 121: 1369–1382.

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

Neji M, Geuna F, Taamalli W, Ibrahim Y, Chiozzottoo R, Abdelly C, Gandour M. 2015. Assessment of genetic diversity and population structure of Tunisian populations of *Brachypodium hybridum* by SSR Markers. *Flora, Morphology, Distribution, Functional Ecology of Plants* 216**:** 42-49.

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

Paszko B. 2007. The differing characteristics of *Brachypodium pinnatum* (L.) P. Beauv. and *B. sylvaticum* (Huds.) P. Beauv. Biodiv. Res. Conserv. 5-8: 11-16.

Peshkova G.A. 1990. Rod Brachypodium (L.) P. Beauv. [The genus Brachypodium (L.) P. Beauv]. Flora Sibiri [The flora of Siberia]. Novosibirsk. Nauka.

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.

Reverdatto, V. V. 1964. Flora of Krasnoyarsk district. Tomsk University Press, Tomsk, Russia. 146 p. [In Russian.]

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

Robertson IH. 1981. Chromosome numbers in *Brachypodium* Beauv. (Gramineae). *Genetica* 56: 55-60.

Saarela, J.M., Wysocki, W.P., Barrett, C.F., Soreng, R.J., Davis, J.I., Clark, L.G., Kelchner, S.A., Pires, J.C., Edger, P.P., Mayfield, D.R., & Duvall, M.R. 2015. Plastid phylogenomics of the cool-season grass subfamily: Clarification of

relationships among early-diverging tribes. *AoB PLANTS* 7: 1–27.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Masood-Quraishi, U., Calcagno, T., Cooke, R., Delseny, M., & Feuillet, C. 2008. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *The Plant Cell* 20: 11–24.

Sancho R, Cantalapiedra CP, López-Álvarez D, Gordon SP, Vogel JP, Catalan P, Contreras-Moreira B. 2018. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist* 218: 1631-1644.

Schippmann U. 1991. Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45: 1-250.

Schneider, J., Winterfeld, G., Hoffmann, M.H., Röser, M. 2011. Duthieeae, a new tribe of grasses (Poaceae) identified among the early diverging lineages of subfamily Pooideae: molecular phylogenetics, morphological delineation, cytogenetics, and biogeography. *Systematics and Biodiversity* 9: 27–44.

Scholthof, K.-B.G., Irigoyen, S., Catalán, P., & Mandadi, K.K. 2018. *Brachypodium*: A monocot grass model system for plant biology. *Plant Cell* 30:1673-1694.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.I., & Morrone, O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53, 117–137.

Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Teisher, J.K., Clark, L.G., Barberá, P., Gillespie, L.J., & Zuloaga, F.O. 2017. A worldwide phylogenetic classification of the Poaceae (Gramineae) II: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution*, 55, 259–290.

Stebbins, G.L. 1949. The evolutionary significance of natural and artificial polyploids in the family Gramineae. *Hereditas*, 35, 461–458.

Stebbins, G.L. 1985. Polyploidy, hybridization and the invasion of new habitats. *Annals of the Missouri Botanical Garden*, 72, 824–832.

Steinwand M.A., Young H.A., Bragg J.N., Tobias C.M., Vogel J.P. 2013. *Brachypodium sylvaticum*, a model for perennial grasses: transformation and inbred line development. *PLoS ONE* 8(9): e75180.

Strömberg, C.A.E. 2011. Evolution of Grasses and Grassland Ecosystems. *Annual Review of Earth and Planetary Sciences* 39: 517–544.

Tsvelev NN. 1976. Grasses of the Soviet Union. St. Petersburg. Nauka (in Russian).

Tyler, L., Lee, S.J., Young, N.D., Deiulio, G.A., Benavente, E., Reagon, M., Sysopha, J., Baldini, R.M., Troìa, A., Hazen, S.P., & Caicedo, A.L. (2016) Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. *Plant Genome* 9: 1–55.

Veldkamp JF, Scheindelen, HJ van. 1989. *Australopyrum*, *Brachypodium* and *Elymus* (Gramineae) in Malesia. *Blumea* 34:61-76.

Vogel JP. 2016. The rise of *Brachypodium* as a model system. *Genetics and genomics of Brachypodium. Plant Genetics and Genomics: Crops Models* (ed. by J.P. Vogel), pp. 1–8. Springer.

Watson L, M. J. Dallwitz R. 1992. *The grass genera of the world*. Oxon: Wallingford, CAB International.

Wolny E, Hasterok R. 2009. Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Annals of Botany* 104: 873–881.

OBJECTIVES

The main goal of the PhD thesis is to study Eurasian and Mediterranean species complexes of the grass model genus *Brachypodium* (Poaceae) to establish their taxonomic boundaries and to infer the evolutionary processes involved in the origins, dispersals and ecological adaptations of their populations. The objective was attained through population genetics and phylogeographic studies of the annual species of the *B. distachyon* complex in the wester Mediterranean region, through environmental niche modelling analysis of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* in the Eurasian-Mediterranean region, and through morphoanatomical analysis of the cryptic taxa and cytotypes of the *B. pinnatum* complex in their Old World range.

The main objective includes the following specific objectives:

- To conduct population-genetics and phylogeographic studies of the three annual species of the *Brachypodium distachyon* complex (*B. distachyon*, *B. stacei*, *B. hybridum*) based on SSR markers to reconstruct their past life history in the Iberian Peninsula hotspot.
- To perform phylogenetic analysis of Mediterranean *B. distachyon*, *B. stacei* and *B. hybridum* representatives based on plastid DNA sequence data to infer their genealogies and to date the potential multiple origins of the hybrid allopolyploids.
- To conduct environmental niche modeling analyses of the perennial Eurasian *B. sylvaticum* and *B. pinnatum* species and of the western Mediterranean *B. phoenicoides* species, to build their current and past niche distribution models and to infer their potential range variations with respect to climatic change and their ecological diversifications.
- To perform a taxonomic study on the perennial species of the Brachypodium *pinnatum* complex (*B. pinnatum*, *B. rupestre*, *B. phoenicoides*) to identify morphoanatomical traits that could serve to separate species, infraspecific ranks and diploid-polyploid cytotypes, and to describe them.

CHAPTER 1

Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands

Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands

Valeriia Shiposha[1,2], Pilar Catalán[1,2], Marina Olonova[2] and Isabel Marques[1]

[1] Department of Agriculture and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain
[2] Department of Botany, Institute of Biology, Tomsk State University, Tomsk, Russia

ABSTRACT

Annual Mediterranean species of the genus *Brachypodium* are promising model plants for energy crops since their selfing nature and short-life cycles are an advantage in breeding programs. The false brome, *B. distachyon*, has already been sequenced and new genomic initiatives have triggered the de-novo genome sequencing of its close relatives such as *B. stacei*, a species that was until recently mistaken for *B. distachyon*. However, the success of these initiatives hinges on detailed knowledge about the distribution of genetic variation within and among populations for the effective use of germplasm in a breeding program. Understanding population genetic diversity and genetic structure is also an important prerequisite for designing effective experimental populations for genomic wide studies. However, population genetic data are still limited in *B. stacei*. We therefore selected and amplified 10 nuclear microsatellite markers to depict patterns of population structure and genetic variation among 181 individuals from 19 populations of *B. stacei* occurring in its predominant range, the western Mediterranean area: mainland Iberian Peninsula, continental Balearic Islands and oceanic Canary Islands. Our genetic results support the occurrence of a predominant selfing system with extremely high levels of homozygosity across the analyzed populations. Despite the low level of genetic variation found, two different genetic clusters were retrieved, one clustering all SE Iberian mainland populations and the island of Minorca and another one grouping all S Iberian mainland populations, the Canary Islands and all Majorcan populations except one that clustered with the former group. These results, together with a high sharing of alleles (89%) suggest different colonization routes from the mainland Iberian Peninsula into the islands. A recent colonization scenario could explain the relatively low levels of genetic diversity and low number of alleles found in the Canary Islands populations while older colonization events are hypothesized to explain the high genetic diversity values found in the Majorcan populations. Our study provides widely applicable information about geographical patterns of genetic variation in *B. stacei*. Among others,

Submitted 11 April 2016
Accepted 4 August 2016
Published 8 September 2016

Corresponding author
Isabel Marques, isabel.ic@gmail.com

Academic editor
Levi Yant

Additional Information and Declarations can be found on page 18

DOI 10.7717/peerj.2407

How to cite this article Shiposha et al. (2016), Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands. PeerJ 4:e2407; DOI 10.7717/peerj.2407

the genetic pattern and the existence of local alleles will need to be adequately reflected in the germplasm collection of *B. stacei* for efficient genome wide association studies.

Subjects Evolutionary Studies, Genetics, Plant Science
Keywords Annual model grass species, *Brachypodium stacei*, SSRs, Genetic diversity and structure, Balearic (Gymnesic) and Canarian islands, Isolation, Western Mediterranean

INTRODUCTION

Approximately one third of Earth's land is covered by grass-dominated ecosystems comprising 600 genera and more than 12,000 species (*Soreng et al., 2015*). Besides their important ecological role, grasses are the core of human nutrition and several genomic efforts have focused on economically important species (e.g., rice: *International Rice Genome Sequencing Project (2005)*; sorghum: *Paterson et al. (2009)*). Among grasses, the genus *Brachypodium*, a member of the Pooideae subfamily, has recently been developed as a new model system to study the evolution of grasses. The genome of the annual *B. distachyon*, commonly known as the false brome, has already been sequenced (*International Brachypodium Initiative, 2010*). This species has several features suitable for the development of a model plant for genomic studies such as a small diploid genome (~355 Mbp), a short annual life-cycle, easily amenable to culture, and a selfing nature (*Gordon et al., 2014*).

The taxonomic identity of *B. distachyon* was recently challenged with the recognition that the three cytotypes attributed to different ploidy levels in this species (e.g., an autopolyploid series of individuals with x = 5 and 2n = 10 (2x), 20 (4x), 30 (6x) chromosomes; *Robertson, 1981*) were in fact three different species: two diploids, each with a different chromosome base number, *B. distachyon* (x = 5, 2n = 10) and *B. stacei* (x = 10, 2n = 20), and their derived allotetraploid *B. hybridum* (x = 5 + 10, 2n = 30) (*Catalán et al., 2012*; *López-Alvarez et al., 2012*). This recent taxonomic split has triggered new genomic initiatives including the re-sequencing of 56 new accessions of *B. distachyon* and the de-novo genome sequencing of *B. stacei* and *B. hybridum*, a project undertaken by the Joint Genome Institute and the International *Brachypodium* Consortium (http://jgi.doe.gov/our-science/science-programs/plant-genomics/brachypodium/). The forthcoming genomes of *B. stacei* and *B. hybridum* will allow the development of several functional genomic analyses on these diploid and polyploid species and their potential transfer to other cereals and forage or biofuel crops. A recent update on phenotypic traits and habitat preferences of the three species has increased the number of discriminant features that distinguish them and has thrown new insights into their respective ecological adaptations (*Catalán et al., 2016a*). However, very scarce genetic information exists for these close relatives of *B. distachyon*, especially for the rarest species of this complex, *B. stacei* (*Catalán et al., 2016b*). It would, therefore, be invaluable to have more information especially because a collection of germplasm reflecting the natural diversity of *B. stacei* is necessary for future genome wide association studies and the creation of reference lines.

Brachypodium stacei is a monophyletic annual diploid species that diverged first from the common *Brachypodium* ancestor, followed consecutively by *B. mexicanum*, *B. distachyon* and the clade of the core perennial taxa (*Catalán et al., 2012*; *Catalán et al., 2016b*). Several studies have revealed it to be distinct from *B. distachyon* and *B. hybridum*: e.g., protein data: *Hammami et al. (2011)*; nuclear SSRs: *Giraldo et al. (2012)*; DNA barcoding: *López-Alvarez et al. (2012)*; isozymes: *Jaaska (2014)*. A recent study using environmental niche models predicted a potential distribution of *B. stacei* in coastal and lowland areas of the circum-Mediterranean region (*López-Alvarez et al., 2015*), concurrent with its known geographic distribution (*Catalán et al., 2016a*). However, a large number of those populations occur in the western Mediterranean region and in Macaronesia (*López-Alvarez et al., 2015*; *Catalán et al., 2016a*). Population genetic studies conducted in its annual congener *B. distachyon* have demonstrated that the genetic structure does not fit a geographic pattern but rather might have resulted from a combination of factors such as long distance dispersal of seeds and flowering time isolation (*Vogel et al., 2009*; *Mur et al., 2011*; *Tyler et al., 2016*).

Here, we studied the patterns of genetic variation in the mainland Iberian Peninsula and the western island populations (continental Balearic Islands and oceanic Canary Islands) of *B. stacei* to unravel the origin and phylogeographic patterns of its populations. From all its range, this area is the best known due to previous studies (*Catalán et al., 2012*; *Catalán et al., 2016a*; *López-Alvarez et al., 2012*; *López-Alvarez et al., 2015*), which can guarantee the correct identification of *B. stacei* since it can be misidentified with its close-relatives (*López-Alvarez et al., 2012*). We specifically addressed the following questions: (1) Is genotypic diversity within populations limited by the prevalence of autogamous pollinations? (2) Do islands (e.g., continental, oceanic) contain less genetic variation than mainland areas? (3) Is there a signature of geographic genetic structure in this self-pollinated plant? Finally, we aim to provide recommendations necessary to establish an efficient germplasm collection of *B. stacei*, with the aim of helping future genomic initiatives in *Brachypodium*.

MATERIAL AND METHODS

Population sampling, DNA extraction and nSSR amplification

A total of 181 individuals were sampled from 19 populations of *B. stacei* covering the whole distribution range of this species within the Iberia Peninsula, plus the continental Balearic (Gymnesic) Islands (Majorca, Minorca) and the oceanic Canary Islands (Gomera, Lanzarote) (Table 1; Fig. 1). Nine populations were sampled in mainland Iberian Peninsula and ten across the two groups of islands (Fig. 1). In each population, ten individuals were collected randomly with a minimum sampling distance of 10 m, with the exception of the Iberian ALI and the Majorcan BANYA populations where only five and six individuals were respectively found. Sampling sizes, locations and geographic coordinates of each population sampled are given in Table 1. Fresh leaves were collected for each individual, dried in silica gel and stored at −20 °C until ready for DNA isolation. The silica samples for all individuals were deposited in the DNA bank of the BioFlora

Table 1 Sampled populations of *Brachypodium stacei* sorted by geographical area. The location, population code, number of plants genotyped (N), mean observed heterozygosity (H_o) and expected heterozygosity (H_e), mean number of alleles (N_a), allelic richness (A_R), inbreeding coefficient (F_{IS}), selfing rate (s), and number of exclusive genotypes (%. between parenthesis) are shown.

Locality	Code	N	Latitude (N)	Longitude (W)	H_o	H_e	N_a	A_R	F_{IS}	s	Exclusive genotypes
Mainland (Iberian Peninsula)											
S Spain: Granada, Moclin	GRA	10	37°19′59″N	3°46′59″W	0.240	0.155	12	1.126	0.667*	0.800	3 (30%)
S Spain: Almeria, Cabo de Gata	ALM	10	36°44′2″N	2°8′35″W	0.170	0.102	11	1.050	0.0001	0.0001	3 (30%)
S Spain: Jaen: Cazorla, Cortijos Nuevos	JAE1	10	38°11′31″N	2°48′14″W	0.120	0.116	12	1.176	0.723*	0.839	4 (40%)
S Spain: Jaen: Quesada, Tiscar	JAE2	10	37°46′5″N	3°1′23″W	0.200	0.100	10	1.000	–	–	1 (10%)
SE Spain: Murcia, Portman	PORT	10	37°34′57″N	0°51′15″W	0.200	0.100	10	1.000	–	–	1 (10%)
SE Spain: Murcia, Calblanque	CALBN	10	37°35′59″N	0°45′29″W	0.140	0.108	14	1.246	0.526*	0.689	4 (40%)
SE Spain: Murcia, Cobaticas	CALBA	10	37°35′59″N	0°45′30″W	0.110	0.105	15	1.339	0.617*	0.763	5 (50%)
SE Spain: Murcia, Cala Reona	CALREL	10	37°36′56″N	0°42′56″W	0.030	0.009	13	1.239	0.520*	0.684	5 (50%)
SE Spain: Alicante, Cabo La Nao	ALI	5	38°45′22″N	0°13′8″E	0.300	0.150	10	1.000	–	–	1 (20%)
Balearic (Gymnesic) Islands											
Spain: Minorca: Es Mercadal, Toro	MEN	10	39°59′6″N	4°6′47″E	0.240	0.173	13	1.203	0.386*	0.556	3 (30%)
Spain: Majorca: Sa Dragonera, Gambes	DRAG	10	39°35′13″N	2°19′37″E	0.111	0.154	16	1.428	0.916*	0.956	5 (50%)
Spain: Majorca: Arta, Peninsula de Llevant	ARTA	10	39°44′10″N	3°20′6″E	0.210	0.128	12	1.126	0.666*	0.799	3 (30%)
Spain: Majorca: Campanet, Coves	CAMPA	10	39°47′31″N	2°58′12″E	0.130	0.138	14	1.434	0.486*	0.654	6 (60%)
Spain: Majorca: Alcudia, Punta Negra	ALCU	10	39°52′48″N	3°10′41″E	0.140	0.108	14	1.200	0.0001	0.0001	2 (20%)
Spain: Majorca: Felenitx, San Salvador	FELEN	10	39°27′4″N	3°11′17″E	0.130	0.109	14	1.200	0.250	0.400	4 (40%)
Spain: Majorca: Petra, Bonany	BONA	10	39°35′38″N	3°5′10″E	0.290	0.391	23	1.992	0.385*	0.5555	9 (90%)
Spain: Majorca: Banyalbufar, Ses Animes	BANYA	6	39°41′6″N	2°30′36″E	0.167	0.239	15	1.496	0.825*	0.904	6 (100%)
Canary Islands											
Spain: Gomera: Agulo	GOM	10	28°10′59″N	17°10′59″W	0.150	0.118	11	1.076	0.891*	0.942	2 (20%)
Spain: Lanzarote: Teguise	LAN	10	29°4′1″N	13°31′1″W	0.230	0.136	11	1.096	1.000*	1	2 (20%)

Notes:
* F_{IS} values deviating from HWE ($P > 0.05$).

group at the University of Zaragoza in Spain and voucher specimens were deposited in the JACA herbarium (Spain).

Total genomic DNA was extracted from fresh leaf tissue or from silica-dried leaf samples using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The 181 samples used in this study were genotyped at 10 variable nuclear microsatellite markers (nSSRs) developed for *B. distachyon* (*ALB006, ALB022, ALB040, ALB050, ALB086, ALB087, ALB139, ALB165, ALB181* and *ALB311*; Vogel et al., 2009). All those microsatellites were selected because during our preliminary studies they displayed good quality and high transferability success in *B. stacei*. The forward primer of each locus was 5-end labeled with a fluorescent dye. Amplifications were carried out in a final volume of 10 µl containing between 0.1 and 0.2 µl of each 10 m diluted primer (forward and reverse), 5 µl PCR Master Mix (QIAGEN) and 2.5 µl DNA. The polymerase chain reactions (PCR) were carried out on a GeneAmp PCR System 9700 thermocycler with a thermal profile consisting of a 4-min initial denaturation step at 95 °C followed by 35 cycles of 30 s at 95 °C, 30 s at 55 °C and 1 min at 72 °C. A final 72 °C extension step of

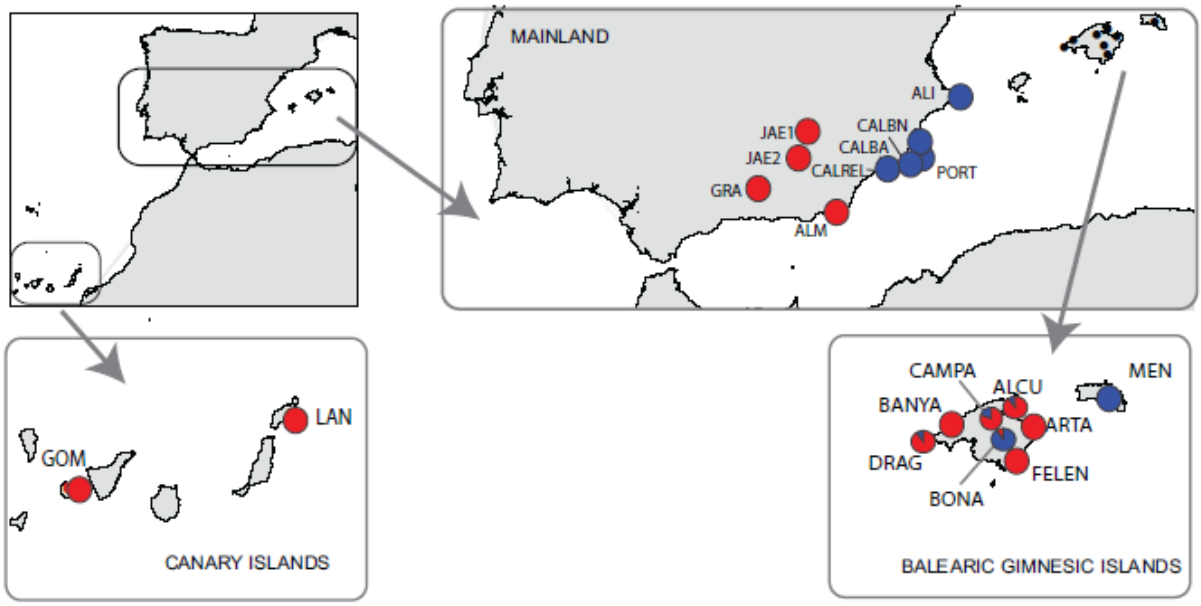


Figure 1 Location of the study area of *Brachypodium stacei*. Collection localities of *Brachypodium stacei* populations in mainland Iberian Peninsula, the continental Balearic (Gymnesic) Islands (Minorca and Majorca) and the oceanic Canary Islands. Pie-charts indicate the proportion of ancestry assigned to individuals of each population by Bayesian clustering analysis using STRUCTURE.

30 min was included to promote non-templated nucleotide addition at the 3′end of the PCR product. Multiplexed PCR products were genotyped on an Applied Biosystems 3130XL Genetic Analyzer using 2 µl of amplified DNA, 12 µl of Hi-Di formamide and 0.4 µl of GeneScan-500 (LIZ) size standard (Applied Biosystem). Allele sizes were determined using Peak Scanner version 1.0 (Life Technologies). Within each population, all loci were checked for the presence of null alleles using MICRO-CHECKER v.2.2.3 (*van Oosterhout et al., 2004*).

Hardy-Weinberg equilibrium, linkage disequilibrium and genetic diversity

Deviation from Hardy-Weinberg Equilibrium (HWE) was tested at both population and locus levels using FSTAT 2.9.3.2 (*Goudet, 2001*). To calculate the extent of linkage disequilibrium between pairs of loci (LD) in each population we set dememorization numbers at 10,000 and performed 100,000 iterations for all permutation tests (exact tests) in Genepop v.4.0.10 (*Raymond & Rousset, 1995*). Significant values were corrected for multiple comparisons by Bonferroni correction (*Rice, 1989*).

For each microsatellite locus and population, genetic polymorphism was assessed by calculating the total number of alleles (N_a, allelic diversity), mean expected heterozygosity (H_e), mean observed heterozygosity (H_o), allelic richness (A_R), and inbreeding coefficient (F_{IS}) using FSTAT 2.9.3.2 (*Goudet, 2001*). The inbreeding coefficient was also estimated using the Bayesian procedure (IIM) implemented in INEst 2.0, which is

robust to the presence of null alleles (*Chybicki & Burczyk, 2009*). Posterior distribution was based on 300,000 steps, sampling every 100 steps and discarding the first 30,000 steps as burn-in. In order to infer the statistical significance of inbreeding we compared the full model (nfb), the model including only the possibility of null alleles and inbreeding (nf), and the model including only null alleles and genotyping failures (nb). The best model was chosen based on the Deviance Information Criterion (DIC; cf. *Chybicki, Oleksa & Burczyk, 2011*).

GenAlEx 6 software was used to estimate the mean expected heterozygosity (H_e) and mean observed heterozygosity (H_o) for each population (*Peakall & Smouse, 2006*). In addition, the selfing rate (s) was also estimated as $s = 2F_{IS}/(1 + F_{IS})$ (*Ritland, 1990*). Spatial patterns of allelic quantity were visualized by mapping variation for the locations across space with the interpolation kriging function in ARCINFO (ESRI, Redlands, CA, USA), using a spherical semivariogram model.

Population genetic structure, genetic differentiation and isolation

The Bayesian program STRUCTURE v.2.3.4 (*Pritchard, Stephens & Donnelly, 2000*) was used to infer the population structure and to assign individual plants to subpopulations. Models with a putative numbers of populations (K) from 1–10, imposing ancestral admixture and correlated allele frequencies priors, were considered. Ten independent runs with 50,000 burn-in steps, followed by run lengths of 300,000 interactions for each K, were computed. The number of true clusters in the data was estimated using STRUCTURE HARVESTER (*Earl & vonHoldt, 2012*), which identifies the optimal K based both on the posterior probability of the data for a given K and the ΔK (*Evanno, Regnaut & Goudet, 2005*). To correctly assess the membership proportions (q values) for clusters identified in STRUCTURE, the results of the replicates at the best fit K were post-processed using CLUMPP 1.1.2 (*Jakobsson & Rosenberg, 2007*). BAPS v.5.2 (*Corander, Marttinen & Mäntyniemi, 2006*) was used to explore population structure further. In contrast to STRUCTURE, BAPS determines optimal partitions for each candidate K-value and merges the results according the log-likelihood values to determine the best K-value. Analyses in BAPS were done at the level group of individuals using the models without spatial information and by selecting 1–10 as possible K-values. Ten repetitions were performed for each K. POPULATION 1.2 (*Langella, 2000*) was used to calculate the Nei's genetic distance (D_A; *Nei, Tajima & Tateno, 1983*) among individuals and to construct an unrooted neighbor-joining tree with 1,000 bootstrap replicates. Nei's genetic distance among individuals was also visualized by Principal Components Analysis (PCoA) with GenAlEx6 (*Peakall & Smouse, 2006*).

We estimated genetic differentiation among locations using an analysis of molecular variance (AMOVA) with ARLEQUIN 3.11 (*Excoffier, Laval & Schneider, 2005*). In addition, molecular variance was also studied (1) between the genetic groups retrieved by STRUCTURE and BAPS, (2) between mainland and island populations, (3) within mainland populations, e.g., S Spain vs. SE Spain, and (4) within island populations, e.g., Balearic vs. Canary Islands. In each analysis, variance was quantified among groups, among locations within groups and within sampling locations. Each AMOVA was run

with 10,000 permutations at 0.95 significance levels. Relationships between genetic and linear geographic distances (isolation-by-distance, IBD) were examined using a Mantel test (*Mantel, 1967*) implemented in ARLEQUIN 3.11 (*Excoffier, Laval & Schneider, 2005*) with 10,000 permutations.

RESULTS

Hardy-Weinberg disequilibrium, non linkage disequilibrium

Deviations from HWE were common in the selfed *B. stacei*. From the 19 populations sampled, only five were at HWE (GRA, MEN, ARTA, FELEN, GOM) at the 5% level after the sequential Bonferroni correction (Table 2). Pairwise comparisons between loci revealed no significant linkage disequilibrium at the $P = 5\%$ suggesting that alleles are assorting independently at different loci.

Genetic diversity and mating system of *Brachypodium stacei*

For each locus, observed heterozygosity values ranged from 0 to 0.058 (respectively for loci *ALB139* and *ALB086*), and expected heterozygosity ranged from 0 to 0.145 (respectively for loci *ALB139* and *ALB087*). F_{IS} values varied between –0.068 and 0.8482 (respectively for loci *ALB311* and *ALB022*; Table 3) across the loci studied. No null alleles were detected. The results from the Bayesian analyses implemented in INEst revealed that only inbreeding contributed to the excessive homozogosity, since this model (DIC_{nf}: 3,300.019) was preferred over the alternative ones (DIC_{nfb}: 4,400.390; DIC_{nb}: 4,400.300) based on the DIC criterion.

From the 19 sampled populations of *B. stacei*, only 37 distinct alleles were found in the 181 individuals studied (Fig. 2; Table S1). Most of the alleles (27 alleles; 73%) were shared between populations while the remaining ones were private to mainland, Majorca, Minorca or the Canary Islands (10 alleles; 27%). Most of the alleles found in the islands were also found in the mainland since only three alleles out of 27 (11%) were not found in the mainland: two alleles were shared between Majorca and Minorca and one allele was shared between Majorca and the Canary Islands (Fig. 2). Out of 37, four alleles were exclusively found only in the mainland (10%; three in SE Spain and one in S Spain), six in Majorca (16%) and one in Minorca (2%) while the Canary Islands had no unique alleles (Fig. 2).

The number of alleles generally increased in the Balearic Islands, most specially in Majorca ($P < 0.0001$) as shown when projected into the geographic space (Fig. 3). Overall, only 38% (69 out of 181) of all genotyped samples exhibited unique multi-locus genotypes, as a consequence of the rampant homozygosis (fixed alleles) observed for most loci in most populations. The observed percentage is lower than one might expect under random mating, where the frequency of multilocus genotypes is expected to be equal to the product of the allelic frequencies. However, a relatively high number of unique multi-locus genotypes were generally found in the populations collected in the island of Majorca, where up to 100% of all the individuals sampled showed unique multi-locus genotypes (Table 1).

Mean observed heterozygosity among the populations of *B. stacei* varied between 0.110 (mainland population CALBA) and 0.290 (Majorcan population BONA) with a

Table 2 Results of the Hardy Weinberg exact tests retrieved by GENEPOP for 19 populations of *Brachypodium stacei*. *P*-value (0.05) associated with the null hypothesis of random union of gametes (or '–' if no data were available, or only one allele was present) estimated with a Markov chain algorithm and the standard error (S.E.) of this estimate.

Population	*P*-value	S.E.
GRA	0.0519	0.0011
ALM	–	
JAE1	0.0259	0.0007
JAE2	–	
PORT	–	
CALBN	0.0077	0.0009
CALBA	0.0007	0.0001
CALREL	0.0249	0.0006
ALI	–	
MEN	0.1016	0.0015
DRAG	0	0
ARTA	0.053	0.0012
CAMPA	0.0361	0.0014
ALCU	–	
FELEN	0.0508	0.0028
BONA	0	0
BANYA	0	0
GOM	0.0515	0.0012
LAN	0.0096	0.0005

Table 3 Characteristics and genetic diversity statistics of the nuclear microsatellite markers used in the genetic study of *Brachypodium stacei*. For each locus, the total number of alleles (N_a), mean expected heterozygosity (H_e), mean observed heterozygosity (H_o), and the fixation index (F_{IS}) obtained from the 181 studied individuals are shown.

Locus	Repeat motif	N_a	H_e	H_o	F_{IS}
ALB006	(GT)15	2	0.016	0.016	0.003
ALB022	(CT)11	2	0.035	0.005	0.848
ALB040	(CTT)8	4	0.129	0.047	0.632
ALB050	(GT)15	4	0.122	0.032	0.717
ALB086	(AAG)7	6	0.119	0.058	0.486
ALB087	(AGC)7	6	0.145	0.032	0.758
ALB139	(AGA)7	1	0.000	0.000	0
ALB165	(ATA)12	4	0.066	0.049	0.298
ALB181	(AC)9	5	0.049	0.037	0.253
ALB311	(GA)6	3	0.025	0.026	−0.069

CI of ± 0.03 at the 95% level, while mean expected heterozygosity varied between 0.090 (mainland population CALREL) and 0.239 (Majorcan population BANYA; Table 1) with a CI of ± 0.04 at the 95% level. The average F_{IS} value was 0.558 (CI: 0.141) varying between 0.0001 (mainland population ALM, Majorcan population ALCU) and 1 (Canary

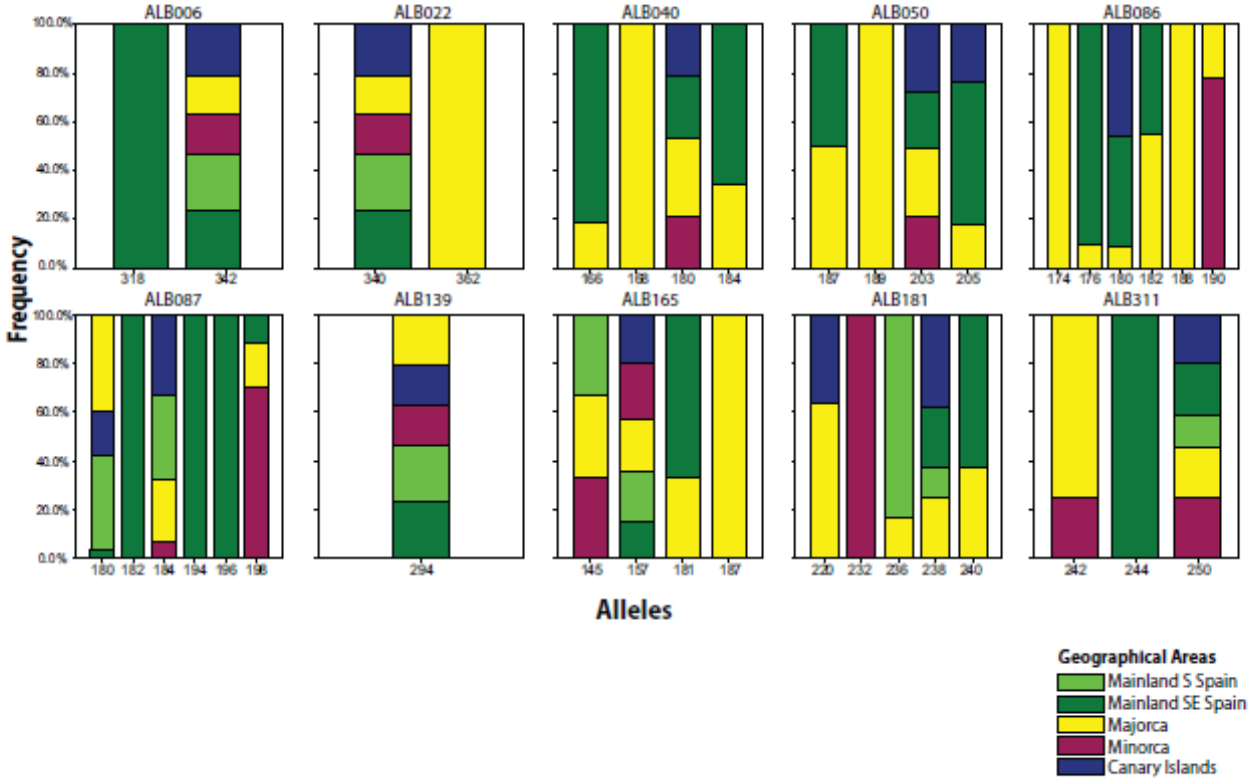


Figure 2 Distribution of *Brachypodium stacei* alleles. Frequency of the alleles found in *Brachypodium stacei* across the geographical area sampled: mainland Iberian Peninsula (SE Spain and S Spain) and the islands of Minorca, Majorca and the Canary Islands. Colors of areas are indicated in the chart.

population of LAN). Therefore, the average rate of self-fertilization in *B. stacei* was estimated to be 71% considering all the populations (Table 1). However, the wide range of F_{IS} values implies that the predicted level of self-fertilization also varies extensively across populations with a CI of ± 0.25 at the 95% level.

Population genetic structure among geographical areas

The Bayesian clustering program STRUCTURE found the highest LnP(D) and ΔK values for $K = 2$ ($P < 0.001$) which differentiated all south (S) Iberian mainland populations from the southeastern (SE) mainland populations. The populations collected in the island of Minorca clustered with the SE mainland populations, whereas samples from the Canary Islands and most of the Majorcan populations were grouped with the S mainland populations, with the exception of the Majorcan population of BONA were most individuals were assigned to the same genetic group found only in the SE mainland populations (Fig. 4A). Some individuals collected in four populations of Majorca showed genetic admixture between the two genetic groups (DRAG, CAMPA, ALCU, BONA; Fig. 4A). These results were also

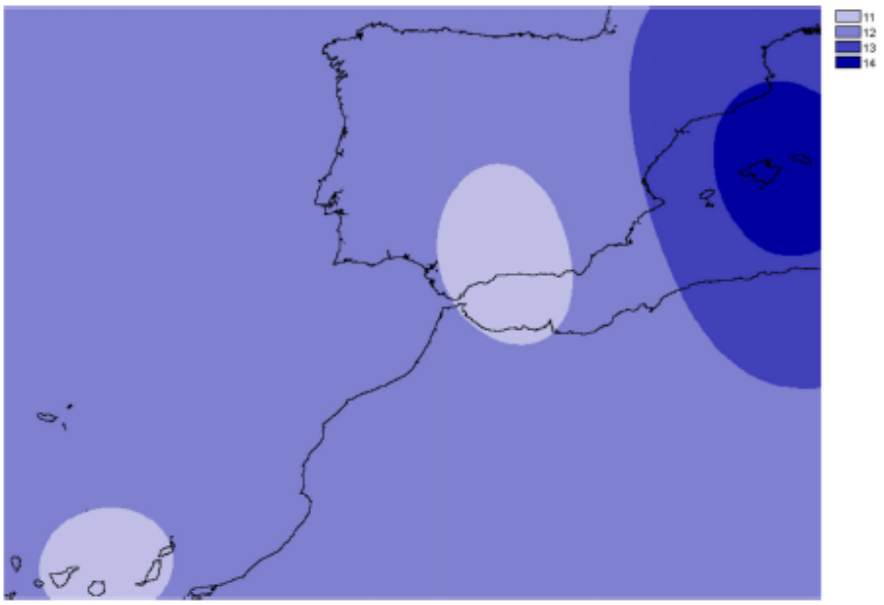


Figure 3 Overall allelic richness of *Brachypodium stacei*. Map of overall allelic richness of *Brachypodium stacei* across the geographic range sampled. Dark areas contain higher richness.

corroborated by BAPS, which retrieved similar results and generally differentiated S mainland populations, Canary Islands and Majorcan populations from all the remaining populations sampled with the exception of the Majorcan population of BONA (Fig. 4B).

The PCoA spatially separated SE Iberian mainland and the Minorca island populations from all remaining populations at both extremes of axis I, which accumulated 44.3% of variance (Fig. 5), partially supporting the genetic boundaries assigned by STRUCTURE and BAPS at $K = 2$. In this two-dimensional plot, the S mainland populations, as well as the SE ones and the Island populations (Minorca, Majorca and the Canary Islands) were well differentiated along the axis 2, which accumulated 26.2% of variance.

The NJ tree separated all SE mainland populations, Minorca and the Majorcan population of BONA from the remaining populations, in a highly supported group (78% bootstrap support (BS) value; Fig. 6A). A similar NJ tree was retrieved when the admixed individuals indicated by STRUCTURE were excluded (Fig. 6B). The remaining sub-divisions found in the NJ trees correspond mainly to the populations sampled although BS values were always very low, with or without admixed individuals (< 43%, Figs. 6A and 6B).

Genetic differentiation and isolation

Overall, genetic differentiation was significantly high (AMOVA FST = 0.748, $P < 0.00001$). The analysis performed over the 19 populations sampled indicated that

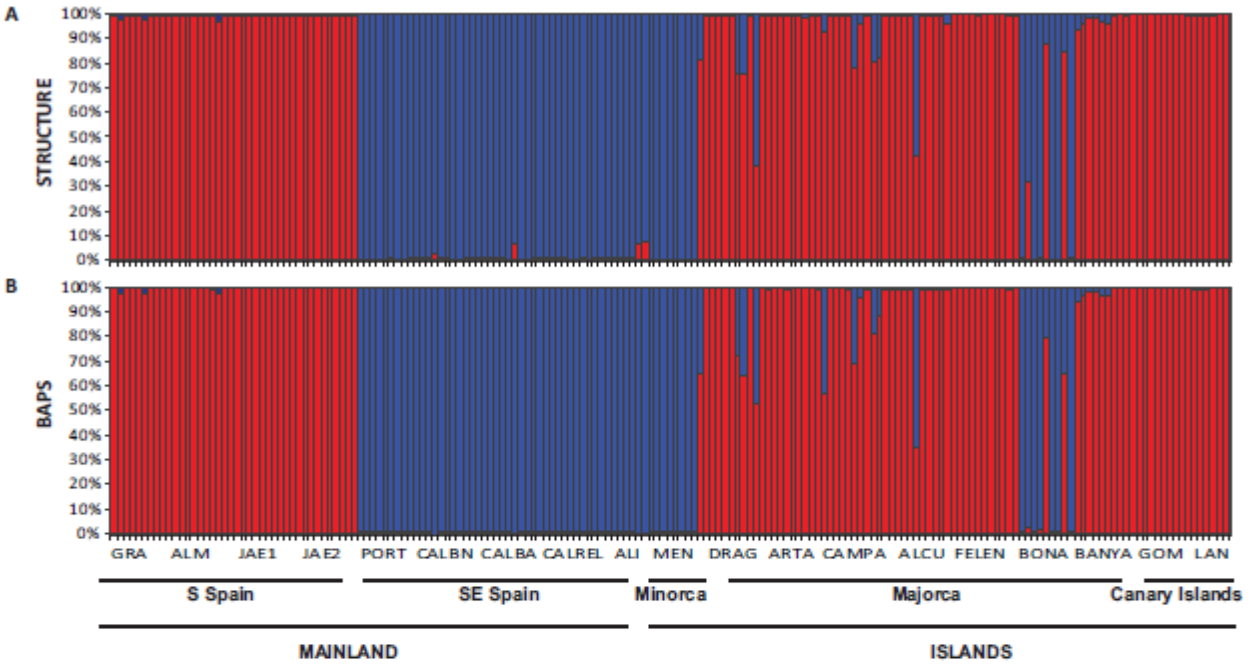


Figure 4 Population structure of *Brachypodium stacei*. Population structure of 181 individuals of *Brachypodium stacei* based on 10 nSSRs and using the best assignment result ($K = 2$) retrieved by STRUCTURE (A) and by BAPS (B) with K from 1 to 10 (replicated 10×) under an admixture model. Each individual is represented by a thin horizontal line divided into K colored segments that represent the individual's estimated membership fractions in K clusters. The different geographic areas are labelled below the graph. Abbreviations of populations follow those indicated in Table 1.

75 and 25% of the genetic variation was attributed to variation among and within populations, respectively ($P < 0.00001$; Table 4). When analyzing the two genetic groups retrieved by STRUCTURE, an independent AMOVA attributed 24, 54 and 22% of the total variation to variation among groups, among populations within groups and within populations (Table 4). Fixation indices of this analysis were FST = 0.779, FSC = 0.710 and FCT = 0.240. To further investigate genetic differentiation between mainland and island populations, an independent AMOVA also attributed the highest percentage of variation among populations within groups (68% of the total variance; FST = 0.758, FSC = 0.737 and FCT = 0.077; Table 4). However, genetic variation was equally partitioned among groups, among populations within groups and within groups when analyzing only island populations (FST = 0.672, FSC = 0.516 and FCT = 0.322), and predominant among groups and among populations within groups when analyzing only mainland populations (FST = 0.884, FSC = 0.783 and FCT = 0.464; Table 4).

The Mantel's test did not detect any significant correlation between the genetic distance $[F_{ST}/(1 - F_{ST})]$ and the geographical distance of the populations studied here ($r^2 = 0.83$, $y = 1.755x - 0.223$, $P = 0.085$).

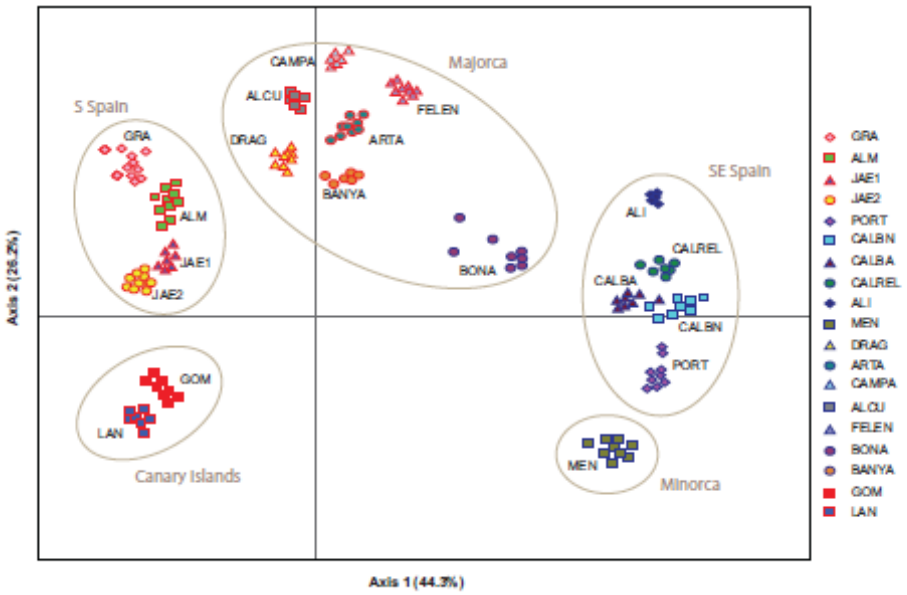


Figure 5 Genetic relationships among *Brachypodium stacei* populations based on Nei's genetic distance. Principal Coordinate analysis (PCoA) samples using the scored nSSRs markers. Percentage of explained variance of each axis is given in parentheses. Population symbols are shown in the chart.

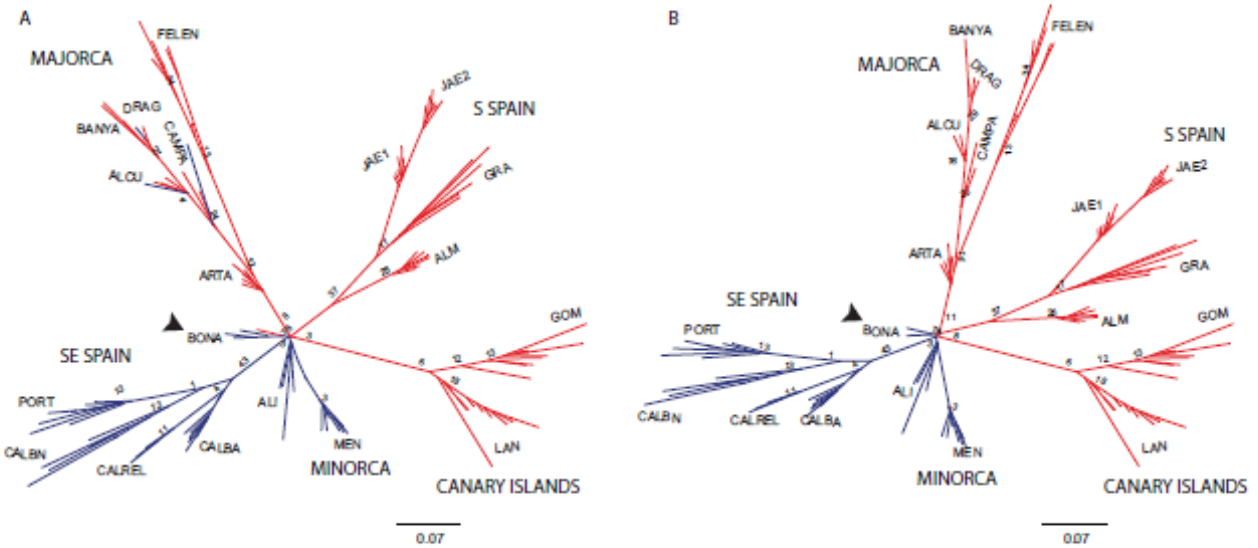


Figure 6 Unrooted neighbor-joining trees of *Brachypodium stacei* populations based on Nei's genetic distance. Unrooted neighbor-joining tree showing relationships among the individuals collected in 19 populations. Numbers associated with branches indicate bootstrap values based on 1,000 replications. Colours followed the ones depicted in Fig. 4 for $K = 2$. Population codes are indicated in Table 1. (A) Genetic relationships among all individuals of *B. stacei*. (B) Genetic relationships without the individuals of *B. stacei* showing admixture in STRUCTURE. Note that the Majorcan population of BONA (arrow) is grouped with SE mainland populations in both NJ trees.

Table 4 Analysis of molecular variance (AMOVA) for 19 populations of *Brachypodium stacei*.

	Source of variance	d.f.	Variance components	% Variance
All populations	Among populations	18	1.011	74.88
	Within populations	343	0.339	25.12
Between genetic groups defined by STRUCTURE and BAPS ($K = 2$)	Among groups	1	0.370	24.05
	Among populations within groups	17	0.831	53.94
	Within populations	343	0.339	22.01
Mainland vs. islands	Among groups	1	0.109	7.77
	Among populations within groups	17	0.954	68.04
	Within populations	343	0.339	24.19
Within mainland populations (S Spain vs. SE Spain)	Among groups	1	0.831	46.41
	Among populations within groups	7	0.752	41.99
	Within populations	161	0.207	11.60
Within island populations (Balearic islands vs. Canary islands)	Among groups	2	0.448	32.26
	Among populations within groups	7	0.486	34.97
	Within populations	182	0.455	32.77

DISCUSSION

Incidence of a highly selfing mating system

From our genetic study, selfing rates of *B. stacei* were estimated as 79% across all populations, even reaching values as high as 95% in some populations (Table 1) although CI values were fairly wide. Even if null alleles have not been detected, the high variation occurring between loci for many of the genetic parameters estimated (Table 3) might influence the reported genetic values. Nonetheless, there was clearly a predominance of homozygous individuals, suggesting that *B. stacei* is primarily a selfing species like its close congener *B. distachyon* (*Draper et al., 2001*; *Gordon et al., 2014*). Although the respective ancestors of these two annual species likely split 16.2 (*B. stacei*) and 10.2 (*B. distachyon*) Mya ago (cf. *Catalán et al., 2016b*), species divergence was not followed by changes in the mating system of *B. distachyon* and *B. stacei*. This analogous mating system is consistent with similarities in floral morphology and floral structure in the two species since they both bear relatively small (mean 7–9 mm) cleistogamous or cleistogamous-type florets having minute (< 0.8 mm) non-exerted anthers (*Catalán et al., 2016a*). Pollination in *B. distachyon* usually occurs in closed flowers leading to extremely high levels of homozygosity (*Vogel et al., 2009*), such as the ones reported here for *B. stacei*. Even more recently diverged species, such as the perennial *B. sylvaticum*, display a predominantly selfing system, although the levels of heterozygosity suggest that this species outcross more often than *B. distachyon* and *B. stacei* (*Steinwand et al., 2013*).

In nature, selfing is thought to be favored due to its inherent transmission advantage, as well as assuring reproduction when pollinators or available mates are scarce (*Marques, Draper & Iriondo, 2014*) and it is expected to evolve whenever these advantages outweigh the costs of inbreeding depression (*Charlesworth & Willis, 2009*).

But contrary to these short-term benefits, selfing might also reduce effective recombination rate leading to frequent genetic bottlenecks (*Goldberg et al., 2010*). Recombination is generally thought to be advantageous because it breaks down associations between alleles (linkage disequilibrium), which might lead to the fixation of deleterious mutations (*Charlesworth & Charlesworth, 2000*). As a result, it has long been argued that the evolutionary potential of highly selfing species is quite limited as a result of reduced genetic diversity and recombination rates (*Lynch, Conery & Burger, 1995*). However, many important crops such as wheat, barley, beans, and tomatoes, are predominantly self-fertilizing species despite the possibility of linkage drag (*Morrell et al., 2005*). Likewise, linkage disequilibrium is absent in *B. stacei* despite being a highly selfing species. There are several possible explanations. The first is that the relatively low levels of linkage disequilibrium results from a recent transition from a strict outcrossing ancestral mating system to a predominantly selfing one, so that the recombination events would still be present (*Lin, Brown & Clegg, 2001*). However, recent phylogenetic studies indicated that *B. stacei* is the earliest extant diverging lineage within the genus and that other early splits also resulted in selfing species (e.g., *B. mexicanum*; *Catalán et al., 2016b*). The second possible explanation is that in a temporal time scale of more than 38 Mya to the common ancestor of the *Brachypodium* stem node (e.g., MRCA of the Brachypodieae/core pooids split; *Catalán et al., 2016b*), even a very low number of outcrossing events might be enough to promote a certain level of recombination. Although plant species might usually mate through selfing, few are strictly selfing (*Igic & Busch, 2013*), creating opportunities for recombination that helps to break down associations between alleles. Large population sizes, which are not uncommon in *B. stacei*, might also reduce linkage disequilibrium. For instance, the near worldwide-distributed *Arabidopsis thaliana* is predominantly a selfing annual species but exhibits a rapid decay in linkage drag in several populations (*Nordborg et al., 2005*).

Origin of island populations

Many plant phylogeographic studies have concluded that genetic diversity erodes across colonization steps, but islands usually exhibit high frequencies of endemism in comparison with large continental areas as a consequence of isolation and habitat diversity (*Kim et al., 1996*; *Sanmartín, van der Mark & Ronquist, 2008*; *Vitales et al., 2014*; see review in *Caujapé-Castells (2011)*). In our study, the genetic structure retrieved by the Bayesian analyses of STRUCTURE or BAPS, or by the results retrieved from the PCoA and the NJ tree suggests a scenario of colonization from the mainland Iberian Peninsula into the islands. Individuals collected in Minorca clustered with SE mainland Iberian populations, whereas individuals from the Canary Islands and most of the Majorcan populations were clustered with S mainland populations (with the exception of BONA which is more related to the SE mainland populations). The large number of alleles (89%) shared between the individuals collected in the Canary Islands and the ones collected in S Spain could support the hypothesis of colonization from the mainland Iberian Peninsula. A recent colonization scenario from a mainland

Iberian source fits well the plausible origin of the oceanic Canary island populations, which show low levels of genetic diversity and multilocus genetic profiles that are a subset of those found in S Spain (Table 1; Fig. 2). Additionally, Canarian populations of *B. stacei* have shown to be phenotypically close to S Spain populations (D. López-Alvarez, P. Catalán, 2016, unpublished data). However, islands could have also been colonized by North African coastal populations of *B. stacei* since ecological niche models predict the existence of conditions suitable for the existence of this species in that area (*López-Alvarez et al., 2015*).

Single vs. multiple colonization scenarios from mainland Iberian sources have been proposed to explain the origins of the Macaronesian plant populations (cf. *Díaz-Pérez et al., 2012*); however, most of them gave rise to new species (*Kim et al., 1996*; *Francisco-Ortega, Jansen & Santos-Guerra, 1996*; *Francisco-Ortega et al., 1997*; *Caujapé-Castells, 2011*). Even if *B. stacei* grows preferentially in relatively stable shady coastal and lowland places along its distribution area (*Catalán et al., 2016a*), seeds of this annual species could be also occasionally dispersed through long distances, as inferred from genetic studies (*López-Alvarez et al., 2012*). The fact that all the studied individuals of the Canarian GOM and LAN populations are morphologically similar to those of the remaining Mediterranean populations (*Catalán et al., 2016a*) indicates that they belong to the same species, suggesting that the introduction of the plant in the Canary isles was probably a very recent one.

Contrastingly, the Balearic populations of *B. stacei* show similar, or even higher genetic diversity values in the case of the Majorcan populations (e.g., BONA, Majorca; Table 1), than the mainland Iberian populations. This scenario could be explained by old colonization events from the mainland followed by insular isolation, which might have favored the appearance and accumulation of new allelic variants and genotypes along time (Fig. 2). Also, admixture after multiple colonization's could have contributed to this scenario, which has been reported to have occurred in other postglacial recolonizations in Europe (e.g., *Lexer et al., 2010*; *Krojerová-Prokešová, Barančeková & Koubek, 2015*). The palaeogeographic configuration of the continental Balearic Islands could have facilitated the migration of coastal SE Spain and S Spain *B. stacei* populations into Minorca and Majorca, and the repeated colonization (and admixture) of the later island from multiple continental sources (Fig. 4). The southern Iberian region together with its eastern Iberian range, the Balearic isles and Provence formed a continuous geological region that split into several microplates during the Oligocene (*Cohen, 1980*). In the late Oligocene (30–28 Ma) the Balearic microplate separated from the eastern proto-Iberian peninsula (*Cohen, 1980*; *Rosenbaum, Lister & Duboz, 2002*) but during the Messinian drought and salinity crisis of the Mediterranean in the late Miocene (c. 6–5 Ma), the Balearic islands formed a single land mass (*Gautier et al., 1994*) and several land bridges re-established the connection with the eastern Iberian Peninsula (*Lalueza-Fox et al., 2005*). Even after the opening of the Gibraltar strait and the refilling of the Mediterranean basin (c. 5 Ma), several land bridges were again created during Middle-Upper Pleistocene that connected the Balearic Gymnesian isles between themselves and between those

islands and the mainland eastern Iberian Peninsula (c. 0.40 Ma; *Cuerda, 1975*), favoring the colonization of the islands from mainland plant populations stocks (*Garnatje et al., 2013*).

Is there a role for ecogeographical isolation in *Brachypodium*?

High values of genetic differentiation and a signature of strong genetic structure were found in *B. stacei*. Though genetic differentiation values obtained for other selfing but more outcrossing species of *Brachypodium* are relatively high (e.g., *B. sylvaticum*: F_{ST} 0.480 ± 0.28 in native Eurasian populations, and 0.446 ± 0.26 in invasive western North American populations; *Rosenthal, Ramakrishnan & Cruzan, 2008*), the high values of genetic differentiation found in *B. stacei* are puzzling. Lower genetic differentiation values than the ones found here usually correspond to different grass species (e.g., *Festuca*, *Díaz-Pérez et al., 2008*). However, all *B. stacei* individuals examined in this study are morphologically similar and correspond to what is considered to be the same species (*Catalán et al., 2016a*). Population turnover is expected to increase genetic differentiation among populations if colonizers are dispersed from different sources (*Pannell & Charlesworth, 2000*). That might only be true for *B. stacei* if wind or other vectors are dispersing seeds across populations, as hypothesized for the also annual and autogamous *B. distachyon* (*Vogel et al., 2009*; *Mur et al., 2011*). That would probably erase the patterns of genetic structure that we have found in STRUCTURE, BAPS, the NJ tree and the PCoA analyses (Figs. 3–5), though the high rates of selfing observed could explain the high levels of genetic differentiation and strong population structure of *B. stacei*, like reported in other primarily selfing plants (*Nybom, 2004*).

Because plants are sessile they experience generations of selection that result in adaptive genetic differentiation to local environmental conditions if there is a strong pressure (*Kremer, Potts & Delzon, 2014*). Although we have no empirical information for *B. stacei*, the distribution of the close relatives, *B. distachyon* and its allopolyploid derivative *B. hybridum*, indicates that they are geographically structured in mesic to arid environments, with *B. distachyon* occurring predominantly in more mesic sites and *B. hybridum* in more aridic sites (*Manzaneda et al., 2012*). *Brachypodium hybridum* is also more efficient in its water usage being significantly more tolerant to drought than *B. distachyon* and behaving as a drought-escapist (*Manzaneda et al., 2015*). Also, environmental niche model analyses indicate a preference of *B. stacei* for warm and arid Mediterranean places (*López-Alvarez et al., 2015*), though its habitat preferences are for shady places, probably as a protection from direct insulation in the aridic environment (*Catalán et al., 2016a*). Therefore, all together, results suggest an important role for ecogeographical differentiation in these lineages of *Brachypodium* (*Manzaneda et al., 2012*; *Manzaneda et al., 2015*; *López-Alvarez et al., 2015*; *Catalán et al., 2016a*; *Catalán et al., 2016b*). More detailed ecological studies are necessary to understand the potential ecological tolerance of *B. stacei* to the arid conditions.

Perspectives towards new genomics initiatives in *B. stacei*

The ongoing de-novo genome sequencing of *B. stacei* led by the Joint Genome Institute and the International *Brachypodium* Consortium (http://jgi.doe.gov/our-science/science-programs/plant-genomics/brachypodium/) will provide significant insights into the mechanisms of polyploid hybrid speciation within the complex *B. distachyon*–*B. stacei*–*B. hybridum*, also allowing comparative studies of genomics and development of functional traits in other crop plants. Biological features, such as a selfing system, a diploid genome and having amenable growing conditions are all advantages for the development of a model system and for genomic resources. All seem to be present in *B. stacei*. Previous studies showed that the species is diploid (*Catalán et al., 2012*) and can easily grow even in laboratory conditions, germinating in less than one week like we have seen in our own laboratory (P. Catalán, 2016, unpublished data).

The results from the present study support the existence of a highly selfing system, which from a practical perspective is an advantage in a model species since it simplifies the process of obtaining pure lines under laboratory conditions (*Gordon et al., 2014*). But plant breeding requires the presence of genetic variability in order to increase the frequencies of favorable alleles and genetic combinations. Populations from SE Spain are genetically different from the ones in S Spain and further differentiation might occur in the islands especially in Majorca and Minorca, where several unique alleles were found. Future studies need to test if population differentiation reflects local adaptation to different environments. Nonetheless, researchers of GWA studies need to be careful to avoid reporting false positive signals (i.e., identifying loci that are not responsible for the variation in the trait), which can be caused by population structure (*Platt, Vilhjálmsson & Nordborg, 2010*; *Brachi, Morris & Borevitz, 2011*). In this sense, several efforts have been raised to address this problem statistically (*Pritchard et al., 2000*; *Price et al., 2006*; *Yu et al., 2006*) and recent GWAS can detect loci that are involved in the natural variation of traits even in highly structure plants like *Arabidopsis* (*Nordborg et al., 2005*).

Thus, to help future genomic initiatives involving *B. stacei* we recommend the following guidelines: (1) a collection of different accessions reflecting different ecological pressures should be generated in order to recover the full genomic diversity of *B. stacei*; (2) the creation of a gene bank collection of these materials constitutes a practical and useful reservoir of genetic variation to avoid uniform cultivars and genetic erosion; (3) collections should be accessible to facilitate the interchange of material useful for breeding and other studies. Finally, there is a lack of information for other areas of the Mediterranean, especially the Eastern Mediterranean–SW Asian area, where *B. stacei* has been also found (*López-Alvarez et al., 2012*; *López-Alvarez et al., 2015*; *Catalán et al., 2016a*). A comprehensive study including populations from other Mediterranean areas is compulsory to fully discover the phylogeographic patterns and genetic diversity of *B. stacei*.

ACKNOWLEDGEMENTS

We thank the Spanish Centro de Recursos Fitogenéticos (CRF-INIA), Consuelo Soler and Antonio Manzaneda for providing us some *B. stacei* seeds, Maria Luisa López-Herranz and Diana López-Alvarez for laboratory and greenhouse assistance, and William Scott for linguistic assistance.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The study has been funded by a Spanish Ministry of Science grant project (CGL2012-39953-C02-01). IM received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement PIOF-GA-2011-301257. VS was funded by a Tomsk State University (TSU, Russia) PhD fellowship. PC and IM were partially funded by a Bioflora grant cofunded by the Spanish Aragon Government and the European Social Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Spanish Ministry of Science grant project: CGL2012-39953-C02-01.
European Union's Seventh Framework Programme: FP7/2007–2013.
REA grant agreement: PIOF-GA-2011-301257.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Valeriia Shiposha performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables.
- Pilar Catalán conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.
- Marina Olonova wrote the paper.
- Isabel Marques performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Data Deposition

The following information was supplied regarding data availability:
The raw data has been supplied as Supplemental Dataset Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.2407#supplemental-information.

REFERENCES

Brachi B, Morris GP, Borevitz JO. 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology* **12(10)**:232 DOI 10.1186/gb-2011-12-10-232.

Catalán P, López-Álvarez D, Bellosta C, Villar L. 2016a. Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei* and *B. hybridum* (Poaceae). *Anales Jardín Botánico Madrid* **73(1)**:e28 DOI 10.3989/ajbm.2428.

Catalán P, López-Álvarez D, Díaz-Pérez A, Sancho R, López-Herranz ML. 2016b. Phylogeny and evolution of the genus *Brachypodium*. In: Vogel J, ed. *Genetics and Genomics of Brachypodium*. New York: Springer, 9–38.

Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, 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 DOI 10.1093/aob/mcr294.

Caujapé-Castells J. 2011. Jesters, red queens, boomerangs and surfers: a molecular outlook on the diversity of the Canarian endemic flora. In: Bramwell D, Caujapé-Castells J, eds. *The Biology of Islands Floras*. Cambridge: Cambridge University Press, 284–234.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biologiaal Sciences* **355(1403)**:1563–1572 DOI 10.1098/rstb.2000.0717.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* **10(11)**:783–796 DOI 10.1038/nrg2664.

Chybicki IJ, Burczyk J. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* **100(1)**:106–113 DOI 10.1093/jhered/esn088.

Chybicki IJ, Oleksa A, Burczyk J. 2011. Increased inbreeding and strong kinship structure in *Taxus bacaata* estimated from both AFLP and SSR data. *Heredity* **107(6)**:589–600 DOI 10.1038/hdy.2011.51.

Cohen CR. 1980. Plate tectonic model for the oligo-miocene evolution of the western Mediterranean. *Tectonophysics* **68(3–4)**:283–311 DOI 10.1016/0040-1951(80)90180-8.

Corander J, Marttinen P, Mäntyniemi S. 2006. A Bayesian method for identification of stock mixtures from molecular marker data. *Fishery Bulletin* **104(4)**:550–558.

Cuerda J. 1975. *Los tiempos cuaternarios en Baleares*. Palma de Mallorca: Instituto de Estúdios Baleáricos.

Díaz-Pérez A, Sequeira MM, Santos-Guerra A, Catalán P. 2008. Multiple colonizations, in-situ speciation, and volcanism-associated stepping-stone dispersals shaped the phylogeography of the Macaronesian red fescues (*Festuca* L. Gramineae). *Systematic Biology* **57(5)**:732–749 DOI 10.1080/10635150802302450.

Díaz-Pérez AJ, Sequeira M, Santos-Guerra A, Catalán P. 2012. Divergence and biogeography of the recently evolved Macaronesian red *Festuca* (Gramineae) species inferred from coalescence-based analyses. *Molecular Ecology* **21(7)**:1702–1726 DOI 10.1111/j.1365-294X.2012.05495.x.

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM. 2001. *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiology* **127(4)**:1539–1555 DOI 10.1104/pp.010196.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4(2)**:359–361 DOI 10.1007/s12686-011-9548-7.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14(8)**:2611–2620 DOI 10.1111/j.1365-294X.2005.02553.x.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**:47–50.

Francisco-Ortega J, Jansen RK, Santos-Guerra A. 1996. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. *Proceedings of the National Academy of Sciences of the United States of America* **93(9)**:4085–4090 DOI 10.1073/pnas.93.9.4085.

Francisco-Ortega J, Santos-Guerra A, Hines A, Jansen RK. 1997. Molecular evidence for a Mediterranean origin of the Macaronesian endemic genus Argyranthemum (Asteraceae). *American Journal of Botany* **84(11)**:1595–1613 DOI 10.2307/2446622.

Garnatje T, Pérez-Collazos E, Pellicer J, Catalán P. 2013. Balearic insular isolation and large continental spread framed the phylogeography of the western Mediterranean *Cheirolophus intybaceus s.l.* (Asteraceae). *Plant Biology* **15(1)**:166–175 DOI 10.1111/j.1438-8677.2012.00632.x.

Gautier F, Caluzon G, Suc J-P, Cravatte J, Violanti D. 1994. Age et durée de la crise de salinité Messinienne. *Comptes rendus de l'Académie des sciences. Série 2. Sciences de la terre et des planètes* **318(8)**:1103–1109.

Giraldo P, Rodríguez-Quijano M, Vázquez JF, Carrillo JM, Benavente E. 2012. Validation of microsatellite markers for cytotype discrimination in the model grass *Brachypodium distachyon*. *Genome* **55(7)**:523–527 DOI 10.1139/g2012-039.

Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igić B. 2010. Species selection maintains self-incompatibility. *Science* **330(6003)**:493–495 DOI 10.1126/science.1194513.

Gordon SP, Priest H, Des Marais DL, Schackwitz W, Figueroa M, Martin J, Bragg JN, Tyler L, Lee C-R, Bryant D, Wang W, Messing J, Manzaneda AJ, Barry K, Garvin DF, Budak H, Tuna M, Mitchell-Olds T, Pfender WF, Juenger TE, Mockler TC, Vogel JP. 2014. Genome diversity in *Brachypodium distachyon*: deep sequencing of highly diverse inbred lines. *The Plant Journal* **79(3)**:361–374 DOI 10.1111/tpj.12569.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). *Available at http://www2.unil.ch/popgen/softwares/fstat.htm.*

Hammami R, Jouve N, Cuadrado A, Soler C, González JM. 2011. Prolamin storage proteins and alloploidy in wild populations of the small grass *Brachypodium distachyon* (L.) P. Beauv. *Plant Systematics and Evolution* **297(1–2)**:99–111 DOI 10.1007/s00606-011-0502-7.

Igic B, Busch JW. 2013. Is self-fertilization an evolutionary dead end? *New Phytologist* **198(2)**:386–397 DOI 10.1111/nph.12182.

International Brachypodium Initiative (IBI). 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463(7282)**:763–768 DOI 10.1038/nature08747.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* **436(7052)**:793–800 DOI 10.1038/nature03895.

Jaaska V. 2014. Isozyme variation and differentiation of morphologically cryptic species in the *Brachypodium distachyon* complex. *Biochemical Systematics and Ecology* **56**:185–190 DOI 10.1016/j.bse.2014.04.017.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23(14)**:1801–1806 DOI 10.1093/bioinformatics/btm233.

Kim SC, Crawford DJ, Francisco-Ortega J, Santos-Guerra A. 1996. A common origin for woody Sonchus and five related genera in the Macaronesian islands: molecular evidence for extensive radiation. *Proceedings of the National Academy of Sciences of the United States of America* **93(15):**7743–7748 DOI 10.1073/pnas.93.15.7743.

Kremer A, Potts BM, Delzon S. 2014. Genetic divergence in forest trees: understanding the consequences of climate change. *Functional Ecology* **28(1):**22–36 DOI 10.1111/1365-2435.12169.

Krojerová-Prokešová J, Barančeková M, Koubek P. 2015. Admixture of eastern and western European red deer lineages as a result of postglacial recolonization of the Czech Republic (Central Europe). *Journal of Heredity* **106(4):**375–385 DOI 10.1093/jhered/esv018.

Lalueza-Fox C, Castresana J, Sampietro L, Marquès-Bonet T, Alcover JA, Bertranpetit J. 2005. Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus*, an extinct endemic Balearic mammal. *BMC Evolutionary Biology* **5(1):**70 DOI 10.1186/1471-2148-5-70.

Langella O. 2000. POPULATIONS 1-2: population genetic software, individuals or population distance, phylogenetic trees. *Available at http://bioinformatics.org/populations/.*

Lexer C, Joseph JA, van Loo M, Barbará T, Heinze B, Bartha D, Castiglione S, Fay MF, Buerkle CA. 2010. Genomic admixture analysis in European Populus spp. reveals unexpected patterns of reproductive isolation and mating. *Genetics* **186(2):**699–712 DOI 10.1534/genetics.110.118828.

Lin J-Z, Brown AHD, Clegg MT. 2001. Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). *Proceedings of the National Academy of Sciences of the United States of America* **98(2):**531–536 DOI 10.1073/pnas.98.2.531.

López-Alvarez D, López-Herranz ML, 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):**e51058 DOI 10.1371/journal.pone.0051058.

López-Alvarez D, Manzaneda AJ, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, Mur LAJ, Caicedo AL, Hazen SP, 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):**1–16.

Lynch M, Conery J, Burger R. 1995. Mutation accumulation and the extinction of small populations. *The American Naturalist* **146(4):**489–518 DOI 10.1086/285812.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27(2 Part 1):**209–220.

Manzaneda AJ, Rey PJ, Anderson JT, Raskin E, Weiss-Lehman C, Mitchell-Olds T. 2015. Natural variation, differentiation, and genetic trade-offs of ecophysiological traits in response to water limitation in *Brachypodium distachyon* and its descendent allotetraploid *B. hybridum* (Poaceae). *Evolution* **69(10):**2689–2704 DOI 10.1111/evo.12776.

Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. 2012. Environmental aridity is associated with cytotype segregation and polyploidy occurrence in *Brachypodium distachyon* (Poaceae). *New Phytologist* **193(3):**797–805 DOI 10.1111/j.1469-8137.2011.03988.x.

Marques I, Draper D, Iriondo JM. 2014. Costs and benefits of the mixed-mating system of *Narcissus serotinus* (Amaryllidaceae) in the conservation of small fragmented populations. *Botany* **92(2):**113–122 DOI 10.1139/cjb-2013-0197.

Morrell PL, Toleno DM, Lundy KE, Clegg MT. 2005. Low levels of linkage disequilibrium in wild barley (*Hordeum* vulgare ssp. *spontaneum*) despite high rates of self-fertilization. *Proceedings of*

the National Academy of Sciences of the United States of America **102(7)**:2442–2447 DOI 10.1073/pnas.0409804102.

Mur LAJ, 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 DOI 10.1111/j.1469-8137.2011.03748.x.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution* **19(2)**:153–170 DOI 10.1007/BF02300753.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R, Jakobsson M, Kim S, Morozov Y, Padhukasahasram B, Plagnol V, Rosenberg NA, Shah C, Wall JD, Wang J, Zhao K, Kalbfleisch T, Schulz V, Kreitman M, Bergelson J. 2005. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology* **3(7)**:e196 DOI 10.1371/journal.pbio.0030196.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* **13(5)**:1143–1155 DOI 10.1111/j.1365-294X.2004.02141.x.

Pannell JR, Charlesworth B. 2000. Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355(1404)**:1851–1864 DOI 10.1098/rstb.2000.0740.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS. 2009. The *Sorghum* bicolor genome and the diversification of grasses. *Nature* **457(7229)**:551–556 DOI 10.1038/nature07723.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6(1)**:288–295 DOI 10.1111/j.1471-8286.2005.01155.x.

Platt A, Vilhjálmsson BJ, Nordborg M. 2010. Conditions under which genome-wide association studies will be positively misleading. *Genetics* **186(3)**:1045–1052 DOI 10.1534/genetics.110.121665.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38(8)**:904–909 DOI 10.1038/ng1847.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155(2)**:945–959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. *American Journal of Human Genetics* **67(1)**:170–181 DOI 10.1086/302959.

Raymond M, Rousset F. 1995. GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86(3)**:248–249.

Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* **43(1)**:223–225 DOI 10.2307/2409177.

Ritland K. 1990. Inferences about inbreeding depression based on changes of the inbreeding coefficient. *Evolution* **44(5)**:1230–1241 DOI 10.2307/2409284.

Robertson IH. 1981. Chromosome numbers in *Brachypodium* Beauv. (Gramineae). *Genetica* **56(1)**:55–60 DOI 10.1007/BF00126930.

Rosenbaum G, Lister GS, Duboz C. 2002. Reconstruction of the tectonic evolution in the western Mediterranean since the Oligocene. *Journal of the Virtual Explorer* **8**:107–130 DOI 10.3809/jvirtex.2002.00053.

Rosenthal DM, Ramakrishnan AP, Cruzan MB. 2008. Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Molecular Ecology* **17(21)**:4657–4669 DOI 10.1111/j.1365-294X.2008.03844.x.

Sanmartín I, van der Mark P, Ronquist F. 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography* **35(3)**:428–449 DOI 10.1111/j.1365-2699.2008.01885.x.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* **53(2)**:117–137 DOI 10.1111/jse.12150.

Steinwand MA, Young HA, Bragg JN, Tobias CM, Vogel JP. 2013. *Brachypodium sylvaticum*, a model for perennial grasses: transformation and inbred line development. *PLoS ONE* **8(9)**:e75180 DOI 10.1371/journal.pone.0075180.

Tyler L, Lee SJ, Young N, Delulio GA, Benavente E, Reagon M, Sysopha J, Baldini RM, Troia A, Hazen SP, Caicedo AL. 2016. Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. *The Plant Genome* **9(2)**: DOI 10.3835/plantgenome2015.08.0074.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4(3)**:535–538 DOI 10.1111/j.1471-8286.2004.00684.x.

Vitales D, Garnatje T, Pellicer J, Vallès J, Santos-Guerra A, Sanmartín I. 2014. The explosive radiation of *Cheirolophus* (Asteraceae, Cardueae) in Macaronesia. *BMC Evolutionary Biology* **14(1)**:118 DOI 10.1186/1471-2148-14-118.

Vogel JP, Tuna M, Budak H, Huo N, Gu Y-Q, Steinwand MA. 2009. Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. *BMC Plant Biology* **9(1)**:88 DOI 10.1186/1471-2229-9-88.

Yu J, Pressoir G, Briggs W, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38(2)**:203–208 DOI 10.1038/ng1702.

CHAPTER 2

Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*

Marques et al. BMC Evolutionary Biology (2017) 17:139
DOI 10.1186/s12862-017-0996-x

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*

Isabel Marques[1*], Valeriia Shiposha[1,2†], Diana López-Alvarez[1,5†], Antonio J. Manzaneda[3], Pilar Hernandez[4], Marina Olonova[2] and Pilar Catalán[1,2]

Abstract

Background: *Brachypodium distachyon* (Poaceae), an annual Mediterranean Aluminum (Al)-sensitive grass, is currently being used as a model species to provide new information on cereals and biofuel crops. The plant has a short life cycle and one of the smallest genomes in the grasses being well suited to experimental manipulation. Its genome has been fully sequenced and several genomic resources are being developed to elucidate key traits and gene functions. A reliable germplasm collection that reflects the natural diversity of this species is therefore needed for all these genomic resources. However, despite being a model plant, we still know very little about its genetic diversity. As a first step to overcome this gap, we used nuclear Simple Sequence Repeats (nSSR) to study the patterns of genetic diversity and population structure of *B. distachyon* in 14 populations sampled across the Iberian Peninsula (Spain), one of its best known areas.

Results: We found very low levels of genetic diversity, allelic number and heterozygosity in *B. distachyon*, congruent with a highly selfing system. Our results indicate the existence of at least three genetic clusters providing additional evidence for the existence of a significant genetic structure in the Iberian Peninsula and supporting this geographical area as an important genetic reservoir. Several hotspots of genetic diversity were detected and populations growing on basic soils were significantly more diverse than those growing in acidic soils. A partial Mantel test confirmed a statistically significant Isolation-By-Distance (IBD) among all studied populations, as well as a statistically significant Isolation-By-Environment (IBE) revealing the presence of environmental-driven isolation as one explanation for the genetic patterns found in the Iberian Peninsula.

Conclusions: The finding of higher genetic diversity in eastern Iberian populations occurring in basic soils suggests that these populations can be better adapted than those occurring in western areas of the Iberian Peninsula where the soils are more acidic and accumulate toxic Al ions. This suggests that the western Iberian acidic soils might prevent the establishment of Al-sensitive *B. distachyon* populations, potentially causing the existence of more genetically depauperated individuals.

Keywords: *Brachypodium*, Environmental isolation, Genetic diversity, Homozygosis, Selfing, Soil pH

* Correspondence: isabelic@gmail.com
† Equal contributors
[1] Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain
Full list of author information is available at the end of the article

Background

During the last decade *Brachypodium distachyon* (L.) P. Beauv. has become one of the most important model systems for functional genomic studies of temperate cereals and forage grasses and for bioenergy crops [1, 2]. The diploid *Brachypodium distachyon* shows a short generation time (annual life cycle), one of the smallest genomes among grass species (272 Mbp in five chromosomes) and it is a highly selfing plant that can easily be grown under controlled conditions [2]. The phylogenetic position of *B. distachyon* reinforces its importance as a model system since it is nested within tribe Brachypodieae (formed by exclusively by the genus *Brachypodium* P. Beauv.), and its sister relationship to the 'core pooids,' a recently evolved lineage of subfamily Pooideae Benth. (Poaceae Barnhart), composed by the four grass tribes that encompass the vast majority of domesticated cool season cereal grain, forage, and turf crops [3, 4].

A high-quality reference genome of *B. distachyon* (based on the diploid inbred line Bd21) is already available [2] and significant investments have been further made in developing and using *Brachypodium* as a model system to learn the genetic mechanisms controlling relevant traits such as cell wall composition, biomass yield, abiotic and biotic stress tolerance, grain development and other features relevant to biomass crop development [3, 5–10]. For instance, candidate genes identified from C_4 grasses that are emerging biomass crops (e.g., maize, sorghum) are being introduced into the temperate C_3-plant *B. distachyon* with the aim of changing its photosynthetic characteristics since the C_4 photosynthetic pathway is generally more efficient under hot and dry conditions [11].

An important key resource essential in any model system is the existence of germplasm collections and inbred lines that reflect traits of interest, as well as its natural genetic variation, which is considered to be the main resource for evolutionary change and for the adaptive potential of a species [3, 12–14]. For instance, in the model plant *Arabidopsis thaliana*, molecular analysis of its natural genetic variation has not only discovered a correlation between the allelic variation of known genes and the phenotypic variation of the species, but has also led to the discovery of novel genes [15]. However, despite all genomic progresses in *B. distachyon* and the fact that it is widely spread across the Mediterranean area [16–19], information about its natural genetic diversity remains scarce. For instance, the first large collection of inbred diploid *B. distachyon* lines was developed from samples collected across the same geographical area (Turkey) but revealed a considerable level of inter-population genetic diversity despite the predominance of homozygous individuals in most populations [20, 21]. A recent study using genotyping-by-sequencing (GBS) of 84 new accessions of *B. distachyon* plus its close relatives (three accessions of *B. stacei* Catalán, Joch. Müll., Mur & Langdon and seven of *B. hybridum* Catalán, Joch. Müll., Hasterok & Jenkins) across its wide circum-Mediterranean native geographic range (e. g., Albania, Armenia, Georgia, Italy, Spain and Turkey) revealed low levels of gene flow, confirming the highly selfing nature of this species and detecting three distinct genetic groups in *B. distachyon* across the Mediterranean populations sampled [22]. Unexpectedly, those genetic groups were not correlated with the geographical origin of the accessions but rather with differences in flowering time, according to the common garden experiment performed [22]. The finding of highly diverged genetic groups is intriguing since individuals clustering to different groups were collected in the same or nearby localities [22]. This would mean that individuals growing in the same locality and under the same environmental conditions could have strong differences in flowering times, creating a barrier to gene flow with their close neighboring individuals [3, 21]. Moreover, it clearly reflects that more studies are needed to understand the natural genetic diversity and genetic structure of *B. distachyon* populations. This information is also crucial to establish efficient germplasm collections and reference lines for the ongoing genomic studies that are being developed by the International *Brachypodium* Initiative (e.g., http://jgi.doe.gov/our-science/science-programs/plant-genomics/brachypodium/brachypodium-t-dna-collection/; http://archive.gramene.org/species/brachypodium/brachypodium_intro.html), especially because there is evidence that annual *Brachypodium* species are ecologically differentiated [12, 23]. The diploid *B. distachyon* usually grows in wet habitats with attenuated summer drought while the allotetraploid *B. hybridum* is generally found in dry habitats with a predictable summer drought period [23]; the allotetraploid is also more efficient in its water use than the close-related diploid *B. distachyon* [12]. Drought-escape strategy (i.e., early flowering) to cope with water stress was found to affect genetic diversity in the studied *B. hybridum* populations but not in *B. distachyon* [13]. However, the potential influence of other environmental factors on the genetic diversity of the annual *Brachypodium* species and populations is still unknown. For instance, soil pH seems to be related to the ecological adaptation of some annual *Brachypodium* populations to acidic substrates [24]. Under acidic conditions (pH < 5.0), the soils can accumulate solubilized Aluminum (Al) ion, mostly as a mononuclear cation (Al^{3+}), which is phytotoxic to most herbaceous plants even at low concentration [24, 25]. *Brachypodium distachyon* is mostly an Al-sensitive plant in contrast to its derived

allotetraploid species *B. hybridum*, which shows both Al-tolerant and Al-sensitive populations [24].

Here, we studied the natural genetic diversity of *B. distachyon* across 14 populations collected in the Iberian Peninsula to create and characterize a future diverse collection of inbred lines, available to *Brachypodium* researchers. This is the best studied Mediterranean area due to several previous works, which allowed us to correctly separate the diploid *B. distachyon* ($2n = 10$) from its close diploid relative *B. stacei* ($2n = 20$) and from their derived allotetraploid *B. hybridum* ($2n = 30$), which were until recently misinterpreted as a single complex species under *B. distachyon* [16–19, 23]. In this study we specifically asked: (1) Is genetic diversity uniform across the Iberian populations of *B. distachyon*? (2) Is homozygosis predominant within populations as expected in a highly selfing plant? (3) How are populations of *B. distachyon* structured genetically? (4) Is there a correlation between genetic diversity and climatic, geographic or other ecological factors such as soil pH?

Methods

Population sampling, DNA extraction and nSSR amplification

A total of 137 individuals were sampled across 14 populations of *B. distachyon* covering the whole distribution range of this species within the Iberian Peninsula (Fig. 1a). In each population, 8–13 individuals were randomly collected with a minimum sampling distance of 10 m. Individual plants from Northeast (NE) Iberian populations LUM, PER, BAN, ARE, YAS, ABI, SAR, CAN and MUR were sampled in the wild, whereas individuals from Northwest (NW) Iberian populations SOB, CMH and ISC and South (S) Iberian populations PLA and GRAZ are first generation seed-germinated selfed plants (S1), each of them obtained from a different wild mother plant. Sampling sizes, locations and geographic coordinates of each population studied are shown in Table 1. Because in some populations *B. distachyon* can morphologically be confounded with the hybrid *B. hybridum*, the identity of the samples was

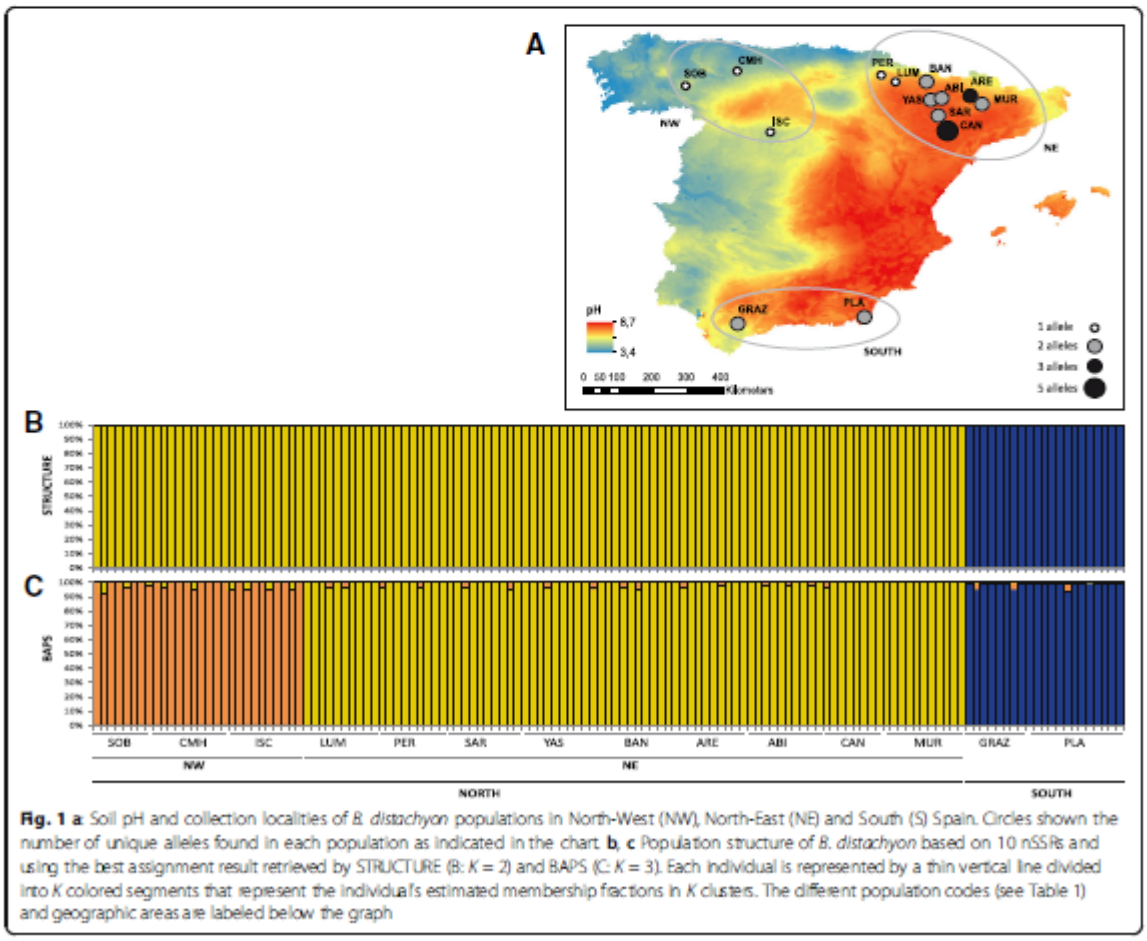


Fig. 1 **a** Soil pH and collection localities of *B. distachyon* populations in North-West (NW), North-East (NE) and South (S) Spain. Circles shown the number of unique alleles found in each population as indicated in the chart. **b**, **c** Population structure of *B. distachyon* based on 10 nSSRs and using the best assignment result retrieved by STRUCTURE (B: $K = 2$) and BAPS (C: $K = 3$). Each individual is represented by a thin vertical line divided into K colored segments that represent the individual's estimated membership fractions in K clusters. The different population codes (see Table 1) and geographic areas are labeled below the graph

Table 1 Sampled populations of *Brachypodium distachyon* sorted by geographical area

Locality	Code	*N*	Latitude (N)	Longitude (W)	Altitude (m.a.s.l)*	H_o	H_e	N_A	A_R	F_{IS}	s	Exclusive genotypes	pH
NW Spain													
Ourense, Sobrado	SOB	9	42.53167	−6.85194	464	0.000	0.000	10	1.000	-	-	1 (11.1%)	5.36
Leon, Campohermoso	CMH	10	42.85472	−5.43472	1074	0.000	0.000	10	1.000	-	-	1 (10.0%)	6.67
Valladolid, Iscar	ISC	8	41.37	−4.53833	828	0.000	0.000	10	1.000	-	-	1 (12.5%)	6.67
NE Spain													
Navarra, Lumbier, Foz	LUM	10	42.63651	−1.30484	434	0.000	0.000	10	1.000	-	-	1 (10.0%)	8.01
Navarra, Puerto del Perdon	PER	10	42.73703	−1.74952	736	0.000	0.000	10	1.000	-	-	1 (10.0%)	8.01
Huesca, Sariñena Laguna	SAR	10	41.78622	−0.18278	292	0.000	0.032	11	1.100	1.000	1.000	2 (20.0%)	8.36
Huesca, Yaso, Sierra de Guara	YAS	10	42.2024	−0.12232	731	0.000	0.032	11	1.100	1.000	1.000	2 (20.0%)	8.36
Huesca, Jaca, Banaguas	BAN	10	42.58063	−0.5799	822	0.000	0.018	11	1.100	1.000	1.000	2 (20.0%)	8.36
Huesca, Aren	ARE	10	42.256944	0.72814	681	0.000	0.060	12	1.200	1.000	1.000	3 (30.0%)	8.36
Huesca, Abizanda, Barranco Mallo	ABI	10	42.195	−0.207	382	0.000	0.018	11	1.100	1.000	1.000	2 (20.0%)	8.36
Zaragoza, Candasnos	CAN	9	41.4651	0.0176	206	0.033	0.104	13	1.300	0.711	0.831	5 (55.5%)	8.27
Lleida, Castillo de Mur	MUR	10	42.09763	0.8775	487	0.000	0.018	11	1.100	1.000	1.000	2 (20.0%)	7.87
S Spain													
Cadiz, Grazalema	GRAZ	8	36.75583	−5.44167	1103	0.000	0.022	11	1.100	1.000*	1.000	2 (25.0%)	8.09
Almeria, Playazo-Rodalquilar	PLA	13	36.8609867	2.0007	11	0.000	0.014	11	1.100	1.000*	1.000	2 (15.4%)	8.59

The location, population code, number of plants genotyped (*N*), latitude, longitude, altitude, soil pH, mean observed heterozygosity (H_o) and expected heterozygosity (H_e), total number of alleles (N_A), mean allelic richness (A_R), inbreeding coefficient (F_{IS}), selfing rate (*s*), and number of exclusive genotypes (%. between parenthesis) are shown. Asterisks indicate F_{IS} values (range 0–1) deviating from Hardy-Weinberg Equilibrium (HWE). Soil pH values were retrieved from [51]

*m.a.s.l: meters above sea level

confirmed through DAPI-staining chromosome counting of the studied materials, coupled with barcoding markers, as indicated in [18]. Fresh leaves were collected for each individual, dried in silica gel and stored at −20 °C until DNA was extracted. Individual samples were stored in the DNA bank of the Bioflora group at the University of Zaragoza in Spain (http://www.bifi.es/bioflora/) and voucher specimens were deposited in the JACA herbarium in Spain (http://herbario.ipe.csic.es/).

Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The 137 samples used in this study were genotyped at 10 polymorphic nuclear simple sequence repeats (nSSRs) previously developed for *B. distachyon* (ALB006, ALB022, ALB040, ALB050, ALB086, ALB087, ALB139, ALB165, ALB181 and ALB311; [21]) and following the procedures outlined in [26]. Based on an initial survey, we selected these ten nSSR markers since they produced robust highly polymorphic amplified bands among the entire collection of our *B. distachyon* samples. Despite being a model system with a variable set of genomic resources, new genomic population methods such as Genotyping-by-Sequencing (GBS, [22]) or restriction site associated DNA sequencing (RAD-Seq; Lopez-Alvarez & Catalan, *unpub. Data*) have only recently started to become available for *B. distachyon*. Although these next-generation sequencing (NGS) techniques will probably be predominant in next years, SSRs still have advantages if they are genetically informative, like previously reported in *Brachypodium distachyon* [21] an in its close annual Mediterranean congeners *B. stacei* [26] and *B. hybridum* [27], as well as in the Eurasian perennial *B. sylvaticum* [28, 29]. Also, the number of biases in a SSR study might be much lower than using NGS methods since each locus can be manually genotyped reducing errors [30].

Amplifications were carried out in a final volume of 10 μl volume containing between 0.1 and 0.2 μl of each 10 M diluted primer (forward and reverse), 5 μl PCR Master Mix (QIAGEN) and 2.5 μl DNA. The polymerase chain reactions (PCR) were carried out in a final volume of 7.5 μl on a GeneAmp PCR System 9700 thermocycler with a thermal profile consisting of a 4-min initial denaturation step at 95 °C followed by 35 cycles of 30 s at 95 °C, 30 s at 55 °C and 1 min at 72 °C. A final 72 °C extension step of 30 min was included to promote non-templated nucleotide addition at the 3'end of the PCR product. Multiplexed PCR products were genotyped on an Applied Biosystems 3130XL Genetic Analyzer using 2 μl of amplified DNA, 12 μl of Hi-Di formamide and 0.4 μl of GeneScan-500 (LIZ) size standard (Applied Biosystem). Allele sizes were determined using Peak

Scanner version 1.0 (Life Technologies) and revised manually. The list of individuals genotyped is shown in Additional file 1: Table S1. Allelic sizes ranged within the expected values obtained for these markers in other genetic studies of *B. distachyon* [21, 26, 31]. Within each population, all loci were checked for the presence of null alleles using MICRO-CHECKER v.2.2.3 [32].

Genetic diversity and selfing

Genetic variation was calculated per locus and population using the following standard genetic indices computed using FSTAT 2.9.3.2 [33]: total number of alleles (Na), allelic richness (A_R), observed within population Nei's heterozygosity (H_o), expected within population Nei's heterozygosity (H_s), expected Nei's heterozygosity within the total population (H_T), Nei's measure of genetic differentiation (G_{st}), and inbreeding coefficient (F_{IS}). F_{IS} was also estimated using the Bayesian procedure implemented in INEst 2.0 [34] that is robust to the presence of null alleles. Posterior distribution was based on 300,000 steps, sampling every 100 steps and discarding the first 30,000 steps as burn-in. In order to understand the importance of inbreeding in our dataset we compared the full model (nfb) with the model including only null alleles (nb). The best model was chosen based on the Deviance Information Criterion (DIC; cf. [35]).

Genetic structure and differentiation

The Bayesian program STRUCTURE v.2.3.4 [36] was used to test whether any discrete genetic structure exists among the populations sampled. The analysis was performed assuming a number of clusters from $K = 1$ to $K = 17$, with 10 repetitions per K. Models were run assuming ancestral admixture and correlated allele frequencies with 50,000 burn-in steps, followed by run lengths of 300,000 interactions for each K. The optimum K was determined using STRUCTURE HARVESTER [37], which identifies the optimal K based both on the posterior probability of the data for a given K and the ΔK [38]. To correctly assess the membership proportions (q values) for clusters identified in STRUCTURE, the results of the replicates at the best-fit K were post-processed using CLUMPP 1.1.2 [39]. BAPS v.5.2 [40] was also used to estimate population structure of *B. distachyon*. In contrast to STRUCTURE, BAPS determines optimal partitions for each candidate K-value and merges the results according the log-likelihood values to determine the best K-value. Analyses in BAPS were done at the level of individuals using the models without spatial information and by selecting 1 to 17 as possible K-values. Ten repetitions were performed for each K. POPULATION 1.2 [41] was used to calculate the Nei's genetic distance (Da; [42]) among individuals and to construct an unrooted neighbor-joining tree with 1000 bootstrap replicates. A Principal Components Analysis (PCoA) was also constructed in GenAlEx6 [43] to detect the genetic relatedness among individuals based on Nei's genetic distance.

Standard and hierarchical analysis of molecular variance (AMOVA) were used to quantify the partitioning of genetic variance within and among the following hierarchal levels: among all populations, between N and S populations, between NE and NW populations, and among NW, NE and S populations. In each analysis, variance was quantified among groups, among locations within groups and within sampling locations. Each AMOVA was run with 10,000 permutations at 0.95 significance levels. The analysis was performed in ARLEQUIN 3.5.1.3 [44].

Pairwise genetic distances between populations were calculated using three metrics in order to cover a range of evolutionary assumptions concerning the relationships between populations [45]. We computed pairwise genetic distances assuming both the infinite alleles model (IAM), e. g., Da distance [42], and the stepwise mutation model (SMM), e. g., Average Square Distance (ASD; [46]), as implemented in POPULATION 1.2 [41], and pairwise linearized FST value distances, e.g., FST/(1 - FST [42] between populations as implemented in GENPOP 3.3 [47]. Isolation by distance (IBD) was assessed through the correlation between the three genetic distance matrices and a matrix of pairwise geographical distances between populations computed with GEOGRAPHIC DISTANCE MATRIX GENERATOR v1.2.3 [48]. Significance was tested with Mantel tests [49] with 1000 permutations using NTSYSpc v. 2.11a [50].

Association between genetic diversity and ecological variables

Using the Pearson correlation coefficient, we tested the degree of association between five genetic diversity parameters (A_R, N_a, H_o, H_e and F_{IS}), and latitude, longitude, altitude plus 19 Bioclimatic variables (Bioclim) previously used in the distribution modeling of annual *Brachypodium* species ([23]; Additional file 2: Table S2) and downloaded from Worldclim- Global Climate Data (http://www.worldclim.org) at a scale of 30 arc-seconds. Soil pH values were retrieved from [51]. A correlation analysis between Bioclim variables was first conducted for avoiding variable redundancy [50]: eight variables were considered correlated (Pearson coefficient: $R \geq 0.95$; Additional file 2: Table S2) and were removed from further analyses. Environmental distances between populations were then estimated to test isolation by environment (IBE) using the 11 Bioclim variables that did not covary significantly. As environmental and geographical distances were significantly correlated ($R = 0.35$, $P = 0.001$), IBE was tested using a partial Mantel test with 10,000 permutations as implemented in R using the '*ecodist*' package [52]. This test allows discriminating unambiguously

between environmental and geographical factors in the correlation structure with genetic variables. Pearson correlation analyses were implemented using the SPSS statistical software package 16.0 (SPSS Inc., Chicago, IL, USA). Holm-Bonferroni corrections were conducted using the R statistical software package (R Development Core Team 2013) to avoid type I error inherent in multiple comparisons [52]. The significance level was $P < 0.05$.

Results

Genetic diversity and selfing

The total number of alleles per locus varied between 14 recorded in 6 of the 10 loci studied (*ALB022, ALB086, ALB087, ALB139, ALB181, ALB311*) and 19 alleles (*ALB050*) (Table 2). Allelic richness per locus varied between 1.983 (*ALB181*) and 4.838 (*ALB050*). Null allele frequencies calculated with INEst were always very low with an average of 0.003 across loci (Table 2) although MICROCHECKER did not detect any null alleles. Nei's observed heterozygosity per locus was recorded as 0 except in locus *ALB024*, which had a value of 0.024, while overall within population Nei's expected heterozygosity varied between 0 and 0.099 (Table 2). However, the expected Nei's heterozygosity within the total population per locus exhibited generally a higher value, reaching an average of 0.507. According to the Nei's measure of genetic differentiation (G_{st}), the estimated divergence of populations per locus varied from 0.819 (*ALB040*) to 1, in 6 of the 10 analyzed loci (Table 2). The inbreeding coefficient F_{IS} had an overall value of 0.922 although it was fixed at 1, in 9 of the 10 loci analyzed (all except locus *ALB040*). Similar values were retrieved when F_{IS} was calculated in INEst though values were higher for locus *ALB040* (Table 2). Results from Bayesian analyses implemented in INEst revealed that only inbreeding contributed to the excessive homozogosity since this model (DICnfb: 2578.593) was preferred over the model that included only null alleles (DICnb: 3683.439) based on the DIC criterion.

Within populations, observed heterozygosity was recorded as 0, in 13 out of the 14 populations analyzed except the NE Iberian population of CAN where it had a value of 0.033, whereas mean expected heterozygosity varied between 0 and 0.104 also in CAN (Table 1). The average number of alleles per population was 11, being a maximum of 13 recorded in CAN, which also showed the maximum value of allelic richness (Table 1). Due to the very low levels of heterozygous individuals found, most populations had a F_{IS} value of 1 (fixed homozygosis) except in the NE Iberian population of CAN where F_{IS} was estimated as 0.711 (Table 1). Therefore, the average rate of self-fertilization estimated for *B. distachyon* was very high, reaching an average of 98% considering all the populations studied (Table 1).

Due to the high level of homozygosis (fixed alleles) observed in most populations, only 27 out of 137 genotyped individuals of *B. distachyon* (19.7%) exhibited a unique multi-locus genotype (Table 1). A relatively high number of unique multi-locus genotypes were found in the NE population of CAN. From the 14 sampled populations of *B. distachyon*, only 33 alleles were found in the 137 individuals studied (Fig. 2). Seven out of the 33 alleles were exclusively found in the southern Iberian populations, 7 only in the NW Iberian populations and 12 only in the NE Iberian populations. Only 7 alleles were shared between geographic regions: 3 between NW and NE Iberian populations, 1 between NW and S Iberian populations, 1 between NE and S Iberian populations, and 2 between the three regions (Fig. 2).

Table 2 Characteristics of the nSSRs markers used in the Iberian populations of *Brachypodium distachyon*

Locus	Repeat motif	Allele size range (bp)	N_a	p_{null}	A_R	H_o	H_s	H_t	G_{st}	F_{IS}	F_{IS}^B
ALB006	(GT)15	360–374	16	0.00399	3.318	0.000	0.051	0.614	0.917	1.000*	1.000
ALB022	(CT)11	354–358	14	0.00464	2.909	0.000	0.000	0.520	1.000	1.000	1.000
ALB040	(CTT)8	176–182	15	0.00416	1.975	0.024	0.031	0.169	0.819	0.221*	0.871
ALB050	(GT)15	217–231	19	0.00311	4.838	0.000	0.072	0.794	0.910	1.000*	1.000
ALB086	(AAG)7	190–198	14	0.00434	2.909	0.000	0.000	0.520	1.000	1.000	1.000
ALB087	(AGC)7	192–202	14	0.00355	2.909	0.000	0.000	0.520	1.000	1.000	1.000
ALB139	(AGA)7	308–310	14	0.00495	1.999	0.000	0.000	0.459	1.000	1.000	1.000
ALB165	(ATA)12	173–201	18	0.00369	4.745	0.000	0.099	0.776	0.873	1.000*	1.000
ALB181	(AC)9	234–238	14	0.00437	1.983	0.000	0.000	0.337	1.000	1.000	1.000
ALB311	(GA)6	244–250	14	0.00152	2.649	0.000	0.000	0.357	1.000	1.000	1.000
Overall	-	-	15	0.00383	3.023	0.002	0.025	0.507	0.952	0.922	0.987

N_a: total number of alleles; p_{null}: average frequency of null alleles across populations; A_R: average allelic richness; H_s: expected within population Nei's heterozygosity; H_o: observed within population Nei's heterozygosity; H_T: expected Nei's heterozygosity within the total population; G_{st}: the Nei's measure of genetic differentiation; F_{IS}: inbreeding coefficient estimated in FSTAT (* indicates values deviating from HWE); F_{IS}^B: inbreeding coefficient estimated using the Bayesian procedure implemented in INEst)

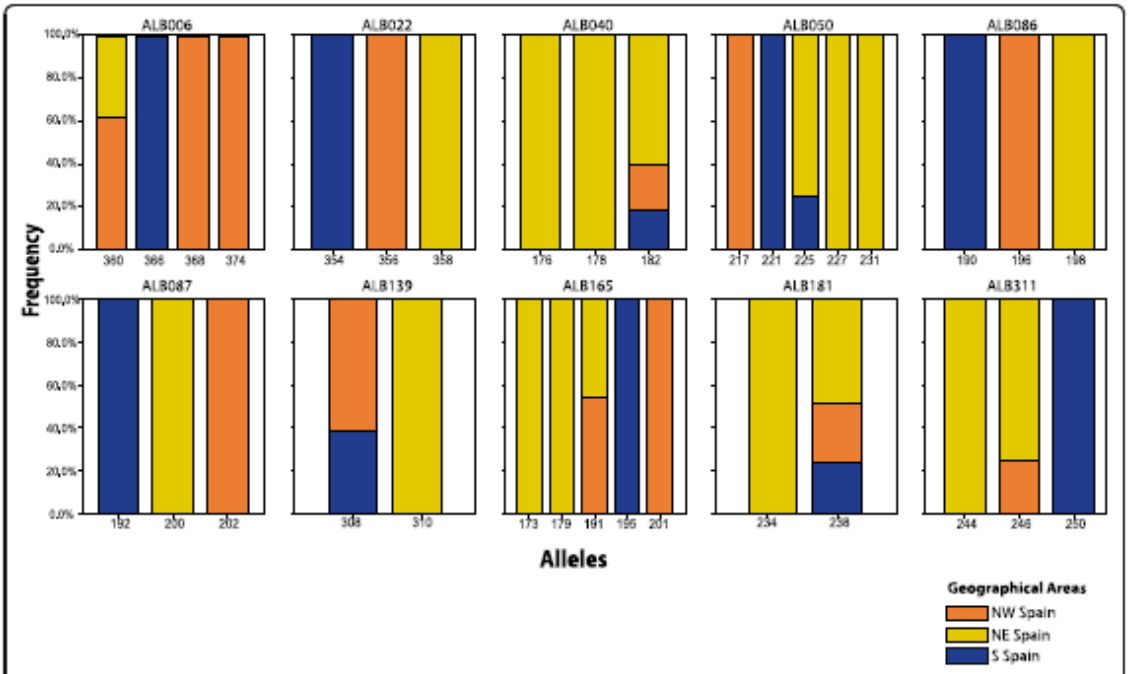


Fig. 2 Frequency of the alleles found in *Brachypodium distachyon* across the three geographical areas sampled: North-West (NW), North-East (NE) and South (S) Spain. The x-axis indicates the allele size (see Additional file 1: Table S1)

Population genetic structure and differentiation

The optimal number of genetic clusters was found to be two by the Bayesian clustering program STRUCTURE that differentiated all North Iberian populations from the South Iberian populations of *B. distachyon* (Fig. 1b; Additional file 3: Fig. S1). This result was partially supported by the Bayesian BAPS analysis that further separated the northern populations into two segregated groups, suggesting an optimal clustering of populations into three genetic groups (NE Iberian, NW Iberian and S Iberian: Fig. 1c). These two programs detected no evidence of genetic admixture between the genetic clusters.

The PCoA spatially separated all populations analyzed into three main groups that clustered NW, NE and S Iberian populations (Fig. 3) being consistent with the genetic structure obtained from BAPS. NE Iberian and NW plus S Iberian populations clustered at both extremes of axis 1, which accumulated 56.1% of variance, while S Iberian populations separated from the NW populations at the negative extreme of axis 2, which accumulated 27.5% of variance (Fig. 3). Due to the high level of fixed alleles within each population, only a very low number of individuals are seen in the PCoA since most individuals bear identical alleles within a population.

The NJ separated all NE Iberian populations, which were grouped in a highly supported clade (94% bootstrap support value, BS), from NW and S populations that clustered in a moderately supported group (65% BS; Fig. 4). Within the last clade, the NW Iberian populations clustered in a group with 64% BS. The remaining sub-divisions found in the NJ tree correspond mainly to the populations sampled although BS values were always very low (<50%, Fig. 4). As mentioned above, only a small number of individuals can be seen in the NJ tree since most individuals within each population share the same alleles.

Genetic differentiation across all 14 populations was significantly high (AMOVA F_{ST} = 0.956, P < 0.00001). Overall, 96.9% and 3.1% of the genetic variation was attributed to variation among and within populations, respectively (Table 4). To further investigate the genetic differentiation between the geographical areas, a hierarchical AMOVA performed between N and S Iberian populations (matching the genetic boundary defined by STRUCTURE) attributed similar percentages of variation among groups (54%) and among populations within groups (43.7%; P < 0.00001, Table 3; F_{ST} = 0.942, F_{SC} = 0.973 and F_{CT} = 0.537). However, it was further exacerbated when performing the hierarchical AMOVA between the NE, NW and S Iberian populations (groups recovered by BAPS), which showed the highest partition of variance among groups (73.9%) and the lowest partition among populations within groups (24%, P < 0.00001,

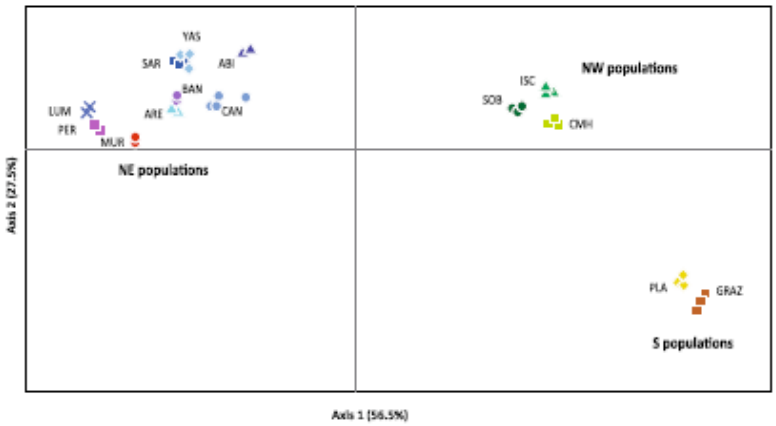


Fig. 3 Principal Coordinate analysis (PCoA) of the studied *Brachypodium distachyon* samples using the scored nSSRs markers. Percentage of explained variance of each axis is given in parentheses. Population symbols and colours are shown in the chart. Population codes are indicated in Table 1. Please note that only a small number of individuals can be seen since most individuals bear identical alleles within each population

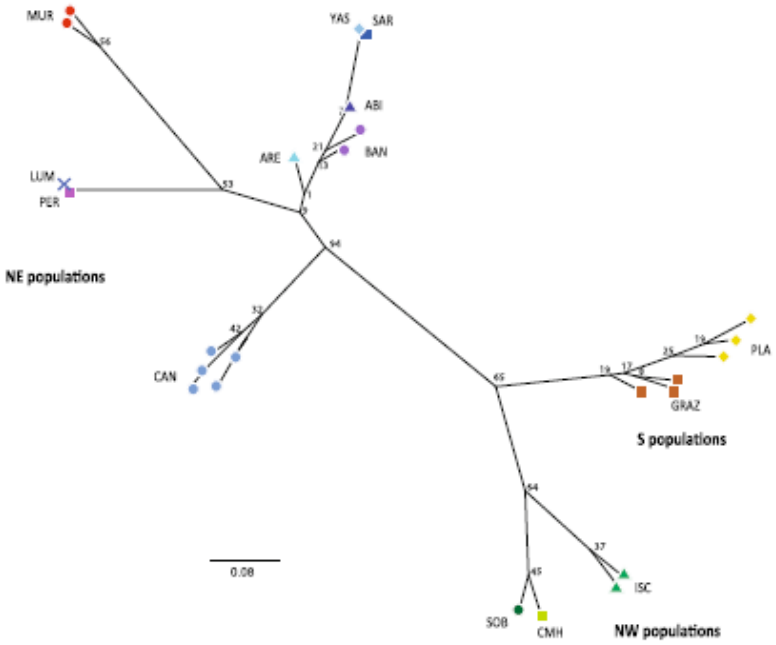


Fig. 4 Unrooted neighbor-joining tree of the studied *Brachypodium distachyon* populations based on Nei's Da genetic distance. Numbers associated with braches indicate bootstrap values based on 1000 replications. Populations symbols and colours followed the ones depicted in Fig. 3. Population codes are indicated in Table 1. Please note that only a small number of individuals can be seen since most individuals bear identical alleles within each population

Table 3 Analysis of molecular variance (AMOVA) for 14 populations of *Brachypodium distachyon*

Source of variance	df.	Variance components	% Variance
All populations			
Among populations	13	2.64	95.92
Within populations	260	0.08	3.08
N vs. S populations			
Among groups	1	2.38	53.76
Among populations within groups	12	1.93	43.57
Within populations	260	0.11	2.67
NE vs. NW populations			
Among groups	1	2.74	70.18
Among populations within groups	10	1.04	26.66
Within populations	220	0.12	3.16
NE vs. NW vs S populations			
Among groups	2	2.99	73.91
Among populations within groups	11	0.97	24.02
Within populations	260	0.08	2.07

Table 4; $F_{IT} = 0.869$, $F_{SC} = 0.921$, $F_{CT} = 0.739$, $F_{IT} = 0.997$). These values were similar to those obtained in a restricted hierarchical AMOVA conducted only with northern Iberian populations (NE vs NW), where 70% and 26.7% of the total variance was distributed among groups and among populations within groups, respectively ($P < 0.00001$, Table 3; $F_{ST} = 0.894$, $F_{SC} = 0.968$ and $F_{CT} = 0.701$).

Correlation between each of the three assayed pairwise genetic distance metrics and pairwise geographical distances revealed significant evidence of IBD between all 14 populations analyzed (DA/geography, $r = 0.843$, $P < 0.001$; ADS/geography, $r = 0.543$, $P < 0.001$; linearized FST values/geography, $r = 0.637$, $P = 0.001$). Genetic distances based on the IAM (Da) showed a clustering of populations more congruent with geography than those based on the SMM (ADS), or the linearized Fst values.

Association between genetic diversity and ecological variables

Two genetic diversity parameters (A_R and N_a) were significantly and negatively associated with the longitude indicating a decrease in genetic diversity towards West (Fig. 1a); no association was found for the remaining genetic diversity indices (Table 4). The level of pH was significantly and positively associated with A_R and N_a since a higher allelic richness and a higher number of alleles were generally found in populations occurring in basic soils (Fig. 1a), located in the East. Of the 11 climatic variables analyzed, two were significantly negatively associated with A_R, Na and F_{IS} (isothermality, BIO3; mean temperature of the driest quarter, BIO9) and one (mean temperature of the wettest quarter, BIO8) was positively associated with those genetic parameters (Table 4; Additional file 4: Fig. S2). Thus, the higher are the values of isothermality and temperature of the driest quarter, the lower is the genetic diversity measured in the populations sampled here; likewise, the higher the value of the temperature in the wettest quarter, the higher is the genetic diversity of the *B. distachyon* populations.

A partial Mantel test confirmed significant IBD among all studied populations ($R = 0.15$, $P < 0.001$), as well as significant IBE ($R = 0.46$, $P < 0.05$). This result indicates that by controlling geographical distance, pairwise differences in soil pH were positively associated to pairwise differences in genetic diversity, which could underlie an adaptive pattern to soil pH and, presumably, to Al sensitiveness. The level of significance in IBE was higher when performing correlation analyses between populations belonging to the NW genetic cluster versus populations from NE and S genetic clusters ($R = 0.56$, $P < 0.001$). No significant IBE was found when only the NE and S genetic cluster were analyzed ($R = 0.38$, $P = 0.482$).

Discussion

Very low heterozygosis across Iberian populations of *Brachypodium distachyon*

Our results indicate that the Spanish populations of *B. distachyon* are characterized by very low levels of genetic

Table 4 Significance differences of the correlation analysis (corrected for multiple comparisons following the Bonferroni procedure) between geographical, soil pH and climatic factors and genetic diversity parameters: mean allelic richness (A_R), mean observed heterozygosity (Ho) and expected heterozygosity (Hs) and inbreeding coefficient (F_{IS})

Parameter	Latitude	Longitude	Altitude	pH	BIO3	BIO8	BIO9
A_R	−0.178	0.558*	−0.353	0.559*	−0.603*	0.699*	−0.560*
H_o	0.002	0.771	−0.346	0.159	−0.240	0.386	−0.225
H_s	−0.050	0.444	−0.313	0.435	−0.451	0.231	−0.278
N_a	−0.178	0.558*	−0.353	0.549*	−0.529*	0.612*	−0.566*
F_{IS}	−0.278	−0.062	0.349	0.023	−0.603*	0.699**	−0.560*

Only soil pH and the three significantly associated biodimatic variables are shown. BIO3: Isothermality; BIO8: Mean Temperature of Wettest Quarter, BIO9: Mean Temperature of Driest Quarter

*$P < 0.05$, **$P < 0.001$

diversity within populations, as a consequence of a high heterozygote deficiency (Table 1; Additional file 1: Table S1). This could be partially a consequence of studying S1 individuals in S and NW Iberian populations; however, similar low genetic diversity and high heterozygote deficiency were observed among wild (non-S1) individuals in some NE Iberian populations (LUM, PER) (Table 1, Fig. 1a). An average of only two multi-locus genotypes was found, exceptionally reaching five in one population (Table 1). The number of alleles per population was also very low since only 33 unique alleles were retrieved among the 137 individuals of *B. distachyon* studied. The extreme low levels of observed heterozygosity ($H_o = 0$ in all populations except one; Table 1) point to high levels of inbreeding ($F_{IS} = 1$ in all populations except one; Table 1) and a strong selfing rate ($s = 1$ in all populations except one; Table 1).

These results are congruent with the highly selfing nature of this species like reported in other studies [3, 10, 21, 22]. Flowers of *B. distachyon* rarely open except under specific environmental conditions (warm, humid and full sun), although even in this case anthers dehisce to the stigmas under the fold of the palea causing primarily self-pollinations [21]. Close-related species, such as the sister species *B. stacei*, are also primarily selfing plants, though genetic diversity values suggest that it might outcross more often than *B. distachyon* [26]. For instance, selfing rates of Iberian, Balearic and Canarian *B. stacei* populations were estimated as 79% [26], which is lower than the ones reported here for Iberian *B. distachyon* populations. Also, the values of heterozygosity detected within populations were slightly higher in the diploid *B. stacei* (e. g., $H_o = 0–0.058$, $H_e = 0–0.145$; [26]) and contrastingly higher in the diploid and predominantly selfing perennial *B. sylvaticum* ($H_o = 0.044–0.438$, $H_e = 0.076–0.592$; [28, 29]) than the ones detected here in *B. distachyon* ($H_o = 0–0.033$, $H_e = 0–0.104$; Table 1).

Primarily selfing plants usually show high genetic differentiation among populations (e. g., [53, 54]). Selfing would explain the high levels of genetic differentiation and the very high fixation index found in *B. distachyon* (averaged $F_{ST} = 0.956$) since it leads to isolation and prevents the efficient flow of genes. Most genetic diversity (96.9%) was observed among populations and only 3.1% of genetic variation within populations (Table 4). Such differences were correlated with geographic distance suggesting the presence of barriers to gene flow between largely distant populations. Selfing indeed inhibits gene flow through pollen and exacerbates genetic differences and genetic structure [55], as found in our analysis. In *B. distachyon*, seed dispersal may also constrain effective gene flow since most seeds land very close to parental plants or are possibly dispersed by ants although within short dispersal distances from the mother plant [56]. The populations studied here are on the edge of the native distribution range of this species that occurs across the Mediterranean - SW Asian region [17–19, 23] but the observed low heterozygosity levels are similar to those found in populations located in other geographical areas, such as Turkey [21], as well as in other Mediterranean areas [22]. Thus, the extreme values of low heterozygosity seem to characterize this model species and it would be invaluable to study other populations to distinguish the influence of selfing from other processes that usually constrain the evolutionary success of populations (e. g., recurrent founder events [57]).

Genetic boundaries in Iberian populations reflect their geographical origin

The results of BAPS (Fig. 1c) and PCoA analyses (Fig. 3), as well as the NJ tree (Fig. 4) and the hierarchical AMOVA with three geographical ranges (Table 3) suggest that the genetic structure of *B. distachyon* in Spain can be grouped (at least) in three clusters congruent with their geographical origin. NE, NW and S Iberian populations all formed separate and homogeneous groups except in the STRUCTURE analysis that clustered all NE and NW populations in one single group (Fig. 1b). Although more populations should be analyzed to verify the existence of further genetic groups in the Iberian Peninsula, our results provide evidence for the existence of a significant genetic structure of *B. distachyon* in Spain like previously suggested by a recent GBS study [22]. This study revealed a significant genetic boundary between NE and S Spanish populations, like the one reported here, although no NW Spanish populations were included in the study [22]. But contrary to the results of the GBS study where the genetic patterns of *B. distachyon* seemed to be primarily explained by differences in flowering time association, our results are better explained by the geographical origin of populations. Although we should keep in mind that PCR-based markers such as the one used here (nSSRs) and GBS techniques might reconstruct similar but slightly different stories [58], several other studies also reported the existence of differentiated genetic clusters in the Iberian geographical areas that we have studied (e. g., *Senecio boissieri* DC.: [59]; *Gentiana alpina* Vill., *Kernera saxatilis* (L.) Rchb. and *Silene rupestris* L.: [60]; *Ferula loscosii* Willk.: [61]; *Cheirolophus intybaceus* (Lam.) Dostál: [62]; *Gentiana lutea* L.: [63]).

Using the same set of nSSRs, we found two main genetic groups in the close relative *B. stacei* both distributed in southern Spain (S and SE Spain groups), from where it colonized the Mediterranean islands of Minorca and Majorca (SE Spanish group) and the oceanic Canary Islands (S Spanish group) apparently through different long distance dispersal (LDD) events [26]. The potential existence of different mechanisms for long distance

dispersal of seeds (not related to geographical distances) was also invoked to explain the unexpected relationships of genetically similar but geographically disjunct *B. distachyon* lineages across the Mediterranean area [3, 18]. Here, the finding of essentially similar low within-population genetic diversities in *B. distachyon*, the low sharing of alleles between geographical regions, and the highly selfing nature of this species support vicariance rather than long-distance dispersal as our preferred explanation for the patterns found in our study.

Ecological adaptations in *Brachypodium distachyon*

Besides historical factors (i.e., demography, glacial refugium), the genetic diversity of *B. distachyon* in the Iberian Peninsula seems to be also shaped by environmental isolation. Isolation by environment, in which genetic differentiation increases with environmental differences, independent of geographic distance, is one of the most important patterns that contribute to genetic divergence in nature [64]. However, a non-zero effect of IBE independent of IBD, like the one reported here has rarely been reported in other studies [65, 66]. Here, we found that the genetic diversity of *B. distachyon* is significantly positively associated with soil pH and the temperature of the wettest quarter, suggesting that the lower these variables are the less would be the genetic variability of the *B. distachyon* in the populations studied. By contrast, the significantly negative association found between the genetic diversity parameters and isothermality, as well as with the temperature of the driest quartersuggests contrary results.

In comparison to other areas, a high number of alleles and even some heterozygosity were found in the NE Iberian populations of *B. distachyon* (Pyrenees, pre-Pyrenees and the Ebro Valley) and in the S Iberian area (Table 1; Fig. 1). Indeed, the Pyrenees and pre-Pyrenees and the Betic ranges probably acted as major glacial refugia in southern Europe where many lineages came into contact [67–70]. These areas experienced several climatic changes given rise to a complex phylogeographic pattern of refugia within refugia (e.g., [69, 71–73]) that might also sustain the diversity of alleles found here. Despite the NE Iberian population were represented by wild individuals that could potentially have higher levels of genetic diversity than the S1 individuals of the S and NW Iberian populations, the high selfing nature of the species makes the sampling effect almost negligible. We should also point that our sampling is limited in the NW and S of the Iberian Peninsula but the geographic area where we are reporting the existence of a higher genetic diversity is congruent with the results found using GBS techniques and a wide sampling of *B. distachyon* throughout the Mediterranean basin [22]. In addition, our study has covered most areas of the Iberian Peninsula where the populations of *B. distachyon* grow and within the Iberian Peninsula our sampling reached novel geographic areas not included in the GBS study [22].

The highest diversity of *B. distachyon* found in the Ebro Valley, in a locality of very low altitude (CAN population), suggests that this area is also an important source of genetic diversity in this species, in accordance with several other phylogeographic studies (e.g., [74–76]). Palynological evidence also supports the Ebro Valley as an important glacial refugium during the last ice age and suggests the existence of a diverse composition of species in these glacial steppes [77–79]. In contrast to the strong topographic feature of the Pyrenees with peaks up to 3404 m a.s.l., and a general Eurosiberian climate (becoming more Mediterranean towards the east), with cold winters and heavily rainfall throughout the year, the Ebro Valley is characterized by a continental Mediterranean climate with low rainfall (300–350 mm/yr), high insolation and evapotranspiration (1000–1500 mm/yr), and the prevalence of strong, dry north-westerly winds where steppe grassland species predominate more often than in the Pyrenees [80]. The number of alleles of *B. distachyon* in the Ebro Valley and its sharing with the populations of the Pyrenees and pre-Pyrenees range suggest that this species could have expanded from this refugium through the NE Spanish Mountains explaining the low number of alleles found in other populations.

The finding of higher allelic richness and higher number of alleles in populations occurring in basic soils suggests that these populations might be better adapted than those occurring in western areas of the Iberian Peninsula were the soils are more acidic (Fig. 1a), and therefore could accumulate Al ions causing toxicity in most plants, including *B. distachyon*, which is mostly an Al-sensitive plant [24]. Thus, this could indicate that western Iberian acidic soils might prevent establishment and expansion of Al-sensitive *B. distachyon* populations, potentially causing the existence of more genetically depauperated individuals. Nevertheless, while low soil pH and the resulting increased Al-induced phytotoxicity could explain the low genetic diversity found here, the data in this study is insufficient to support a causal connection and this hypothesis should be tested experimentally. It is also worth noting that our soil pH levels were not measured during our sampling but rather taken from a publication. Although pH levels might be stable over time, this argues for caution when interpreting the accuracy of the statistical correlations found here.

Progress towards new genomic initiatives in *Brachypodium distachyon* and current limitations

The finding of a higher level of genetic variability and adaptation of *B. distachyon* to basic soils is promising within an agricultural context where tree iron chlorosis

is a problem in some basic soils, which could be alleviated by grass covers, like those of annual *Brachypodium* species [81]. This species and *B. hybridum* can protect the soil from being eroded [81] and are therefore suitable grass cover crops to olive grooves, vineyards and dry fruit croplands [81, 82]. Due to the high degree of homozygosity, obtaining inbreeding lines of *B. distachyon* can be easily done even under laboratory conditions allowing the rapid generation of reference and cultivated lines. A large and diverse germplasm collection of *B. distachyon* has now been assembled and it is freely available for the research community but new genetic studies continue to demonstrate novel unsuspected geographical areas of genetic diversity like the ones reported here. This demonstrates that more population genetic studies are needed to fully uncover the genetic diversity of this species. For instance, our knowledge concerning Central and Eastern Mediterranean and SW Asian populations is still limited (but see [14, 21]) and more studies are necessary to understand the genetic structure of this species across its native Mediterranean distribution. It would also be important to compare the native Mediterranean populations of *B. distachyon* to the ones introduced in other areas (e. g., Australia, J. Borewitz & J. Streitch, *pers. com.*) since although genetic diversity is lower in introduced areas, invasiveness might have also triggered the activation of new allelic variants in this species.

Additional files

Additional file 1: Table S1. Genotypes of the studied Spanish *Brachypodium distachyon* individuals based on nSSR analysis (TXT 12 kb)

Additional file 2: Table S2. Description of the bioclimatic worldclim layers (http://www.worldclim.org/bioclim) used in the correlation analysis (XLSX 10 kb)

Additional file 3: Figure S1. STRUCTURE analysis of *Brachypodium distachyon* in Spain. (A) Mean log probability of data LnP(D) over 10 runs for each K value as a function of K (error bars represent standard deviation). (B) Evanno's ad hoc statistic; DK as a function of K (PDF 376 kb)

Additional file 4: Figure S2. Bioclimatic variables significantly associated to genetic diversity in *Brachypodium distachyon*. A. BIO3. B. BIO8. C. BIO9 (PDF 9255 kb)

Acknowledgements
We thank María Luisa López-Herranz and Dimitry Feoktistov for laboratory assistance.

Funding
The study has been funded by two consecutive Spanish Ministry of Economy and Competitiveness grant projects (CGL2012–39953-C02–01, CGL2016–79790-P) and one Aragon Government and European Social Fund Bioflora grant to PC and IM. DL-A and VS were funded by a Spanish Ministry of Science and Innovation and a Tomsk State University PhD fellowship, respectively. DL-A was also supported by a Spanish Instituto de Estudios Altoaragoneses grant. IM received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement PIOF-GA-2011-301,257. PH was funded by Spanish Junta de Andalucía project AGR-0482.

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' contributions
PC conceived and designed the experiments, IM, VS, PC analyzed the data, PC, DLA, contributed reagents/materials/analysis tools, IM, PC, AM, PH, MO wrote the paper, and all authors reviewed drafts of the paper. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

Publisher's Note

Author details
[1]Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain. [2]Department of Botany, Institute of Biology, Tomsk State University, Lenin Av. 36, Tomsk 634050, Russia. [3]Departamento de Biología Animal, Biología Vegetal y Ecología, Universidad de Jaén, Paraje Las Lagunillas s/n, 23071 Jaén, Spain. [4]Instituto de Agricultura Sostenible (IAS-CSIC), Alameda del Obispo s/n, 14004 Córdoba, Spain. [5]Present address: Centro de Bioinformática y Biología Computacional de Colombia, BIOS, Parque los Yarumos, Manizales, Colombia.

Received: 2 January 2017 Accepted: 8 June 2017
Published online: 15 June 2017

References

1. Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, et al. *Brachypodium distachyon*. A new model system for functional genomics in grasses. Plant Physiol. 2001;127:1539–55.
2. International Brachypodium Initiative (IBI). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature. 2010;463:763–8.
3. Mur LA, Allainguillaume J, Catalán P, Hasterok R, Jenkins G, Lesniewska K, et al. Exploiting the *Brachypodium* tool box in cereal and grass research. New Phytol. 2011;191:334–47.
4. Catalan P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LAJ, Pecchioni N, Rasmussen SK, Vogel JP, Voxeur A) 2014. Update on genomics and basic biology of *Brachypodium*. Trends Plant Sci 19:414–418.
5. Watt M, Schneebeli K, Dong P, Wilson IW. The shoot and root growth of *Brachypodium* and its potential as a model for wheat and other cereal crops. Funct Plant Biol. 2009;36:960–9.
6. Bevan MW, Garvin DF, Vogel JP. *Brachypodium distachyon* Genomics for sustainable food and fuel production. Curr Opin Biotechnol. 2010;21:211–7.
7. Guillon F, Bouchet B, Jamme F, Robert P, Quéméner B, Barron C, et al. *Brachypodium distachyon* Grain: characterization of endosperm cell walls. J Experimental Botany. 2011;62:1001–15.
8. Opanowicz M, Hands P, Betts D, Parker ML, Toole GA, Mills ENC, et al. Endosperm development in *Brachypodium distachyon*. J Exp Bot. 2011;62:735–48.
9. Wang K, Han X, Dong K, Gao L, Li H, Ma W, et al. Characterization of seed proteome in *Brachypodium distachyon*. J Cereal Sci. 2011;52:177–86.
10. Gordon SP, Priest H, Des Marais DL, Schackwitz W, Figueroa M, Martin J, Bragg JN, Tyler L, Lee C-R, Bryant D, Wang W, Messing J, Manzaneda AJ, Barry K, Garvin DF, Budak H, Tuna M, Mitchell-Olds T, Pfender WF, Juenger TE, Mockler TC, Vogel JP. 2014. Genome diversity in *Brachypodium distachyon*: deep sequencing of highly diverse inbred lines. Plant J 79: 361–374.
11. Brkljacic J, Grotewold E, Scholl R, Mockler T, Garvin DF, Vain P, et al. *Brachypodium* as a model for the grasses: today and the future. Plant Physiol. 2011;157:3–13.
12. Manzaneda AJ, Rey PJ, Bastida JM, Weiss-Lehman C, Raskin E, Mitchell-Olds T. Environmental aridity is associated with cytotype segregation and

polyploidy occurrence in *Brachypodium distachyon* (Poaceae). New Phytol. 2012;193:797–805.
13. Manzaneda AJ, Rey PJ, Anderson JT, 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;69:2689–704.
14. Tyler L, Fangel JU, Fagerström AD, Steinwand MA, Raab TK, Willats WGT, et al. Selection and phenotypic characterization of a core collection of *Brachypodium distachyon* inbred lines. BMC Plant Biol. 2014;14:25. doi:10.1186/1471-2229-14-25.
15. Shindo C, Bernasconi G, Hardtke CS. Natural genetic variation in *Arabidopsis*: tools, traits and prospects for evolutionary ecology. Ann Bot. 2007;99(6):1043–54.
16. Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, Langdon T, et al. Evolution and taxonomic split of the model grass *Brachypodium distachyon* (L.) P. Beauv. Ann Bot. 2012;109:385–405.
17. Catalán P, López-Alvarez D, Bellosta C, Villar L. 2016a. Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei* and *B. hybridum* (Poaceae). *Anales Jardín Botánico Madrid* 73(1): e028.
18. López-Alvarez D, López-Herranz ML, 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): e51058.
19. López-Alvarez D, Zubair H, Beckmann M, Draper J, Catalán P. 2016. Diversity and association of phenotypic and metabolomic traits in the close model grasses *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Annals of Botany* (early on-line) doi:101093/aob/mcw239.
20. Filiz E, Ozdemir BS, Budak F, Vogel JP, Tuna M, Budak H. Molecular, morphological, and cytological analysis of diverse *Brachypodium distachyon* inbred lines. Genome. 2009;52:876–90.
21. Vogel JP, Tuna M, Budak H, Huo N, Gu Y-Q, Steinwand MA. Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. *BMC Plant Biol*. 2009;9(1):88.
22. Tyler L, Lee SJ, Young N, Delulio GA, Benavente E, Reagon M, et al. Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. The Plant Genome. 2016; doi:10.3835/plantgenome2015.08.0074.
23. López-Alvarez D, Manzaneda AJ, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, et al. Environmental niche variation and evolutionary diversification of the *Brachypodium distachyon* grass complex species in their native circum-Mediterranean range. Am J Bot. 2015;102:1–16.
24. Contreras R, Figueiras AM, Gallego FJ, Benito C. *Brachypodium distachyon*: a model species for aluminium tolerance in Poaceae. Funct Plant Biol. 2014;41:1270–83.
25. Von Uexküll HR, Mutert E. Global extent, development and economic impact of acid soils. Plant Soil. 1995;171:1–15.
26. Shiposha V, Catalán P, Olonova M, Marques I. 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;4:e2407. doi:10.7717/peerj.2407.
27. Neji M, Geuna F, Taamalli W, Ibrahim Y, Chiozzottoo R, Abdelly C, et al. Assessment of genetic diversity and population structure of Tunisian populations of *Brachypodium hybridum* by SSR Markers. *Flora-Morphology, Distribution, Functional Ecology of Plants*. 2015;216:42–9.
28. Rosenthal DM, Ramakrishnan AP, Cruzan MB. Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. In North America. Mol Ecol. 2008;17:4657–69.
29. Mo X, Gao J, Gao L. Characterization of microsatellite markers and their application to genetic diversity analysis of *Brachypodium sylvaticum* Var. *breviglume* from Yunnan, China. Am J Plant Sci. 2013;4:1427–34.
30. Hodel RGJ, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, et al. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. Applications in Plant Sciences. 2016;4:1600025.
31. Giraldo P, Rodriguez-Quijano M, Vazquez JF, Carrillo JM, Benavente E. Validation of. Microsatellite markers for cytotype discrimination in the model grass *Brachypodium distachyon*. Genome. 2012;55:523–7.
32. van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Resour. 2004;4:535–8.
33. Goudet J. FSTAT, version 2.9. 3, A program to estimate and test gene diversities and fixation indices. Lausanne: Lausanne University; 2001.
34. Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. J Hered. 2009;100:106–13.
35. Chybicki IJ, Oleksa A, Burczyk J. Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data. Heredity. 2011;107:589–600.
36. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
37. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4:359–61.
38. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–20.
39. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23:1801–6.
40. Corander J, Marttinen P, Mantyniemi S. Bayesian identification of stock mixtures from molecular marker data. Fish Bull. 2006;104:550–8.
41. Langella O. POPULATIONS 1-2: population genetic software, individuals or population distance, phylogenetic trees. 2000. http://bioinformatics.org/~tryphon/populations/.
42. Nei M, Chesser RK. Estimation of fixation indices and gene diversities. Ann Hum Genet. 1983;47:253–9.
43. Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6:288–95.
44. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinformatics Online. 2005;1:47–50.
45. Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. Genetics. 1995;139:457–62.
46. Goldstein DB, Pollock DD. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J Hered. 1997;88:335–42.
47. Raymond M, Rousset F. GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. J Hered. 1995;86(3):248–9.
48. Ersts PJ. 2016. Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Available from http://biodiversityinformatics.amnh.org/open_source/gdmg. Accessed from internet on 2016–9-2.
49. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967;27:209–20.
50. Rohlf FJ. 2002. NtSYSpc, numerical taxonomy and multivariate analysis system. Version 2.11a, user guide. Exeter software, New York.
51. López M, Grau JM. 2004. Metales pesados, materia orgánica y otros parámetros de la capa superficial de los suelos agrícolas y de pastos de España peninsular. Resultados por provincias II. Instituto Nacional de Tecnología Agraria y Alimentaria (INIA). Ministerio de Educación y Ciencia de España.
52. Goslee SC, Urban DL. The ecodist package for dissimilarity-based analysis of ecological data. J Stat Softw. 2007;22(7):1–19.
53. Hamrick JL, Godt MJW. Effects of life history traits on genetic diversity in plant species. Philosophical Transactions of the Royal Society of London Series B. 1996;351:1291–8.
54. Nybom H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol. 2004;13:1143–55.
55. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution. 1965;19:395–420.
56. Catalán P, López-Alvarez D, Díaz-Pérez A, Sancho R, López-Herranz ML. Phylogeny and evolution of the genus *Brachypodium*. In: Vogel J, editor. Genetics and genomics of *Brachypodium*. Series Plant Genetics and Genomics: Crops Models. Springer. New York; 2016b. p. 9–38.
57. Shirk RY, Hamrick JL, Zhang C, Qiang S. Patterns of genetic diversity reveal multiple introductions and recurrent founder effects during range expansion in invasive populations of *Geranium carolinianum* (Geraniaceae). Heredity. 2014;112:497–507.
58. Gimode D, Odeny DA, de Villiers EP, Wanyonyi S, Dida MM, Mneney EE, et al. Identification of SNP and SSR markers in finger millet using next generation sequencing technologies. PLoS One. 2016;11(7):e0159437.
59. Peredo EL, Angeles Revilla M, Jimenez-Alfaro B, Bueno A, Fernández Prieto JA, Abbott RJ. Historical biogeography of a disjunctly distributed, Spanish alpine plant, *Senecio boissieri* (Asteraceae). Taxon. 2009;58(3):883–92.

60. Kropf M, Comes HP, Kadereit JW. Long-distance dispersal vs vicariance: the origin and genetic diversity of alpine plants in the Spanish sierra Nevada. New Phytol. 2006;172:169–84.
61. Pérez-Collazos E, Sánchez-Gómez P, Jiménez JF, Catalán P. The phylogeographic history of the Iberian steppe plant *Ferula loscosii* (Apiaceae): a test of the abundant-centre hypothesis. Mol Ecol. 2009;18:848–61.
62. Garnatje T, Pérez-Collazos E, Pellicer J, Catalán P. Balearic insular isolation and large continental spread framed the phylogeography of the western Mediterranean *Cheirolophus intybaceus* s.L. (Asteraceae). Plant Biol. 2013;15:165–75.
63. González-López O, Polanco C, Gyorgy Z, Pedryc A, Casquero PA. Genetic variation of the endangered *Gentiana lutea* L. Var. *aurantiaca* (Gentianaceae) in populations from the northwest Iberian Peninsula. International Journal of Molecular Science. 2014;2014:10052–66.
64. Wang IJ, Bradburd GS. Isolation by environment. Mol Ecol. 2014;23:5649–62.
65. Shafer ABA, Wolf JBW. Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. Ecol Lett. 2013;16:940–50.
66. Sexton JP, Hangartner SB, Hoffmann AA. Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution. 2014;62:1–15.
67. Hewitt GM. The genetic legacy of the quaternary ice ages. Nature. 2000;405:907–13.
68. Hewitt GM. Genetic consequences of climatic oscillations in the quaternary. Philosophical Transactions of the Royal Society of London Series B. 2004; 358:183–96.
69. Nieto FG. Southern European glacial refugia: a tale of tales. Taxon. 2011;60:365–72.
70. Theissinger K, Bálint M, Feldheim KA, Haase P, Johannesen J, Laube I. Pauls SU. 2013. Glacial survival and post-glacial recolonization of an arctic–alpine freshwater insect (*Arcynopteryx dichroa*, Plecoptera, Perlodidae) in Europe. J Biogeogr 40: 236–248.
71. Charrier O, Dupont P, Pornon A, Escaravage N. Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the Pyrenees. PLoS One. 2014;9(3):e92976. doi:10.1371/journal.pone.0092976.
72. Liberal IM, Burrus M, Suchet C, Thébaud C, Vargas V. 2014. The evolutionary history of *Antirrhinum* in the Pyrenees inferred from phylogeographic analyses. *BMC Evolutionary Biology* 2014, 14:146.
73. Bidegaray-Batista L, Sánchez-García A, Santulli G, Malorano L, Guisan A, Vogler AP, et al. Imprints of multiple glacial refugia in the Pyrenees revealed by the phylogeography and paleodistribution modelling of an endemic spider. Mol Ecol. 2016;25:2046–64.
74. Sánchez-Goñi MF, Hannon G. High-altitude vegetational pattern on the Iberian Mountain chain (north-central Spain) during the Holocene. The Holocene. 1999;9:39–57.
75. Cottrella V, Krystufekb, HE, Tabbenera AD, Milnera T, Connollye L, Singa S, Fluchb K, Burgb F, Lefèvrec P, Achardc S, Bordâcsd K, Gebhardte B, Vornamf MJM, Smuldersg AH, Vanden Broedkh J, Van Slyckenh V, Stormei W, Boerjani S, Castiglionei T, Fossatij N, Albak D, Agúndezk C, Maestrol E, Notivoli J, Bovenschenm BC, van Damm. 2005. Postglacial migration of *Populus nigra* L.: lessons learnt from chloroplast DNA. JE. For Ecol Manag 206: 71–90.
76. Pérez-Collazos E, Segarra-Moragues JG, Catalán P. Two approaches for the selection of relevant genetics units for conservation (RGUCs) in the narrow European endemic steppe plant *Boleum asperum* (Brassicaceae). Biological Journal Linnean Society. 2008;94:341–54.
77. González-Sampériz P, Valero-Garcés B, Carrión García JS. Was the Ebro valley a refugium for temperate trees? An Biol. 2004;26:13–20.
78. González-Sampériz P, Valero-Garcés BL, Carrión JS, Peña-Monné JL, García-Ruiz JM, Martí-Bono C. Glacial and lateglacial vegetation in northeastern Spain: new data and a review. Quat Int. 2005;140(141):4–20.
79. Gómez A, Lunt DH. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, editors. Phylogeography in southern European refugia: evolutionary perspectives on the origin and conservation of european biodiversity. Dordrecht: Springer; 2007. p. 155–88.
80. Cuadrat JM. 1999. El Clima de Aragón. CAI-100. Zaragoza: Institución "Fernando El Católico."
81. Cañasveras JC, del Campillo MC, Barrón V, Torrent J. Intercropping with grasses helps to reduce iron chlorosis in olive. J Soil Sci Plant Nutr. 2014; 14(3):554–64.
82. González Moreno A, Casanova Pena C, Gascó A, Rodríguez Martín JA. *Brachypodium hybridum* plant cover improves water infiltration in Mediterranean crop soils. Journal of Plant Chemistry and Ecophysiology. 2016;1(2):1008.

CHAPTER 3

Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot

Valeriia Shiposha[1,2§], Isabel Marques[1†§], Diana López-Alvarez[1], Antonio J. Manzaneda[3], Pilar Hernandez[4], Marina Olonova[2], Pilar Catalán[1,2,5,*]

1. Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain.
2. Department of Botany, Institute of Biology, Tomsk State University, Lenin Av. 36, Tomsk 634050, Russia.
3. Departamento de Biología Animal, Biología Vegetal y Ecología, Universidad de Jaén, Paraje Las Lagunillas s/n, 23071-Jaén, Spain
4. Instituto de Agricultura Sostenible (IAS-CSIC), Alameda del Obispo s/n, 14004 Córdoba, Spain
5. Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza E-50059, Spain
†current address: cE3c - Centre for Ecology Evolution and Environmental Changes, Faculty of Sciences, University of Lisbon, Lisbon, Portugal

§ Both authors contributed equally to this work
*Corresponding author: pcatalan@unizar.es

Short running title: Multiple origins of the model allotetraploid grass *Brachypodium hybridum*

Abstract

Background and Aims: It is accepted that contemporary allopolyploid species have originated recurrently, but very few cases have been documented using multiple natural formations of the same species. To extend our knowledge, we have investigated the multiple origins, genetic variation, and structure of the allotetraploid grass *Brachypodium hybridum* with respect to its progenitor diploid species *B. distachyon* (D genome) and *B. stacei* (S genome). For this, our primary focus is the Iberian Peninsula, an evolutionary hotspot for the genus *Brachypodium*.

Methods: We analysed 342 *B. hybridum* individuals from 36 populations using 10 nuclear SSR loci and two plastid loci. The *B. hybridum* genetic profiles were compared with those previously reported for *B. stacei* and *B. distachyon*. In addition, phylogenetic analysis of the plastid data was performed for a reduced subset of individuals.

Key Results: The nuclear SSR genetic analysis detected medium to high genetic diversity, with a strong south-to-north genetic structure cline, and a high selfing rate in *B. hybridum*. Comparative genetic analysis showed a close relatedness of current *B. hybridum* D allelic profiles with those of *B. distachyon*, but a lack of similarity with those of *B. stacei*. Plastid analysis detected three different bidirectional allopolyploidization events: two involved distinct *B. distachyon*-like ancestors and one involved a *B. stacei*-like ancestor. The Southeastern (SE) Iberian Peninsula *B. hybridum* populations were more genetically diverse and could have originated from at least two hybridization events whereas Northeaster-Northwestern (NE-NW) Iberian Peninsula *B. hybridum* populations were less diverse and may have derived from at least one hybridization event.

Conclusions: The genetic and evolutionary evidence support the plausible *in situ* origin of the SE and northern Iberian Peninsula *B. hybridum* allopolyploids from their respective local *B. distachyon* and unknown *B. stacei* ancestors. The untapped multiple origins and genetic variation detected in these *B. hybridum* populations opens the way to future evolutionary analysis of allopolyploid formation and genomic dominance and expression in the *B. hybridum – B. distachyon – B. stacei* grass model complex.

Keywords: allopolyploidy, *Brachypodium hybridum– B. distachyon – B. stacei*, genetic structure and diversity cline, Iberian hotspot, multiple origins, population genetics.

INTRODUCTION

In the plant kingdom polyploid species almost equal the number of current diploid species (Barker *et al.,* 2016; Doyle and Sherman-Broyles, 2017; Marques *et al.,* 2018). Polyploidy is considered the primary driver of diversity in several families (Soltis *et al.,* 2016; Van der Peer *et al.,* 2017). It is a recurrent phenomenon that has lasted since the origin of the angiosperms, whose proto-ancestors experienced one or more whole genome duplication (WGD) events that converted most current flowering plants in descendants of early palaeopolyploids (Jiao *et al.,* 2011). Some paleopolyploids returned to a diploid state through various mechanisms of downsizing genomes, such as massive gene losses and large genomic and chromosomal rearrangements (te Beest *et al.,* 2012; Marques *et al.,* 2016). Recent polyploids emerged as the result of new WGD events which were inferred to have occurred in the Oligocene-Miocene or in the Quaternary, producing meso- and neo-polyploids, respectively (Stebbins 1985; Soltis *et al.,* 2016).

Although recent studies are revealing a higher than expected frequency of autopolyploidy events in angiosperms (Spoelholf *et al.,* 2017; Doyle and Sherman-Broyles, 2017; Baduel *et al.,* 2018), the large majority are allopolyploids species that originated after hybridization of diploid or lower-ploidy progenitor species bearing homeologous genomes (Doyle and Sherman-Broyles, 2017; Soltis *et al.,* 2016). The effective reproductive isolation of the allopolyploid from its progenitor species has been recognized as the main factor driving rapid speciation (te Beest *et al.,* 2012), thus avoiding the potential influx of gene flow from them through repeated backcrossing and introgression. Nevertheless, the origin of most allopolyploids is still unknown except for the intensively studied cultivated plants (e. g., *Brassica*, *Gossypium*, *Triticum*) and the tracked contemporary origins of some wild experimental species (e. g., *Arabidopsis*, *Senecio*, *Tragopogon*) (Soltis *et al.,* 2016). The 'multiple origins' evolutionary scenario of allopolyploids (Doyle and Sherman-Broyles, 2017) still remains largely unexplored. Only a few studies have reported on large population samplings of both extant progenitor species and their derived allopolyploids (Soltis *et al.,* 2016).

Brachypodium has emerged as a model system for temperate cereals and bioenergy grasses (Vogel *et al.,* 2010; Mur *et al.,* 2011; Catalán *et al.,* 2014; Scholthof *et al.,* 2018). Contrary to other model plants, the annual *B. distachyon* has a rich combination of desirables

attributes such as a short life cycle with simple growth requirements, is highly homozygous and can be easily transformed (Scholthof *et al.*, 2018). *Brachypodium distachyon* (x = 5, 2n = 10; genome size 0.36-0.39 pg/272 Mbp), the first fully sequenced Pooideae genome (reference genome: accession Bd21; Vogel *et al.*, 2010) has remarkable similarity to the genome composition of other temperate grasses. Also, contrary to other grasses where crop domestication has created a genetic bottleneck compared with wild ancestors (Buckler *et al.*, 2001), *Brachypodium* was never domesticated. It has retained its maximum genetic variability in wild populations, which can be used to decipher gene functions for improving agronomic traits and for comparative ecological and evolutionary studies (Gordon *et al.*, 2017; Scholthof *et al.*, 2018). Nuclear SNPs from resequenced *B. distachyon* lines (Gordon *et al.*, 2017) and genotyping-by-sequencing (GBS) data (Tyler *et al.*, 2016) together with whole plastome analyses (Sancho *et al.*, 2018) have detected two main diverged lineages in *B. distachyon*, a mostly Extremely Delayed Flowering (EDF+) clade and a mostly Spanish (S+) – Turkish (T+) clade. Interestingly, these clades are not primarily connected with geography, but with flowering time phenotypic traits, although counterbalanced by introgression between them (Tyler *et al.*, 2016; Sancho *et al.*, 2018).

Besides *B. distachyon*, reference genomes of the diploid *B. stacei* (x = 10, 2n = 20; 0.564 pg/2C 234 Mbp) and of their derived allotetraploid *B. hybridum* (x = 5 + 10, 2n = 30; 1.265 pg/2C 509 Mb) are also available (Scholthof *et al.*, 2018). Phylogenetic analyses estimated that *B. hybridum* could have arisen ca. 1 Ma (Catalan *et al.*, 2012), almost contemporarily with its progenitor *B. distachyon* species (Sancho *et al.*, 2018). Genetic studies based on barcoding nuclear loci indicated that the *B. distachyon*-type (D) and *B. stacei*-type (S) subgenomes of *B. hybridum* were overall highly intact compared to the studied genomes of current progenitor species, whereas the maternally inherited plastid markers showed that *B. hybridum* originated from bidirectional crosses (Lopez-Alvarez *et al.*, 2012). Artificial crosses have corroborated these findings through the creation of a synthetic fertile allotetraploid, which phenotypically resembles *B. hybridum* after the hybridization of *B. distachyon* and *B. stacei* species (Dinh Thi *et al.*, 2016).

Traditional population genetic studies based on nuclear microsatellites, as well as Genotyping-by-Sequencing (GBS) and plastome data have identified the Iberian Peninsula as an important source of genetic variation either in *B. stacei* (Shiposha *et al.*, 2016) or in *B.*

distachyon (Tyler *et al.*, 2016; Marques *et al.*, 2017; Sancho *et al.*, 2018). Remarkably, the genetic diversity and the origins of *B. hybridum*, the stable allotetraploid hybrid species, has been scarcely studied, although considered an invasive species outside its native circum-Mediterranean range (Bakker *et al.*, 2009; López *et al.*, 2012). A main drawback for the investigation of the multiple origins hypothesis of allopolyploids has been the lack of sample numbers (Soltis *et al.*, 2016). Therefore, in this study we have analysed a large number of populations of *B. hybridum* and of its progenitor species (*B. stacei* and *B. distachyon*) across the Iberian Peninsula using nuclear and plastid data. Specifically, we aimed to answer the following questions: 1) Is the genetic diversity of *B. hybridum* geographically structured in the Iberian Peninsula?; 2) How many founder events have contributed to it?; 3) Can we track the parental origin of the populations of *B. hybridum*?; and, 4) Does the center of genetic diversity of *B. hybridum* coincide with the genetic diversity centers of the progenitor species or has a shift occurred?

MATERIAL AND METHODS

Population sampling, DNA extraction and nSSR amplification

A total of 342 individuals of *B. hybridum* were collected across 36 populations covering the whole distribution range of this species within the Iberian Peninsula (Fig. 1). Sampling sizes, locations and geographic coordinates of each population sampled are given in Table S1. Because *B. hybridum* can be morphologically confused with the parental species *B. distachyon*, the identity of the samples was first confirmed through DAPI-stained chromosomes, coupled with barcoding markers (López-Alvarez *et al.*, 2012). Fresh leaves were collected from each individual, dried in silica gel and stored at −20°C until DNA was extracted. Total genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. Samples of *B. hybridum* were genotyped at ten polymorphic nuclear simple sequence repeats (SSRs) developed for Turkish populations of *B. distachyon* (ALB006, ALB022, ALB040, ALB050, ALB086, ALB087, ALB139, ALB165, ALB181 and ALB311; Vogel *et al.*, 2009) and applied previously to Iberian populations of *B. distachyon* (Marques *et al.*, 2017) and *B. stacei* (Shiposha *et al.*, 2016). SSR amplifications in *B. hybridum* were carried out as described in Shiposha *et al.* (2016). Multiplexed PCR products were

genotyped on an Applied Biosystems 3130XL Genetic Analyzer using 2 μl of amplified DNA, 12 μl of Hi-Di formamide and 0.4 μl of GeneScan-500 (LIZ) size standard (Applied Biosystems). Allele sizes were determined using Peak Scanner version 1.0 (Life Technologies) and revised manually. Because of the confirmed allotetraploidy and disomic inheritance of *B. hybridum* (Catalán *et al.,* 2012; Díaz-Pérez *et al.*, 2018), the scored SSR alleles were assigned to the parental *B. stacei*-type (S subgenome) and *B. distachyon*-type (D subgenome) genotypes of the sampled individuals by decoupling each locus into two subgenomic loci, following the procedures indicated in Catalán *et al.* (2006). Two of the 10 loci (ALB87, ALB181) showed overlapping allelic sizes in both parents and in *B. hybridum* and were encoded as a single locus each, one locus (ALB139) showed single genetic dosage from one of the subgenomes (D), and seven loci (ALB006, ALB22, ALB40, ALB50, ALB86, ALB165, ALNB311) showed alleles from the two subgenomes. Individual genotypes from a total of 17 loci were encoded as for conventional diploids in *B. hybridum* (Table S2). Single diploid genotypes from 181 individuals of *B. stacei* (19 populations; Shiposha *et al.*, 2016) and 148 individuals of *B. distachyon* (16 populations; 2 from the current study, 14 from Marques *et al.*, 2017) were also incorporated into this study (Table S3).

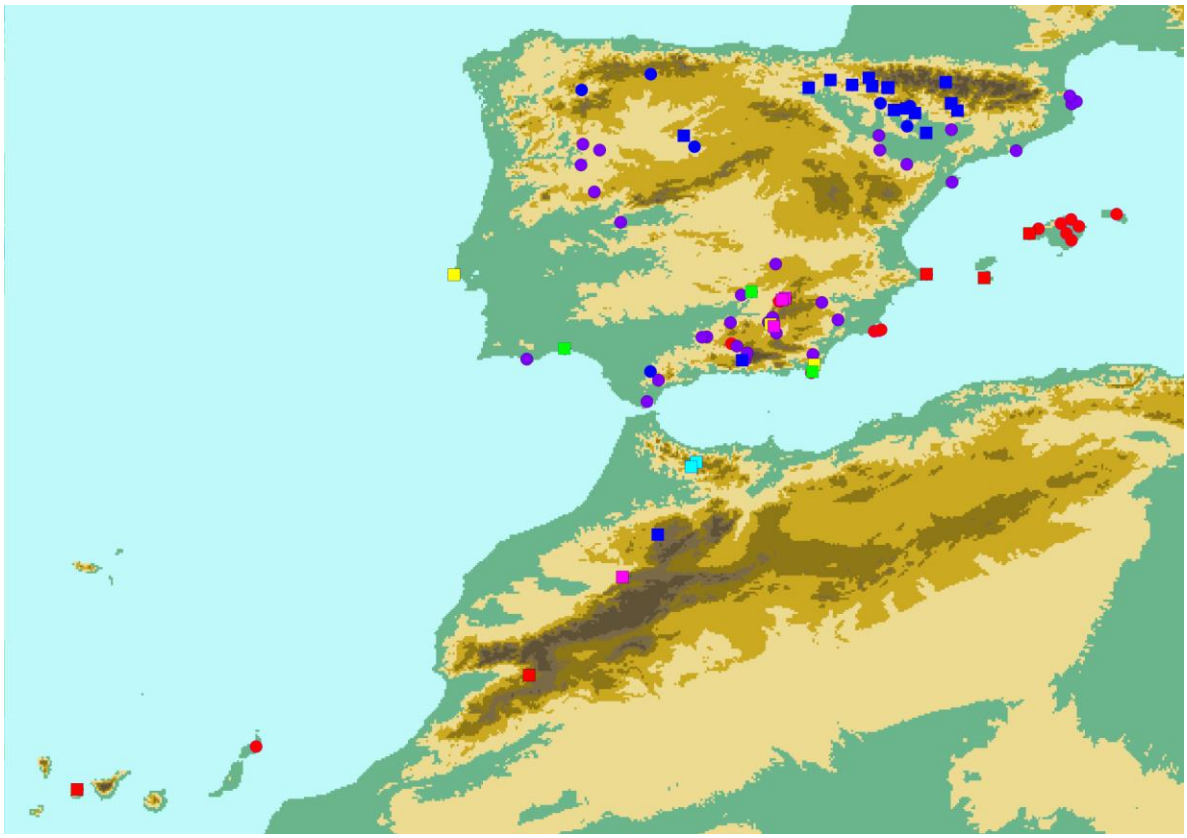


Figure 1: Geographic distribution of the allotetraploid *Brachypodium hybridum* and the diploid *B. stacei* and *B. distachyon* progenitor species used in genetic and evolutionary analyses of the allopolyploid model complex. The circles indicate populations used in the SSR analysis (*B. hybridum*, violet; *B. stacei*, red; *B. distachyon*, dark blue). The squares indicate samples used in the plastid DNA sequence analysis (*B. hybridum* S plastotypes, yellow; *B. hybridum* D1 plastotypes, light green; *B. hybridum* D2 plastotypes purple; *B. stacei*, red; *B. distachyon* early diverging lineages, light blue; *B. distachyon* recently diverging lineages, dark blue).

Genetic diversity in *Brachypodium hybridum*

We calculated genetic diversity and structure of the *B. hybridum* populations using 14 SSR loci. Genetic variation based on total number of alleles (Na), allelic richness (A_R), observed within population Nei's heterozygosity (H_o), expected within population Nei's heterozygosity (H_s), expected Nei's heterozygosity within the total population (H_T), Nei's measure of genetic differentiation (G_{st}), and inbreeding coefficient (F_{IS}) per locus and population was estimated using FSTAT 2.9.3.2 (Goudet *et al.*, 2001). F_{IS} was also estimated using the Bayesian

procedure implemented in INEst 2.0 (Chybicki and Burczyk, 2009), which is robust for the presence of null alleles. Posterior distribution was based on 300,000 steps, sampling every 100 steps and discarding the first 30,000 steps as burn-in. In order to understand the importance of inbreeding in our dataset we compared the full model (nfb) with the model including only null alleles (nb). We choose the best model based on the Deviance Information Criterion (DIC; Chybicki *et al.*, 2011).

Genetic relationships,population structure and differentiation in *Brachypodium hybridum*

We used POPULATION 1.2 (Langella *et al.*, 2000) to calculate Nei's genetic distance (Da; Nei and Chesser 1983) among individuals and to construct an unrooted neighbor-joining tree with 1000 bootstrap replicates. We also constructed a Principal Components Analysis (PCoA) in GenAlEx6 (Peakall and Smouse, 2006) to detect the genetic relatedness among individuals based on Nei's genetic distance. To understand the genetic structure of *B. hybridum* in the sampled area, we used the Bayesian program STRUCTURE v.2.3.4 (Pritchard *et al.*, 2000). Analyses were performed from K=1 to the number of genetic groups detected in the previous NJ and PCoA searches plus 2 (K=19), with 10 repetitions per K. We ran models assuming ancestral admixture and correlated allele frequencies with 50,000 burn-in steps, followed by run lengths of 300,000 iterations for each K. We selected the optimum K using StructureSelector (Li & Liu 2017), which besides the commonly used Ln Pr(X|K) and ΔK statistics (Evanno *et al.*, 2005) also uses four alternative statistics (medmedk, medmeak, maxmedk and maxmeak) to infer the optimal K (Li and Liu, 2017). The results of the replicates at the best-fit K were post-processed using CLUMPP 1.1.2 (Jakobsson and Rosenberg, 2007).

We used standard and hierarchical analysis of molecular variance (AMOVA) to quantify the partitioning of genetic variance within and among the following hierarchal levels: among all populations and between several geographical groups that also showed genetic differentiation in the NJ and PCoA analyses. In each analysis, we quantified variance among groups, among locations within groups and within sampling locations. We ran each AMOVA with 10,000 permutations at 0.95 significance levels in ARLEQUIN 3.11 (Excoffier *et al.*, 2005). The eelationships between population pairwise Nei's Da genetic distances and linear geographic distances (isolation-by-distance, IBD) were examined using a Mantel test (Mantel, 1967) implemented in ARLEQUIN 3.11 (Excoffier *et al.*, 2005) with 10,000 permutations.

Genetic relationships in the *B. stacei-B. distachyon-B. hybridum* complex

In order to decipher the potential relationships and the genome splitting-and-merging evolutionary history of the Iberian Peninsula diploid progenitors *B. stacei* and *B. distachyon* and allotetraploid *B. hybridum* populations, we analyzed the genetic structure and the phylogeny of the allotetraploid and diploid individuals using homologous nuclear SSR alleles. Microsatellite alleles from a total of 18 decoupled SSR loci (ALB006S, ALB006D, ALB022A, ALB022B, ALB040S, ALB040D, ALB050S, ALB050D, ALB086S, ALB086D, ALB087A, ALB139S, ALB139D, ALB165S, ALB165D, ALB181A, ALB311S, ALB311D) were used to encode S and D diploid-like genotypes from 342 individuals of *B. hybridum* (36 populations, current study), and single diploid genotypes from the 181 *B. stacei* and 148 *B. distachyon* individuals (Table S3). The SSR alleles were then recorded by their presence/absence into a binary data matrix consisting of 973 *B. stacei* and *B. distachyon* individual phenotypes and *B. hybridum* S and D subgenomic phenotypes and 98 alleles after discarding a few outlier samples (Table S3). The genetic relationships among the diploid *B. stacei* and *B. distachyon* individuals and the diploidized S and D subgenomes of the allotetraploid *B. hybridum* individuals were visualized using a multivariate PCO analysis with pairwise Nei and Li genetic distances in NTSYSPC v. 2.11a (Rohlf, 2002). The genetic structure of the complex was investigated using STRUCTURE v.2.3.4, searching for *K*=1-15 potential genetic groups identified in the PCoA analysis and imposing the non-admixture ancestry model and the non-correlated allele frequency model. Each search consisted of an initial burn-in of 50,000 MCMC steps followed by 500,000 MCMC additional steps, running 10 replicated per each *K*. The number of genetic groups was estimated using STRUCTURE HARVESTER (v. 0.9.94) (Earl and vonHoldt, 2012) which identifies the optimal *K* based on both Ln Pr(X|K) and ΔK (Evanno *et al.*, 2005) statistics. The phylogeny was reconstructed using a maximum likelihood search for binary data in IQTREE (Nguyen *et al.*,2014), imposing the best-fit nucleotide substitution model GTE2+FO+G4 that was automatically selected by the ModelFinder option of the program according to the Bayesin Information Criterion (BIC), and the automated computation of 20 Maximum Likelihood (ML) starting trees from 98 alternative randomized Maximum Parsimony (MP) trees, searching for best-scoring ML trees and estimating branch support for the best tree from 1,000 bootstrap replicates (BS) using the ultrafast bootstrap option implemented in the software. A second IQTREE search run with the same parameters plus the

ascertainment bias correction model for non-constant data (Nguyen *et al.*, 2014) did not provide a better resolution than the previous search and is not discussed further.

Phylogeny of the *B. stacei-B. distachyon-B. hybridum* complex We aimed to corroborate the different S and D genetic lineages detected with the SSR markers within *B. hybridum*, with respect to those of its progenitor species, and to reconstruct their phylogeny. For this, we used two DNA barcoding loci, the plastid trnLF and ndhF regions, which had proved to be useful to discriminate between *B. stacei* and *B. distachyon* and the S and D subgenomes of *B. hybridum*, as well as to identify divergent clades among them (López-Alvarez *et al.*, 2012; Díaz-Pérez *et al.*, 2018). A subset of samples from the three studied species were used in the DNA sequencing analyses, together with a representation of other DNA sequences sampled across the Iberian Peninsula and other geographical ranges retrieved from our previous work (Díaz-Pérez *et al.*, 2018; Sancho *et al.*, 2018; Fig. 1, Table S4). The procedures for DNA amplification and sequencing, data processing and alignments of the trnLF and ndhF loci were described previously (López-Alvarez *et al.* 2012; Díaz-Pérez *et al.* 2018). For some *B. distachyon* and *B. hybridum* accessions sequence information on the two loci were retrieved from whole plastome data (Sancho *et al.*, 2018; Catalán *et al.*, unpub. data). The new trnLF and ndhF sequences have been deposited in Genbank under accessions codes xxxx to xxxx (Table S4). The aligned trnLF and ndhF data matrices were used to build a ML plastid tree with IQTREE. For this we used the same parameters stated above, but for DNA sequence data, and imposed the TPM2u+F+G4 nucleotide substitution model chosen by the program based on BIC. Three close outgroup grasses (*Oryza sativa*, *Melica ciliata*, *Glyceria fluitans*) were incorporated into the analysis, and *O. sativa* was used to root the tree. Separate phylogenetic analyses of the two loci gave congruent topologies with that recovered for the concatenated trnLF+ndhF haploid data matrix; only results from the latter analysis will be explained further.

RESULTS

Genetic diversity in *Brachypodium hybridum*

B. stacei, *B. distachyon* and *B. hybridum* individuals showed overlapping allelic sizes for locus ALB22A, whereas some *B. hybridum* individuals showed additional alleles from a second, likely duplicated, locus ALB22B, and their respective genotypes were encoded as for diploids (Tables S2, S3). Three of the 17 loci that showed nullisomic alleles in several *B. hybridum* individuals (ALB22B, ALB50S, ALB86S) were discarded from population genetic analysis of this species though they were used for comparative genetic and evolutionary analysis of the allotetraploid and its progenitor species. The fourteen loci showed high variability between populations with the number of alleles varying from 34 (ALB006D) to 68 (ALB181) (Table S5). Null allele frequencies calculated with INEst were always very low with a maximum value of 0.026 in locus ALB86S. Allelic richness per locus was always very low, varying from 0.944 (*ALB006D*) to 1.889 (*ALB181*). Five of the 10 studied loci had observed heterozygosity and F_{IS} values of 0 and 1, respectively (ALB86S, ALB86D, ALB165S, ALB311S, ALB311D). One locus showed an exceptionally high Ho value of 0.361 (ALB50D). The expected heterozygosity for loci varied from 0.006 (ALB311S) to 0.251 (ALB50D), being generally higher in the subgenomic loci derived from *B. distachyon* (Table S3). The estimated divergence of populations per locus (G_{St}) varied from 0.700 (*ALB181*) to 0.983 (ALB86S). F_{IS} had a minimum value of -0.439 for locus ALB050D, although values were even higher when F_{IS} was calculated in INEst (Table S5). Our results from Bayesian analyses implemented in INEst revealed that only inbreeding contributed to the excessive homozygosity. This model (DICnfb: 17395.576) was preferred over the model that included only null alleles (DICnb: 20094.100) based on the DIC criterion. In assessing HWE, we found that only one locus showed a significant deviation after the Bonferroni correction (ALB006D: $P<0.05$; Table S5).

Within populations, observed heterozygosity varied between 0 and 0.214 in the NE population (Table S1). The mean expected heterozygosity varied between 0 and 0.107 in the same populations. The total number of alleles per population ranged from 13 and 23 and the maximum value of allelic richness was also found in CUBI. F_{IS} values varied between 1 (fixed homozygosis) and some negative values (see Table S1). The rate of self-fertilization estimated

for *B. hybridum* varied considerably between populations in that some, but not all, plants were highly selfing (Table S1). Due to the high level of homozygosis (fixed alleles) observed in some populations, of the 342 individuals of *B. hybridum* genotyped only 120 (35%) exhibited a unique multi-locus genotype. The percentage of unique genotypes varied between 12% and 100% (Table S1).

Population genetic differentiation and structure in *Brachypodium hybridum*

The PCoA separated all populations into five main groups, clustering all Southeast (SE) populations in the left extreme of axis 1 and all Northwest (NW) and Northeast (NE) populations in the right side of this same axis (Fig. 2A). The Southwest (SW) populations were fragmented into different groups, one group was spatially closer to the NE and NW populations that included the populations of ALGA, ALGE, FARO, LEPE and MONF. The second group, which clustered close to all SE populations included ALCA and CABR (Fig. 2A). The very low variance was explained by the two axes since axis 1 only accumulated 17.26% and axis 2 only 11.51%. In accordance to PCoA results, the NJ tree also separated all SE Iberian Peninsula populations (plus SW ALCA and CABR) from all NE and NW populations and from the remaining SW populations of ALGA, ALGE, FARO, LEPE and MONF (Fig. 2B). The BS values of branches were always very low (<43%; Fig. 2B).

All STRUCTURE statistics [Ln Pr(X|K), ΔK, medmedk, medmeak, maxmedk and maxmeak] found K=2 as the optimal number of genetic clusters in the sampling area of *B. hybridum* followed by K=4 (Fig. 2C) and K=14 (Figs. S1). At K=2, all SE populations plus two SW populations (ALCA and CABR) were separated from all of the remaining populations. However, at K=4 a homogeneous genetic group of northern Iberian NE+NW populations separated from three genetic groups in southern Iberia that included populations from both the SW and the SE (Fig.2C). At K=14, populations matched the previous PCoA and NJ results, showing the segregation of two NW, four NE, two SW and seven SE distinct population groups (Fig. S1). There was no evidence of genetic admixture between clusters, except for the introgressed individuals of population ESPU with the MORA and RUID genetic group for K=14 (Fig. S1).

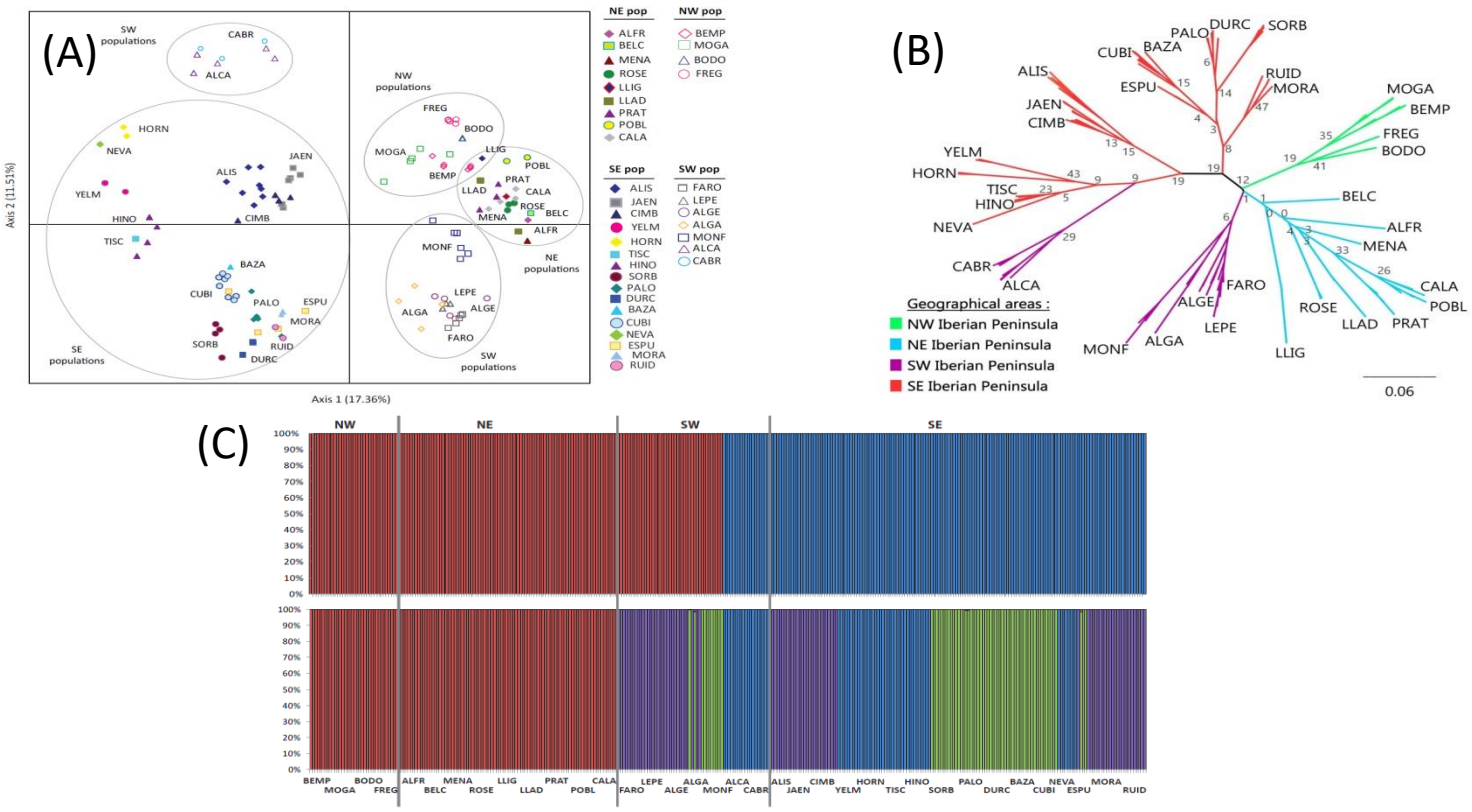


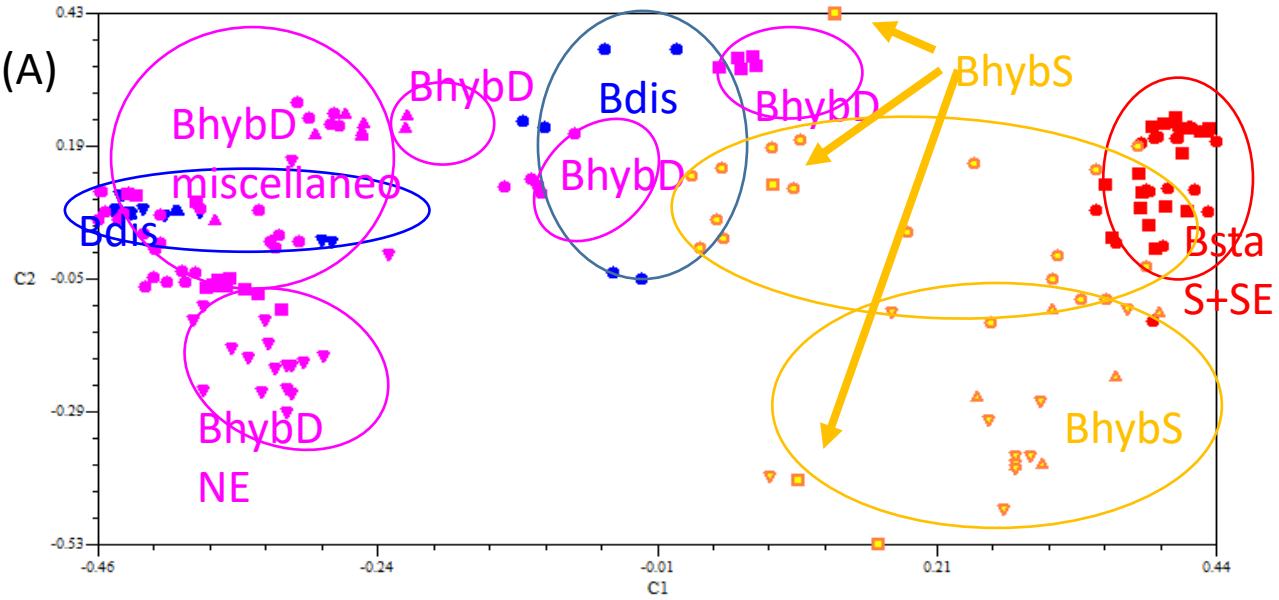
Figure 2A: Principal Coordinate analysis (PCoA) of *Brachypodium hybridum* samples. The percentage of variance of each axis is given in parentheses. Population symbols and colors are shown in the chart. **2B:** Unrooted neighbor-joining tree based on Nei's Da genetic distance of the studied *Brachypodium hyridum* populations. Numbers associated with the branches indicate bootstrap values based on 1000 replications. Note that only a small number of individuals can be seen in the PCoA and in the NJ tree since most individuals share identical alleles within each population. **2C:** Population structure of *B. hybridum in the Iberian Peninsula* at *K*=2 and K=4, following the best assignments retrieved by STRUCTURE. Each individual is represented by a thin vertical line divided into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Major geographic areas are labeled below the graph. Population codes follow Table S1.

Genetic differentiation was significantly high across all 36 populations (AMOVA F_{ST}=0.871, P<0.00001). Overall, 83.1% and 12.9% of the genetic variation was attributed to variation among and within populations, respectively (Table S6). A hierarchical AMOVA performed between the four geographical regions found the highest percentage of variance among populations within groups (61.3%). However, this was further exacerbated when performing the hierarchical AMOVA between populations of the K=2 boundary defined by STRUCTURE. This showed the highest partition of variance among populations within groups (P<0.00001, Table S6; 62.4%). We found no correlation between genetic distance and geographic distance as indicated by a Mantel-test (r=0.011, P=0.105).

Genetic relationships of *B. hybridum* and its progenitor species

The analysis of *B. stacei*, *B. distachyon* and *B. hybridum* (S and D) phenotypes based on SSR alleles (Table S3) showed evidence of a close genetic relatedness between the *B. hybridum* S and D subgenomic phenotypes and some phenotypes from the progenitor species. The PCO plot showed a clear separation of the *B. stacei* and *B. distachyon* phenotypes at the respective positive and negative sides of axis C1, which accumulated 32.19% of variance (Fig. 3A). The *B. stacei* phenotypes clustered close to each other at the positive extreme of C1. In contrast, the *B. distachyon* phenotypes spread through the left half of the plot, showing clusters of S and N Iberian population phenotypes in the middle and the left of the plot, respectively. The *B. hybridum* S and D phenotypes showed subgenomic clusterings congruent with those of their progenitors' clusters along axes C1 and C2 (15.06% of variance) (Fig. 3A). Interestingly, the plot distribution of the geographic *B. hybridum* D phenotypes matched relatively well with the geographic *B. distachyon* phenotypes. Specifically, *B. hybridum* D from SE and SW Iberia were close to *B. distachyon* from S Iberia, and *B. hybridum* D from NE and NW Iberia were close to *B. distachyon* from N Iberia. By contrast, the *B. hybridum* S phenotypes from the SE (and some SW) and from NE+NW Iberia were not close to the *B. stacei* phenotypes, except for a few SE Iberia *B. hybridum* S phenotypes (Fig. 3A).

(A)


BhybD
miscellaneo
Bdis
BhybD
Bdis
BhybD
BhybD
BhybS
Bsta
S+SE
BhybD
NE
BhybS
C2
C1
0.43
0.19
-0.05
-0.29
-0.53
-0.46
-0.24
-0.01
0.21
0.44

(B)

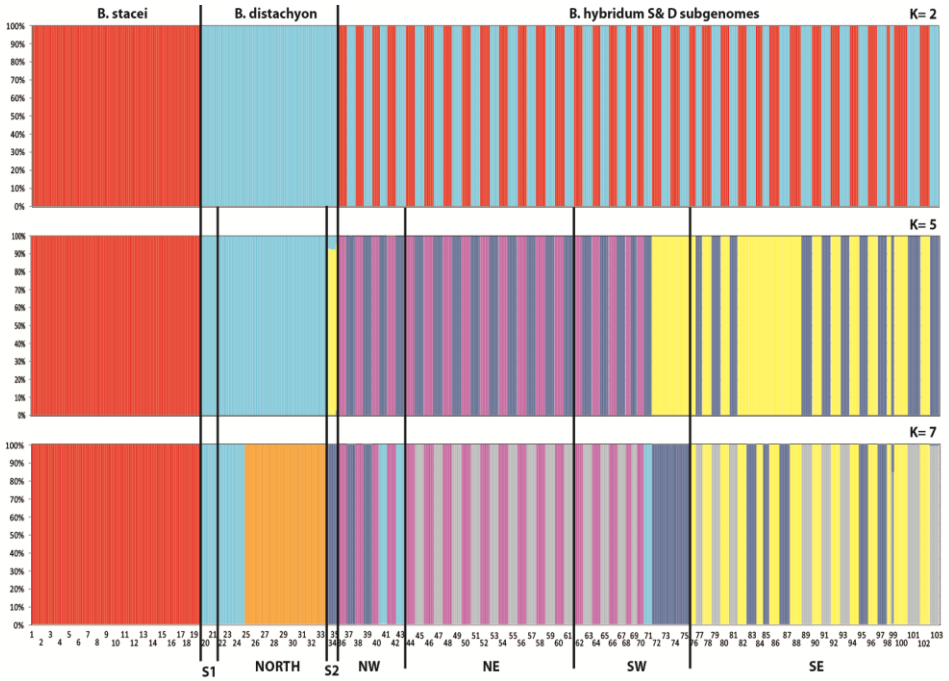

B. stacei
B. distachyon
B. hybridum S& D subgenomes
K= 2
K= 5
K= 7
S1
NORTH
S2
NW
NE
SW
SE

Figure 3A. Principal Coordinate Analysis (PCoA) of the *Brachypodium hybridum-B. stacei-B. distachyon* complex. The percentage of variance of each axis is given in parentheses. Colors and symbols of species and genomes and of geographic areas of populations are as follows: *B. stacei*: Iberian Peninsula SE + Balearic islands (red dot), Iberian Peninsula S + Canary Islands (red square); *B. distachyon*: Iberian Peninsula S (blue dot), Iberian Peninsula NW (blue upper triangle), Iberian Peninsula NE (blue lower triangle); *B. hybridum* S subgenome: Iberian Peninsula SE (yellow dot), Iberian Peninsula SW (yellow square), Iberian Peninsula NW (yellow upper triangle), Iberian Peninsula NE (yellow lower triangle); *B. hybridum* D subgenome: Iberian Peninsula SE (purple dot), Iberian Peninsula SW (purple square), Iberian Peninsula NW (purple upper triangle), Iberian Peninsula NE (purple lower triangle). Note that populations with the same PCoA coordinates overlap. **3B:** Genetic structure analysis of the *B. hybridum-B. stacei-B. distachyon* complex. The plots show the percentage of membership of the individual SSR profiles to the assigned genetic groups at K=2, K=5 and K=7. The SSR phenotypes from the *B. hybridum* individuals were separated into the respective S (*B. stacei*-type) and D (*B. distachyon*-type) subgenomes. Population codes correspond to those indicated in Table S3. Major geographic areas are labeled below the graph.

The Structure analysis selected *K*=2 as the best genetic grouping based on ΔK, followed by *K*=5 and *K*=7, whereas the latter groups were selected as optimal groups based on Ln Pr(X|K) values (Fig. 3B). *K*=2 grouped the *B. hybridum* S subgenomic phenotypes with the *B. stacei* group and the D subgenomic phenotypes with the *B. distachyon* group, confirming the hybrid allotetraploid nature of all the studied *B. hybridum* samples. At *K*=5 the *B. stacei* phenotypes formed a group and the *B. distachyon* phenotypes were segregated into three genetic groups, one from the north and two from the south (S1-ALM, GRAZ; S2-YELMO, HORNOS). Notably, i) all the *B. hybridum* D phenotypes from the NE, NW and SW regions were identified with the *B. distachyon* S2 group, whereas none of their respective S phenotypes corresponded to the *B. stacei* group, clustering instead into a fifth genetic group, and ii) all *B. hybridum* D phenotypes from the SE also corresponded to the *B. distachyon* S2 group, whereas the respective *B. hybridum* S phenotypes from ALCA and CABR aligned with the same *B. distachyon* S2 group and the others with the *B. distachyon* S1 group (Fig. 3B). *K*=7 identified a single and unique genetic group of phenotypes for *B. stacei* and three genetic groups (N, S1, S2) for *B. distachyon*. Regarding *B. hybridum*, the S phenotypes from the NE, NW and SW formed a fifth group, the D phenotypes from the NE a sixth group, and the D phenotypes from the NW, SW and most SE a seventh group. *B. hybridum* SE individuals from ALCA and CABR had both S and D phenotypes assigned to the *B. distachyon* S2 group. In addition, SE individuals from YELMO, TISC and HINO S phenotypes were assigned to the *B. distachyon* S1 group, D phenotypes were assigned to the *B. distachyon* S2 group, and the remaining populations individuals S phenotypes were assigned to the *B. distachyon* S1 group (Fig. 3B).

The unrooted ML phylogenetic tree was congruent with the PCO and Structure results and showed the divergence of the main *B. stacei*+S and *B. distachyon*+D clades although a major split also separated a clade of largely divergent *B. distachyon* S Iberian phenotypic lineages (ALME, GRAZ) from the rest. Noticeably, within the *B. stacei*+S clade the *B. hybridum* S lineages from the SE diverged before the *B. stacei* lineages, whereas those from the NW, NW and SW diverged after *B. stacei* (Fig. 4). By contrast, within the *B. distachyon*+D clade the *B. hybridum* D lineages from the NW were closely related to the *B. distachyon* NW lineages, the D lineages from the NE, SW and SE were relatively closely related to the *B. distachyon* NE lineages. Finally, some D lineages from the SE were very close (Jaen) to relatively close (ALIS, HINO, TISC, YELM, CIMB) to the *B. distachyon* S lineages (Fig. 4).

Surprisingly, both *B. hybridum* D and S lineages from the SE (ALCA and CABR) were nested within the *B. distachyon*+D clade though they fell into different subclades. Branch support was high to moderate for the more ancestral lineages of the *B. stacei*+S and *B. distachyon*+D clades and low for most recently evolved lineages within both groups (Fig. 4).

Evolutionary relationships of *B. hybridum* and its progenitor species

The rooted ML plastid tree showed a main split for the highly supported *B. stacei* (S) and *B. distachyon* (D) clades, with both clades showing *B. hybridum* plastotypes nested within (Fig. 5). Most of the analysed *B. hybridum* individuals fell within the S clade. A few individuals fell within the D clade, indicating their respective maternal S and D plastid inheritances. Interestingly, two types of *B. hybridum* D plastotypes could be identified in each of the two highly supported *B. distachyon* diverging lineages. One type corresponded to a clade of southern Iberian plastotypes (D1 plastotypes: Bhyb26-CIMB, Bhyb118-5, Bhyb30-2-LEPE) that were resolved as sister to a clade of early diverging Moroccan *B. distachyon* plastotypes (Bdis52, Bdis53). The second type corresponded to a polyphyletic group of SE Iberian and Moroccan plastotypes (D2 plastotypes: Bhyb54, Bhyb123-1, Bhyb130-1, Bhyb162-1) that fell into different subclades of a more recent and large clade of circum-Mediterranean *B. distachyon* plastotypes (Fig. 5). By contrast, only one clear type of *B. hybridum* S plastotypes could be identified within the S clade, although they showed polyphyletic relationships to the studied *B. stacei* plastotypes none of those divergences were well supported. One of the polyphyletic *B. hybridum* D2 plastotypes (Bhyb54) was nested within the *B. distachyon* EDF+ group, whereas the others were nested within the *B. distachyon* T+S+ group (Sancho *et al.*, 2018), although their support was low to moderate (Fig. 5).

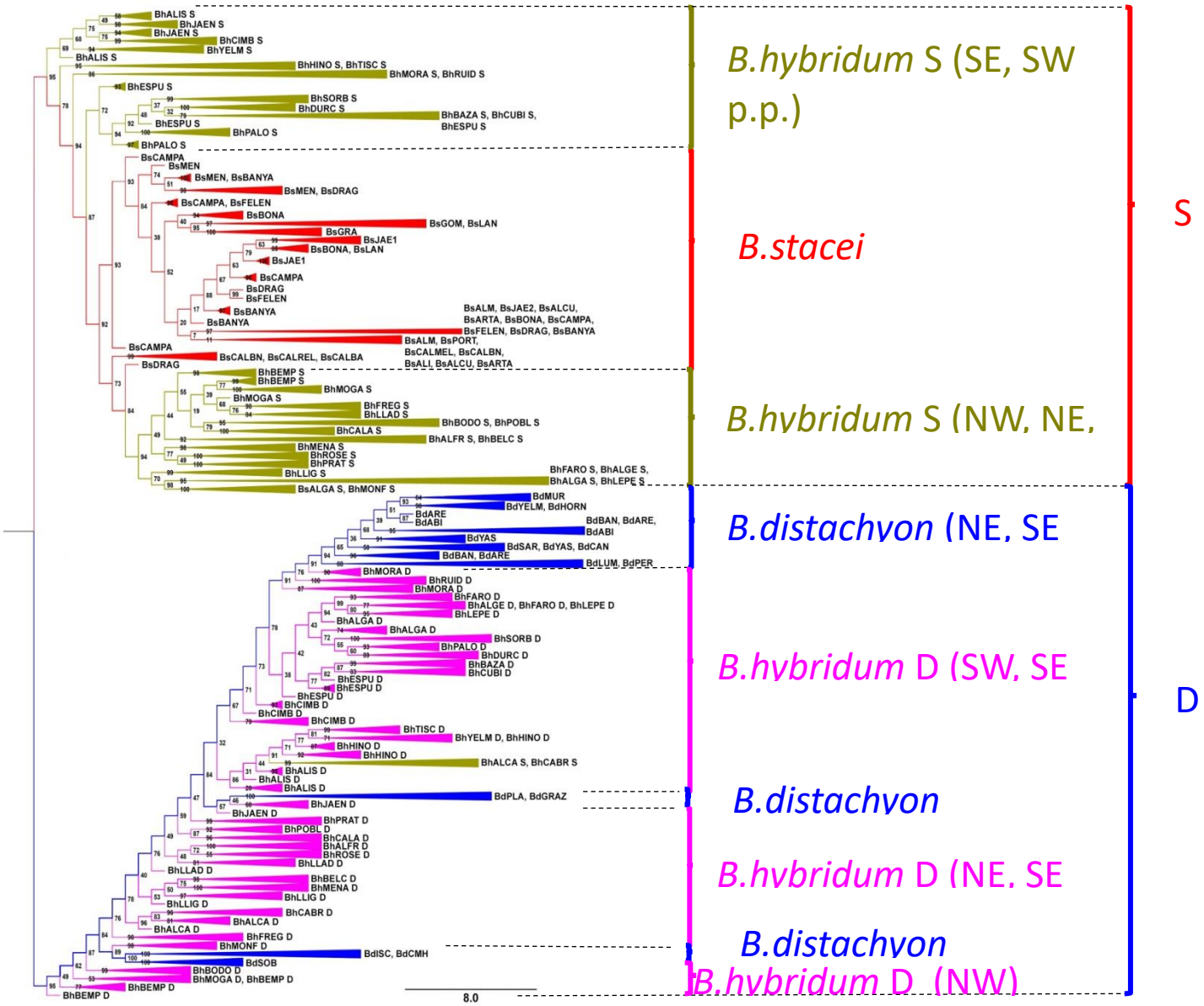


Figure 4. Unrooted Maximum Likelihood SSR phylogenetic tree of the *Brachypodium hybridum-B. stacei-B. distachyon* complex based on 98 alleles. The tree is rooted at the *B. stacei*/*B. distachyon* mid-point split. Branches of zero length and monophyletic populations are collapsed for clarity. Boostrap support values are high to moderate only for the main lineages of each major clade. Population and individual codes correspond to those indicated in Table S3.

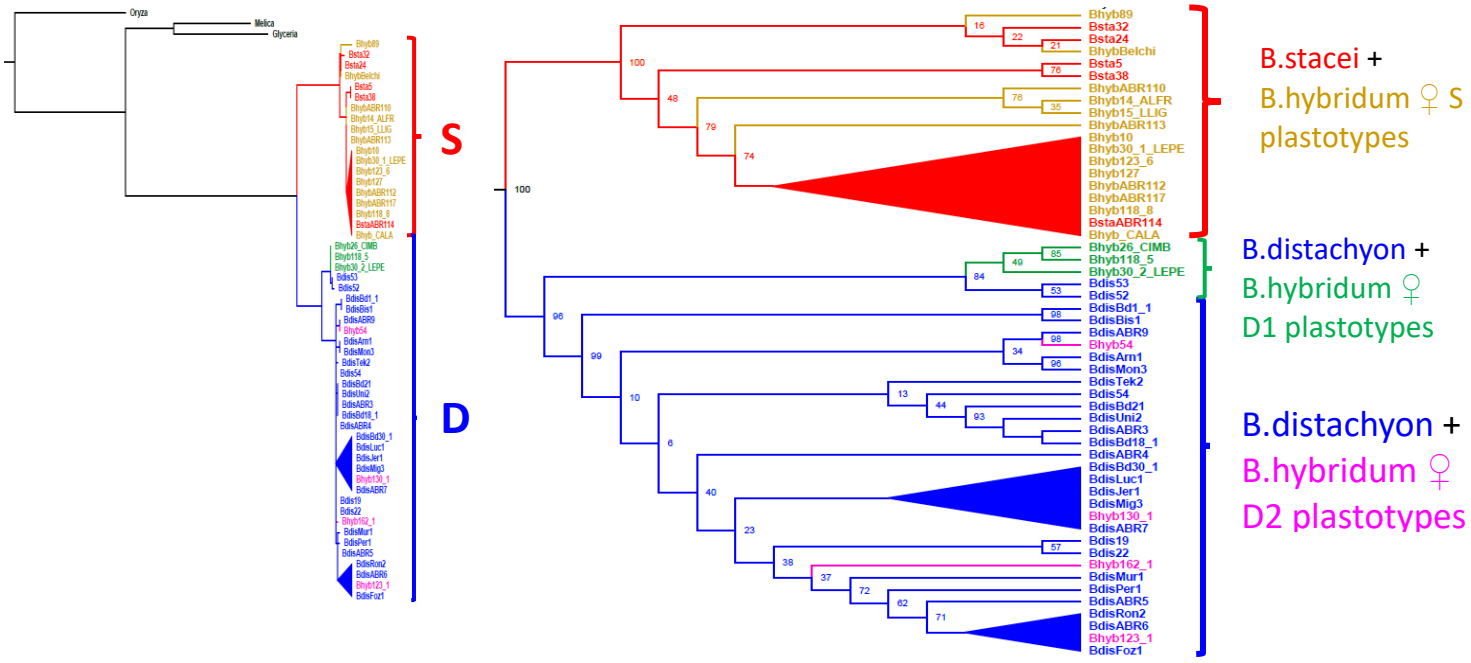


Figure 5. Maximum Likelihood plastid phylogenetic tree of the *Brachypodium hybridum-B. stacei-B. distachyon* complex based on trnLF and ndhF DNA sequences. a) Phylogram showing the main S (*B. stacei*-type) and D (*B. distachyon*-type) *Brachypodium* clades. b) Cladogram showing the three main origins of the *B. hybridum* S, D1 and D2 plastotypes. Collapsed branches correspond to accessions showing the same plastotype. Bootstrap support is indicated on branches. Color codes: *B. stacei* red, *B. distachyon* blue, *B. hybridum* S plastotypes olive, *B. hybridum* D1 plastotypes green, *B. hybridum* D2 plastotypes purple. Individual codes correspond to those indicated in Table S4.

DISCUSSION

Genetic variation of Iberian Peninsula B. hybridum populations is highly structured and follows a South to North cline

Population genetic analysis of allopolyploid species through the separate genotyping of their homeologous subgenomic loci permits a more precise analysis of individuals and populations as well as an accurate reconstruction of their evolutionary history (Catalán *et al.*, 2006). The allotetraploid *B. hybridum* constitutes an exceptional case study due to the possession of largely divergent parental S and D genomes (Catalan *et al.*, 2014; Díaz-Pérez *et al.*, 2018) and a marked subgenomic integrity (López-Alvarez *et al.*, 2012, Giraldo *et al.*, 2012). Most of the S and D loci showed high levels of homozygosity and five of them were fixed in one or the other subgenome, though He was slightly higher in the D loci (Table S5), indicating a more diverse

genetic background inherited from the parental *B. distachyon* genomes. These results agree with the extremely high levels of homozygosis found in the diploid progenitor *B. stacei* (Shiposha *et al.,* 2016) and *B. distachyon* (Marques *et al.,* 2017) lending support to the hypothesis that the current allelic SSR composition of the individual *B. hybridum* S and D subgenomes could be very close to that acquired after the hybridization and genome doubling events. The high levels of inbreeding attributed to *B. hybridum* (Table S1) corroborate that the allotetraploid, like its two progenitor species (Shiposha *et al.,* 2016, Marques *et al.,* 2017) is a selfing plant. Although *B. hybridum* shows open flowers, the rate of outcrossing detected in laboratory and greenhouse experiments (Vogel *et al.,* 2009) or through the analysis of nucleotide diversity in weediness genes (Bakker *et al.,* 2009) was very low. Altogether the evidence supports the predominant selfing nature of the species.

Despite the importance of *B. hybridum* as a model grass for allopolyploidy (Catalán *et al.,* 2014; Gordon *et al.* 2016) and the availability of its reference genome (ABR113 line; Scholthof *et al.,* 2018), to date only two studies, in Tunisia and California, have revealed the genetic variation of its populations (Neji *et al.,* 2015; Bakker *et al.,* 2009). Our survey, conducted in the Iberian Peninsula, one of the native hotspots of genetic diversity for *B. hybridum* (López-Alvarez *et al.,* 2012), has detected considerable genetic variation across the 36 searched populations although most populations are highly homogeneous and exhibit few distinct genotypes (Fig. 1, Table S1). The level of heterozygosis is among the lowest in most populations from the SE and the SW; homozygosis is fixed in 11 of them. However, in these regions the Na is the highest, indicating that the individuals within the populations are genetically diverse, even though they don't cross hybridize (Table S1). Mixed, but not genetically admixed, *B. hybridum* populations occur in southern Iberia, probably caused by multiple founder effects or multiple introductions coupled with a low mating rate due to selfing. Bakker *et al.* (2009) also observed a lack of out-crossing among different *B. hybridum* genotypes in California after several decades of invasions; however, in native southern Iberian Peninsula selfing may have lasted several thousand years (Catalán *et al.,* 2012, 2016). By contrast, the highest levels of heterozygosis were found in the NE Iberian populations of *B. hybridum* (e. g., MENA and CALA, Ho=0.240, Table S1), which have fewer alleles than the SE populations, indicating that sporadic crosses were more frequent in this northern region resulting in more admixed populations.

Genetic relationships and the overall SSR genetic variation in *B. hybridum* are highly structured in the Iberian Peninsula, showing a clear cut separation of southern and northern populations in all the PCoA, NJ and Structure analyses (Figs. 2). Noticeably, genetic substructuring is much higher in the South than in the North. Thus, whereas most of the NE and NW individuals and populations are genetically similar to each other, individuals and populations from SE are more divergent, showing at least three subclusters of populations [1 (ALIS, CIB, JAEN); 2 (HINO, HORN, NEVA, TISC, YELM); 3 (BAZA, CUBI, DURC, ESPU, MORA, PALO, RUID, SORB)] that are more separated in the PCoA genetic space than the NE and NW clusters (Fig. 2). Similar conclusions were obtained from the NJ tree with better supported branches for the three SE lineages than for the NE and NW lineages, and sister relationships of SW1/NE and SW2/ SE2 (Fig. 2). It is also coincident with the Structure plots that identify two main genetic northern+SW1 and SE+SW2 groups for *K*=2 (Fig. 2), four (NW, NE, SE2+SW2 and SE1+SE3+SW1) for *K*=4 and a large substructuring of small groupings coincident with the PCoA subclusters (Fig. 2) for *K*=14 (Fig. S1). The strong geographically structured genetic differentiation observed for *B. hybridum* in the Iberian Peninsula is in agreement with the strong structure of invader *B. hybridum* populations detected in California (Bakker *et al.*, 2009). The absence of significant isolation-by-distance in the Iberian *B. hybridum* populations agrees with similar findings from Neji *et al.* (2015). Nonetheless, in our case the non-significant IBD is a consequence of the high genetic divergence observed between spatially close population clusters in the SE (Figs. 1, 2).

A descendant cline of genetic diversity from South to North and a parallel strong geographic genetic structure has been observed in several Iberian plants and animals (Gómez and Lunt, 2007; Perez-Collazos *et al.*, 2009; Nieto-Feliner, 2011). This has been associated with the abundance of warmer glacial refugia in SE Iberia during the Last Glacial Maximum (LGM) compared to the less abundant or cooler refugia in the North (Perez-Collazos *et al.*, 2009). Nonetheless, detailed phylogeographic and palaeoecological studies have identified new LGM refugia in other Iberian regions (Nieto-Feliner, 2011), such as those from the four main quadrants (SE, SW, NE, NW) plus central Iberia for the model dicot *Arabidopsis thaliana* (Picó *et al.*, 2008). The suitability of these refugia for plants depends on the ecological adaptability of each species and its dispersal capabilities. Thus, for the mesic Mediterranean diploid progenitor species *B. distachyon*, NE Iberia probably constituted a glacial refugium (Marques *et al.*, 2017).

For the warm Mediterranean diploid progenitor species *B. stacei*, two close SE Iberian ranges (coastal, inland) have been identified as potential glacial shelters (Shiposha *et al.,* 2016). *B. hybridum* shows environmental niche preferences intermediate between those of its progenitor species, but closer to those of *B. stacei* (López-Alvarez *et al.,* 2015). Our population genetic study supports the existence of continuous or temporarily larger refugia in the SE Iberian Peninsula for the allotetraploid. The existence of complex *B. hybridum* mixed populations with distinct genotypes is likely a consequence of the easy long distance dispersal of seeds and subsequent colonization and establishment of this annual species (Vogel *et al.,* 2009; Catalan *et al.,* 2016). However, the predominance of mixed *B. hybridum* populations in southern Iberian Peninsula could have also resulted from more favorable climatic conditions in the past, which, in time, contributed to the increased genetic differentiation in the South when compared with the North.

Multiple origins of B. hybridum in the Iberian Peninsula

The recurrent origin of allopolyploid plants is a known phenomenon that has been extensively documented in some native and introduced angiosperms. In most cases the polyploid has undergone profound genomic rearrangements (Soltis *et al.,* 2016), whereas in other, usually young allopolyploid cases, the genomes have remained almost intact (e. g., *Triticum aestivum*, IWGC, 2018), although subgenomic expression changes occurred almost immediately after the allopolyploidization event (Zhang *et al.,* 2014). The hypothesis of the multiple origins of an allopolyploid constitutes a paradox as it is questionable if different hybridizations from distinct parental genotypes of the same progenitor species would led to the same allopolyploid species (Doyle and Sherman-Broyles, 2017). Except for a few exceptions where it has been observed that evolution repeats itself both in natural and in synthetic plants (Soltis *et al.,* 2012), no other recurrently formed allopolyploids have been extensively studied in nature. Our genetic and evolutionary study of the model allopolyploid grass *B. hybridum* has demonstrated that the species was formed at least three different times in nature (Figs. 4, 5) and that more origins probably could be uncovered due to the polyphyletic nature of the D2 and S plastotype groups (Fig. 5). Furthermore, our data show that all of the currently known *B. hybridum* D plastotypes are present only in the western Mediterranean area (Figs. 1, 5) whereas the *B. hybridum* S plastotypes are distributed along the Iberian Peninsula and elsewhere in the circum-Mediterranean region (López-Alvarez *et al.,* 2012). Our population genetics and phylogenetic

analyses support the western Mediterranean origin of at least two *B. hybridum* D1 and D2 plastotypes whose plastomes were likely inherited from ancestors close to current Moroccan and Iberian *B. distachyon* lines (Figs. 5) and the existence of at least a third *B. hybridum* S plastotype whose plastome could have been acquired from a widespread ancestor close to current Mediterranean *B. stacei* lines (Fig. 5). The earlier divergence of the *B. distachyon* Moroccan clade indicates that the *B. hybridum* D1 plastotypes are more ancestral than the *B. hybridum* D2 and S plastotypes, corroborating that the western Mediterranean region is one of the places with the highest accumulation of plant genetic variation in the northern hemisphere across recent Cenozoic evolutionary times (Jakob and Blattner, 2006).

Our plastid findings have been confirmed by the combined analysis of progenitor and allotetraploid species using nuclear SSR data. Our PCoA analysis shows the multi-locus SSR D profiles of some of the *B. hybridum* D genotypes from SE Iberian clustering with the profiles of the *B. distachyon* SE Iberian genotypes, whereas other *B. hybridum* D genotypes profiles from SE Iberia cluster with *B. distachyon* genotypes from N Iberia (Fig. 3). The Structure analysis also shows the clustering of these *B. hybridum* D profiles with the *B. distachyon* S2 and S2-S1 profiles for *K*=5 and *K*=7, respectively, but there is not a close connection with the *B. distachyon* N Iberian profiles (Fig. 3).

The comparative analyses of the *B. hybridum* S and D lineages and progenitor *B. stacei* and *B. distachyon* SSR profiles sheds further light into other potential multiple origins of the allotetraploid. One of the most noticeable findings is that, in contrast to the close genetic relatedness found between the geographic *B. hybridum* D profiles and their corresponding *B. distachyon* profiles, there are no close relationships between the *B. hybridum* S profiles and *B. stacei*. Both the PCoA and the Structure analyses reveal the isolation and uniqueness of the highly homogeneous *B. stacei* genetic group, which is shared by geographically spread populations from S Iberia and the Balearic and Canary Islands, but is absent in almost all the 342 studied Iberian *B. hybridum* individuals (Figs. 1, 4, 5). Interestingly, the ML SSR phylogenetic tree shows an earlier divergence of *B. hybridum* S lineages from SE Iberia, before the *B. stacei* splits, followed by a more recent divergence of *B. hybridum* S lineages from NE, NW and SW Iberia (Fig. 4). This is also corroborated by their respective separations in the PCoA and the Structure *K*=5 and *K*=7 (Fig. 3) plots. These data suggest that two alternative hybridizations may have occurred regarding the contribution of the progenitor *B. stacei*

genome, one of which is apparently more ancestral, resulting in the SE Iberian S subgenomes, and other more recent connected with the origins of the NE, NW and SW Iberian S subgenomes.

Frequent seed-mediated long distance dispersals (LDD) have been inferred for the widely studied *B. distachyon* and *B. hybridum* annuals (Vogel *et al.*, 2009; Catalán *et al.*, 2016) and have been evidenced by the colonization success of *B. hybridum* on non-native continents (e. g., North America; Bakker *et al.*, 2009; Catalán *et al.*, 2012, 2016). Our plastid analysis has proved the existence of mixed *B. hybridum* populations in southern Iberia which contain individuals with different hybrid origins (e. g., LEPE and Bhyb118 (Almeria, Cabo de Gata) with D1 and S plastotypes; Figs. 1, 5). Nonetheless, the strong genetic relationships observed between the SE Iberian *B. hybridum* D genotypes with the endemic SE Iberian and Moroccan (Fig. 5) *B. distachyon* genotypes would favor the hypothesis of a local origin of their *B. hybridum* ancestor in the southern Iberian Peninsula-western Mediterranean area. This requires an assumption of a potential extinction of an old (or as yet unsampled) *B. stacei* lineage which apparently has only survived in the allopolyploid hybrid descendants. The evolutionary fate of the newer *B. stacei* lineage that gave rise to the NE, NW and SW Iberian *B. hybridum* S genotypes (Fig. 5) is less clear. One possibility is that the multiple ancestors of the current NE, NW and SW Iberian *B. hybridum* individuals originated in northern Iberia from crosses of *B. distachyon* parents similar to current northern Iberian *B. distachyon* genotypes with a potentially extinct (or as yet unsampled) recent *B. stacei* lineage. However, this outcome could not be elucidated with the current data.

Genotypic diversity is considered to be the substrate of phenotypic diversity. Yet the impact of merging distinct parental genotypes in the allopolyploid outcome can have dramatically opposite speciation consequences. Bidirectional crosses or even the same directional cross of progenitor species have originated different allotetraploids in *Aegylops* (Meimberg *et al.*, 2009). For *Tragopogon*, recurrent crosses in either directional have created the same allotetraploid species (Soltis *et al.*, 2012). Our study of *B. hybridum* aligns with the latter case (Catalán *et al.*, 2016). Moreover, comparative phenotypic analyses of *B. hybridum* and its progenitor *B. distachyon* and *B. stacei* parents across their respective native Mediterranean regions reveals discriminant phenotypic differences within each of the diploids, but not within the allotetraploid for the same set of morphological traits (López-Alvarez *et al.*,

2017). Nonetheless, phenotypic analyses separate *B. hybridum* individuals collected in southern and northern Spain, with the former being slightly taller and having more culm nodes and a larger caryopsis than the latter (López-Alvarez *et al.*, 2017). These features reflect the distinct genetic composition of the geographic *B. hybridum* groups detected in this study. Our study has demonstrated the occurrence of at least three natural multiple origins of the allopolyploid *B. hybridum*; all the different genotypes are currently present in the Iberian Peninsula, and one of these plausible hybridization and genome doubling events took part in the western-Mediterranean region. The detection of a large genetic variation within the Iberian Peninsula *B. hybridum* populations has untapped an invaluable source of diverse germoplam for future analysis of allopolyploid formation and genomic expression in the *B. hybridum – B. distachyon – B. stacei* grass model complex.

FUNDING

This work was supported by the Spanish Ministry of Economy and Competitivity (Mineco) CGL2016-79790-P and University of Zaragoza UZ2016_TEC02 grant projects. VS was funded by a Russian Ministry of Education and Science fellowship. PC and IM were partially funded by a European Social Fund and Aragón Government Bioflora grant.

ACKNOWLEDGEMENTS

We thank Karen Scholthof for proofreading this manuscript, Maria Luisa López-Herranz and Ruben Sancho for laboratory and plastome assembling assistance, respectively, and John Vogel and the Department of Energy-Joint Genome Institute (DOE-JGI, USA) for facilitating access to some of the *Brachypodium hybridum* genome data.

LITERATURE CITED

Baduel P, Bray S, Vallejo-Marin M, Kolár F, Yant L. 2018. The "Polyploid Hop": shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Frontiers in Ecology and Evolution* **6**:117. doi: 10.3389/fevo.2018.00117.

Bakker EG, Montgomery B, Nguyen T, Eide K, Chang J, Mockler TC, Liston A, *et al*. 2009. Strong population structure characterizes weediness gene evolution in the invasive grass species *Brachypodium distachyon* . *Molecular Ecology* **18**: 2588-2601.

Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* **210**: 391–398.

Buckler ES 4th, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. *Genetic Resources* **77(3):** 213-8.

Catalán P, López-Alvarez D, Díaz-Pérez A, Sancho R, López-Herranz ML. 2016. Phylogeny and evolution of the genus *Brachypodium*. In: Vogel J ed. *Genetics and genomics of Brachypodium.* pp. 9-38. Series Plant Genetics and Genomics: Crops Models. Springer. New York, 9-38.

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

Catalán P, Segarra-Moragues JG, Palop-Esteban M, Moreno C, Gonzalez-Candelas F, 2006. A Bayesian approach for discriminating among alternative inheritance hypotheses in plant polyploids: the allotetraploid origin of genus *Borderea* (Discoreaceae). *Genetics* 172:1939-1953.

Catalán P, Chalhoub B, Chochois V, Garvin DF, Hasterok R, Manzaneda AJ, Mur LAJ, Pecchioni N, Rasmussen SK, Vogel JP, Voxeur A, 2014. Update on the genomics and basic biology of *Brachypodium. Trends in Plant Science* **19:** 414-418.

Chybicki IJ, Burczyk J. 2009. Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity* **100:** 106-13.

Chybicki IJ, Oleksa A, Burczyk J. 2011. Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data. *Heredity* **107:** 589-600.

Díaz-Pérez A, López-Álvarez D, Sancho R, Catalan P. 2018. Reconstructing 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.

Dinh Thi VH, Coriton O, Le Clainche I, Arnaud D, Gordon SP, Linc G, Catalan P, Hasterok R, Vogel JP, Jahier J, *et al.*2016. Recreating stable Brachypodium hybridum allotetraploids by uniting the divergent genomes of *B. distachyon* and *B. stacei*. PLoS One 11:e0167171.

Doyle J, Sherman-Broyles S. 2017. Double trouble: taxonomy and definitions of polyploidy. *New Phytologist* **213(2):** 487-493.

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* **4:** 359-361.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14:** 2611-2620.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1:** 47–50.

Giraldo P, Rodriguez-Quijano M, Vazquez JF, Carrillo JM, Benavente E. 2012. Validation of microsatellite markers for cytotype discrimination in the model grass *Brachypodium distachyon. Genome* **55:** 523-537.

Gómez A, Lunt DH. 2007. Refugia within refugia: Patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss, S. and Ferrand, N. (eds.), *Phylogeography of southern European refugia*. Dordrecht: Springer, 155-188.

Gordon SP, Contreras-Moreira B, Daniel Woods D, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin A, Schackwitz W, Tyler L, Martin J, Lipzen A, Dochy N, Phillips J, Barry K, Geuten K, Juenger TE, Amasino R, Caicedo AL, Goodstein D, Davidson P, Mur L, Figueroa M, Freeling M, Catalan P, Vogel JP. 2017. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with phenotypic variation. *Nature Communications* **19; 8(1):** 2184.

Goudet J. 2001. FSTAT, version 2.9. 3. A program to estimate and test gene diversities and fixation indices. Lausanne: Lausanne University.

IWGSC. 2018. Shifting the limits in wheat research and breeding using a fully annotated reference genome**.** *Science* Aug **17; 361**(6403).

Jacob SS, Blattner FR. 2006. A chloroplast genealogy of *Hordeum* (poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution* **23(8):** 1602-1612.

Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:** 1801–1806.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali S, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al.*, 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473:** 97–100.

Langella O. 2000. POPULATIONS 1·2: population genetic software, individuals or population distance, phylogenetic trees. http://bioinformatics.org/~tryphon/populations/.

Li YL, Liu JX. 2018. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods.*Molecular Ecology Resources* **18(1):** 176-177.

López-Alvarez D, López-Herranz ML, Betekhtin A, Catalán P. 2012. A DNA barcoding method to discriminate between the model plant *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). *PLoS One* **7:** e51058.

López-Alvarez D, Manzaneda A J, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, Mur LAJ, Caicedo AL, Hazen SP, 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: 1-16.

Madlung A, Wendel JF. 2013. Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenetic and Genome Research* **140:** 270-285.

Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27:** 209-220.

Marques I, Montgomery SA, Barker MS, Macfarlane TD, Conran JG, Catalán P, Rieseberg LH, Rudall PJ, Graham SW. 2016. Transcriptome-derived evidence supports recent polyploidization and a major phylogeographic division in Trithuria submersa (Hydatellaceae, Nymphaeales). *New Phytologist* **210:** 310-323.

Marques I, Shiposha V, López-Álvarez D, Manzaneda A, Hernández P, Olonova M, Catalan P. 2017. Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*. *BMC Evolutionary Biology* **17:** 139.

Marques I, Loureiro J, Draper D , Castro M, Castro S.2018. How much do we know about the frequency of hybridisation and polyploidy in the Mediterranean region? *Plant Biology* **20 Suppl 1:** 21-37.

Meimberg H, Rice KJ,Milan NF, Njoku CC,McKay JK. 2009. Multiple origins promote the ecological amplitude of allopolyploid Aegilops (Poaceae). *American Journal of Botany* **96:** 1262-1273.

Mur LAJ, Allainguillaume J, Catalán P, Hasterok R, Jenkins G, Lesniewska K, Thomas I, Vogel J. 2011. Exploiting the *Brachypodium* tool box in cereal and grass research. *New Phytologist* **191:** 334-347.

Nei M, Chesser RK. 1983. Estimation of fixation indices and gene diversities. *Annals Human Genetics* **47:** 253-259.

Neji M, Geuna F, Taamalli W, Ibrahim Y, Chiozzottoo R, Abdelly C, *et al.*,2015. Assessment of genetic diversity and population structure of Tunisian populations of Brachypodium hybridum by SSR Markers. *Flora, Morphology, Distribution, Functional Ecology of Plants* **216:** 42-49.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Nieto Feliner G. 2011.Southern European glacial refugia: A tale of tales. *Taxon* 60: 365-372.

Paterson AH, Wendel JF. 2015. Unraveling the fabric of polyploidy. *Nature Biotechnology* **33:** 491-493.

Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Molecular Ecology Notes **6:** 288–295.

Pérez-Collazos E, Sánchez-Gómez P, Jiménez JF, Catalán P. 2009. The phylogeographic history of the Iberian steppe plant *Ferula loscosii* (Apiaceae): a test of the Abundant-centre hypothesis. *Molecular Ecology* **18:** 848-861.

Picó FX, Méndez-Vigo B, Martínez-Zapater JM, Alonso- Blanco C. 2008. Natural genetic variation of *Arabidopsis thaliana* is geographically structured in the Iberian Peninsula. *Genetics* **180:** 1009-1021.

Pritchard JK, Stephens M, Donnelly P.2000. Inference of population structure using multilocus genotype data. Genetics **155:** 945-959.

Rohlf FJ. 2002. NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.11a. Applied Biostatistics Inc., New York.

Sancho R, Cantalapiedra CP, López-Álvarez D, Gordon SP, Vogel JP, Catalan P, Contreras-Moreira B. 2018. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist* **218:** 1631-1644.

Scholthof KB, Irigoyen S, Catalan P, Mandadi KK. 2018. *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* **30:** 1673-1694.

Shiposha V, Catalan P, Olonova M, Marques I. 2016. Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands. *PeerJ* **4:** e2407.

Soltis DE, Buggs R, Barbazuk B, Chamala S, Chester M, Gallagher JP, Schnable PS, Soltis PS. 2012. The early stages of polyploidy: Rapid and repeated evolution in *Tragopogon.* In: Soltis PS, Soltis DE (eds.), Polyploidy and genome evolution. Springer, Berlin, Germany, 271–292.

Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: Pitfalls and paths to a paradigm. *American Journal of Botany* **103(7):** 1146-1166.

Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution* **53:** 117-137.

Spoelhof JP, Soltis PS, Soltis DE. 2017. Pure polyploidy: closing the gaps in autopolyploid research. *Journal of Systematics and Evolution* **55:** 340–352.

Stebbins GL. 1985. Polyploidy, hybridization, and the invasion of new habitats. *Annals Missouri Botanical Garden* **72:** 824-832.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M,Pysek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* **109:** 19-45.

Tyler L, Fangel JU, Fagerström AD, Steinwand MA, Raab TK, Willats WGT. 2014. Selection and phenotypic characterization of a core collection of *Brachypodium distachyon* inbred lines. *BMC Plant Biology* **14:** 25. doi:10. 1186/1471-2229-14-25

Van de Peer Y, Mizrachi E, Marchal K.2017. The evolutionary significance of polyploidy. *Nature Review Genetics* **18(7):** 411-424.

Vogel JP, Tuna M, Budak H, Huo NX, Gu YQ, Steinwand MA. 2009. Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. *BMC Plant Biology* **9:** 88.

Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, *et al.*,2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463:** 763-768.

Zhang H, Zhu B, Qi B, Gou X, Dong Y, Xu C, Zhang B, *et al.* 2014. Evolution of the BBAA component of bread wheat during its history at the allohexaploid level. *Plant Cell* **26:** 2761-2776.

CHAPTER 4

Glacial refugia and postglacial range shifts of the widespread Eurasian *Brachypodium sylvaticum* and *B. pinnatum* and the western Mediterranean *B. phoenicoides* grasses: insights from environmental niche modeling analysis and conservation status of relict Siberian populations of *B. sylvaticum*.

Valeriia Shiposha[1,2], Dimitry S. Feoktistov[1], Isabel Marques[2,§], Pilar Catalán[1,2,3,*], Marina Olonova[1,*]

[1.] Department of Botany, Institute of Biology, Tomsk State University, Lenin Av. 36, Tomsk 634050, Russia.
[2.] Department of Agriculture and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain.
[3.] Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza E-50059, Spain
[§.] current address: cE3c - Centre for Ecology Evolution and Environmental Changes, Faculty of Sciences, University of Lisbon, Lisbon, Portugal

*Corresponding authors: olonova@list.ru, pcatalan@unizar.es

Abstract

Species distribution models are powerful analytical tools to investigate the environmental dynamics of organismic niche distributions within spatio-temporal and evolutionary frameworks. They have been also used to predict past glacial refugia and range distributions shifts associated to climate and anthropogenic changes during the Late Quaternary glacial and postglacial phases and to infer niche features, niche inheritance and niche competition among closely related taxa. Here we reconstruct the environmental niche models (ENM) of three perennial species of the grass model genus *Brachypodium* that originated in the Ionian-Pleistocene transition, the widespread Eurasian *B. sylvaticum* and *B. pinnatum* and the western Mediterranean *B. phoenicoides*. Using current occurrence records from 1656 sites and data from 19 bioclimatic variables we built ENMs for each species at three geoclimatic scenarios, the Last Glacial Maximum (LGM, ~20 kya), the Mid-Holocene (MH, ~6 kya) and the present, with Maxent. Climatic envelope data indicate that *B. pinnatum* is a cold-adapted continental species showing a large distribution range in Eurasia, although it has a unique niche in the central-northern Asian range, *B. sylvaticum* is a temperate oceanic species showing a main distribution in the half western humid Eurasia and in the East Asia Pacific coast, and *B. phoenicoides* is a warm Mediterranean species showing a restricted distribution if the western range of the basin. The LGM models predicted larger distributions ranges for the three species than MH and current models. The analyses detected cryptic glacial refugia for *B. pinnatum* in central-northern Europe and in central-northern Siberia, continuum glacial refugia for *B. sylvaticum* in coastal, islands and lowland areas of western Europe (Atlantic ocean) and eastern Asia (Pacific ocean), and lowered coastal Mediterranean refugia for *B. phoenicoides* in the western Mediterranean region. The niche distribution of the recently evolved allotetraploid *B. phoenicoides* is smaller than that of its putative diploid *B. sylvaticum*-type progenitor species, indicating that the allopolyploid does not outcompete its parent in its native range. Moreover, the *B. phoenicoides* niche partially overlaps with that of the Eurosiberian *B. sylvaticum* although its features are typically Mediterranean, suggesting that *B. phoenicoides* could have inherited its Mediterranean-type niche from its second putative Mediterranean *B. distachyon*-type progenitor species. The widespread *B. sylvaticum* is currently a rare and protected species in western Siberia. Niche distribution modelling predicts a range contraction shift of the plant from the LGM to the present in the area; however, the decreased range distribution could have resulted from either climatic or anthropogenic change. The longevity of the species and its capability to propagate could lower its potential risk of extinction if the Siberian black taiga and broad-leaved forests that constitute its natural niche in the region are preserved.

Keywords: *Brachypodium sylvaticum – B. pinnatum – B. phoenicoides*, climatic scenarios, environmental niche modeling, glacial refugia, range inheritance, Siberian relict habitats.

Running title: ENMs, glacial refugia and relict habitats of Eurasian *Brachypodium* grasses

Introduction

Environmental niche modeling (ENM) analysis has been extensively used to predict the potential distribution ranges of a large number of organisms and to test the suitability of their environmental niches under different ecological and evolutionary hypotheses (Waltari et al. 2007; Araujo & Peterson 2012). ENMs have contributed, among other issues, to discern speciation processes caused by ecological adaptation and niche divergence from those caused by evolutionary history (McCormack et al., 2010), to identify potential hybridization areas for allopolyploid formation (López-Alvarez et al. 2015), to detect changes in the distribution ranges of threatened species (Guisan et al. 2005), and to reveal shifts in niche distribution ranges through current and past climatic scenarios (Peterson and Nyari, 2008). Spatially explicit environmental data and models allow for large-scale tests of whether speciation is associated with niche divergence or whether closely related species tend to be similar ecologically (niche conservatism) (Smith and Donoghue, 2010; McCormack et al., 2010). Compared time course niche modelling analyses further allow us to estimate the influence of past climatic events in the inherited ecological tolerances and the survival of plant species (Geber and Griffen, 2003; McIntyre, 2012).

Historical environmental scenarios have been used to predict the fate of populations and species in recent Quaternary times (Fordham et al. 2014). The distribution ranges of Palaearctic species experienced a series of contractions and expansions according to the prevailing environmental conditions of the glacial and interglacial Pleistocene cycles and their adaptations to these environmental changes (Hewitt 2004; Steward et al. 2010). A number of isotopic, paleontological, genetic and evolutionary studies have corroborated the existence of southern European and western Asian plant glacial refugia (e. g. Iberian, Italian and Balkan peninsulas, Caucasus) (Taberlet et al. 1998, Feliner 2011) and also of the less known Central Asian refugia (e. g., Khazasthan Highlands, Altai mountains) (Arkhipov et al. 1999; Friesen et al. 2016). Amounting evidence suggests, however, that cryptic glacial refugia for cold temperate plants also occurred in glaciated North Eurasian latitudes (e. g., Scandinavia, Northeastern Siberia (Yakutia) as well as a oceanic/continental longitudinal gradient from more humid to more drier vegetation landscapes during the glacial phases (Stewart & Lister 2001; Steward et al. 2010).

Environmental niche modeling has also been used to disentangle the potential expansions of recently formed neopolyploid species and their niche divergence with respect to those of their progenitor taxa (McIntyre, 2012; Lopez-Alvarez et al. 2015). It has been hypothesized that plant polyploid species have broader ecological niches, occupy broader distributions and could be more invasive than diploid or lower-ploidy species due to their higher genomic heterogeneity and increased ecological amplitude (te Beest et al., 2012). By contrast, other studies have shown that the niche breath of plant polyploids does not exceed that of their related diploids (Martin & Husband, 2009, McIntyre, 2012) or could be even smaller (Lopez-Alvarez et al. 2015). Although plant invasiveness is 20% more likely for polyploids compared to diploids (Pandit et al. 2011) and is facilitated by their higher adaptive plasticity (Hanh et al. 2012) it also depends on habitat suitability for any particular species (Linder at al. 2013).

Simple ecological models have suggested that niche differentiation is required for long term co-existence of diploids and polyploids (Levin 1975), whereas others indicate that niche overlap is more frequent between diploids and recently derived neopolyploids than between distinct diploids (McIntyre, 2012). New data support the utility of the ENM approaches for testing the tempo and the environmental niche availability for multiple founder events of plant allopolyploids (Lopez-Alvarez et al. 2015).

Brachypodium has been recognized as a model system genus for functional genomics of grasses and monocots (Scholthof et al. 2018). This small genus contains ca. 20 species of which three are annuals and the rest perennials, with half of them being diploids and the other half allopolyploids (Catalán et al. 2016; Diaz-Perez et al. 2018). *B. sylvaticum*, the best-known perennial species of the genus, was recently selected as a model plant for perenniality (Scholthof et al. 2018). Genomic and transcriptomic resources are available for *B. sylvaticum*, including its reference genome (*B. sylvaticum* Ain1) and a second resequenced line (*B. sylvaticum* Sin1) (Scholthof et al. 2018). The species shows the largest native distribution of all *Brachypodium* taxa, being spread across Eurasia. All cytogenetically studied populations indicate that it is a diploid species (Catalan et al. 2016; Diaz-Perez et al. 2018). Other phylogenetically close perennial species are *B. pinnatum* and *B. phoenicoides*. *B. pinnatum* has the second largest native range distribution among the *Brachypodium* taxa, covering most of the Eurasian supercontinent. It includes diploid and tetraploid cytotypes, though all of them have been classified under the same species (Schippmann 1991; Catalán et al. 2016) and are, apparently, undistinguisable, based on gross morphological traits. *B. phoenicoides* is a tetraploid species with a native distribution range restricted to the western Mediterranean region (Catalan et al. 2016; Diaz-Perez et al. 2018). *B. sylvaticum* is the only invasive plant of the three species that has been introduced and is spread in western N America (Rosenthal et al. 2008). The three taxa differ from each other in their phenotypic, karyotypic and ecological features (Betekhtin et al. 2014; Catalan et al. 2016). *B. sylvaticum* is a nemoral species, characterized by its slender habit and rhizomes, nodding panicle, densely hairy indumentum, long-awned lemma, and selfing reproductive system. By contrast, the other two species are strong rhizomatose outbreeders. The mesic and open grassland *B. pinnatum* shows flat leaves, erect panicles and spikelets and short awns whilst *B. phoenicoides*, adapted to dry places, is glabrous and presents partially inrrolled leaves, semi-patent twisted spikelets and awnless lemmas (Catalan et al. 2016). Dated phylogenetic studies indicate that the three species belong to a recently evolved clade of core perennial *Brachypodium* taxa, and that they could have originated in the Ionian-Upper Pleistocene transition (Diaz-Pérez et al. 2018).

Despite the considerable knowledge gained in the genomic study of the perennial *Brachypodium* model grasses, environmental modelling analyses have not been performed on them yet. In this study we have constructed current and past niche distribution models of the three taxa in their respective native ranges aiming to test: 1) whether the Eurasian *B. sylvaticum* and *B. pinnatum* had similar overlapping distribution niches and niche identities in current conditions, and similar niche contractions and expansions in current, Mid-Holocene (MH) and Last Glacial Maximum (LGM) scenarios; 2) if the oceanic *B. sylvaticum* and the continental *B.*

pinnatum could have survived in cryptic northern Eurasian glacial refugia and the xeric *B. phoenicoides* in western Mediterranean refugia; 3) if the diploid-polyploid *B. pinnatum* and polyploid *B. phoenicoides* could have had greater niche expansions than the diploid *B. sylvaticum*; 4) if the historical niche amplitudes of the three species could be related to their inferred dates of origin; and 5) if the widespread *B. sylvaticum* is a rare glacial relict in threatened northern Central-Asian populations.

Material and Methods

Species, study areas and data collection

Environmental niche modelling (ENM) analyses were conducted for the two widespread Eurasian species *Brachypodium sylvaticum* and *B. pinnatum* and for the western Mediterranean species *B. phoenicoides*. *B. sylvaticum* is known as a diploid cytotype (2n=2x=18); its native Palearctic area ranges from Macaronesia (West) to New Guinea (East) and from Scandinavia and Siberia (North) to northern Africa and Malaysia (South), although some of the East Asian and Malaysian populations may correspond to different infraspecific or close taxa (Catalán et al., 2016; Scholthof et al. 2018). *B. pinnatum* s. l. is a large aggregate that includes two diploid [2n=16 (x=8); 2n=18 (x=9)] and one tetraploid (2n=28) cytotypes, that have not been separated taxonomically (Schippmann 1991; Catalan et al. 2016). Its distribution range greatly overlaps with that of *B. sylvaticum* though it is absent in both SW Europe–NW Africa–Macaronesia and Far East Asia but reaches higher latitudes in Scandinavia and central Siberia than *B. sylvaticum* (Catalán et al. 2016). *B. phoenicoides* is known as a tetraploid cytotype (2n=28) and its distribution range is circumscribed to the W Mediterranean region (Schippmann 1991; Catalan et al. 2016). A total of 907 Old World records of *B. sylvaticum*, 658 of *B. pinnatum* and 91 of *B. phoenicoides*, taken from studied herbarium specimens (AAU, B, C, FI, G, JACA, LE, MA, MW, NS, NSK, S, SEV and TK) and from reliable literature sources (Tsvelev 1964; Hulten 1971; Probatova 1985; Olonova 1990; Schippmann 1991) were used in the niche modeling study. *B. sylvaticum* occurrence data from alien populations of NW America and other regions were not used in the analysis.

In order to reduce the likely effects of spatial autocorrelation of localities due to uneven sampling bias of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides*, both the GIS data and the environmental data from 19 Bioclim climatic variables were filtered using the program SDMtoolbox v.1.1 (Brown, 2014). It was accomplished by retaining single occurrence records of each species in a range of 1 to 20 km, according to climate heterogeneity, using the Spatially Rarefy Occurrence Data tool implemented in SDMtoolbox in ArcGIS 9 by imposing 5 heterogeneity classes (Brown, 2014).

Environmental layers, current and past envelopes and occurrence data, and niche modelling

Current and past ENMs were based on environmental layers that consisted of 19 temperature and precipitation Bioclim variables (Table 1), downloaded from the WorldClim dataset

(http://www.worldclim.org). Current climatic scenarios were reconstructed at a scale of 30 arc seconds (c. 1 Km^2), whereas past climate scenarios for mid-Holocene (MH; ~6 ka) were reconstructed at the same scale and those for Last Glacial Maximun (LGM; ~21 ka) at 2.5 arc minutes (c. 5 X 5 Km) scale (Hijmans et al., 2005b). Although the climatic niche variables describe environmental conditions rather than resources they show less spatial heterogeneity and higher correlations between close cells than ecological variables (Soberón, 2010). Additionally, they could be measured over broad geographic areas and are more predictive that ecological variables because they have lower variability at small GIS scales (McCormack et al., 2010; Nakazato et al., 2010; Lopez-Alvarez et al. 2015).

The occurrence data of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* under current conditions were projected to the past MH and LGM climatic envelopes to test the hypothesis that niche contractions and expansions across different temporal Quaternary windows could reveal the glacial refugia occupied by each species during the last glacial phases and the potential expansions during postglacial times. We also tested if the last events could have shaped the current distribution of these species and how their distinct biological features influenced their current and past niches. We used past general circulation models (GCMs) to explore the potential variability of ENM scenarios during those periods. The MH envelope was simulated from the Community Climate System Model version 4 (CCSM4) (Collins et al., 2006) and the LGM envelope from the Model for Interdisciplinary Research on Climate (MIROC-ESM) (K-1_model_developers, 2004). The LGM-MIROC-ESM climatic simulation was selected over other LGM climatic models as it provides a more convincing pattern of the LGM climate (Tarkhnishvili et al., 2012).

Current and past climate data from each occurrence point were extracted using DIVA-GIS v7.5 (Hijmans et al., 2005a). To avoid overfitting of the data (Phillips, et al., 2006), we removed highly correlated variables (R>0.7; Table S1) and defined species-specific geographic backgrounds using SDMtoolbox v1.1. The 19 bioclimatic variables were reduced to seven variables (Bio2, 7, 8, 9, 15, 18, 19) for *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* (Table 1). The estimated background areas of each species for each temporal envelope were generated using the buffered local adaptive convex-hull tool of SDMtoolbox v1.1, imposing a buffer of 150 km and an alpha parameter value of 18 (Brown, 2014). These figures appropriately represent the potential dispersal distance of the studied species (Catalán et al. 2016) and the search distance used to define the convex hull shape and size around each set of locality data of each species-climatic envelope, respectively.

We used Maxent v.3.3.3 (Phillips et al., 2006) to generate the species distribution models of the three species. The program estimates the optimal potential distribution from a maximum entropy probability distribution from presence-only data (Elith et al., 2006). We selected Maxent because it performs better than other alternative modelling methods for this type of data (Elith et al., 2006). The selected variables of each species and climatic envelope were used as predictors to calibrate the distribution models in Maxent. The occurrence data in each case were randomly split into training (75%) and test (25%) data for model evaluation. Maxent was

run according to the following settings: auto features (feature types were automatically selected depending on the training sample size), logistic output format, regularisation multiplier set to 1, maximum iterations set to 1000, convergence threshold set to 0.00001 and maximum number of background points set to 10,000. The performance of the models was evaluated by a threshold-independent Receiver-Operating Curve (ROC) analysis, where the Area Under the ROC Curve for presence only data (AUC) measures the ability of the model to identify presences more accurately than random prediction. Ten subsample replicates were performed for each model using default settings. We used the average prediction from all the model replicates to construct the species ENM distribution maps. We also ran a jackknife analysis to measure the percentage of contribution and the importance of the variables to each of the assayed models (Phillips et al., 2006).

Comparing environmental niches and niche overlaps

We compared the current environmental conditions of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* aiming to find significant differences among the three species, and particularly within the two widespread Eurasian taxa, for the 19 bioclimatic variables studied plus three additional geographic variables (Latitude. Longitude, Altitude). We also analysed the potential intraspecific environmental variation within the largely spread *B. pinnatum* using the same 22 variables. We performed the analyses using one-way ANOVA tests and subsequent Tukey multiple comparisons tests using STATISTICA v. 9 (StatSoft Inc.; http://www.statsoft.com/). Principal Component Analyses (PCA) of extracted data were also performed to discriminate the three species in the environmental space and to identify the variables that most contributed to their differentiation, and to discriminate potential intraspecific groups within *B. pinnatum*.

We used a method for niche model comparisons that uses niche overlap and niche identity test. We quantified the amounts of niche overlap among the three studied species imposing a threshold of 10% of presence value for the occurrence of the species in a particular range. Given that each of the three species showed a different specific background area but with overlapping zones, we united them and created a common area to run the comparative niche pairwise tests (Eurasia). We employed ENMtools v1.4.2 (Warren et al. 2010) to calculate the niche overlap and to assess significant niche differences for all species pairs under current climatic conditions. Pairwise niche overlap was assessed among the three species using Schoener's D metric (Schoener, 1968), which measures the proportional similarity of two distributions for inhabiting particular regions (cells) as an indicator of niche overlap based upon species abundance in those locations. Tests for niche differentiations were done through the niche identity test, which tests the null hypothesis of niches being identical by randomly generating a distribution of niche overlap values with unknown species identities, generating a Hellinger'-based I index. Both indices range from 0 (complete divergence/no overlap) to 1 (high similarity/complete overlap). The compared niches are considered significantly different if the observed value of niche overlap (Schoener's D) is less than the niche overlap value from 95 ($P < 0.05$) or 99 ($P < 0.01$) of the niche overlap values estimated from the random pseudoreplicates (Hellinger's I) (Warren et al. 2010).

Results

Environmental features of B. sylvaticum, B. pinnatum and B. phoenicoides

Our statistical analysis found significant environmental differences for the three studies species in seven of the 22 analysed variables ($P<0.001$; Tables 1, S2) under current climatic conditions (Annual Mean Temperature, Temperature Seasonality , Min Temperature of Coldest Month, Mean Temperature of Driest Quarter, Mean Temperature of Warmest Quarter, Mean Temperature of Coldest Quarter, Longitude) (Fig. 1). There was a major differentiation between the Mediterranean *B. phoenicoides* and the Eurasian *B. sylvaticum* and *B. pinnatum*, but also a clear distinction between the latter species, discerning the more oceanic *B. sylvaticum* from the more continental *B. pinnatum*. Six variables significantly discriminated *B. phoenicoides* from *B. sylvaticum* and *B. pinnatum* (Isothermality, Max Temperature of Warmest Month, Precipitation of Driest Month, Precipitation of Driest Quarter, Precipitation of Warmest Quarter, Latitude), four significantly discriminated *B. sylvaticum* from *B. pinnatum* and *B. phoenicoides* (Mean Diurnal Range, Annual Precipitation, Precipitation of Wettest Month, Precipitation of Wettest Quarter), and three significantly discriminated *B. pinnatum* from *B. sylvaticum* and *B. phoenicoides* (Temperature Annual Range, Mean Temperature of Wettest Quarter, Precipitation of Coldest Quarter), indicating that each species has unique niche preferences (Table 1, Fig. 1). The most extreme significantly different environmental values were found for high temperatures, low precipitations and low latitude in *B. phoenicoides*, for high precipitations *B. sylvaticum*, and for low temperatures and high longitude in *B. pinnatum* (Table 1; Fig. 1).

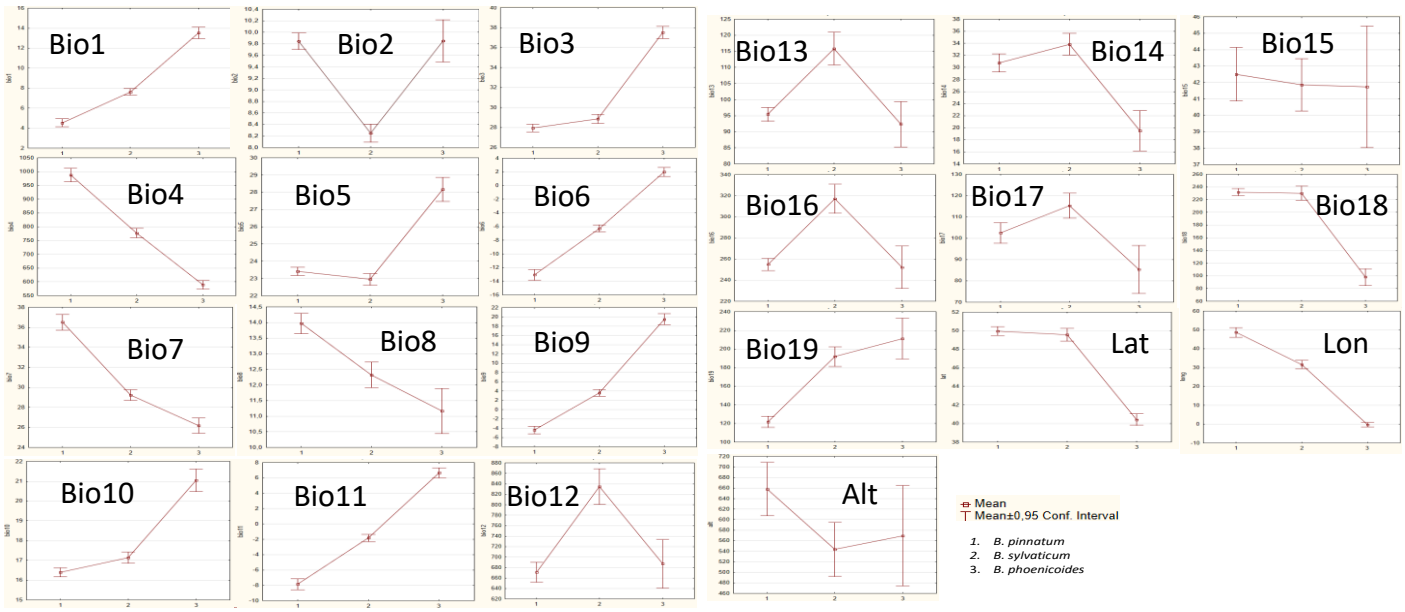


Figure 1. Environmental variation of the *B. pinnatum* (1), *Brachypodium sylvaticum* (2) and *B. phoenicoides* (3) species for the 19 climatic variables studied plus latitude, longitude and altitude. The graphs show the species' mean and 0.95 coefficient interval values for each variable.

Consistently with these results, the bidimensional environmental PCA plot showed a relatively clear differentiation of *B. phoenicoides* from *B. sylvaticum* and *B. pinnatum* samples on both extremes of the second axis which accumulated 24.46 % of variance, whereas those of *B. pinnatum* and *B. sylvaticum* overlapped in most of their environmental space though a group of populations from the former species separated from the rest along the positive extreme of the first axis (38.76% variance) (Figs. 2A). The variables that most contributed to the first PCA axes were Temperature Seasonality, Min Temperature of Coldest Month, Temperature Annual Range, Mean Temperature of Driest Quarter, Mean Temperature of Coldest Quarter, Annual Precipitation, Precipitation of Coldest Quarter and Longitude (PCA1), Max Temperature of Warmest Month and Mean Temperature of Warmest Quarter (PCA2), and Altitude (PCA3) (Table S3; Fig. 2B).

Table 1. Mean values of the 19 bioclimatic and 3 geographic variables analyzed for environmental niche modeling of *Brachypodium sylvaticum*, *B. pinnatum*, and *B. phoenicoides* in their Palaeartic native regions. These variables were used for comparative interspecific statistical analysis. Values of the bioclimatic variables selected to construct the environmental niche model (ENM) of each species with Maxent are highlighted in bold face. Superscripts letters denote Tukey pairwise comparisons among species; values with the same letter do not differ significantly (P<0.001). PC and PI indicate estimates of percentage contribution and of permutation importance of the variables used in the ENM of each secies, respectively. N, sampling size.

Enviromental Variable	**CCode**	***B. pinnatum* (N=658)**		***B. sylvaticum* (N=907)**		***B. phoenicoides* (N=91)**	
		MMean	**PPC (PI)**	**MMean**	**PPC (PI)**	**MMean**	**PC (PI)**
Annual Mean Temperature	BBio1	44.54 A		77.63 B		113.54 C	
Mean Diurnal Range	**BBio2**	99.85 A	10.3(14)	88.25 B	6.8(6.4)	99.85 A	3.3(5.3)
Isothermality	BBio3	227.94 A		228.87 A		337.54 B	
Temperature Seasonality	BBio4	9987.33 A		7776.72 B		5589.38 C	
Max Temperature of Warmest Month	BBio5	223.42 A		222.96 A		228.17 B	
Min Temperature of Coldest Month	BBio6	--13.08 A		--6.28 B		11.98 C	
Temperature Annual Range	**BBio7**	336.50 A	12.7(12)	229.24 B	14.2(11.6)	226.19 B	21.1(80.90)
Mean Temperature of Wettest Quarter	**BBio8**	113.97 A	18.5(19.7)	112.33 B	9.3(14.2)	111.17 B	10.8(2.5)
Mean Temperature of Driest Quarter	**BBio9**	--4.38 A	8.5(6.5)	33.61 B	17.6(43.6)	119.51 C	27.7(0.2)
Mean Temperature of Warmest Quarter	BBio10	116.40 A		117.14 B		221.05 C	
Mean Temperature of Coldest Quarter	BBio11	--7.86 A		--1.83 B		66.67 C	
Annual Precipitation	BBio12	6671.03 A		8834.55 B		6687.22 A	
Precipitation of Wettest Month	BBio13	995.41 A		1115.80 B		992.34 A	
Precipitation of Driest Month	BBio14	330.77 A		333.84 A		119.53 B	
Precipitation Seasonality	**BBio15**	442.51	0.9(0.7)	441.84	11.1(1.6)	441.73	2.2(5.6)
Precipitation of Wettest Quarter	BBio16	2254.92 A		3316.94 B		272.143 A	
Precipitation of Driest Quarter	BBio17	1102.47 A		1115.31 A		85.34 B	
Precipitation of Warmest Quarter	**BBio18**	2231.96 A	27.7(27.8)	2230.16 A	7.6(2.6)	97.95 B	15.3(3.3)
Precipitation of Coldest Quarter	**BBio19**	1121.73 A	21.4(19.2)	1191.77 B	43.4(19.9)	211.12 B	19.7(2)
Altitude	AAlt	6657.78		5543.06		569.33	
Latitude	LLat	449.96 A		449.58 A		40.43 B	
Longitude	LLon	448.73 A		331.68 B		-0.32 C	

The most influential variables on the current ENMs based on permutation importance were Precipitation of Coldest Quarter and Mean Temperature of Driest Quarter, for *B. sylvaticum*,

Mean Temperature of Wettest Quarter and Precipitation of Warmest Quarter for *B. pinnatum* and Mean Temperature of Driest Quarter and Temperature Annual Range for *B. phoenicoides* (Table S4). Additionally, the jackknife pseudoreplicate tests indicated that these variables showed the highest training gain for current ENMs when used in isolation for each of these species (Figs. S1, S2, S3). The same analysis but considering only intraspecific *B. pinnatum* populations confirmed the segregation of a group of populations from Siberia and Middle and Central Asia that separated from the remaining European populations at the positive extreme of the first axis (Fig. 2C). Similarly, the intraspecific analysis of *B. sylvaticum* populations not only separated the Siberian and Central Asian populations from the rest at the positive extreme of the first axis but also different segregations of Mediterranean and Scandinavian groups of populations in the upper left and lower left quadrants of the space defined by the first and second axes, respectively (Fig. 2D).

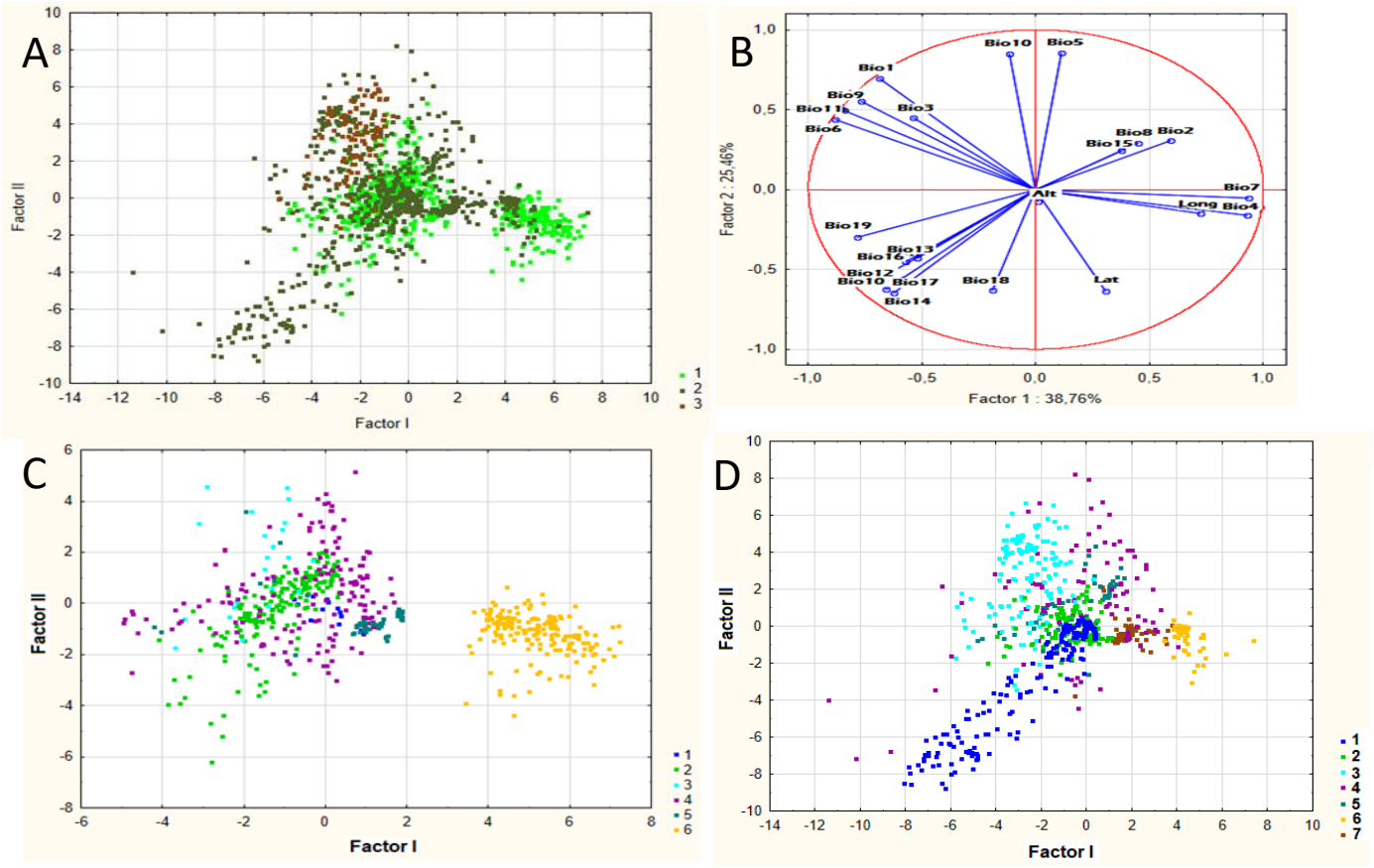


Figure 2. Bidimensional principal component analysis (PCA) plot of *B. pinnatum* (1, light green), *B. sylvaticum* (2, dark green), and *B. phoenicoides* (3, brown) records based on data from 19 bioclimatic variables plus latitude, longitude and altitude (A) and the contribution vector of each variable (B). PCA 1 and PCA 2 accounted for 38.76% and 24.46% of the variance, respectively. Intraspecific PCA plots of *Brachypodium pinnatum* (C) *and B.*

sylvaticum (D) records based on data from the same analysis. Numbers and color codes of current distribution ranges of samples are indicated in the charts: 1. Scandinavia (dark blue), 2. Western Europe (light green), 3.Mediterranean (light blue), 4. Caucasus (purple), 5. Eastern Europe (dark green), 6. Siberia and middle and Central Asia (yellow).

Environmental niche models of B. sylvaticum, B. pinnatum and B. phoenicoides under current and past climatic conditions

The niche distribution models of *B. sylvaticum and B. pinnatum* were spread along their entire native Eurasian region (Figs. 3A, 4A) and that of *B. phoenicoides* along its native western Mediterranean region (Fig. 5A). All the environmental niche models showed high AUC values (>0.88), being highest for *B. phoenicoides* (0.986, training data) and *B. pinnatum* (0.904test data) (Table S4). The species distribution models included all the sampled localities of each species (Fig. S4), supporting the high predictive power of these models. The predicted ENM of *B. sylvaticum* ranged from Iceland and Madeira in the west to Japan and the Kuril islands in the east, and from northern Scandinavia in the north to Taiwan and the Canary isles in the south (Figs. 3A, S4A) and that of *B. pinnatum* from England in the west to Central-Eastern Siberian in the east, and from Siberian Kola and Taymir in the north to Caucasus and Sicily in the south (Figs. 4A; S4B). The ENM of *B. phoenicoides* ranged from the western Iberian Peninsula and Morocco in the west to the Apennines in the east, and from the Jura in the north to the High Atlas in the south (Figs. 5A, S4C). Comparative analysis of overlapping areas for the *B. sylvaticum* and *B. pinnatum* environmental distribution models indicated ranges of potential shared occupancy by the two species across a wide geographical range of relatively humid zones in central-eastern and northern Europe, the Caucasus and the Himalayas (Table S5; Figs. 3A, 4A, S4). By contrast, the *B. sylvaticum* model was distributed exclusively in high precipitation coastal and subcoastal western Atlantic and eastern Pacific ranges (Fig. 3A), and the *B. pinnatum* model in colder and drier continental northern Central Asian ranges (Fig. 4A). The *B. phoenicoides* model shared some distribution occupancy with the predicted *B. sylvaticum* ENM in its warmer SW range and less of it with the predicted distribution of *B. pinnatum* (Table S5; Figs. 3A, 4A, 5A, S4).

The projection of the current presence data of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* to the respective MH and LGM envelopes showed different temporal range shifts for their respective ENMs (Figs. 3B-3C, 4B-4C, 5B-5C). All the past climatic models supported the potential distributions of the three species in their respective Eurasian and western Mediterranean areas. The MH distribution models of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* (Figs. 3B, 4B, 5B) were highly similar to the current climate models. The major differences were observed for the LGM-MIROC models, which showed the highest niche distributions for *B. sylvaticum* and *B. pinnatum* (Figs. 3C, 4C) and to a lesser extent for *B. phoenicoides* (Fig. 5C) with respect to current niche projections. The observed decreases in the distribution niches of *B. sylvaticum* and *B. pinnatum* from the LGM to the present could be a

consequence of the shift from northern Eurasian glacial refugia to respectively humid and cold sheltered latitudes, longitudes and altitudes due to the postglacial retrait. Similarly, the moderately decrease observed in the *B. phoenicoides* distribution range shift could have resulted from a large expansion to lower altitudinal areas in the LGM, such as a lowered coastal line of approx. 200 m below current sea level, as inferred for other Mediterranean *Brachypodium* species (Lopez-Alvarez et al. 2015), and its subsequent climbing to current sea border limits without any posterior altitudinal gain.

Overlaps and niche identity tests of the B. sylvaticum, B. pinnatum and B. phoenicoides ENMs

Niche overlap analysis under current climatic conditions showed a pattern of moderate to relatively high overlap for the ENMs of *B. sylvaticum* and *B. pinnatum* (D = 0.6096), and of low overlap for those of *B. sylvaticum* and *B. phoenicoides* (D = 0.2625) and of *B. pinnatum* and *B. phoenicoides* (D = 0.1480) (Table 2; Fig. S5). The niche identity tests found significant environmental divergence in the three species-pair comparisons, indicating that the niches of *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* are different ($P < 0.01$). The *B. phoenicoides* niche showed a slightly higher overlap with the *B. sylvaticum* niche than with the *B. pinnatum* niche.

Table 2. Observed niche overlap values and results of the niche identity tests for different species-pair comparisons under current environmental conditions. Overlap values (Schoener's D) smaller that the null distribution (Hellinger'I) support niche divergence (D) and values larger than it niche conservatism; when niche overlap values are similar to the null distribution background the results are inconclusive. Asterisks denote significance at ** $P < 0.01$.

Species pair (A-B)	**Observed overlap (Schoener's D value)**	**Identity test (Hellinger'I value)**
Current conditions		
B. sylvaticum-B. pinnatum	0.6096	0.8264**D
B. sylvaticum-B. phoenicoides	0.2625	0.5454**D
B. pinnatum- B. phoenicoides	0.1480	0.3291**D

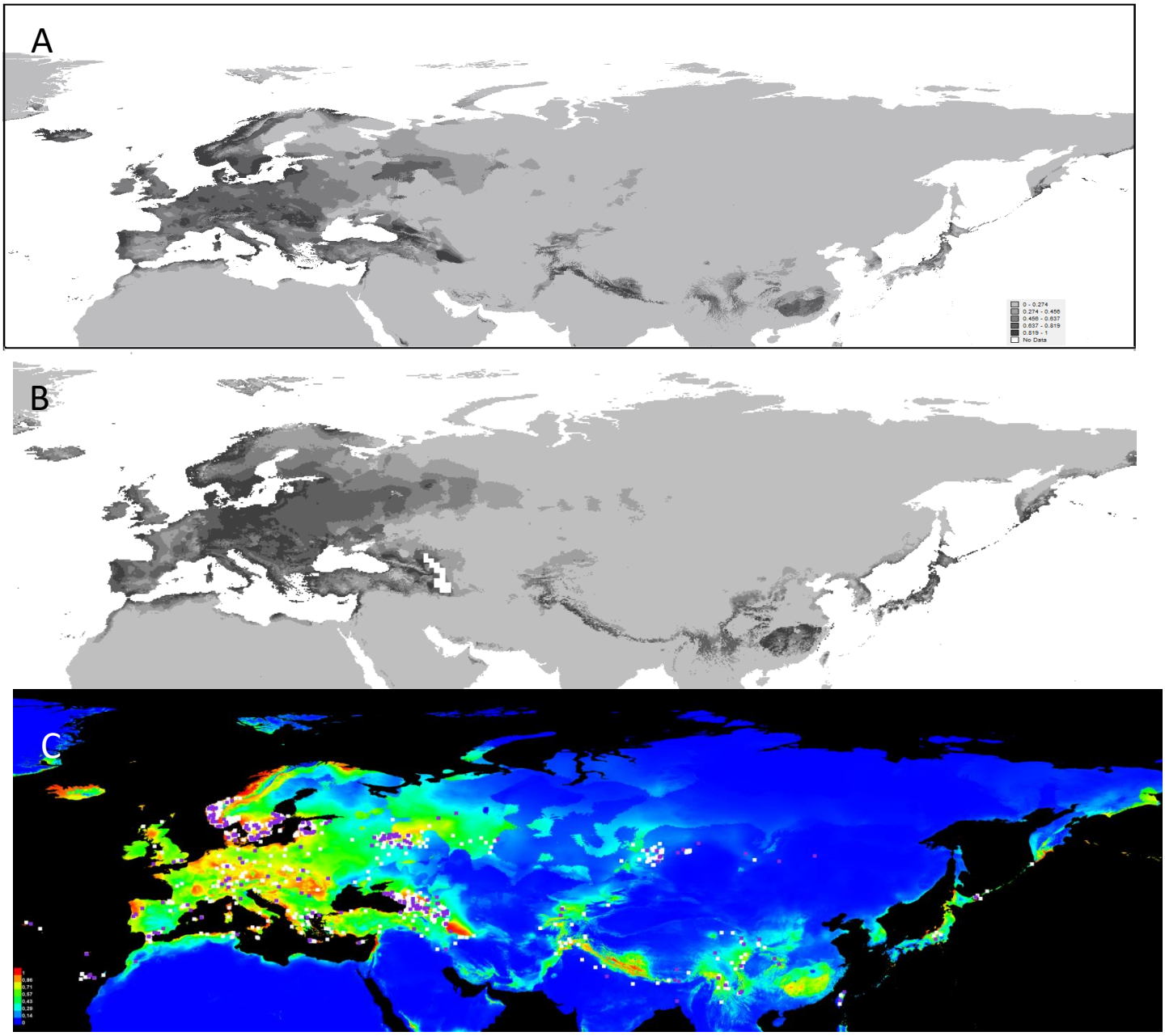


Figure 3. Environmental niche models constructed with Maxent for*Brachypodium sylvaticum* under (A) current (B) Mid-Holocene (MH) and (C) Last Glacial Maximum (LGM) climatic conditions. Scales indicate suitability scores.

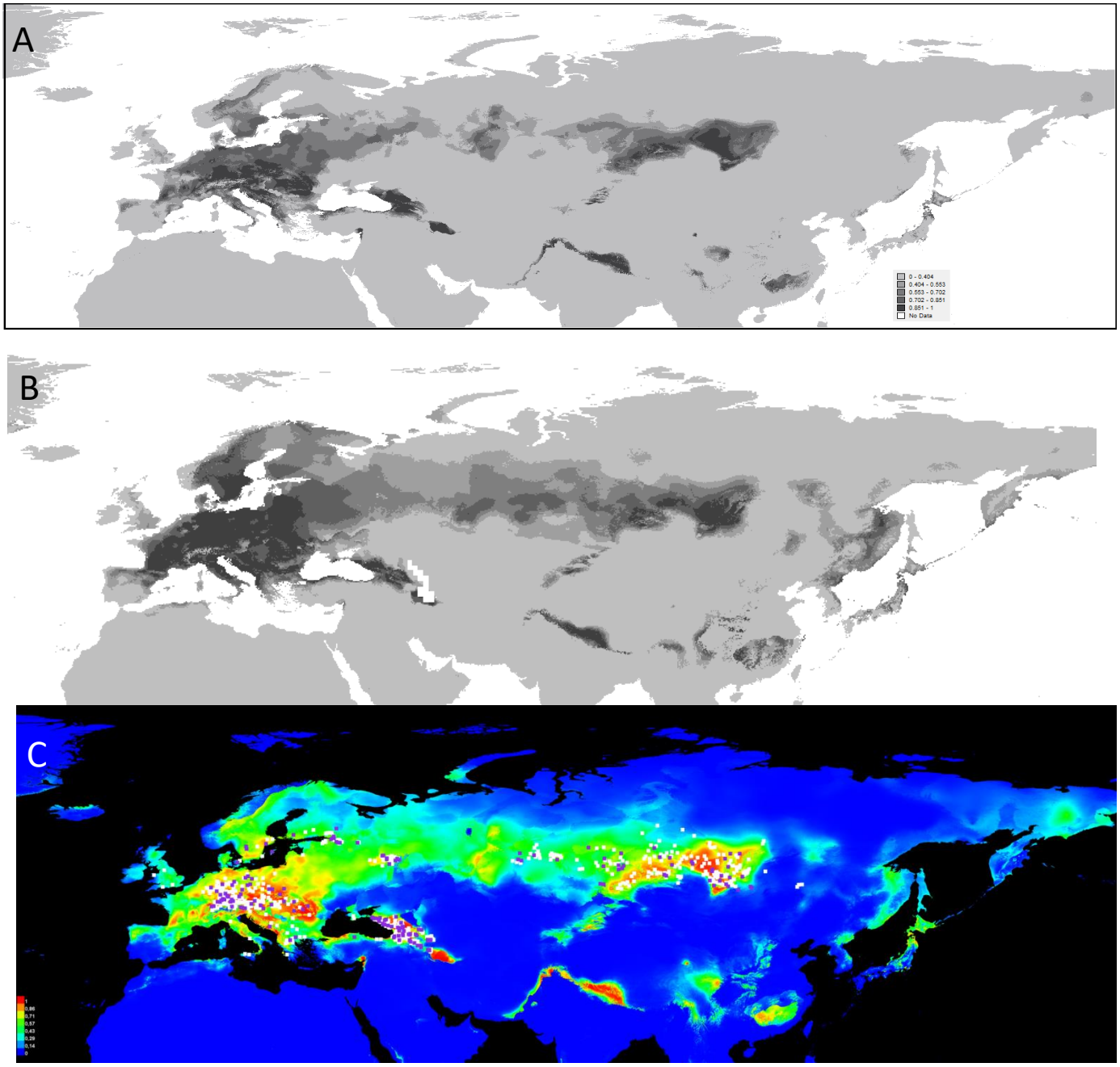


Figure 4. Environmental niche models constructed with Maxent for*Brachypodium pinnatum* under (A) current (B) Mid-Holocene (MH) and (C) Last Glacial Maximum (LGM) climatic conditions. Scales indicate suitability scores.

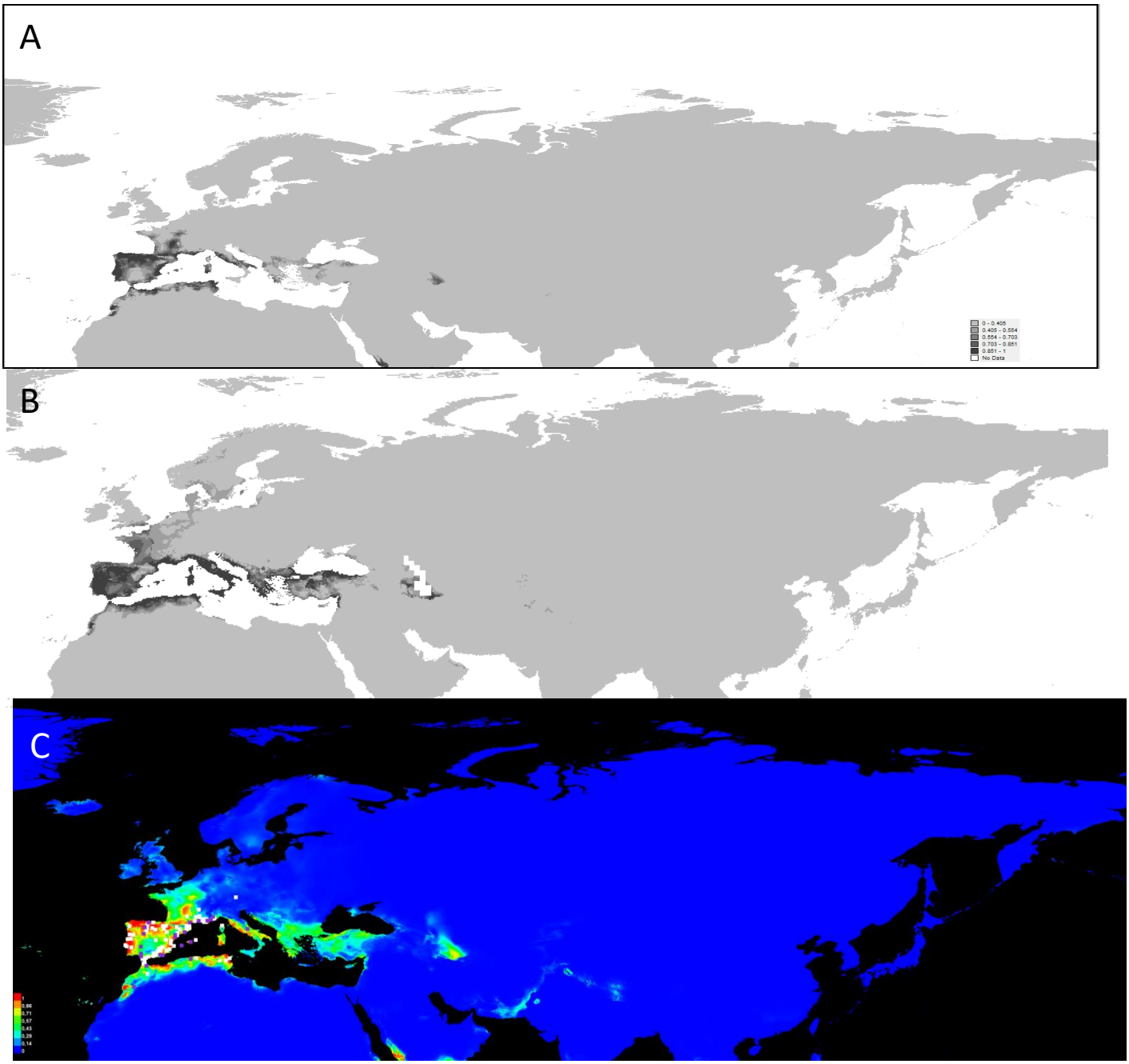


Figure 5. Environmental niche models constructed with Maxent for*Brachypodium phoenicoides* under (A) current (B) Mid-Holocene (MH) and (C) Last Glacial Maximum (LGM) climatic conditions. Scales indicate suitability scores.

Discussion

Northern Eurasian glacial refugia of B. pinnatum, ample Atlantic refugia of B. sylvaticum and western Mediterranean in situ survival of B. phoenicoides

Our environmental niche modeling studies of *Brachypodium sylvaticum*, *B. pinnatum* and *B. phoenicoides* have allowed us to reconstruct their potential niches since the LGM to the present; these niches depict different distribution scenarios and refugia for each species (Figs. 3, 4, 5). The statistical and niche identity test analyses conducted show a clear differentiation of the environmental niches of each of the three species (Tables 1, 2). Overlap values (Schoener's D) were smaller that the null distributions (Hellinger'I values) in every case, supporting niche divergence among the three species (Table 2; Fig. S5). Although the three species could be described as temperate Holarctic grasses (Catalán et al. 2016), current climate variables indicate that the diploid-to-tetraploid *B. pinnatum* is the most cold-tolerant plant, showing the lowest values for mean temperatures in the coldest month, coldest quartet and driest quartet (Table 1, Fig. 1). Noticeably, *B. pinnatum* also shows the highest values for mean temperatures in the annual range and the wettest quarter (Table 1; Fig. 1), supporting its adaptation to strong continental climates of cold winters and hot summers, which is connected to the highest longitudinal and high latitudinal values of its European and central Asian populations (Figs. 4, S4B). The adaptation of *B. pinnatum* to cold and continental latitudinal and altitudinal places is coupled with its adaptation to moderately humid habitats, as indicated by its intermediate values in most precipitation parameters (Table 1; Fig. 1), and a separation of its populations along the positive extreme of PCA axis 1 determined by the temperature seasonality and temperature annual range variables (Table 1; Figs. 1, 2A, 2B). The cold continental imprinting of *B. pinnatum* is reflected in its current ENM that shows its highest probability of occupancy not only in central-northern Europe and western Asia, but most remarkably in a large portion of the strong continental central-northern Siberia region (Fig. 4A), an almost exclusive niche distribution area for this species. It has been further corroborated by the clear divergent separation of these Siberian and middle-central Asian populations from the remaining European populations of the species in the intraspecific PCA plot (Fig. 2C). Both diploid and tetraploid cytotypes of the recently evolved Upper Pleistocene species *B. pinnatum* (Diaz-Perez et al. 2018) have been found across Eurasia (Schippmann 1991); our ENM study suggests the potential survival of cold-adapted diploid cytotypes in northern glacial refugia (Fig. 4C) and in southern interglacial refugia (Figs. 4A, 4B), and the expansion of continental climate-adapted tetraploid cytotypes across the whole distribution range (Figs. 4, S4B).

By contrast, the largely spread *B. sylvaticum* is best adapted to wet climate conditions, having the highest values for current mean values of annual precipitation, wettest month and wettest quarter variables (Table 1; Fig. 1), and to mild to moderately cold places, as shown in its intermediate values for most of the low and high temperature variables (Table 1; Fig.1). Populations of *B. sylvaticum* are scattered across the environmental space determined by the first two axes of the PCA plot, overlapping with populations of both *B. pinnatum* and *B. phoenicoides* though more largely with those of the former species (Fig. 2A). The oceanic nature of *B. sylvaticum* is explained by its current ENM that shows its highest probability of distribution in the Atlantic and Pacific continental coastal areas and islands, as well as in humid places of Europe, western and southern Asia, and northern Africa (Figs. 3A, S4). Even though

it is only a strict diploid species, its current distribution range is the largest of the three species (Figs. 3A, S4). However, a large environmental segregation of different eco-regional *B. sylvaticum* groups could be observed in the intraspecific PCA plot that shows the separation of a more continental Siberian and central Asian group at the positive extreme of axis 1 and of a warm Mediterranean group and a humid western Scandinavian group at the respective positive and negative extremes of axis 2 (Fig. 2D). These distinct environmental adaptations could be related to the different eco-regional lineages detected within this relatively old Mid-Pleistocene species (Diaz-Perez et al. 2018). Our ENM analysis supports the likely existence of western and northern Eurasian glacial refugia for the different diploid lineages of *B. sylvaticum* in the LGM (Fig. 3C) and of interglacial refugia in the most humid places of its current range (Fig. 3A). In contrast to its congeners, *B. phoenicoides* is the warmest and driest tolerant species, showing the highest values for current mean annual temperature and for temperatures of the warmest month, driest quartet, warmest quartet and coldest quartet variables (Table 1; Fig.1). These values together with its lowest latitude and longitude values separate its tightly clustered populations from those of the other species towards the positive extreme of the second axis (Fig. 2A). *B. phoenicoides* also shows low niche overlaps with the niches of its two congeners, especially with that of *B. pinnatum* (Table 2; Fig. S5). The warm-temperate features of *B. phoenicoides* explain its current ENM which shows its highest niche occupancy in the western Mediterranean region (Fig. 5A). The restricted Mediterranean distribution area of this recently evolved Upper Pleistocene Tarantian allotetraploid species is influenced not only by warm climatic conditions but also by its place of origin and evolutionary history (Catalán et al., 2016; Diaz-Pérez et al. 2018).

The *Brachypodium* core perennial species have had a very recent Late Quaternary origin (Catalán et al. 2016; Diaz-Perez et al. 2018). The single known fossil remains of *Brachypodium* correspond to starch grains preserved at the Bilancio II Paleolithic site (central Italy) ~28 Kya (Revedin et al. 2010). This finding confirms that *Brachypodium* plants existed at least during the LGM in this southern Mediterranean refugium although it is not possible to identify the species. Phylogenetic Bayesian dating analysis have inferred the splits ages of the *B. sylvaticum*, *B. pinnatum* and *B. phoenicoides* lineages at 600 kya, 300 kya and 125 kya, respectively (Diaz-Pérez et al. 2018). Even data from more stringent coalescence dating analysis indicate a LGM age (48 kya) for the recentmost hybrid allotetraploid *B. phoenicodes*. All the summed evidence supports the presence of the three perennial *Brachypodium* species in Eurasia since the LGM and, consequently, their respective niche distribution changes during the last Quaternary climatic oscillations. The reconstructed ENMs of *B. pinnatum*, *B. sylvaticum* and *B. phoenicoides* clearly display the relatively broad niche distribution shifts experienced by their populations in the last Würm-Holocene glacial-interglacial phase (Figs. 3, 4, 5).

The concept and the circumscriptions of Quaternary refugia have changed widely during the last decades (Steward et al. 2010). It is broadly recognized today that the warm circum-Mediterranean peninsulas and the Caucasus acted as glacial refugia for temperate biotas at places located below the 1000m altitudinal ice cap limit (Taberlet et al. 1998; Ehlers & Gibbard 2004), whereas the highest alpine an subalpine mountain ranges in these same areas acted, and currently act, as interglacial refugia for cold biotas (Hewitt 2004). Conversely, some

of the currently cold central and northern Eurasian areas served as nunatak glacial refugia for some cold-adapted biotas while they may have acted and act as interglacial refugia for many temperate elements (Stewart & Lister 2001; Steward et al. 2010). Recent discoveries of DNA fossil records of 10-40 kya angiosperms, including several Poaceae species, in the arctic Siberian Kolyma region (Willerslev et al. 20013), and the regeneration of fertile plants from LGM seeds preserved in the NE Siberian permafrost (Yashina et al. 2012), have confirmed the existence of glacial refugia for cold and temperate plants in extremely high Eurasian latitudes. Our LGM niche model analysis indicates that the cold-adapted *B. pinnatum* had a potential distribution range that covered some of the purportedly glaciated areas in central-northern Siberia and in central-northern Asia (Fig. 4C). Its plausible survival in those places could be associated with the existence of the glacial refugia reported for Central Asia (Arkhipov et al. 1999; Friesen et al. 2016) and for Central Europe, Scandinavia and the Caucasus (Steward et al. 2010). The species' niche apparently also experienced range expansions to southern warmer Mediterranean and western Asian areas and to Eastern Siberia (Fig. 4C), a long-term refugia for cold-adapted plants (Stewart & Dalén 2008), covering a potential niche broader than its current niche (Figs. 4A, 4C). The *B. pinnatum* niche projection to the MH shows a contraction in its distribution range (Fig. 4B) with respect to that of the LGM (fig. 4C) and a similar distribution than the current niche (Fig. 4A). The species' range contraction was probably caused by the climatic change trend of increasing temperatures and precipitations from the last glacial phase to the interglacial, making the southern and lowland Eurasian regions less hospitable for this cold continental plant.

In contrast to the multiple documented cases of latitudinal migrations of temperate Eurasian plants, consisting of southwards glacial contractions and northwards interglacial expansions of populations, fewer examples have evidenced the longitudinal migrations of plants during the Quaternary. Except for steppe plants of Central Asia that could have colonized western Europe during the glacial phases (Steward et al. 2010), the migrations of oceanic plants are less known. Leipold et al. (2017) reconstructed the LGM niche of the oceanic *Hyppocrepis comosa* and its postglacial expansion, identifying potential cryptic glacial refugia in central-western Europe and both latitudinal and longitudinal postglacial range expansion. Our environmental niche models of *B. sylvaticum* indicate that its LGM distribution range (Fig. 3C) was larger than that of the MH (Fig. 3B) and than that of today (Fig. 3A). The LGM niche of this oceanic species expanded through the Macaronesian and the central and north Atlantic isles, Europe, NW Africa, and the western Pacific central and northern islands and coastal area. Noticeably, it also included the western and southern Asian mountain regions but very small areas in the Central-northern Siberia (Fig. 3C). It can be assumed that the species was likely sheltered in a continuum of humid coastal and subcostal northwestern and northeastern glacial refugia during the LGM. A longitudinal migrating scenario from Central Asia to the west is inferred for this species as the climate became drier in Central Siberia since the MH to the present (Fig. 3).

The niche distributions of Mediterranean temperate plants also experienced contraction and expansion steps within the Quaternary climatic fluctuations (Lopez-Alvarez et al. 2015); however they were less pronounced than those of Mediterranean alpine and subalpine plants

(Boucher et al. 2012) due to their broader realized niche. Lowland Mediterranean plants potentially colonized emerged areas below the current sea level during the LGM, enlarging thus their potential distribution ranges (Lopez-Alvarez et al. 2015). The reconstructed LGM niche of *B. phoenicoides* (Fig. 5C) is also larger than those from the MH (Fig. 5B) and today (Fig. 5A); however it shows less contraction through time than those of its congeners. *B. phoenicoides* is adapted to warm-and-dry to sub-humid climatic conditions but also to moist soils (Catalan et al. 2016). This ecological amplitude could explain its wide adaptation and distribution in both coastal and inland areas in the western Mediterranean region, as well as its small variation along the temporal distributions (Fig. 5A, B, C). The western Mediterranean region is recognized as a main hotspot of taxonomic and genetic diversity for plants. An important part of such diversity originated in the area as a consequence of the secondary glacial and interglacial contacts and hybridizations of species and populations (Nieto Feliner 2011). The allotetraploid *B. phoenicoides* shows more reduced LGM, MH and current niches than those of *B. sylvaticum* and *B. pinnatum* (Figs. 3, 4, 5; Table S5). It has been hypothetised that *B. phoenicoides* could have originated from a cross of a *B. sylvaticum*-type (2n=2x=18) progenitor species and an as yet unkown parent (Catalán et al. 2016; Diaz-Perez et al. 2018). Karyotypic data further suggest the potential participation of a second *B. distachyon*-type (2n=2x=10) progenitor taxon in the origin of this 2n=4x=28 allotetraploid species (Betekhtin et al. 2014). *B. phoenicoides* shows a much more restricted niche distribution range than current *B. sylvaticum* (Tables 2, S5; Figs. 3, 5) and *B. distachyon* (López-Alvarez et al. 2015) species. Its recent hybrid origin and its potential adaptive competition with the putative progenitor species' niches would explain its still limited distribution range, corroborating that the hybrid polyploid does not outperform its progenitor diploid species in their environmentally stable native ranges (McIntyre, 2012; Lopez-Alvarez et al. 2015). Lopez-Alvarez et al (2015) found that the annual Mediterranean allotetraploid *B. hybridum* shared intermediate niche features and an overlapping niche distribution with those of its progenitor *B. stacei* and *B. distachyon* species. The *B. phoenicodes* niche overlaps with that of the current nemoral Eurosiberian *B. sylvaticum* in some parts of the western Mediterranean region (Tables 2, S5; Figs. 2A, 3A, 5A); however it is clearly a plant preferentially adapted to the typical Mediterranean climate (Fig. 5). Comparative niche distribution modeling suggests that *B. phoenicoides* could have inherited its predominant Mediterranean niche features from its purported Mediterranean *B. distachyon*-type parent (Table 1; Fig. 5; López-Alvarez et al. 2015) although it would have acquired its perennial habit from its purported Eurosiberian *B. sylvaticum*-type parent (Catalán et al. 2016; Diaz-Pérez et al. 2018).

Conservation of rare and endangered Central Asian populations of the widespread B. sylvaticum

Widespread species that are abundant in most of their distribution ranges could be rare, and even endangered, in some parts of their broad range. Despite its current wide Palaearctic distribution, *Brachypodium sylvaticum* is an uncommon species in the forested taiga and open grassland ecosystems of Siberia and is absent from the strong continental and aridic central Asian grasslands and steppes (Fig. 3A). It is also a rare plant in the European part of Russia. *B.*

sylvaticum has been included in the catalogue of "Rare and Endangered Plants of Siberia" (Semenova et al. 2017), in the list of protected species of Siberia (Malyshev & Peshkova, 1979; Semenova, 2007), and in the Red Data Books of the Russian Buryatia, Altai and Krasnoyarsk Territories, the Novosibirsk Region, and the Vladimir, Kirov and Yaroslavl provinces, being classified under state III of rare species (https://www.rferl.org/a/russia-endangered-species/24915307.html). Relict species are known to be the most sensitive and vulnerable to climatic and global change; when the environmental conditions vary, they could be the first to go extinct (Hampe & Jump 2011). Conservation strategies are thus necessary for the protection of their endangered populations. Environmental niche modeling studies are useful tools to investigate the effect of rapid climatic changes in the environmental distribution and niche suitability of threatened trailing edge populations (Pearson & Dawson 2003).

Brachypodium sylvaticum has a largely disjunct current distribution range (Figs. 3A, S4A). Its main abundance center in located in the forested areas of western and central-northern Europe, as well as in the European part of Russia. It covers a broad area west of the Urals, and also in the Caucasus, Turkey and some dispersed setting in North Africa and the Middle East. In addition, isolated distribution ranges of the species are also found in the mountains of Central and Middle Asia, and of the Far East, in China, Japan and the Kuril islands (Tsvelev 1976; Probatoba 1984; Hultén & Fries 1986; Krapivkina, 2009). In Siberia *B. sylvaticum* also has a discontinuous distribution range. The species is found in isolated Western Siberian areas that are scattered across the Novosibirsk, Kemerovo and Tomsk provinces, the Altai mountains, the Krasnoyarsk Territory (Reverdatto, 1964), Buryatia and Transbaikalia (Sergievskaya, 1969). Such distribution, as well as the confinement of the species to relict areas of broad-leaved and “black taiga” Siberian forest, considered to be the remains of the now extinct Tertiary coniferous-broadleaf forests (Koropachinskii 1996, Krapivkina, 2009), prompted researches to classify this species in Siberia as a nemoral Pliocene relict (Krapivkina 2009). Lashchinsky (2009) even suggested that “black taiga” communities with dominance of *B. sylvaticum* could correspond to one of the oldest relict segments of the ancestral Oligocene Turgai temperate vegetation. This unlike scenario does not match the young Late Quaternary age of *B. sylvaticum* (Diaz-Perez et al. 2018), though it could indicate its establishment in this natural humid ecosystem. The persistence of *B. sylvaticum* in the black taiga and broad-leaved Siberian forest is associated to its biological attributes. This selfing perennial plant (Scholthof et al. 2018) shows a high seed production and vegetative tussock renewal that leads to the formation of stable, full-grown populations dominating the under-canopy of these relict forests (Lashchinsky 2009).

We have examined our environmental niche characterization and niche predictions for *B. sylvaticum* in the north-central Siberian region to provide clues about the global environmental changes that affected the niche contraction of this species since the LGM to the present and to address initiatives for its conservation. Our data indicate that the Siberian and central Asian populations of *B. sylvaticum* show distinct environmental features that separate them from other *B. sylvaticum* populations along the first axis of the PCA plot (Fig. 2D). These populations are adapted to broader temperature seasonality and to more contrasted continental-

type temperature changes than the western and southern Eurasian populations. The historical contraction trend of the niche distribution range indicates that the *B. sylvaticum* range may have experienced a reduction from the LGM niche to its current niche (Figs. 3A, 3C). This range contraction could be related to a decrease in mean annual temperature and in mean annual precipitation since the LGM to the present, according to the LGM and current bioclimatic envelopes estimates. It is also possible, however, that the purported extinction of the *B. sylvaticum* LGM ranges was caused by anthropogenic activity, since large areas of black taiga and broad-leaved Siberian forests experienced a considerable reduction in the last decades caused by lodging and agricultural practices (Achard et al. 2006; Bergen et al. 2008; Hansen et al. 2013). The ability of *B. sylvaticum* to propagate actively, both sexually and vegetatively, together with its noticeable longevity, with individuals lasting for more than 20 years, makes the plant a strong stress-tolerant competitor (Haeggström & Skytén 1996). Even under less favorable climatic conditions, the survival of the threatened Siberian *B. sylvaticum* populations could be guaranteed and the species preserved from risk of extinction if their natural forested ecosystems are maintained. Conservation efforts should focus on the preservation of the relict Siberian black taiga and broad-laved forest for the safeguard of these environmentally distinct *B. sylvaticum* populations.

Acknowledgements

We thank the curators of the AAU, B, C, FI, G, JACA, LE, MA, MW, NS, NSK, S, SEV and TK herbaria for loans of specimens and facilities for the study of the *Brachypodium sylvaticum*, *B. pinnatum* and *B. phoenicoides* materials. This work was supported by the Spanish Ministry of Economy and Competitivity (Mineco) CGL2016-79790-P and University of Zaragoza UZ2016_TEC02 grant projects. VS was funded by a Russian Ministry of Education and Science fellowship. PC and IM were partially funded by a European Social Fund and Aragón Government Bioflora grant.

References

Achard F, Mollicone D, Stibig HJ, Aksenov D, Laestadius L, Li Z, Popatov P, Yaroshenko A. 2006. Areas of rapid forest-cover change in boreal Eurasia. Forest Ecology and Management 237: 322–334.

Arkhipov, S.A., Volkova, V.S., Zolnikov, I.D., Zyklina, V.S., Krukover, A.A. & Kulkova, I.A.n 1999. West Siberia. Climate and environment changes during the last 65 million years; Cenozoic: from Palaeocene to Holocene (ed. by A.A. Velichko and V.P. Nechaev), pp. 84–109. GEOS, Moscow (In Russian).

Araujo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology, 93(7): 1527–1539.

Bergen K, Zhao T, Kharuk V, Blam Y, Brown D, Peterson L, Miller N. 2008. Changing regimes: forested land cover dynamics in Central Siberia 1974–2001. Photogrammetric Engineering and Remote Sensing 74: 787–798.

Betekhtin, A., Jenkins, G., Hasterok, R., 2014. Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. PLoS ONE 9: e115108.

Boucher, F.C., Thuiller, W., Roquet, C., Douzet, R., Aubert, S., Alvarez, N. et al. 2012. Reconstructing the origins of high-alpine niches and cushion life form in the genus Androsace S.L. (Primulaceae). Evolution 66: 1255–1268.

Brown, J. L. 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* 5: 694-700

Catalán P, López-Alvarez D, Díaz-Pérez A, Sancho R, López-Herranz ML. 2016. Phylogeny and evolution of the genus *Brachypodium*. *In* Vogel J (ed.). Genetics and genomics of *Brachypodium*. pp. 9-38. Series Plant Genetics and Genomics: Crops Models. Springer. New York

Díaz-Pérez A, López-Álvarez D, Sancho R, Catalan P. 2018. Reconstructing 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

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier MS, Guisan Hijmans ARJ, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1: 330-342.

Fordham, D.A., Akcakaya, H.R., Araujo, M.B., Elith, J., Keith, D.A., Pearson, R. et al. 2012. Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming? *Glob. Change Biol.*, **18**: 1357–1371.

Fordham DA, Brook BW, Moritz C, Nogués-Bravo D. 2014. Better forecasts of range dynamics using genetic data. Trends in Ecology and Evolution 29: 436-443.

Friesen N, German DA, Hurka H, Herden T, Oyuntsetseg5 B, Neuffer B. 2016. Dated phylogenies and historical biogeography of Dontostemon and Clausia (Brassicaceae) mirror the palaeogeographical history of the Eurasian steppe. Journal of Biogeography 43: 738–749

Geber MA, Griffen LR. 2003. Inheritance and natural selection on functional traits. *International Journal of Plant Sciences* 164: S21-S42.

Guisan A, Broennimann O, Engler R, Vust M, Yoccoz NG, Lehmann A, Zimmermann NE. 2005. Using niche-based models to improve the sampling of rare species. Conservation Biology 20: 501–511.

Haeggström CA, Skytén R. 1996. Flowering and individual survival of a population of the grass *Brachypodium sylvaticum* in Nåtö, Åland Islands, SW Finland. *Annales Botanici Fennici* 33: 1-10.

Hampe A, Jump AS. 2011. Climate Relicts: Past, Present, Future. Annual Review of Ecology Evolution and Systematics 42:313-333.

Hansen MC, Potapov PV, Moore R, Hancher M, Turubanova S, Tyukavina A, Thau D, et al. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342: 850–853.

Hahn MA, van Kleunen M, Muller-Scharer H. 2012. Increased phenotypic plasticity to climate may have boosted the invasion success of polyploid *Centaurea stoebe*. PLoS ONE 7(11): e50284.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 359: 183-195.

Hijmans R J, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hultén E, Fries M. 1986. Atlas of North European vascular plants: north of the Tropic of Cancer I–III. Königstein: Koeltz Scientific Books.

Koropachinskii I (ed.) 1996. Green nook of Siberia: rare and endangered plant communities. Novosibirsk: Nauka. pp. 104–107 (in Russian).
Krapivkina ED. 2009. Nemoral relicts in the flora of blackish taiga in Gornaya Shoriya. Novosibirsk. Izdatel'stvo SORAN (in Russian).
Lashchinskiy NN. 2009. Vegetation of Salair ridge. Novosibirsk (in Russian).

Leipold M, Tausch S, Poschlod P, Reisch C. 2017. Species distribution modeling and molecular markers suggest longitudinal range shifts and cryptic northern refugia of the typical calcareous grassland species *Hippocrepis comosa* (horseshoe vetch). Ecol Evol: 1919–1935.

Levin DA. 1975. Minority cytotype exclusion in local plant populations. *Taxon* 24: 35-43.

Linder HP, Antonelli A, Humphreys AM, Pirie MD, Wuest RO. 2013. What determines biogeographic ranges? Historical wanderings and ecological constraints in the danthonioid grasses. Journal of Biogeography 40: 821–834.

López-Alvarez D, Manzaneda A J, Rey PJ, Giraldo P, Benavente E, Allainguillaume J, Mur LAJ, Caicedo AL, Hazen SP, 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: 1-16.

Malyshev LI, Peshkova GA. (ed.) 1979. Flora of Central Siberia. v. 2. Nauka, Novosibirsk (in Russian).

Mandle L, Warren DL, Hoffmann MH, Peterson AT, Schmitt J, von Wettberg EJ. 2010. Conclusions about niche expansion in introduced *Impatiens walleriana* populations depend on method of analysis. *Plos One* 5:e15297.

Martin SL, Husband BC. 2009. Influence of phylogeny and ploidy on species ranges of North American angiosperms. *Journal of Ecology* 97: 913-922.

McCormack JE, Zellmer AJ, Knowles LL. 2010. Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution* 64: 1231-1244.

McIntyre PJ. 2012. Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany* 99: 655-662.

Nakazato T, Warren DL, Moyle LC. 2010. Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany* 97: 680-693.

Nieto Feliner GN. 2011. Southern European glacial refugia: a tale of tales. *Taxon* 60: 365-372.

Pandit MK, Pocock MJO, Kunin WE. 2011. Ploidy influences rarity and invasiveness in plants. *Journal of Ecology* 99: 1108-1115.

Pearson RG, Dawson TP. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Glob. Ecol. Biogeogr. 12: 361-371.

Peterson AT, Nyari AS. 2008. Ecological niche conservatism and pleistocene refugia in the thrush-like mourner, Schiffornis sp., in the neotropics. *Evolution* 62: 173-183.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.

Probatova NS. 1984. New taxa in the family Poaceae from the Far East of the U.S.S.R. Bot. Zh. 69 (2): 251-259 (in Russian).

Revedin A, Aranguren B, Becattini R, Longo L, Marconi E, Lippi MM, Skakun N, Sinitsyn A, Spiridonova E, Svoboda J. 2010 Thirty thousand-year-old evidence of plant food processing. Proc Natl Acad Sci U S A. 107(44):18815-9.

Rosenthal DM, Ramakrishnan AP, Cruzan MB. 2008. Evidence for multiple sources of invasion and intraspecific hybridization in Brachypodium sylvaticum (Hudson) Beauv. in North America. Mol Ecol. 17(21):4657-69.

Reverdatto VV. 1964. Flora of Krasnoyarsk Region. Tomsk State University, Tomsk (in Russian).

Schippmann U. 1991. Revision der europäischen arten der gattung *Brachypodium* Palisot de Beauvois (Poaceae), vol. 45. Conservatoire et jardin botaniques de Genève, Genève.

Scholthof KB, Irigoyen S, Catalan P, Mandadi KK. 2018. *Brachypodium*: A monocot grass model genus for plant biology. *Plant Cell* 30:1673-1694.

Semenova GP. 2007. Rare and endangered species of Siberian flora: Biology, conservation. Geo Academic Publishing. Novosibirsk.

Semenova NM, Volkova II, Amelchenko VP, Volkov IV, Kolesnichenko LG. 2017. Rare and endangered plants of specially protected areas of West Siberian boreal coniferous forests (inside Tomsk oblast), International Journal of Environmental Studies, 74:5, 732-743.

Sergievskaya L.P. *1969. Flora* of Transbaikal. Tomsk State University. Tomsk.

Soberón JM. 2010. Niche and area of distribution modeling: a population ecology perspective. *Ecography* 33: 159-167.

Stewart JR, Lister AM. 2001 Cryptic northern refugia and the origins of modern biota. Trends Ecol. Evol. 16: 608–613

Stewart, J. R. & Dalén, L. 2008 Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. Clim. Change 86, 1–2.

Stewart JR, Lister AM, Barnes I, Dalen L. 2010. Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences* 277: 661-671.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453–464.

Tarkhnishvili D, Gavashelishvili A, Mumladze L. 2012. Palaeoclimatic models help to understand current distribution of Caucasian forest species. *Biological Journal of the Linnean Society* 105: 231-248.

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesova M, Pysek P. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109: 19-45.

Tsvelev NN. 1976. *Flora of the URSS*. Nauka. Leningrad (in Russian).

Waltari E, Hijmans RJ, Peterson AT, Nyari AS, Perkins SL, Guralnick RP. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. PLoS ONE 2: e563.

Warren DL, Glor RE, Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33: 607-611.

Willerslev E, Hansen AJ, Binladen J, Brand TB, Gilbert MTP, et al. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* 300:791–95.

Yashina S, Gubin S, Maksimovich S, Yashina A, Gakhova E, Gilichinsky D. 2012. Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost. Proc. Natl. Acad. Sci. U. S. A., 109: 4008-4013

CHAPTER 5

Morphoanatomical study of the *Brachypodium pinnatum* complex species (*B. pinnatum, B. rupestre, B. phoenicoides*) (Poaceae) and their diploid and polyploid cytotypes

Valeriia Shiposha[1,2], Isabel Marques[2,§], Pilar Catalán[1,2,3,*], Marina Olonova[1,*]

[1.] Department of Botany, Institute of Biology, Tomsk State University, Lenin Av. 36, Tomsk 634050, Russia.

[2.] Department of Agriculture and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Ctra. Cuarte km 1, 22071 Huesca, Spain.

[3.] Grupo de Bioquímica, Biofísica y Biología Computacional (BIFI, UNIZAR), Unidad Asociada al CSIC, Zaragoza E-50059, Spain

[§.] current address: cE3c - Centre for Ecology Evolution and Environmental Changes, Faculty of Sciences, University of Lisbon, Lisbon, Portugal

*Corresponding authors: olonova@list.ru, pcatalán@unizar.es

Abstract

Taxonomically complex groups are often composed of easily identifiable species but also of undercovered cryptic species. The detection of cryptic species depends on the availability of phenotypic traits and the relationship between phenotypic disparity and genetic divergence, but also on levels of gene flow and reproductive isolation. In plants, highly hybridogenous groups with sexually and asexually propagating species and several levels of interconnected ploidies are a source of new cryptic species. Here we analyse populations of the *Brachypodium pinnatum* complex, a grass group formed by three recognized species, *Brachypodium pinnatum, B. rupestre* and *B. phoenicoides*, and several cytotypes and ecotypes of unclear taxonomic adscription. Sampling included a large representation of 509 individuals covering the whole native distribution range of each species in their respective Eurasian (*B. pinnatum*, 273), Eurosiberian (*B. rupestre*, 90) and western Mediterranean (*B. phoenicoides*, 146) areas. Seventeen potentially diagnostic morphoanatomical traits were measured in all the samples and the data were used in subsequent statistical analyses. Pollen grain morphology was examined under scanning microscopy in a subset of representative samples. Thirteen morphological traits significantly discriminate groups of species and four of them significantly separate the three species. *B. pinnatum* and *B. phoenicoides* are the most distinct species, differing in 13 discriminant traits and clustering in opposite extremes of the PCA plot. Noticeably, *B. rupestre,* formerly classified within *B. pinnatum*, shows intermediacy between *B. pinnatum* and *B. phoenicoides* in several traits and also separates from them in three phenotypically discriminant traits. Pollen grain morphology is similar in the three species though exine ornamentation varies between the commonest granular pattern and a verrucose pattern observed only in some individuals. Infraspecific phenotypic analyses detect a large morphoanatomical variation within each of the three studied species. Up to two, three and three morpho-groups that have been respectively identified within *B. pinnatum*, *B. rupestre* and *B. phoenicoides* could correspond to cryptic taxa. Each group separates from its conspecifics in several discriminant traits and some of them show a specific geographic distribution. These infraspecific groups may be associated to population sheltered in different glacial and interglacial refugia or to potential interspecific fertile crosses between similar ploidy-level cytotypes. Stomata guard cell length measurements indicate the potential existence of different ploidies within *B. pinnatum*, *B. rupestre* and *B. phoenicoides* that could also correspond to different cryptic taxa. The *B. pinnatum* complex could be considered a compilospecies group of successfully interbreeding taxa with different chromosome base numbers and ploidies.

Keywords: *Brachypodium pinnatum – B. rupestre – B. phoenicoides*, cryptic taxa, Eurasian-Mediterranean grasses, morphological characters, statistics, taxonomy.

Running title: Morphological variation of cryptic *Brachypodium pinnatum* complex grasses

Introduction

Cryptic species may constitute a large portion of the still uncovered biodiversity on earth (Struck et al. 2017). The widespread use of barcoding and genome sequencing approaches is exposing many morphologically similar but genetically diverse potential cryptic species. It has been estimated that up to 10-20% of the current single morphological species may turn out to be two or more species, especially in some megadiverse and still poorly studied organismic groups (Janzen et al. 2017), but also across all branches of the tree of life (Struck et al. 2017). However, the cryptic species concept is also a scientific paradox. Taxonomists have traditionally used morphological characters to determine species boundaries, though speciation and establishment of the diagnostic features are not always directly correlated, except for some notorious examples (Stace 1991). Cryptic species are often viewed as such when new genomic and evolutionary approaches detect potential conspecifics through non apparent visual cues (Gustaffsson et al. 2014; Vigalondo et al. 2015). Methods to identify and quantify cryptic species have been based on detailed analysis of phenotypic disparity related to genetic divergence, levels of gene flow and reproductive isolation (Struck et al. 2017), and on statistical deviations of phenotypic traits from neutral expectations (Harmon et al. 2003). However, the counter-effect of hybridization could complicate the dissection of cryptic species, although it would also depend on the degree of isolation of the hybrid from the progenitor species (Stebbins 1969).

In plants, cryptic species are often associated to highly hybridogenous complexes of sexually reproducing species (Harlan & de Wet 1963), sometimes showing apomictic components (de Wet & Harlan 1966) or different chromosomal cytotypes and ploidies (Stebbins & Dremann 1998). The size and taxonomic diversity of these complexes vary enormously depending on the evolutionary history of the group. Within highly reticulate groups, genetically aggressive compilospecies often incorporate portions of the genomes of closely related sympatric species through extensive introgression (Harlan & de Wet 1963), subsuming them onto their taxonomic limits (de Wet & Harlan 1966) and conferring them a transclade phylogenetic nature (Fuertes Aguilar et al. 1999). It is generally assumed, however, that species should be distinguished by stable and prominent phenotypic features, and that the interbreeding individuals could leave fertile descendants (Mayr 1942). 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 1990; Holstein & Luebert 2017), and therefore those biological entities should be treated as infraspecific ranks or ecotypes (Stace 1991).

Brachypodium (L.) P. Beauv. is a largely isolated genus of the subfamily Pooideae that has been classified within its own tribe Brachypodieae (Catalán et al. 2016a). Phylogenomic analyses have resolved it in an intermediate position within the Pooideae tree and sister to the recently evolved core pooid clade (Sancho et al. 2018). A characteristic feature of the genus is its shortly pedicellate spikelet, which differentiate it from its close sessile spikelet (Triticeae) or longly pedicellate spikelet (Bromeae, Poeae, Aveneae) relatives (Catalán et al. 1997). *Brachypodium* also separate from them by a number of private biological (embryo development), biochemical (exclusive seed storage proteins, seed globulins, seed storage polysaccharides and stem and leaf fructosans) (Schippmann 1991), karyotype (Robertson 1981;

Khan 1984) and genetic landmark (Catalán et al. 2016a) characters. According to the most recent taxonomic and evolutionary updates, it includes approximately 18 recognized species; three of them are annual species, and 15 are perennial taxa (Schippmann 1991; Catalán et al. 2012; Diaz-Perez et al. 2018). It has been recently demonstrated that the three annuals (*B. distachyon, B. stacei, B. hybridum*) have a wide distribution in their native circumMediterranean region (Catalán et al. 2012; López-Alvarez et al. 2012, 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 isolated or restricted 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. bolusii*), tropical and South Africa (*B. flexum*), Madagascar (*B. madagascariense*), Taiwan (*B. kawakamii*), SE Asia – New Guinea (*B. sylvaticum* var. *pseudodistachyon*), and America (*B. mexicanum*)] (Schippmann 1991; Catalán et al. 2016a; Diaz-Pérez et al. 2018). Approximately, half of the recognized *Brachypodium* species are diploids and the other half are confirmed or purported allopolyploids (Diaz-Perez et al. 2018). Taxonomic uncertainty still persists among some poorly known *Brachypodium* taxa, and particularly within some cryptic complex taxa (Schippmann 1991; Catalán & Olmstead 2008; Catalán et al. 2016a)

The genus shows a remarkable disploidy, with chromosome base numbers of diploids ranging from the presumably more ancestral x=10 (*B. stacei*), through x= 9 (*B. arbuscula, B. sylvaticum, B. pinnatum, B. rupestre*) and x=8 (*B. glaucovirens*), to x=5 (*B. distachyon*) (Robertson 1981; Betekhtin et al. 2014; Catalán et al. 2016a). A recent minimum evolution-based phylogenetic study suggested there were six homeologous diploid genomes that could have participated in allopolyploidization events within *Brachypodium* leading to current extant polyploids and spanning several levels of phylogenetic depth (Diaz-Perez et al. 2018). 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. 2012, 2016b; López-Alvarez et al. 2012, 2017). The demonstration that the model plant was not one but three species (Catalán et al. 2012) opened the way to a thoroughly comparative genomic and morphoanatomical study of this diploid-polyploid complex. Eight quantitative (Catalán et al. 2012; Lopez-Alvarez et al. 2017) and one qualitative (Catalán et al. 2016b) morphological traits were found to discriminate among the diploid *B. stacei* (2n=2x=20) and *B. distachyon* (2n=2x=10) progenitor species and their derived allotetraploid *B. hybridum* (2n=4x=30) species.

Despite these advances, *Brachypodium* is still in lack of deep systematic studies for some of its species. Cryptic taxa also exist among the perennial *Brachypodium* species, namely among the diploid and tetraploid cytotypes (2n=2x=18; 2n=4x=28) of *B. pinnatum* and *B. rupestre* distributed across Eurasia. However, the cytogenetic hypothesis about the potential origin of these Eurasian tetraploids and of the close western Mediterranean tetraploid *B. phoenicoides* (2n=4x=28) from putative crosses of x=9 perennial diploids and an unknown ancestral x=5 diploid (Khan & Stace 1999; Wolny & Hasterok 2009; Betekhtin et al. 2014; Catalán et al. 2016a) has not yet been confirmed. Dated phylogenetic analysis of the *Brachypodium pinnatum* complex species inferred a very recent origin for the diploid lineages

(*B. rupestre*/*B. pinnatum*, 0.3 Ma), and coalescent-based analysis indicated that different allotetraploid *B. pinnatum*, *B. rupestre* and *B. phoenicoides* cytotypes showed the participation of late Quaternary *B. sylvaticum*-type, *B. arbuscula*-type, to *B. glaucovirens*-type and *B. pinnatum*-type diploid genomes (Diaz-Pérez et al., 2018). The long-rhizomatose *B. pinnatum*, *B. rupestre* and *B. phoenicoides* show erect panicles. *B. phoenicoides*, adapted to dry places, is glabrous and presents partially inrrolled leaves, semi-patent twisted spikelets and awnless lemmas, whereas the mesic *B. pinnatum* and *B. rupestre* have short awns and bright green colored leaves. *B. rupestre*, considered until recently a subspecies of *B. pinnatum*, differs from it in its glabrous leaves and spikelets and in leaf epidermal traits (Schippmann 1991; Catalán et al. 2016a).

Our study aims to clarify the taxonomy of the *Brachypodium pinnatum* complex species investigating the morphoanatomical variation of *B. pinnatum*, *B. rupestre* and *B. phoenicoides* across their respective geographic distributions and at the population and cytotypic levels. Our specific goals are to: i) identify phenotypic traits that significantly discriminate the three species; ii) identify potential infraspecific morphotypic groups within each species and the phenotypic traits that discriminate them; iii) establish the potential relationships of infraspecific morphotypic groups to cytotypic, environmental or geographical variables; iv) infer the potential existence of new cytotypes within the studied samples; v) contribute to clarify the nature of the cryptic species.

Materials and Methods

Sampling

A large sampling was performed in order to cover the representation of infraspecific ranks, cytotypes, populations and individuals of the three *B. pinnatum* s. l. complex species in their respective native regions. Few studies have been conducted on the Asian populations, except for those connected with some regional Floras. Even the largely distributed Eurasian species *B. pinnatum* has mostly been studied in the European part of its Holarctic range (Schippmann 1991; Paszko 2007, 2008), although Tsvelev (1976, 2015) also studied it in the former USSR territory. This author recognized five independent species corresponding to taxa formerly classified within *B. pinnatum* (Tzvelev 2015). Our aim was to cover as much intraspecific phenetic and environmental variability as possible, considering that geographic distribution and environmental variability might affect phenetic variability. A total of 509 samples of *B. pinnatum* (273 individuals, 91 populations), *B.rupestre* (90 individuals, 31 populations) and *B. phoenicoides* (146 individuals, 55 populations) were used in the phenotypic analysis (Table S1). The samples were collected in the field or were obtained from herbaria (B, C, FI, G, JACA, LD, LE, MW, NS, NSK, SEV, TK, Unizar). All the studied materials were analysed phenotypically in mature samples. Herbarium vouchers of the newly collected materials have been deposited in the JACA, TK, and Unizar (University of Zaragoza) herbaria.

Phenotypic analysis

Phenotypic analysis was performed using 16 morphological characters and two anatomical character employed to separate and to identify the species of the *B. pinnatum* s. l. complex or

other species of *Brachypodium* in previous studies (Schippmann, 1991; Paszko, 2007, 2008; Lopez-Alvarez, 2017). Fourteen of the morphological characters [Culm Lenght (CL); Leaf Sheath length (LSL), Closing part of Leaf Sheath Length (CLSL), Leaf Blade Length (LBL), Leaf Blade Width (LBW), Inflorescence Length (IL), Distance between 1st and 2nd Spikelet (DS), Spikelet Length (SL), Lower Glume Length (LGL), Upper Glume Length (UGL), Lemma Length (LL), Palea Length (PL), Awn Length (WL), Stomata Guard Cell length (SGCL)] and one of the anatomical characters [Stomata Guard Cell Length (SGCL)] were quantitative, and three morphological characters were discrete [Number of Nodes of Culm (NNC); Number of Spikelets per Inflorescence (NSI); Number of Flowers per Spikelet (NFS)] (Tables 1, S1).

Macromorphological characters were measured with a hard ruler under a dissecting microscope. The microanatomical character SGCL was measured on a subset of 100 samples from the three species under a microscope using an ocular micrometer. For measurements of the stomata guard cell length, abaxial and adaxial epidermises were peeled off from dried leaves which were pre-treated in a 90% lactic acid solution for 8 hours. The leaf blade was softened in boiling water, placed on a glass slide with the abaxial or adaxial side up, and the outer epidermis and mesophyll were scraped off using a scalpel. The resulting adaxial and abaxial epidermis was left in glycerin for several hours to lighten. 10 stomata guard cells were measured per each epidermis and sample at 10X magnifications.

Pollen grain morphology was examined in the same subset of 100 samples from the three species. Anthers were boiled in water and pollen grains were mildly acetolysed with a 9:1 solution of acetic acid (CH3COOH) and sulfuric acid (SO4H2), transferred to glycerin and stained with iodine. Pollen grain samples with large concentrations of silicates required an additional treatment with hydrofluoric acid (HF) before staining. These samples were used for scanning microscopy analysis.

Scanning microscopy

The morphological features of pollen grain were analysed through scanning electron microscopy (SEM). The mature pollen grains were fixed in 2.5 % glutaraldehyde/ 0.1 M PBS (phosphate-buffered saline), pH 7.0, dehydrated in a graded ethanol series and critical point dried. Then, the pollen grains were mounted on aluminum stubs, tweezed to release pollen, sputter coated with gold and photographed using a SNE-4500M (Korea) scanning electron microscope. Up to ten individuals per population were used for assessment of morphological characters.

Statistical analyses

Exploratory descriptive analyses were applied to all samples to eliminate outliers. Simple statistic descriptors of the inter-species and intra-species phenetic diversity (mean, range, standard deviation, box plots of median, range and percentiles) were calculated from the data. The analysis of the inter-species response variables was estimated through one way ANOVA chi-square tests as the variables complied with requirements of normality (Table S1). Multiple pairwise comparisons of means were based on Tukey's tests for groups with unequal samples

sizes. Inter-species and intra-species response variables were evaluated through a multicollinearity analysis in order to determine which characters were correlated with each other in the common data set and in each of the species data sets; a matrix of Spearman correlation coefficients was obtained by averaging the values from each individual in each case. All the statistical analyses were conducted with the program STATISTICA v. 9 (StatSoft Inc.; http://www.statsoft.com/).

Multivariate Principal Component Analysis (PCA) of the 16 morphological variables was performed to examine the structure of the taxa, to assess if the observed groupings were consistent with the taxonomic circumscriptions proposed for the three species of the *Brachypodium pinnatum* s. l. complex, and to evaluate the level of covariation in variables. 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 using STATISTICA v. 9. Intraspecific PCAs were conducted in each of the *B. pinnatum*, *B. rupestre* and *B. phoenicoides* subgroup samples using the same data. These separate analyses would allow estimation of the intraspecific substructure of the taxa.

Results

Morphological differentiation between Brachypodium pinnatum, B. rupestre and B. phoenicoides

The sixteen quantitative and discrete morphological characters complied with the condition of normality after the Kolmogorov-Smirnov tests (Tables 1, S1) and were used to characterize the morphological variation of the *Brachypodium pinnatum* complex species at both interspecific and intraspecific levels. Thirteen of the 16 traits (CL, NNC, LSL, CLSL, LBW, IL, NSI, DS, SL, NFS, LGL, UGL, AL) were useful in significantly discriminating different groups of taxa from each other at $p < 0.05$ (Table 1). The remaining traits (LBL, LL, PL) did not significantly differentiate between the taxa. Four of the discriminant traits (CL, SL, NFS, AL) differentiated the three species from each other, three (LSL, LBW, UGL) differentiated *B. phoenicoides* from *B. pinnatum* + *B. rupestre*, three (NNC, CLSL, LGL) *B. pinnatum* from *B. rupestre* + *B. phoenicoides*, and three (IL, NSI, DS) *B. pinnatum* from *B. phoenicoides* (Table 1, Fig. 1). The individuals of *B. pinnatum* showed values significantly higher for culm length ($\bar{x}$ =66.65 ± 18.16 cm), number of nodes of culm ($\bar{x}$ =3.3 ± 0.74), closing part of the sheath length ($\bar{x}$ =5.92 ± 2.25 cm), lower glume length ($\bar{x}$ =6.32 ± 1.42 mm) and awn length ($\bar{x}$ =3.72 ± 1.31 mm) than those of *B. rupestre* and *B. phoenicoides*, and values significantly higher for number of spikelets per inflorescence ($\bar{x}$ =0.04 ± 1.83) than those of *B. phoenicoides* (Table 1). By contrast, the individuals of *B. phoenicoides* showed values significantly higher for leaf sheath length ($\bar{x}$ =8.63 ± 2.42 cm), spikelet length ($\bar{x}$ =27.43 ± 5.97 mm) and number of florets per spikelet ($\bar{x}$ =13.03 ± 3.38) than those of *B. pinnatum* and *B. rupestre*, and values significantly higher for inflorescence length ($\bar{x}$ =11.47 ± 3.11 cm) and distance between 1st and 2nd spikelet ($\bar{x}$ =19.59 ± 6.65 mm) than those of *B. pinnatum* (Table 1, Fig. 1). The individuals of *B. rupestre* showed significantly different intermediate values between those of *B. pinnatum* and *B. phoenicoides* for awn length, spikelet length and number of florets per spikelet, and non-significantly different intermediate values between those of *B. pinnatum* and *B. phoenicoides* for inflorescence length, number of spikelets per inflorescence, and distance between 1st and 2nd

spikelet (Table 1, Fig. 1). They also showed values higher for upper glume length ($\bar{x}$ =7.63 ± 1.38 mm) than those of *B. phoenicoides* and *B. pinnatum*. Individuals of both *B. pinnatum* and *B. rupestre* showed values significantly higher for leaf blade width than those of *B. phoenicoides* (Table 1, Fig. 1).

The lengths of the stomata guard cells measured in stomata from both adaxial and abaxial epidermises (SGCL_abax, SGCL_adax) did not show significant differences among the three studied species (Fig. 2A, 2B). The SGCL_abax values ranged between 25-30 µm and the SGCL_adax values between 20-30 µm in *B. pinnatum*, *B. rupestre* and *B. phoenicoides*. These values, however, showed broad intervals and extremes, which could reflect the participation of different cytotypes (ploidies) within each species.

Table 1. Statistics of 16 morphoanatomical traits and significance tests of their mean values analysed in individuals from 509 samples of *Brachypodium pinnatum*, *B. rupestre* and *B. phoenicoides*. Underlined variables are those that significantly discriminate between the three species. N, number of samples analysed. ANOVA tests of variables used for comparisons between species. Superscripts denote Tukey pairwise comparisons between species; means with the same letter do not differ significantly ($p<0.05$). See text and Table S2 for abbreviations of variables.

Species	CL (cm)	NNC	LSL (cm)	CLSL (cm)	LBL (cm)	LBW (mm)	IL (cm)	NSI	DS (mm)	SL (mm)	NFS	LGL (mm)	UGL (mm)	LL (mm)	PL (mm)	AL (mm)
B. pinnatum																
N	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273	273
Minimum	30	2	3.4	0.1	3.1	2	1.5	4	2	10	5	2.3	3	2.2	2.2	0.3
Maximum	123.5	9	16	15	34.5	11	18.5	15	35	38	21	10.6	12.4	11.5	11.5	10
Mean	66.65^{A}	3.3^{A}	8.12^{A}	5.92^{A}	16.40	5.82^{A}	10.39^{A}	8.04^{A}	17.35^{A}	21.86^{A}	9.72^{A}	6.32^{A}	7.19^{A}	8.85	8.05	3.72^{A}
Std. Deviation	18.16	0.74	1.86	2.25	5.43	1.64	2.57	1.83	5.22	4.86	2.63	1.42	1.55	1.03	1.19	1.31
B. rupestre																
N	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90	90
Minimum	31.2	2	4.2	0.1	6.7	3	4.6	3	3	13	6	3	4.8	6	4.7	1.5
Maximum	101.2	4	14.5	11.6	31.5	9	23.2	11	43	46	22	11.5	15	14	10	5.8
Mean	47.27^{B}	2.66^{B}	7.80^{A}	4.59^{B}	15.41	5.68^{A}	11.2	7.48	19.08	24.30^{B}	10.93^{B}	5.77^{B}	7.63	8.80	7.90	3.26^{B}
Std. Deviation	11.98	0.55	2.44	2.79	5.77	1.23	3.94	1.84	7.89	6.34	3.54	1.32	1.38	1.27	0.96	1.08
B. phoenicoides																
N	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146
Minimum	26.3	2	4.8	0.1	4.6	2.5	6	4	2	14	6	3.5	4.9	6.8	4.7	0.1
Maximum	105	7	18.2	10.9	41	8	25	14	36	55	22	9	10.5	12.2	11.3	8
Mean	57.47^{C}	2.90^{B}	8.63^{B}	4.04^{B}	15.10	4.24^{B}	11.47^{B}	7^{B}	19.59^{B}	27.43^{C}	13.08^{C}	5.93^{B}	7.57^{B}	9.06	8.22	1.58^{C}
Std. Deviation	14.38	0.84	2.42	2.77	6.03	1.06	3.11	1.98	6.65	5.97	3.38	1.06	1.03	0.98	0.93	1.37

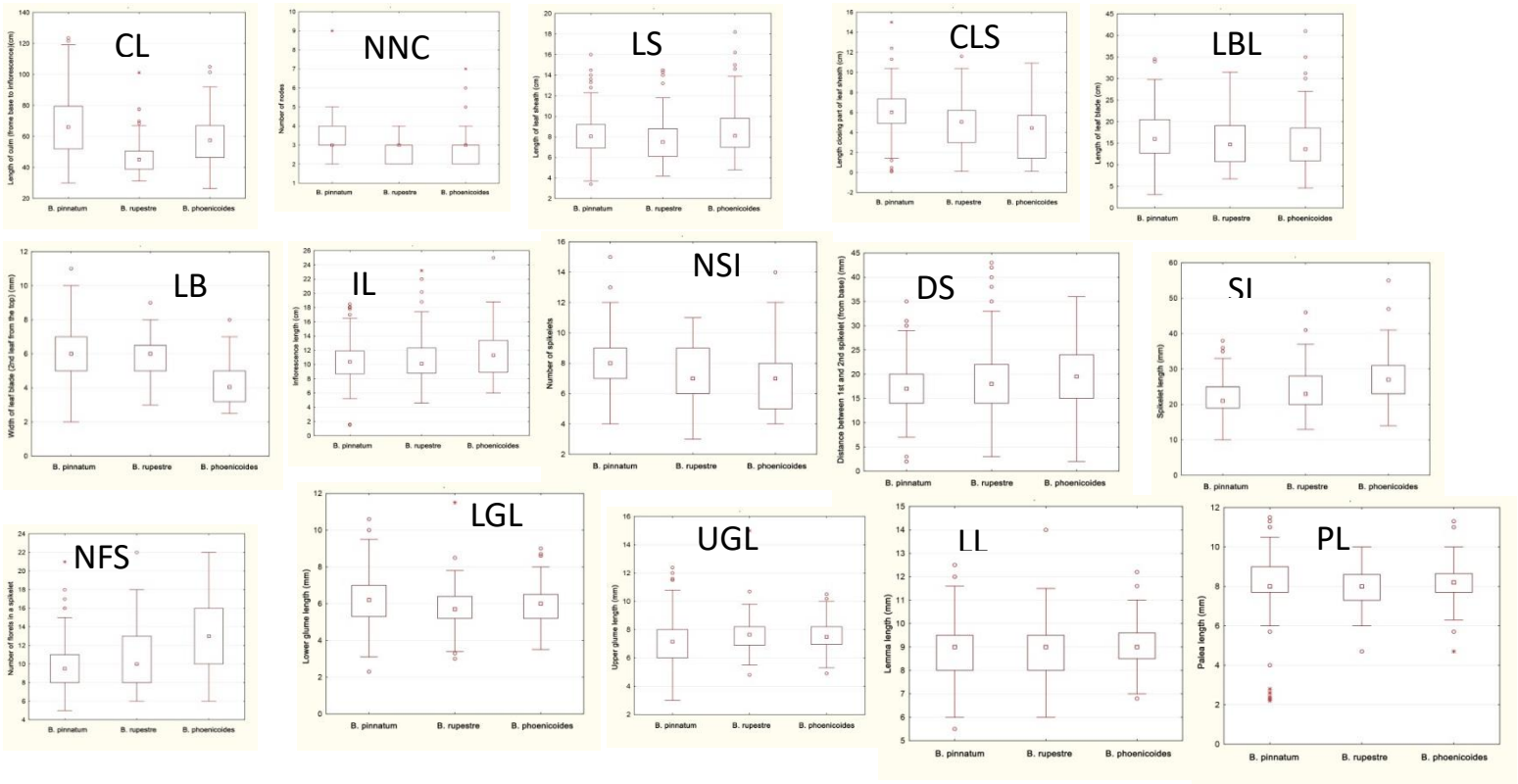
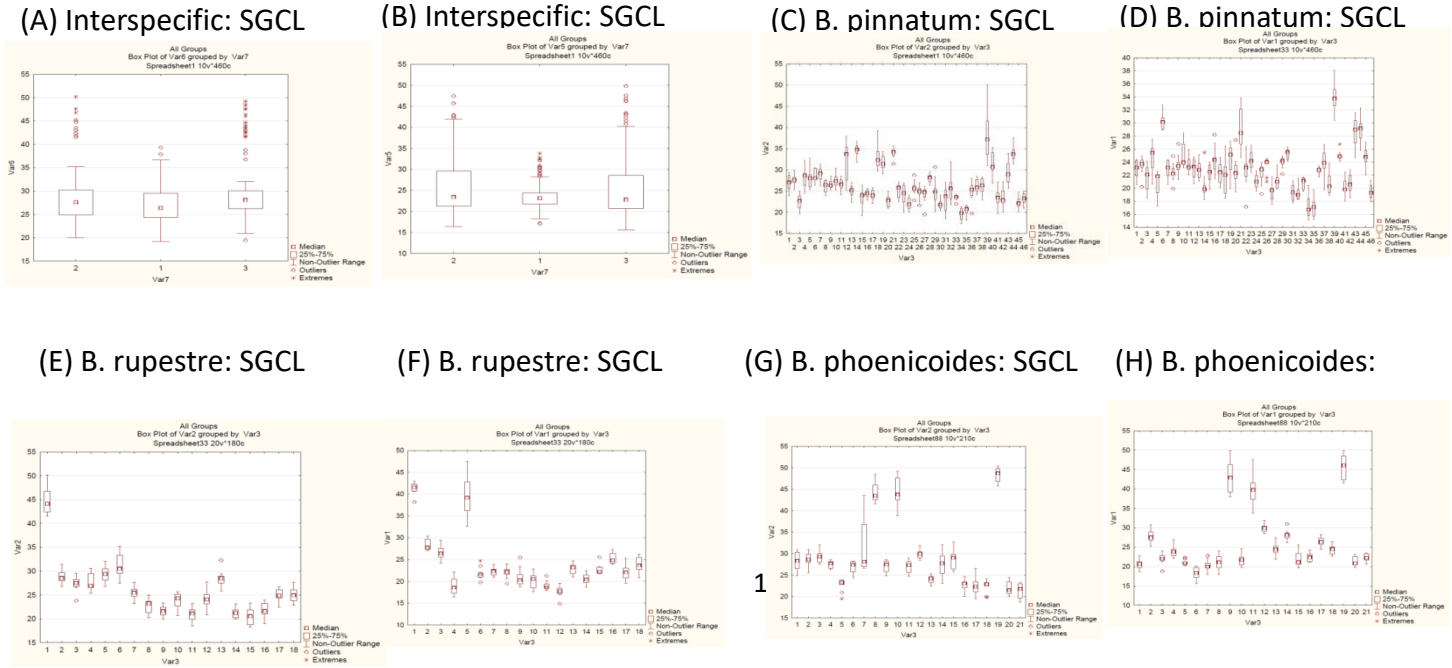


Figure 1. Box and whisker plots of 16 morphoanatomical variables analysed across the studied *Brachypodium pinnatum* (1), *B. rupestre* (2) and *B. phoenicoides* (3) samples.

Figure 2. Box and whisker plots of the microanatomical Stomata Guard Cell length (SGCL) variable analysed across abaxial and adaxial epidermes of individual samples from the three studied *Brachypodium* species (A, B), and of their respective populations: *B. pinnatum* (C, D), *B. rupestre* (E, F), *B. phoenicoides* (G, H).



The PCA identified two principal components which altogether accounted for 42.02% of the observed morphological variation detected by the 16 analysed morphological characters (C1: 25.83%, C2: 16.19%; Table S2, Fig. 3). Five characters were identified as the most important contributors to the positive and negative extremes of the first components of the PCA. LSL (-0,755738) and IL (-0,798985), SL (-0,609746) and NFS (-0,591156), and LL (-0,633729) showed the highest contributions to the first, second and third components, respectively (Table S2). The impact of these components on the hierarchical structure of the three species was visualized in a bidimensional PCA plot (Fig. 3). Population samples of the three species overlapped in the two-dimensional space created by the first two components. Samples of *B. pinnatum* and of *B. phoenicoides* tended to cluster along the opposite positive and negative extremes of component C2, respectively, whereas those of *B. rupestre* overlapped with samples from one and the other species in the middle part of C2 (Fig. 3).

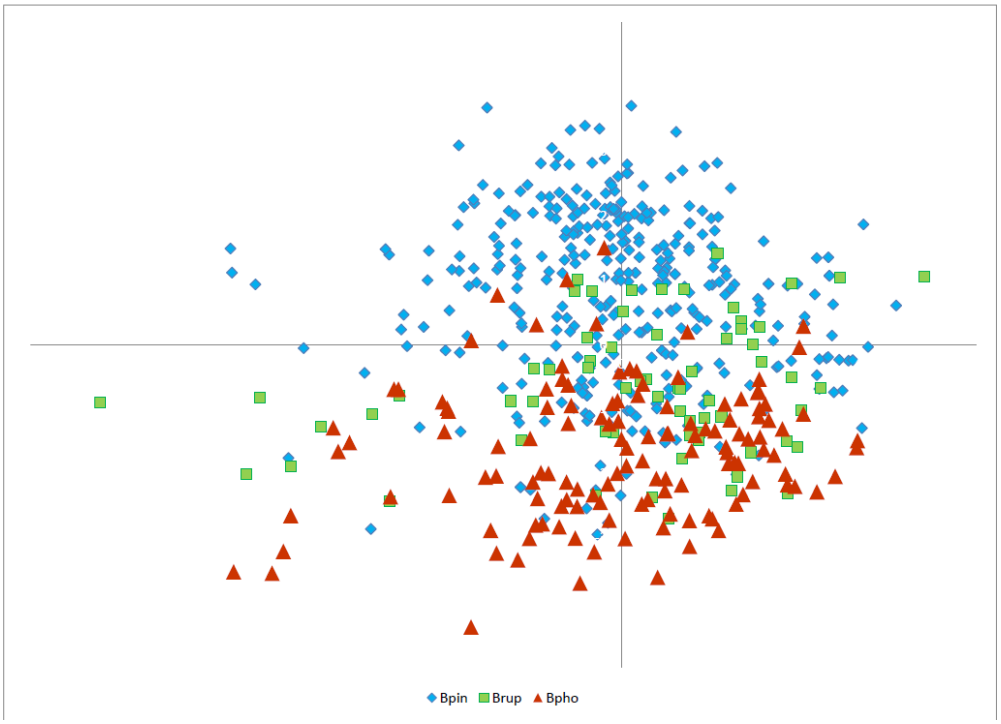


Figure 3. Principal Component Analysis of 509 samples of *Brachypodium pinnatum* (blue diamonds), *B. rupestre* (green squares) and *B. phoenicoides* (red triangles) based on 16 morphoanatomical variables (see Tables 1, S2). Species symbols and colors are indicated in the chart.

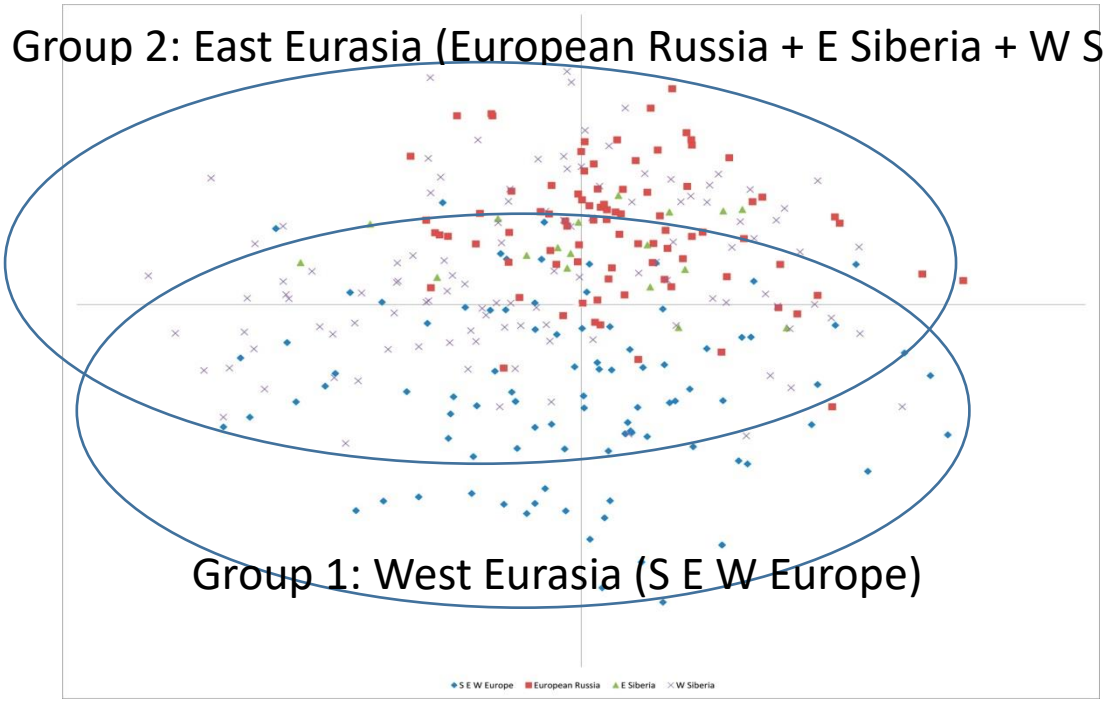


Figure 4. Principal Component Analysis of 273 samples of *Brachypodium pinnatum* based on 16 morphoanatomical variables. Two main West Eurasia and east Eurasia morpho-groups could be differentiated. Other more inclusive groups are indicated in the chart.

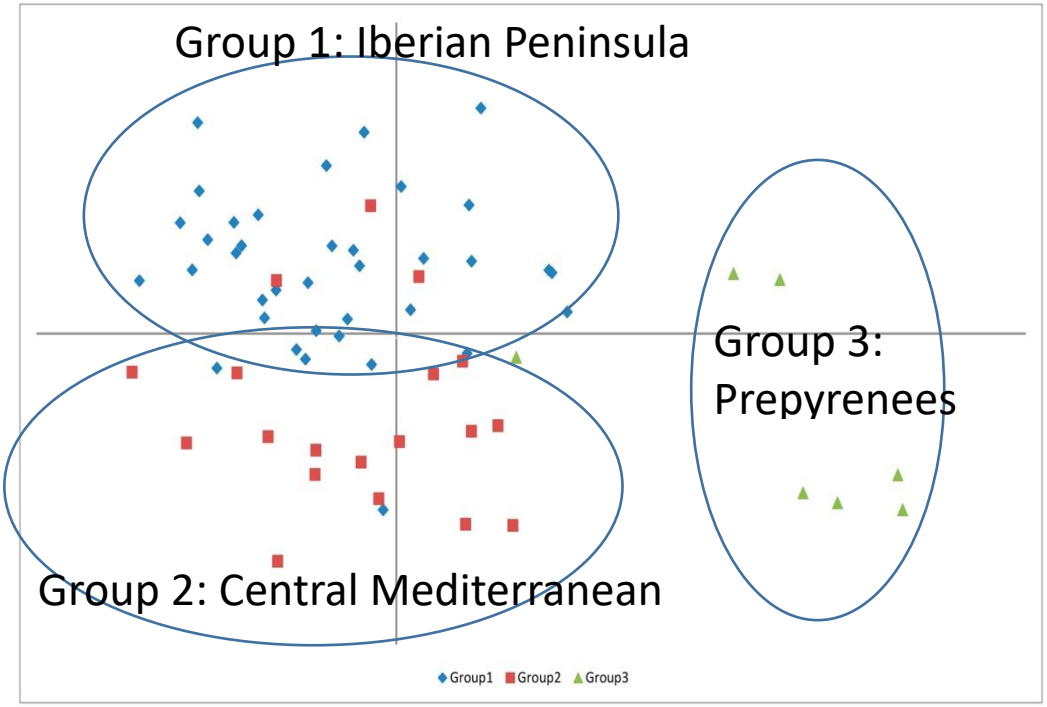


Figure 5. Principal Component Analysis of 90 samples of *Brachypodium rupestre* based on 16 morphoanatomical variables. Three main Iberian Peninsula, Central Mediterranean and Prepyrenees morpho-groups could be differentiated. Symbols and color codes of these groups are indicated in the chart.

Intraspecific differentiation within Brachypodium pinnatum, B. rupestre and B. phoenicoides

The PCAs conducted at the intraspecific level in *B. pinnatum*, *B. rupestre* and *B. phoenicoides* using the same data from the 16 morphological traits showed different geographical groupings of samples and populations in one and another studied species. In *B. pinnatum*, the two-dimensional plot constructed with the first two components indicated the separation of two groups of populations (Fig 4). These two large groups of populations from West Eurasia (E, W and S Europe) and from East Eurasia (European Russian, E Siberia and W Siberia) separated, respectively, at the negative and positive extremes of C2, showing some overlapping of populations around the central part of this component (Fig. 4). No clear subclusters of populations could be further differentiated, neither within the western Eurasian group nor within the eastern Eurasian group (Fig. 5).

In *B. rupestre* the bidimensional PCA plot classified three groups of Iberian Peninsula (western Mediterranean), Central Mediterranean and Prepyrenees populations that separated along C1 and C2 (Fig 6). The Prepyrenean population separated from the rest at the positive extreme of C1, whereas the Iberian Peninsula and the Central Mediterranean populations clustered, respectively at the positive and negative extremes of C2.

In *B. phoenicoides* the first two components of the two-dimensional PCA plot showed the segregation of three groups of North Spain+ (mostly northern Spain populations plus those from other close regions), Central Spain+ (predominantly central Spain populations plus those form other places), and South Spain+ (mostly southern Spain populations plus those form other localities) populations along the first component (Fig. 6). The South Spain+, North Spain+ and Central Spain+ populations clustered, respectively, at the positive extreme, central range, and negative extreme of C1.

The stomata guard cell lengths taken from stomata present in both adaxial and abaxial epidermises (SGCL_abax, SGCL_adax) showed marked differences within each of the three studied species (Fig. 2). In *B. pinnatum* the SGCL_Abax mean values ranged between 19-37 μm and the SGCL_Adax mean values 17-34 μm (Fig. 2C, 2D). Two main groups of populations based on SGCL_Abax mean values and their respective predominantly non-overlapping ranges could be differentiated (<30 μm, ≥30 μm) (Fig. 2C); by contrast, the variation was less structured in the SGCL_Adax (Fig. 2D). In *B. rupestre* the SGCL_Abax mean values ranged between 20-44 μm and the SGCL_Adax mean values 17-42 μm (Fig. 2E, 2F). The SGCL_Abax mean values also separated two clear groups of populations with non-overlapping values (<30 μm, ≥30 μm) (Fig. 2E); it was also observed for the SGCL_Adax values (<30 μm, ≥30 μm) (Fig. 2F) though stomata guard cells with SGCL≥30 μm in both epidermises were only found in one population (no. 1). In *B. phoenicoides* the SGCL_Abax mean values ranged between 21-48 μm and the SGCL_Adax mean values 18-46 μm (Fig. 2G, 2H). Similarly, the SGCL_Abax and SGCL_Adax traits distinguished two groups of populations with non-overlapping values (<30 μm, ≥30 μm), that were coincident for both cases in one population (no. 19) (Fig. 2G, 2H).

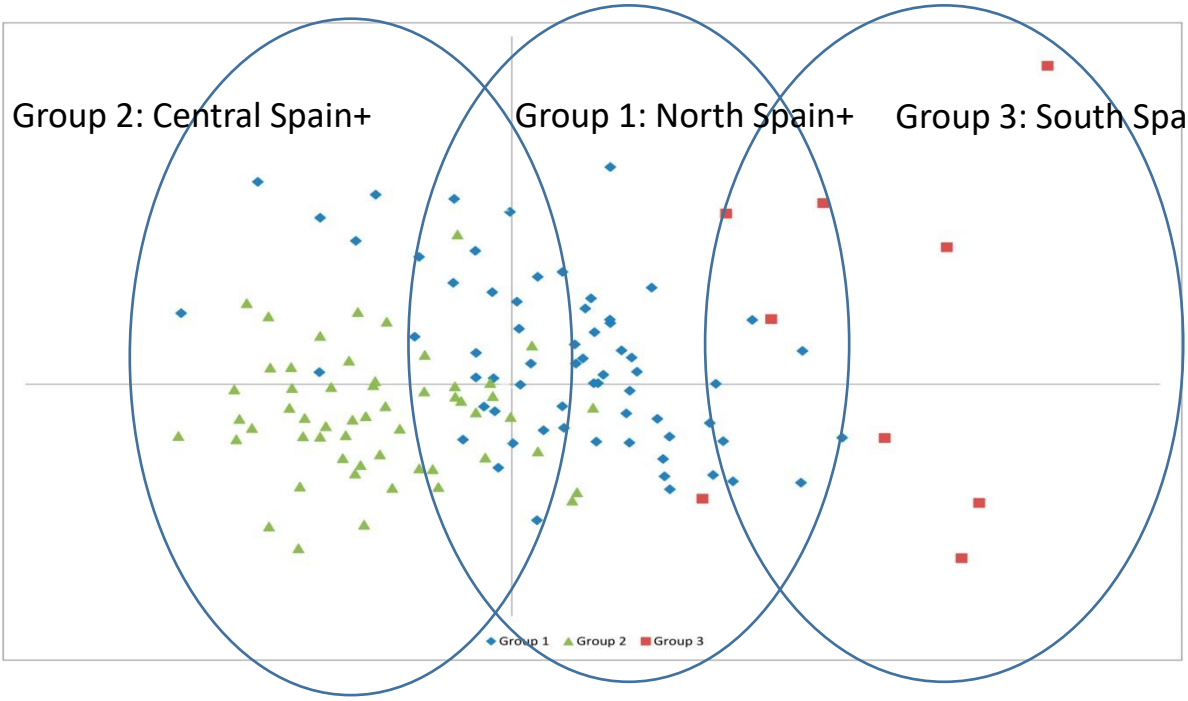


Figure 6. Principal Component Analysis of 146 samples of *Brachypodium phoenicoides* based on 16 morphoanatomical variables. Three main North Spain+, Central Spain+ and South Spain+ morpho-groups could be differentiated. Symbols and color codes of these groups are indicated in the chart.

Pollen morphology

The morphological SEM analysis of the pollen grain showed that mature pollen grains of *B. pinnatum*, *B. rupestre* and *B. phoenicoides* were spheroidal in shape and contained a single operculate, annulate pore (Fig. 7). The exine was predominantly smooth and uniform with granular ornamentation in the pollen grains of the three species (Fig. 7A, 7B, 7C), although some samples of *B. pinnatum* (Fig. 7D) and especially of *B. phoenicoides* (Fig. 7E) also showed a verrucose ornamentation pattern. Some pollen grains were covered with Ubish bodies (Fig. 7E).

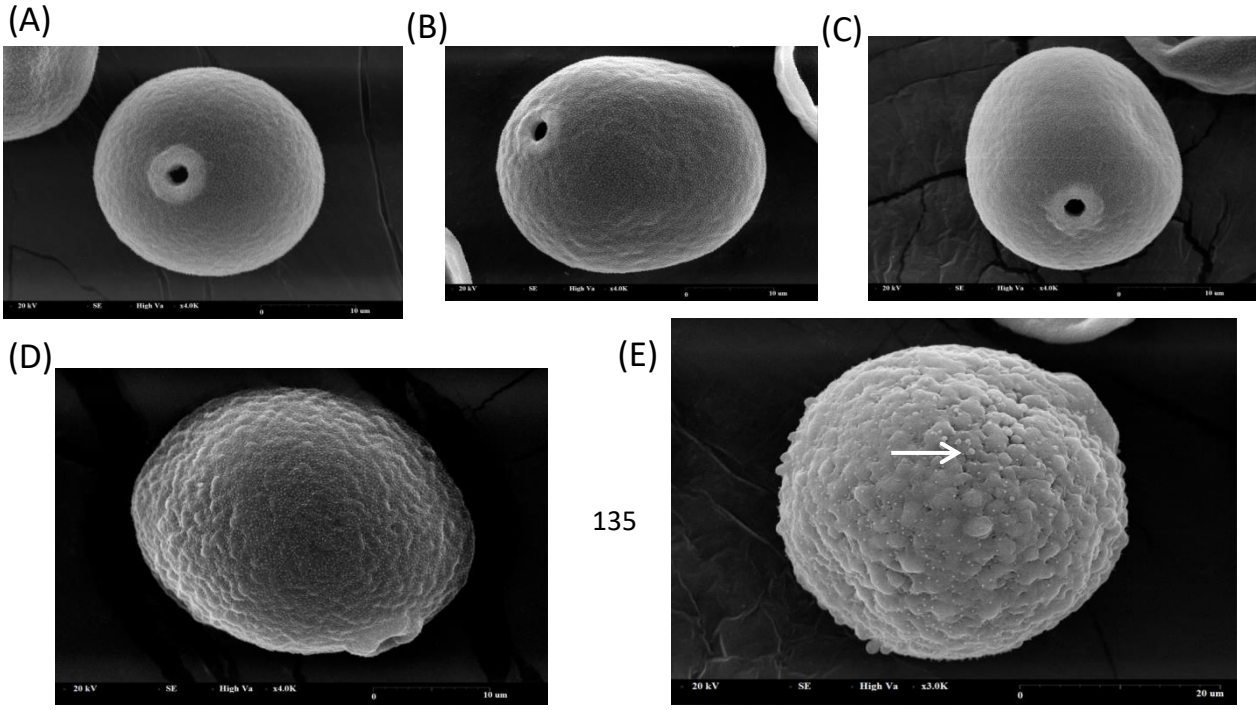


Figure 7. Pollen morphology under scanning electron microscope. Pollen grains of *Brachypodium pinnatum* (A), *B. rupestre* (B) and *B. phoenicoides* (C) showing the operculum, the annulus, and a granular exine ornamentation pattern. Pollen grain of *B. pinnatum* (D) showing a verrucose exine ornamentation pattern. Pollen grain of *B. phoenicoides* (E) showing a verrucose exine pattern and covered with Ubish bodies (arrow).

Discussion

Phenotypic variation and taxonomic differentiation within the Brachypodium pinnatum complex

Our morphoanatomical study has demonstrated that the *Brachypodium pinnatum* complex is an aggregate of at least three phenotypically close species, *B. pinnatum, B. rupestre* and *B. phoenicoides* (Table 1; Figs. 1, 2, 3). Our results concur with the taxonomic treatment of Schippmann et al. (1991), who classified the three taxa at the specific rank, but have also provided the first statistical analysis of morphological and anatomical traits across the broadest geographical representation of the three species studied so far in their respective native ranges. The multivariate PCA analysis shows overlapping but segregated distributions for the two most distinct species, *B. pinnatum* and *B. phoenicoides*, along the opposite extremes of component C2 (Fig. 3). These species also differ significantly in 13 out of the 16 analysed morphological traits (Table 1; Fig. 1). *B. pinnatum* shows overall larger vegetative traits, e. g. taller culms and wider leaves, than *B. phoenicoides* (Table 1; Fig. 1), probably as a consequence of the adaptation of the former species to humid Eurasian habitats and of the second species to dry Mediterranean habitats (Catalán et al. 2016a). By contrast, *B. phoenicoides* shows overall larger reproductive traits, e. g. longer inflorescence and spikelet, and a higher number of florets per spikelet, than *B. pinnatum* (Table 1; Fig. 1), though these traits could be biased by heterosis in the full allopolyploid *B. phoenicoides* with respect to the diploid-allotetraploid *B. pinnatum* (Catalán et al. 2016a). Interestingly, the third species of the complex *B. rupestre* shows an intermediate position between *B. pinnatum* and *B. phoenicoides* in the PCA morphospace (Fig. 3) and also intermediacy in most significantly different and non-significantly different trait values between the other species (Table 1; Fig. 1). Traditionally classified as a variety or as subspecies of *B. pinnatum* in the European and Asian Floras (Tzvelev 1976; Smith 1980), *B. rupestre* was recognized as an independent but close species to *B. pinnatum* by Schippmann (1991) and in some recent Floras (Stace 2019; Tzvelev 2015). Our phenotypic data, however, indicate that *B. rupestre* is as close to *B. pinnatum* as to *B. phoenicoides* (Table 1; Figs. 1, 3), with some *B. rupestre* individuals showing more phenotypic affinities to *B. phoenicoides* individuals and others to *B. pinnatum* individuals (Fig. 3). Like *B. pinnatum, B. rupestre* also shows diploid-allotetraploid ploidies and a predominant distribution in humid Eurosiberian habitats, but similarly to *B. phoenicoides, B. rupestre* also grows in mesic Mediterranean loam edaphic substrates (Catalán et al. 2016a).

The taxonomic boundaries of the three species, and especially those of *B. rupestre* vs *B. pinnatum* and of *B. rupestre* vs *B. phoenicoides* often blur as a consequence of the presumably gene flow experienced by these phylogenetically close lineages (Diaz-Perez et al. 2018). Khan & Stace (1999) demonstrated that *B. pinnatum* and *B. phoenicoides* (*B. rupestre* was integrated in *B. pinnatum*) were outbreeders that could hybridize with other species of the core perennial

clade, like *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 that crosses of diploid and tetraploid individuals of *B. pinnatum* (Khan & Stace 1999). These authors also demonstrated that the 2n=28 chromosomes of the tetraploid *B. pinnatum* and *B. phoenicoides* cytotypes included two parental genomes of x=9 and x=5 chromosomes each. Recent phylogenetic studies indicated that the genome donor of the x=5 chromosome set could be an ancestor that split before the MRCA of the core perennial clade, whereas the genome donors of the x=9 chromosome sets could be recent core perennial diploid ancestors (Sancho et al. 2018). Diaz-Perez et al. (2018) identified *B. sylvaticum*-type and *B. arbuscula*-type diploid ancestors as the genome donors of the x=9 subgenome of tetraploid *B. phoenicoides* and *B. glaucovirens*-type and *B. pinnatum*-type diploid ancestors as those of tetraploid *B. pinnatum* and *B. rupestre*. The potential occurrence of distinct hybridization processes involved in the origins of the different allotetraploid cytotypes of *B. phoenicoides, B. pinnatum* and *B. rupestre* would partially explain the statistically significant morphoanatomical differences observed among the three species in our phenotypic study (Table 1; Figs. 1, 3). Nonetheless, the likely existence of extensive gene flow among the species, especially between the interspecific tetraploid cytotypes, would counterbalance their genetic isolation, impeding the fixation of phenotypic traits. This would explain the overlapping distribution of samples from the three species in the morphoanatomical PCA space (Fig. 3) and the difficulties to properly classify some individuals based solely on morphoanatomical traits.

The above evidences suggest that the *Brachypodium pinnatum* complex could be considered a compilospecies group (*Harlan & Wet, 1963*) formed by three main morphoanatomically differentiable species and several putative hybrid swarms of fertile plants. Pollen grain morphology shows remarkably similar features in the three species (Fig. 7) and could not help to separate them, as indicated for most grasses (Kellogg 2015). *Brachypodium pinatum, B. rupestre* and *B. phoenicoides* show a predominant granular ornamentation pattern in their pollen grain exines, like the one observed in *B. distachyon* (Sharma et al. 2014). Some individuals of *B. pinnatum*, and more frequently of *B. phoenicoides*, also show a verrucose ornamentation pattern of their pollen grain cell wall, though it could be associated with certain ecotypes. The occasional presence of Ubisch bodies on the pollen grains is physiologically related to the production of sporopollenin and structural Raftin protein by the tapetum for the exine coat (Wang et al. 2003; Sharma et al. 2014).

Infraspecific variation and cryptic taxa within the B. pinnatum, B. rupestre and B. phoenicoides

The phenotypic analyses conducted separately at infraspecific level have untapped the enormous infraspecific morphological diversity of each of the three species of the complex. The separate PCA analyses have corroborated the existence of at least two morpho-groups within *B. pinnatum*, three within *B. rupestre* and three within *B. phoenicoides* (Figs. 4, 5, 6). Noticeably, *B. pinnatum* populations from the more restricted western Eurasian region separate from populations from the largest eastern Eurasian region, which clustered together in a

heterogeneous but compact group (Fig. 4). These findings support the hypothesis of the potential existence of long term glacial and interglacial refugia in both regions and along a longitudinal cline, which could have sheltered phenotypically and genetically diverse populations in both sides of their distribution range. Our data support the differentiation of two phenotypic groups which could correspond to cryptic species. However, it does not support the identification of all the five species recognized within the *B. pinnatum* complex in Russia by Tzvelev (2015), though our analysis in only based on 16 quantitative traits.

The infraspecific morphotypic differentiation is even more evident within the more geographically restricted *B. rupestre* and *B. phoenicoides*, where three morpho-groups could be distinguished in each species (Figs. 5, 6). Most populations of *B. rupestre* from the Iberian Peninsula cluster in the same morphotypic group 1, but those from the Prepyrenees differentiate from the former based on leaf sheath lenght and inflorescence length, and those from the Central Mediterranean region separate from the Iberian group based on spikelet length and number of florets per spikelet (Table S2; Fig. 5). In *B. phoenicoides* the three phenotypic groups detected in the PCA (Fig. 6) have an overall geographic component, including in each case populations originated predominantly from North, Central and South Spain, which separate from each other based on traits related to leaf sheath lenght and inflorescence length (Table S2). These morpho-groups could have resulted from isolated populations sheltered in different glacial refugia or from recent postglacial secondary contacts that have melted distinct ecotypes and cytotypes (Diaz-Perez et al. 2018).

One important source of intraspecific variation is based on differences in stomata guard cell lengths. The SGCL-Abax and SGCL-Adax traits have detected diverse population groups within *B. pinnatum*, *B. rupestre* and *B. phoenicoides* (Figs. 2C-2H). This non-endoreduplicating plant cell type trait shows significantly larger measurements in polyploids than in diploids (Catalán et al. 2012). Significant correlation between its increasing values and increasing ploidy levels has been corroborated in the annual *Brachypodium* species (Catalán et al. 2012; López-Alvarez et al. 2017) and in other pooids (Bretagnolle and Lumaret, 1995). Our analyses indicate the existence of at least two SGCL-abax and SGCL_adax groups within each *B. pinnatum*, *B. rupestre* and *B. phoenicoides* species (Figs. 2C-2H). The overlapping smallest SGCL sizes in *B. pinnatum* and *B. rupestre* could correspond to diploid cytotypes and the largest SGCL sizes to low ploidy (4x) cytotypes, though these assignments have to be confirmed though cytotypic studies. *B. phoenicoides* also shows an intermediate range of SGCL values and a large range that could correspond to low a high ploidy cytotypes, respectively (Figs. 2G, 2H). The detection of various infraspecific morphogroups showing significant differences in several traits (Figs. 4, 5, 6) and significantly different SGCL values (Fig. 2) is indicative of the existence of cryptic species within the *B. pinnatum* complex. Some of the cryptic taxa correspond to the 2x and 4x cytotypes of *B. pinnatum* and *B. rupestre* (Betekhtin et al. 2914; Catalán et al. 2016a). A new SGCL group has been detected in *B. phoenicoides* (Fig. 2), that may correspond to an as yet uncovered cytotype. Ongoing genomic studies of these representative samples could help to clarify the origins and nature of these cryptic species.

Acknowledgements

We thank the curators of the B, C, FI, G, JACA, LD, LE, MW, NS, NSK, SEV, TC, Unizar and TK herbaria for loans of specimens and facilities for the study of the *Brachypodium pinnatum*, *B. rupestre* and *B. phoenicoides* materials. This work was supported by the Spanish Ministry of Economy and Competitivity (Mineco) CGL2016-79790-P and University of Zaragoza UZ2016_TEC02 grant projects. VS was funded by a Russian Ministry of Education and Science fellowship. PC and IM were partially funded by a European Social Fund and Aragón Government Bioflora grant.

References

Betekhtin, A., Jenkins, G., Hasterok, R., 2014. Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. *PLoS ONE* 9: e115108.

Bretagnolle F, Lumaret R. 1995. Bilateral polyploidization in *Dactylis glomerata* L. subsp. *lusitanica*: occurrence, morphological and genetic characteristics of first polyploids. *Euphytica* 84: 197–207.

Catalán, P., Kellogg, E.A., Olmstead, R.G. 1997. Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequencing. *Molecular Phylogenetics and Evolution* 8: 1-18.

Catalán P, Olmstead RG. 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.

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

Catalán, P., López-Alvarez, D., Díaz-Pérez, A., Sancho, R., López-Herranz, M.L. 2016a. Phylogeny and evolution of the genus *Brachypodium*. In Vogel J (ed.). *Genetics and genomics of Brachypodium*. *Series Plant Genetics and Genomics: Crops Models*. 9-38. Springer.

Catalan P, López-Álvarez D, Bellosta C, Villar L. 2016b. Updated taxonomic descriptions, iconography and habitat preferences of *Brachypodium distachyon*, *B. stacei* and *B. hybridum* (Poaceae). *Anales del Real Jardín Botánico de Madrid* 73 (1): e028

De Wet JMJ, Harlan JR. 1966. Morphology of the compilospecies *Bothriochloa intermedia*. *American Journal of Botany* 53(1):94-98.

Díaz-Pérez A, López-Álvarez D, Sancho R, Catalan P. 2018. Reconstructing 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.

Fuertes Aguilar J, Rossello JA, Nieto Feliner G. 1999. *Molecular evidence for the compilospecies model of reticulate evolution in Armeria (Plumbaginaceae)*. *Syst. Biol.* 48, 735–754.

Gustafsson ALS, Skrede I, Rowe HC, Gussarova G, Borgen L, Rieseberg LH, et al. (2014) Genetics of cryptic speciation within an arctic mustard, *Draba nivalis*. *PLoS ONE* 9(4): e93834

Harlan JR, Wet JMJ de. 1963. The Compilospecies Concept. Evolution 17 (4): 497–501.

Harmon, L.J. et al. (2003) Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301: 961–964.

Holstein N, Luebert F. 2017. Taxonomy: stable taxon boundaries. *Nature* 548(7666):158.

Janzen DH, Burns JM, Cong Q, Hallwachs W, Dapkey T, Manjunath R, Hajibabaei M, Hebert PDN, Grishin NV. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proc Natl Acad Sci USA* 114:8313–8318.

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

Khan MA. 1984. Biosystematic studies in *Brachypodium* (Poaceae). PhD thesis. University of Leicester.

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

López-Alvarez D, López-Herranz ML, 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): e51058.

López-Alvarez 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

Mayr E 1942. *Systematic and the origin of species*. Columbia University Press, New York.

Paszko, B. 2007. The differing characteristics of *Brachypodium pinnatum* (L.) P. Beauv. and *B. sylvaticum* (Huds.) P. Beauv. *Biodiv. Res. Concervation* 5(8): 11-16.

Paszko, B. 2008. The variability of natural populations of *Brachypodium pinnatum* and *B. sylvaticum* based on morphological features. *Acta Societalis Botanicorum Poloniae* 77(3): 255-262.

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

Sancho R, Cantalapiedra CP, López-Álvarez D, Gordon SP, Vogel JP, Catalan P, Contreras-Moreira B. 2018. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist* 218: 1631-1644.

Schippmann U. 1991. Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). *Boissiera* 45: 1-250.

Sharma A, Singh M, Bhalla P. 2014. Cytochemistry of pollen development in *Brachypodium distachyon*. *Plant Syst Evol* 300: 163–1648.

Smith PM. 1980. *Brachypodium* (L.) P. Beauv., in Tutin TG. Et al. (eds.), *Flora Europaea* 5: 189-190. Cambridge University Press.

Stace CA. 1991. *Plant Taxonomy and Biosystematics*. Cambridge University Press. Cambridge.

Stace CA. 2019. *New Flora of the British Isles*. 4th edition. pp. 1-1266. C & M Floristics. Suffolk.

Stebbins GL. 1969. The significance of hybridization for plant taxonomy and evolution. *Taxon* 18: 26-35.

Stebbins, GL, Dremann CC. 1998. Jepson Manual chromosome numbers may indicate new "cryptic" native grass species. *Grasslands* 8 (3): 4-5.

Stevens PF. 1990. Nomenclatural stability, taxonomic instinct, and flora writing — a recipe for disaster?. In: Baas P., Kalkman K., Geesink R. (eds) *The Plant Diversity of Malesia*. Springer, Dordrecht

Struck TH, Feder JL, Bendiksby M, Birkeland S, Cerca J, Gusarov VI, Kistenich S, Larsson KH, Liow LH, Nowak MD, Stedje B, Bachmann L, Dimitrov D. 2018. Finding evolutionary processes hidden in cryptic species. *Trends Ecol Evol*. 33(3):153-163.

Tsvelev NN. 1976. *Flora of the URSS*. Leningrad. Nauka (in Russian).

Tsvelev NN. 2015. On the genus *Brachypodium* P. Beauv. (Poaceae) in Russia. Novosti sistematiki vysshikh rasteniy 46: 91–97.

Vigalondo, B. et al. (2015) Unmasking cryptic species : morphometric and phylogenetic analyses of the Ibero-North African *Linaria incarnata* complex. *Bot. J. Linn. Soc*. 177: 395–417.

Wang, A., Xia, Q., Xie, W., Datla, R., and Selvaraj, G. 2003. The classical Ubisch bodies carry a sporophytically produced structural protein (RAFTIN) that is essential for pollen development. *Proc. Natl. Acad. Sci. USA* 100: 14487–14492

Wolny E, Hasterok R. 2009. Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Annals of Botany* 104: 873–881.

CONCLUSIONS

1- The population genetic and phylogeographic study of the highly selfing species *Brachypodium stacei* in the western Mediterranean region using SSR markers has detected two main genetic groups, one including SE Iberian +Minorca populations and the other S Iberian + Canary Islands + Majorcan populations, and different colonization routes from the mainland Iberian Peninsula into the islands. A recent colonization scenario could explain the relatively low levels of genetic diversity and low number of alleles found in the Canary Islands populations while older colonization events are hypothesized to explain the high genetic diversity values found in the Majorcan populations. Our study provides widely applicable information about geographical patterns of genetic variation in *B. stacei* to be adequately reflected in the germplasm collection of this model grass species.

2- The genetic study of the highly selfing species *Brachypodium distachyon* in the Iberian Peninsula using SSR markers has found three genetic clusters and several hotspots of genetic diversity. Statistically significant Isolation-By-Distance (IBD) and statistically significant Isolation-By-Environment (IBE) among all studied populations reveals the presence of environmental-driven isolation. Populations growing on basic soils are significantly more diverse than those growing in acidic soils. The higher genetic diversity in eastern Iberian populations occurring in basic soils suggests that these populations can be better adapted than those occurring in western areas of the Iberian Peninsula where the soils are more acidic and accumulate toxic Al ions. The western Iberian acidic soils might prevent the establishment of Al-sensitive *B. distachyon* populations, potentially causing the existence of more genetically depauperated individuals.

3- The population genetic study of the hybrid allotetraploid and highly selfing species *Brachypodium hybridum* in the Iberian Peninsula using SSR markers has detected medium to high levels of genetic diversity and a strong south-to-north genetic structure cline. Genetic substructuring is much higher in the South, where three subclusters of populations have been found, than in the North, where most populations are genetically similar to each other. The absence of significant isolation-by-distance in the Iberian *B. hybridum* populations is a consequence of the high genetic divergence observed between spatially close population clusters in southern Iberia. Our population genetic study supports the existence of long term refugia in the SE Iberian Peninsula for the species. The existence of southern Iberian Penisnula *B. hybridum* mixed populations with distinct genotypes is likely a consequence of the easy long distance dispersal of seeds and subsequent colonization and establishment of this annual species, although they could have also resulted from more favorable climatic conditions in the past, which, in time, contributed to the increased genetic differentiation in the South when compared with the North.

4- The comparative phylogenetic analysis of the allotetraploid species *B. hybridum* and its progenitor diploid species *B. distachyon* and *B. stacei* in the western Mediterranean region using plastid and nuclear data has detected three different bidirectional allopolyploidization

events: two involved distinct *B. distachyon*-like ancestors and one involved a *B. stacei*-like ancestor. The nuclear data indicate a close relatedness of current *B. hybridum* D allelic profiles with those of *B. distachyon*, but a lack of similarity with those of *B. stacei*. The Southeastern Iberian Peninsula *B. hybridum* populations could have originated from at least two hybridization events whereas Northeaster-Northwestern Iberian Peninsula *B. hybridum* populations may have derived from at least one hybridization event. The genetic and evolutionary evidence supports the plausible *in situ* origin of the southeastern and northern Iberian Peninsula *B. hybridum* allopolyploids from their respective local *B. distachyon* and unknown *B. stacei* ancestors. The untapped multiple origins and genetic variation detected in these *B. hybridum* populations opens the way to future evolutionary analysis of allopolyploid formation in the *B. hybridum – B. distachyon – B. stacei* grass model complex.

5- The environmental niche modeling analysis of the perennial species *Brachypodium sylvaticum*, *B. pinnatum* and *B. phoenicoides* using occurrence data, 19 bioclimatic variables and Maxent has reconstructed the realized niches of the three species for current and past (Last Glacial Maximum, Mid Holocene) scenarios. *B. pinnatum* is a cold-adapted continental species showing a large potential distribution range in Eurasia, although it has a unique niche in the central-northern Asian range, *B. sylvaticum* is a temperate oceanic species showing a main niche distribution in the half western humid Eurasia and in the East Asia Pacific coast, and *B. phoenicoides* is a warm Mediterranean species showing a restricted niche distribution in the western range of the basin. The analyses detected cryptic glacial refugia for *B. pinnatum* in central-northern Europe and in central-northern Siberia, continuum glacial refugia for *B. sylvaticum* in coastal, islands and lowland areas of western Europe and eastern Asia, and lowered coastal Mediterranean refugia for *B. phoenicoides* in the western Mediterranean region. The widespread *B. sylvaticum* is currently a rare and protected species in western Siberia. Niche distribution modelling predicts a range contraction shift of the plant from the LGM to the present in the area. The longevity of the species and its capability to propagate could reduce its potential risk of extinction if the Siberian black taiga and broad-leaved forests that constitute its natural niche in the region are preserved.

6- The morphoanatomical analysis of the perennial species *Brachypodium pinnatum*, *B. rupestre* and *B. phoenicoides* and its cytotypes and ecotypes across their respective native ranges has detected significant differences among the species and a large phenotypic diversity within them. *B. pinnatum* and *B. phoenicoides* are the most distinct species, differing in 13 discriminant traits. *B. rupestre,* formerly classified within *B. pinnatum*, shows intermediate features between those of *B. pinnatum* and *B. phoenicoides* for several traits and also separates from them in three phenotypically discriminant traits. Up to three, four and three morpho-groups have been respectively identified within *B. pinnatum*, *B. rupestre* and *B. phoenicoides*. Each group separates from its conspecifics in several discriminant traits and some of them show a specific geographic distribution. These infraspecific groups may be associated to different glacial and interglacial population refugia or to potential interspecific fertile crosses between similar ploidy-level cytotypes. Stomata guard cell length

measurements indicate the potential existence of different ploidies within *B. pinnatum*, *B. rupestre* and *B. phoenicoides* that could correspond to different cryptic taxa.

APPENDICES

Appendix I: Supporting Information of Chapter 1

Supplemental Information 1**:** List of alelles genotyped in *Brachypodium stacei*. peerj-04-2407-s001.pdf (13K). DOI: 10.7717/peerj.2407/supp-1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018678/

Appendix II: Supporting Information of Chapter 2

Additional file 1: Table S1. Genotypes of the studied Spanish *Brachypodium distachyon* individuals based on nSSR analysis (TXT 12 kb). https://static-content.springer.com/esm/art%3A10.1186%2Fs12862-017-0996-x/MediaObjects/12862_2017_996_MOESM1_ESM.txt

Additional file 2: Table S2. Description of the bioclimatic worldclim layers (http://www.worldclim.org/bioclim) used in the correlation analysis (XLSX 10 kb). http://www.worldclim.org/bioclim

Additional file 3: Figure S1. STRUCTURE analysis of *Brachypodium distachyon* in Spain. (A) Mean log probability of data LnP(D) over 10 runs for each K value as a function of K (error bars represent standard deviation). (B) Evanno's ad hoc statistic; DK as a function of K (PDF 376 kb). https://static-content.springer.com/esm/art%3A10.1186%2Fs12862-017-0996-x/MediaObjects/12862_2017_996_MOESM3_ESM.pdf

Additional file 4: Figure S2. Bioclimatic variables significantly associated to genetic diversity in *Brachypodium distachyon*. A. BIO3. B. BIO8. C. BIO9 (PDF 9255 kb). https://static-content.springer.com/esm/art%3A10.1186%2Fs12862-017-0996-x/MediaObjects/12862_2017_996_MOESM4_ESM.pdf

Appendix III: Supporting Information of Chapter 3

Table S1. Sampled populations of *Brachypodium hybridum* sorted by geographical areas. The location, population code, number of plants genotyped (N), latitude, longitude, mean ± SE observed heterozygosity (H_o) and mean ± SE expected heterozygosity (H_e), total number of alleles (N_A), mean allelic richness (A_R), inbreeding coefficient estimated in FSTAT(F_{IS}), inbreeding coefficient estimated using the Bayesian procedure implemented in INEst (F_{IS})*, selfing rate (s), and number of exclusive genotypes within populations (% between parenthesis) are shown. (ς) High negative artefactual F_{IS} values (-1, -0.773, -0.429) were retrieved from FSTAT analysis of predominant homomorphic homozygous genotypes in the populations (treated as a single genotype) for most loci except for one or few heterozygous individuals at some particular loci.

Locality	**Code**	**N**	**Latitude**	**Longitude**	**H_o**	**SE**	**H_e**	**SE**	**N_A**	**A_R**	**F_{IS}**	**F_{IS}***	**s**	**Exclusive genotypes**
North-Western (NW)														
Portugal: Trás-os-Montes, Bemposta	**BEMP**	9	41° 18' 2" N	6° 28' 41" W	0	0	0.081	0.046	18	1.286	1	1	1	6 (67%)
Portugal: Trás-os-Montes, Mogadouro	**MOGA**	9	41° 25' 6" N	6° 49' 20" W	0.063	0.049	0.074	0.036	18	1.286	0.2	0.963	0.333	5 (55%)

Spain: Salamanca, El Bodón	**BODO**	8	40° 26' 22" N	6° 35' 27" W	0.071	0.071	0.036	0.036	15	1.071	-1	0.02	-	1 (12%)
Spain: Salamanca, La Fregeneda	**FREG**	10	40° 59' 39" N	6° 51' 50 " W	0	0	0	0	13	0.929	-	-	-	1 (10%)
North-Eastern (NE)														
Spain: Zaragoza, Alfranca	**ALFR**	10	41° 36' 1" N	0° 45' 36" W	0.143	0.097	0.071	0.049	16	1.143	-1	0	-	1 (10%)
Spain: Zaragoza, Belchite	**BELC**	10	41° 17' 57" N	0° 44' 30" W	0.071	0.071	0.036	0.036	15	1.071	-1	0.006	-	1 (10%)
Spain: Lleida, Menarguens	**MENA**	10	41° 43' 12" N	0° 43' 28" E	0.214	0.114	0.107	0.057	17	1.214	-1	0	-	1 (10%)
Spain: Girona, Roses, Castell de la Trinitat	**ROSE**	10	42° 14' 51" N	3° 11' 34" E	0.143	0.097	0.08	0.054	17	1.214	-0.773	0.0001	0	3 (30%)
Spain: Girona, Cadaqués, Port Lligat	**LLIG**	10	42° 17' 41" N	3° 17' 15" E	0.143	0.097	0.097	0.049	17	1.214	-0.429	0.019	0	2 (20%)
Spain: Girona, Lladró, Cap Lladró	**LLAD**	10	42° 24' 34" N	3° 9' 9" E	0.071	0.071	0.061	0.038	17	1.214	-0.111	0.521	0	2 (20%)
Spain: Barcelona, El Prat Del Llobregat	**PRAT**	10	41° 17' 12" N	2° 3' 25" E	0	0	0.026	0.017	16	1.143	1	1	1	3 (30%)

Spain: Tarragona, Poble Nou del Delta	**POBL**	10	40° 38' 26" N	0° 44' 22" E	0.071	0.071	0.07	0.048	16	1.143	0.032	0.723	0.062		2 (20%)
Spain: Teruel, Calaceite, Poblado S. Antonio	**CALA**	10	41° 0' 17" N	0° 11' 6" W	0.214	0.114	0.167	0.06	19	1.357	-0.233	0.09	0		4 (40%)
North-Western (SW)															
Portugal: Algarve, Faro	**FARO**	10	37° 0' 51" N	7° 58' 35" W	0.136	0.087	0.102	0.058	19	1.357	-0.286	0.053	0		5 (50%)
Spain: Huelva, Lepe	**LEPE**	9	37° 13' 44" N	7° 11' 54" W	0.111	0.074	0.086	0.049	18	1.286	-0.244	0.121	0		3 (33%)
Spain: Cádiz, Algeciras	**ALGE**	9	36° 8' 36" N	5° 30' 42" W	0.175	0.099	0.202	0.07	21	1.5	0.193	0.949	0.323		5 (55%)
Spain: Málaga, Algatocín	**ALGA**	6	36° 34' 54" N	5° 16' 36" W	0.083	0.072	0.133	0.062	20	1.429	0.449	0.998	0.619		4 (67%)
Spain: Cáceres,Monfrague	**MONF**	9	39° 49' 19" N	6° 2' 52" W	0.032	0.032	0.102	0.061	20	1.429	0.719	1	0.836		6 (67%)
Spain: Córdoba, Carabuey, Sierra Alcaide	**ALCA**	10	37° 31' 1" N	4° 12' 35" W	0	0	0.104	0.045	19	1.357	1	1	1		5 (50%)
Spain:Córdoba, Cabra	**CABR**	10	37° 27' 24" N	4° 22' 32" W	0	0	0.036	0.025	16	1.143	1	1	1		3 (30%)
South-Eastern (SE)															
Spain: Jaén, La Aliseda	**ALIS**	8	38° 19' 47" N	3° 34' 44" W	0.009	0.009	0.182	0.065	22	1.571	0.957	1	0.978		8 (100%)

Spain: Jaén , Jaén	**JAEN**	10	37° 45' 34" N	3° 47' 53" W	0.029	0.029	0.114	0.051	19	1.357	0.772	1	0.871	6 (60%)
Spain: Jaén, La Cimbarra	**CIMB**	9	38° 23' 27" N	3° 22' 13" W	0.19	0.103	0.158	0.07	19	1.357	-0.15	0.203	0	4 (44%)
Spain: Jaén, El Yelmo	**YELM**	10	38°15'14" N	2°39'50" W	0	0	0.013	0.013	15	1.071	1	1	1	2 (20%)
Spain: Jaén, Hornos, Sierra Segura	**HORN**	10	38°12'52" N	2°42'50" W	0	0	0.034	0.034	15	1.071	1	1	1	2 (20%)
Spain: Jaén, Quesada, Sierra Cazorla	**TISC**	7	37°50'09" N	3°05'05" W	0	0	0	0	14	1	-	-	-	1 (14%)
Spain: Jaén, Hinojares, Sierra Cazorla	**HINO**	11	37°43'08.8" N	2°59'58.9"W	0	0	0.068	0.04	17	1.214	1	1	1	4 (36%)
Spain: Almería, Sorbas	**SORB**	12	37° 6' 19" N	2° 6' 34" W	0.018	0.018	0.061	0.051	18	1.286	0.73	1	0.844	5 (42%)
Spain: Cazorla, Pto de las Palomas	**POBL**	10	37° 52' 11" N	2° 56' 23" W	0.157	0.096	0.158	0.057	21	1.5	0.055	0.745	0.104	5 (50%)
Spain: Granada, Nigüelas	**DURC**	10	36° 59' 26" N	3° 30' 30" W	0.071	0.071	0.07	0.048	15	1.071	0.032	0.728	0.062	2 (20%)
Spain: Granada, Baza	**BAZA**	10	37° 32' 38" N	2° 51' 29" W	0.071	0.071	0.036	0.036	15	1.071	-1	0.006	-	1 (10%)
Spain: Granada, Cubillas	**CUBI**	10	37° 16' 37" N	3° 40' 10" W	0.079	0.071	0.118	0.066	23	1.643	0.381	0.999	0.552	8 (80%)
Spain: Granada, Sierra Nevada	**NEVA**	8	37°02'12" N	3°18'53" W	0	0	0	0	12	0.857	-	-	-	1 (12%)

Spain: Murcia, Sierra Espuña	**ESPU**	4	37° 49' 14" N	1° 35' 58" W	0	0	0.17	0.055	20	1.429	1	1	1	4 (100%)
Spain: Murcia, Moratalla	**MORA**	14	38° 10' 37" N	1° 55' 38" W	0	0	0.087	0.046	17	1.214	1	1	1	2 (14%)
Spain: Ciudad Real, Ruidera	**RUID**	10	38° 57' 42" N	2° 52' 17" W	0.007	0.007	0.007	0.007	15	1.071	0	1	-	2 (20%)

Table S2. Genotypes of the 342 studied *B. hybridum* individuals from 36 Iberian populations encoded for 17 SSR loci. The *B. hybridum* genotypes were encoded separately for the respective S (*B. stacei*-type) and D (*B. distachyon*-type) subgenomes. Upon request to the authors.

Table S3. SSR phenotypes of 973 individuals of *B. hybridum* (322 individuals from 35 populations), *B. stacei* (181 individuals from 19 populations), and *B. distachyon* (148 individuals from 16 populations) encoded for 98 binary (presence/absence) alleles. The *B. hybridum* phenotypes were encoded separately for the respective S (*B. stacei*-type) and D (*B. distachyon*-type) subgenomes. Upon request to the authors.

Table S4. *Brachypodium hybridum*, *B. distachyon* and *B. stacei* accessions used in the plastid phylogenetic analysis of the multiple origins of *B. hybridum* plastotypes using plastid trnLF and ndhF DNA sequences. Taxa, accessions codes, locatity of origin, Genbank trnLF and ndhF codes and ENA plastomes codes are indicated for each accession. New plastid sequences obtained in this study are indicated in bold. trnLF and ndhF plastid data from plastomes were obtained from total genomic sequences of resequenced *B. distachyon* and *B. hybridum* lines at the Joint Genome Institute, following the procedures indicated in Sancho et al. (2018). Missing ndhF sequences are indicated with a hyphen.

Species	Code	Locality	trnLF	ndhF	Plastome
B.hybridum	BhybABR113	Portugal: Lisboa	JN187658	JN187632	
B.hybridum	Bhyb10	Portugal: Lisboa	JN187664	JN187639	
B.hybridum	Bhyb14_ALFR	Spain: Zaragoza, La Alfranca	JN187667	JN187642	
B.hybridum	Bhyb15_LLIG	Spain: Girona, Cadaques, Port Lligat	JN187668	JN187643	
B.hybridum	Bhyb26_CIMB	Spain: Jaen,La Cimbarra	JN187662	JN187637	
B.hybridum	Bhyb30_1_LEPE	Spain: Huelva, Lepe	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb30_2_LEPE	Spain: Huelva, Lepe	JX665920	-	
B.hybridum	Bhyb54	Morocco: High Atlas, El-Ksiba-Imilchil.	JX665887	-	
B.hybridum	Bhyb89	Morocco: Middle Atlas, Taza, Tahala-Tissa	JX665977	-	
B.hybridum	Bhyb118_5	Spain: Almeria: Cabo de Gata: Genoveses	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb118_8	Spain: Almeria: Cabo de Gata: Genoveses	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb123_1	Spain: Jaen: Srra Segura: Cortijos Nuevos	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb123_6	Spain: Jaen: Srra Segura: Cortijos Nuevos	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb127	Spain: Jaen: Hinojares2	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	Bhyb130_1	Spain: Jaen: Fontanar: Guadiana	**xxxxxxxx**	**xxxxxxxx**	

		Menor			
B.hybridum	Bhyb162	Spain: Jaen: Srra Segura, La Vaquilla	**xxxxxxxx**	**xxxxxxxx**	
B.hybridum	BhybBelchi	Spain: Zaragoza, Belchite	JX665848	-	
B.hybridum	Bhyb_CALA	Spain: Teruel, Calaceite, Poblado	JX665898	KP709773	
B.hybridum	BhybABR110	France: Aude	JN187657	JN187633	
B.hybridum	BhybABR112	France: Corsica, Bonifacio	JN187662	JN187637	
B.hybridum	BhybABR117	Afghanistan: USDA PI219965	JN187635	JN187660	
B. distachyon	BdisABR3	Spain: Huesca: Aisa			LT558584
B. distachyon	BdisABR4	Spain: Huesca: Aren			LT558585
B. distachyon	BdisABR5	Spain: Huesca, Jaca			LT558586
B. distachyon	BdisABR6	Spain: Navarra, Los Arcos			LT222229
B. distachyon	BdisABR7	Spain: Valladolid, Otero			LT558587
B. distachyon	BdisArn1	Spain: Huesca, Aren			LT558591
B. distachyon	BdisBd30_1	Spain: Granada, Dilar			LT558599
B. distachyon	BdisFoz1	Spain: Navarra: Foz de Lumbier			LT558613
B. distachyon	BdisJer1	Spain: Huesca, Adahuesca, San Jerónimo			LT558614
B. distachyon	BdisLuc1	Spain: Huesca, Berdún, Santa Lucía			LT558619
B. distachyon	BdisMig3	Spain: Huesca, Ibieca, San Miguel de Foces			LT558620
B. distachyon	BdisMon3	Spain: Zaragoza, Castejón de Monegros			LT558621
B. distachyon	BdisMur1	Spain: Lérida, Castillo de Mur			LT558622
B. distachyon	BdisPer1	Spain: Navarra, Puerto del Perdón			LT558623
B. distachyon	BdisRon2	Spain: Navarra, Roncal			LT558625
B. distachyon	BdisUni2	Spain: Huesca, Escuela Politécnica Superior			LT558630
B. distachyon	Bdis19	France: Herault, Octon	JN187661	JN187636	
B. distachyon	Bdis22	Spain: Huesca, Guasillo	JN187666	JN187641	
B. distachyon	Bdis52	Morocco:Rif mountains,10km from Ketama.	JX665885	-	
B. distachyon	Bdis53	Morocco: Rif mountains 17km from Ketama	JX665886	-	
B. distachyon	Bdis54	Morocco: Middle Atlas, 14km from Azrou	JX665888	-	
B. distachyon	BdisABR9	Slovenia: Lubjana	JN187663	JN187638	
B. distachyon	BdisBd1_1	Turkey: Manisa, Soma			LT558592
B.	BdisBd18_1	Turkey: Kiresehir, Kaman			LT558595

distachyon					
B. distachyon	BdisTek2	Turkey: Tekirdag			LT558629
B. distachyon	BdisBis1	Turkey: Bismil			LT558612
B. distachyon	BdisBd21	Iraq: Salah ad Din- Salahuddin			LT558597
B. stacei	BstaABR114	Spain: Formentera, Torrent	JN187659	JN187634	
B. stacei	Bsta5	Spain: Alicante, Cabo de La Nao	JX666003	KP709772	
B. stacei	Bsta24	Spain: Canary Isles: Gomera, Agulo_TE4.3	JX666022	-	
B. stacei	Bsta32	Morocco: Anti-Atlas mountains,Taliouine	JX666028	-	
B. stacei	Bsta38	Spain: Mallorca, Sa Dragonera, Gambes	JX666033	-	
Outgroups					
O. sativa	Oryza	cultivar	KP711163.1	X15901	
M. ciliata	Melica	Spain; Ebro valley, PC 1793	JN187650.1	U71049.1	
G. declinata	Glyceria	Spain: Picos de Europa, PC 2893,	JN187651.1	U71048	

Table S5. Characteristics of the SSRs markers used in the population genetic study of Iberian Peninsula *Brachypodium hybridum* populations. Loci were encoded separately for the respective S and D subgenomes. Total number of alleles (N_A), proportion of null alleles (pnull), mean allelic richness (A_R), observed heterozygosity (H_o), expected heterozygosity (H_e), total heterozygosity (H_T), coefficient of genetic differentiation (G_{st}), inbreeding coefficient estimated in FSTAT (F_{IS}), inbreeding coefficient estimated using the Bayesian procedure implemented in INEst (F_{IS})*, and deviation from Hardy-Weinberg Equilibrium (HWE), are show for each locus.

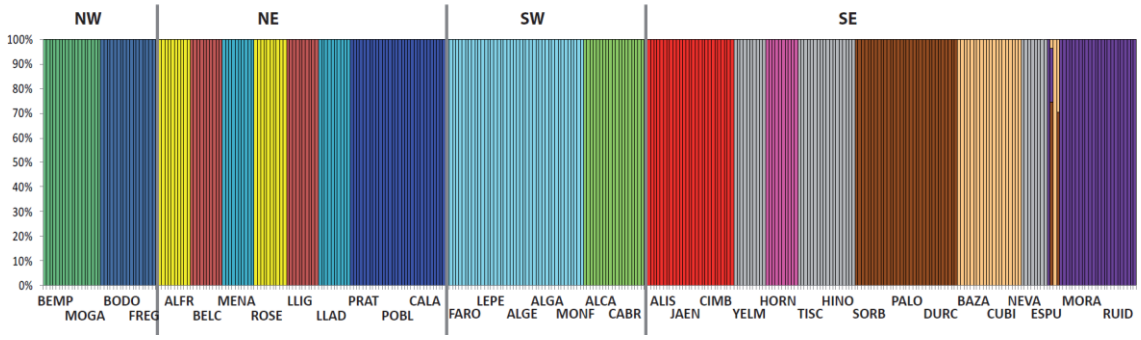
Locus	**Allele size range (bp)**	**N_a**	**p_{null}**	**A_R**	**H_o**	**H_e**	**H_T**	**G_{st}**	**F_{IS}**	**F_{IS}***	**HWE**
ALB006S	338-346	37	0.0243	1.028	0.075	0.066	0.557	0.882	-0.346	0.864	0.820
ALB006D	350-392	34	0.0136	0.944	0.103	0.077	0.803	0.904	-0.341	0.868	0.909*
ALB040S	178-192	44	0.0003	1.222	0.003	0.068	0.701	0.903	0.946	0.996	0
ALB040D	178-184	43	0.0110	1.194	0.028	0.075	0.723	0.896	0.534	0.959	0
ALB050D	209-293	64	0.0001	1.778	0.361	0.251	0.903	0.722	-0.439	0.594	0.999
ALB086S	174-194	37	0.0268	1.028	0.000	0.012	0.721	0.983	1.000	1.000	0.031
ALB086D	188-206	39	0.0003	1.083	0.000	0.022	0.627	0.966	1.000	1.000	0
ALB087A	190-202	49	0.0002	1.361	0.056	0.136	0.701	0.806	0.577	0.916	0

ALB139D	302-336	51	0.0186	1.417	0.137	0.156	0.776	0.799	0.151	0.839	0
ALB165S	157-181	40	0.0003	1.111	0.000	0.043	0.532	0.919	1.000	1.000	0
ALB165D	157-195	42	0.0003	1.167	0.028	0.063	0.775	0.918	0.573	0.966	0
ALB181	230-280	68	0.0002	1.889	0.166	0.256	0.852	0.700	0.341	0.802	0
ALB311S	246-252	37	0.0003	1.028	0.000	0.006	0.468	0.987	1.000	1.000	0.059
ALB311D	240-246	39	0.0003	1.083	0.000	0.027	0.403	0.934	1.000	1.000	0

Na: total number of alleles; p_{null}: average frequency of null alleles across populations; A_R: average allelic richness; H_s: expected within population Nei's heterozygosity; H_o: observed within population Nei's heterozygosity; H_T: expected Nei's heterozygosity within the total population; G_{st}: the Nei's measure of genetic differentiation; F_{IS}: inbreeding coefficient estimated in Fstat; F_{IS}*: inbreeding coefficient estimated using the Bayesian procedure implemented in INEST). HWE: * Empty cells indicate no deviation from HWE values; *($P < 0.05$) indicate significant deviation from HWE after sequential Bonferroni correction.

Supplementary Figures:

Figure S1. Population structure of *Brachypodium hybridum* in the Iberian Peninsula at *K*=14, following the best assignments retrieved by STRUCTURE. Each individual is represented by a thin vertical line divided into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Population codes correspond to those indicated in Table S1. Major geographic areas are labeled below the graph.



Appendix IV: Supporting Information of Chapter 4

Table S1. Correlation values between the 19 bioclimatic and 3 geographic variables analyzed for environmental niche modeling of *Brachypodium sylvaticum*, *B. pinnatum*, and *B. phoenicoides*. Significant correlations at $p<0.05$ are indicated in bold. Upon request to authors.

Table S2. Environmental niche modeling analysis of the studied *Brachypodium* species. Significance of Tukey pairwise comparison tests (p values) of mean values for 19 bioclimatic and 3 geographic variables (see Table 1). Missing values correspond to non significant tests ($p>0.05$).

Variable Code	*B. pinnatum* – *B. sylvaticum*	*B. pinnatum* – *B. phoenicoides*	*B. sylvaticum* – *B. phoenicoides*
Bio1	**0.00002**	**0.00002**	**0.00002**
Bio2	**0.00002**		**0.00002**
Bio3	**0.006**	**0.00002**	**0.00002**
Bio4	**0.00002**	**0.00002**	**0.00002**
Bio5		**0.00002**	**0.00002**
Bio6	**0.00002**	**0.00002**	**0.00002**
Bio7	**0.00002**	**0.00002**	**0.04691**
Bio8	**0.00002**	**0.00025**	
Bio9	**0.00002**		**0.00002**
Bio10	**0.000257**	**0.00002**	**0.00002**
Bio11	**0.00002**	**0.00002**	**0.00002**
Bio12	**0.00002**		**0.01783**
Bio13	**0.00002**		**0.00503**
Bio14	**0.002780**	**0.00052**	**0.00002**
Bio15			
Bio16	**0.00002**		**0.00634**
Bio17	**0.002976**		**0.000829**
Bio18		**0.00002**	**0.00002**

Bio19	**0.00002**	**0.00002**	
Lat		**0.00002**	**0.00002**
Lon	**0.00002**	**0.00002**	**0.00002**
Alt	**0.004**		

Table S3. Correlation analysis between 19 bioclimatic variables plus latitude, longitude and altitude and the first tree axes of the environmental Principal Component Analysis (PCA) showing the percentages of variance explained by each variable to each axis. Values of variables that most contributed to each axis are highlighted in bold.

Variables	C**ode**	**Axis 1**	**Axis 2**	**Axis 3**
Annual Mean Temperature	Bio1	-0.68702	0.694079	0.001864
Mean Diurnal Range	Bio2	0.593693	0.308281	-0.5731
Isothermality	Bio3	-0.54174	0.450503	-0.51243
Temperature Seasonality	Bio4	**0.930853**	-0.16446	-0.02116
Max Temperature of Warmest Month	Bio5	0.108797	**0.85547**	-0.08155
Min Temperature of Coldest Month	Bio6	**-0.87986**	0.436449	0.119319
Temperature Annual Range	Bio7	**0.934421**	-0.05399	-0.1568
Mean Temperature of Wettest Quarter	Bio8	0.451017	0.288483	0.09595
Mean Temperature of Driest Quarter	Bio9	**-0.76761**	0.55214	-0.00133
Mean Temperature of Warmest Quarter	Bio10	-0.11734	**0.849496**	0.019739
Mean Temperature of Coldest Quarter	Bio11	**-0.8537**	0.499919	0.029178
Annual Precipitation	Bio12	**-0.71359**	-0.57939	-0.28875
Precipitation of Wettest Month	Bio13	-0.52201	-0.42897	-0.62326
Precipitation of Driest Month	Bio14	-0.62624	-0.64524	0.171125
Precipitation Seasonality	Bio15	0.37252	0.24469	-0.71891
Precipitation of Wettest Quarter	Bio16	-0.57321	-0.45089	-0.58324
Precipitation of Driest Quarter	Bio17	-0.65834	-0.62553	0.15025
Precipitation of Warmest Quarter	Bio18	-0.19238	-0.63248	-0.46523
Precipitation of Coldest Quarter	Bio19	**-0.78574**	-0.29721	-0.02377
Latitude	Lat	0.307205	-0.63379	0.626216
Longitude	Long	**0.725145**	-0.15223	-0.49787
Altitude	Alt	0.006048	-0.07657	**-0.73325**
	% Variance	**38.76**	**25.46**	**2.07**

Table S4. Environmental niche models of the studied *Brachypodium* species. AUC values of training (tr) and test (tst) data obtained from Maxent for ENMs of *B. pinnatum*, *B. sylvaticum* and *B. phoenicoides* under current conditions. Jackknife rating and permutation importance of variables that most contribute to the ENM of each species. N, sampling size.

	B. pinnatum N=658		*B. sylvaticum* N=907		*B. phoenicoides* N=91	
N tr / N tst	494/164		681/226		69/22	
AUCtr/AUCtst	0.911/0.904		0.897/0.884		0.986/0.886	
Standard deviation	0.007		0.009		0.002	
Logistic threshold	0.404		0.274		0.405	
Evaluation of the most significant contributing variables	Jackknife rating	Permutation importance	Jackknife rating	Permutation importance	Jackknife rating	Permutation importance
	Bio8	Bio18=27.8	Bio19	Bio9=43.6	Bio9	Bio7=80.9
	Bio18	Bio8=19.7	Bio9	Bio19=19.9	Bio7	Bio15=5.6
	Bio9	Bio19=19.2	Bio7	Bio8=14.2	Bio19	Bio2=5.3
	Bio15	Bio2=14.0	Bio2	Bio7=11.6	Bio18	Bio18=3.3

Table S5. Environmental niche models of the studied *Brachypodium* species showing the overlapping niche distributions between pairs of species for current climatic conditions.

B. sylvaticum – B. pinnatum	***B. pinnatum – B. phoenicoides***	***B. sylvaticum – B. phoenicoides***
B. sylvaticum 551002-cells	*B. pinnatum* 473180 cells	*B. sylvaticum* 551002 cells
B. pinnatum 473180 cells	*B. phoenicoides* 78030 cells	*B. phoenicoides* 78030 cells
Overlapping (I) 8278786 cells	Overlapping (I) 8278786 cells	Overlapping (I) 8278786 cells

Supplemental Figures

Figure S1. *Brachypodium sylvaticum*. Results of jackknife pseudoreplicate tests performed with Maxent indicating the variables that showed the highest training gain, test gain and AUC values for ENMs when used in isolation (blue), or decreased the gain when omitted alone (green), compared with values from all the variables (red), for current conditions.

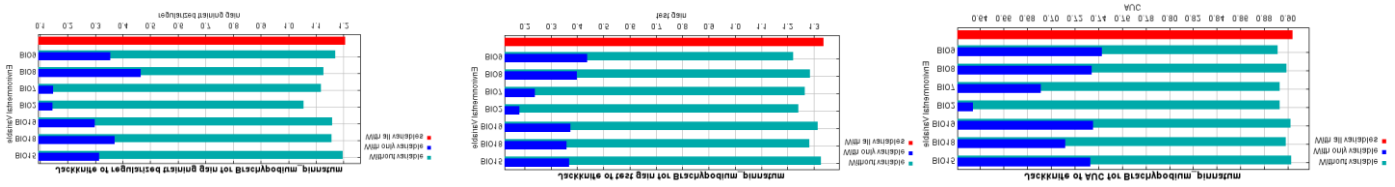


Figure S2. *Brachypodium pinnatum*. Results of jackknife pseudoreplicate tests performed with Maxent indicating the variables that showed the highest training gain, test gain and AUC values for ENMs when used in isolation (blue), or decreased the gain when omitted alone (green), compared with values from all the variables (red), for current conditions.

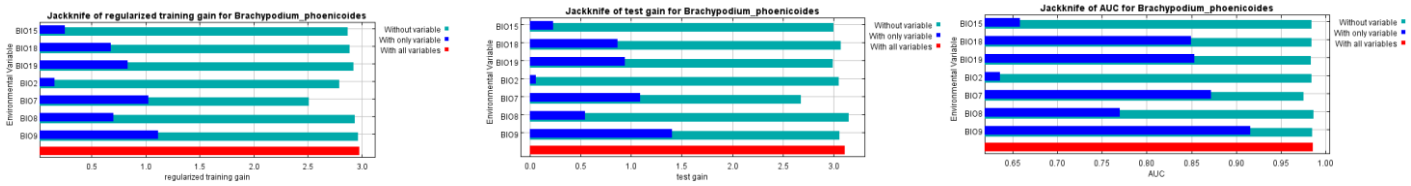


Figure S3. *Brachypodium phoenicoides*. Results of jackknife pseudoreplicate tests performed with Maxent indicating the variables that showed the highest training gain, test gain and AUC values for ENMs when used in isolation (blue), or decreased the gain when omitted alone (green), compared with values from all the variables (red), for current conditions.

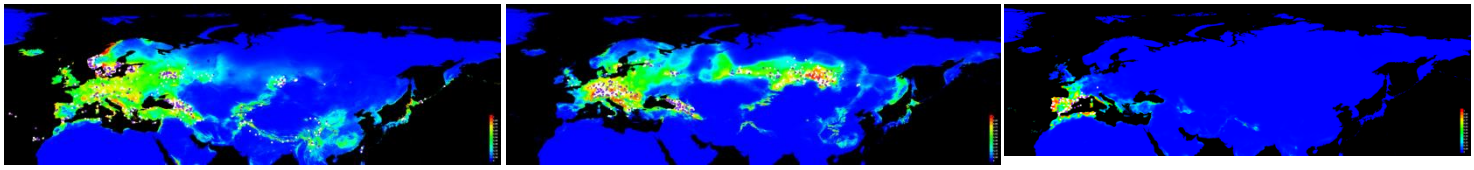


Figure S4. Current ENMs of *Brachypodium sylvaticum* (A), *B. pinnatum* (B) *and B. phoenicoides* (C). Dots correspond to occurrence data.

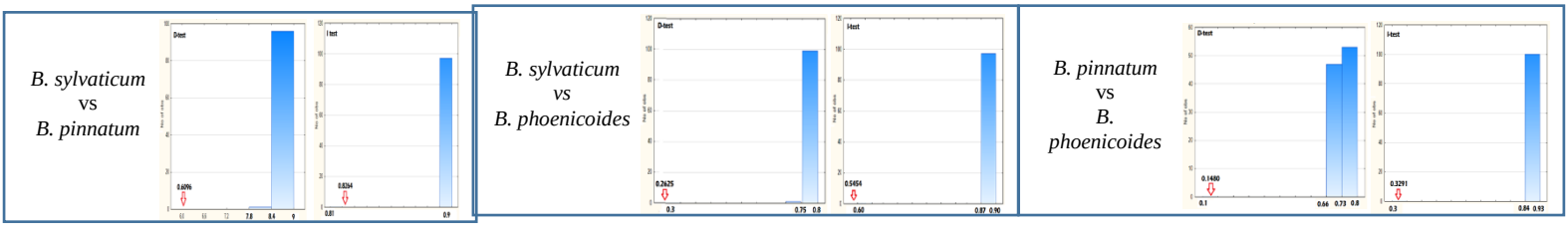


Figure S5. Results of the pairwise niche overlaps and pairwise niche identity tests performed with ENMtools for the environmental niches of *Brachypodium sylvaticum* vs *B. pinnatum* (A), *B. sylvaticum vs B. phoenicoide* (B), and *B. pinnatum vs B. phoenicoides* (C). Observed niche overlap ranges and Schoener's D overlap values, and randomly simulated niche overlap ranges and Hellinger'I niche identity test values are shown for each pairwise comparison (see Table 2).

Appendix V: Supporting Information of Chapter 5

Table S1. Morphoanatomical variables used in the phenotypic analysis and their respectives codes. All the variables complied with the Kolmogorov-Smirnov normality test at $p<0.5$.

Trait	Code
Culm length (cm)	CL
Number of nodes of culm	NNC
Leaf sheath length (cm)	LSL
Closing part of leaf sheath length (cm)	CLSL
Leaf blade length (cm)	LBL
Leaf blade length width (mm)	LBW
Inflorescence length (cm)	IL
Number of spikelets per inflorescence	NSI
Distance between 1st and 2nd spikelet (mm)	DS
Spikelet length, without awns (mm)	SL
Number of flowers per spikelet	NFS
Lower glume length (mm)	LGL
Upper glume length (mm)	UGL
Lemma length, from the basal floret (mm)	LL
Palea length (mm)	PL
Awn length (mm)	AL
Stomata guard cell length (µm)	SGCL

Table S2. Loadings of phenotypic traits onto the first PCA components of the interspecific multivariate analysis (*Brachypodium pinnatum*, *B. rupestre*, *B. phoenicoides*). Highest contributions of characters to each component are highlighted in bold.

Phenotypic trait	**C1**	**C2**	**C3**
CL	-0,422898	0,409973	0,302213
NNC	-0,161407	0,522167	0,288528
LSL	**-0,755738**	0,013981	0,166254
CLSL	-0,428394	0,570032	0,005579
LBL	-0,721793	0,200705	0,132298
LBW	-0,485690	0,447054	-0,069803
IL	**-0,798985**	-0,211387	0,215074
NSI	-0,489330	0,433306	0,137258
DS	-0,573409	-0,343061	0,289036
SL	-0,506146	**-0,609746**	0,212017
NFS	-0,365255	**-0,591156**	0,353493
LGL	-0,486160	0,110150	-0,565300
UGL	-0,375426	-0,398448	-0,409033
LL	-0,530064	-0,221890	**-0,633729**
PL	-0,394411	-0,070130	-0,558205
AL	-0,083863	0,536791	-0,246828
SGCL	-0,422898	0,409973	0,302213

PUBLICATIONS OF THE PhD THESIS

The studies conducted in the PhD thesis have produced two published articles, one article under review and two draft manuscripts.

1- Research paper: Shiposha V, Catalan P, Olonova M, Marques I. 2016. Genetic structure and diversity of the selfing model grass *Brachypodium stacei* (Poaceae) in Western Mediterranean: out of the Iberian Peninsula and into the islands. *PeerJ* 4:e2407; DOI 10.7717/peerj.2407.

2- Research paper: Marques I, Shiposha V, López-Álvarez D, Manzaneda A, Hernández P, Olonova M, Catalan P. 2017. Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*. *BMC Evolutionary Biology* 17: 139. DOI 10.1186/s12862-017-0996-x.

3- Research paper: Shiposha V, Marques I, López-Álvarez D, Manzaneda A, Hernández P, Olonova M, Catalan P. 2017. Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot. *Annals of Botany* (under review).

4- Research paper: Shiposha V, Feoktistov DS, Marques I, Catalan P, Olonova M. 2019. Glacial refugia and postglacial range shifts of the widespread Eurasian *Brachypodium sylvaticum* and *B. pinnatum* and the western Mediterranean *B. phoenicoides* grasses: insights from environmental niche modeling analysis and conservation status of relict Siberian populations of *B. sylvaticum* (draft manuscript).

5- Research paper: Shiposha V, Marques I, Olonova M, Catalan P. 2019. Morphoanatomical study of the *Brachypodium pinnatum* complex species (*B. pinnatum, B. rupestre, B. phoenicoides*) (Poaceae) and their diploid and polyploid cytotypes (draft manuscript)