Citação:

Valeriia Shiposha, Isabel Marques, Diana López-Alvarez, Antonio J Manzaneda, Pilar Hernandez, Marina Olonova, Pilar Catalán, Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot, *Annals of Botany*, Volume 125, Issue 4, 13 March 2020, Pages 625–638, https://doi.org/10.1093/aob/mcz169

DOI: https://doi.org/10.1093/aob/mcz169

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

**For correspondence. E-mail pcatalan@unizar.es*

§ Both authors contributed equally to this work

- **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*, suggesting another *B. stacei* source for the *B. hybridum* S alleles. 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.

Key words: 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). This 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 (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 allopolyploid species originated after hybridization of diploid or lower-ploidy progenitor species (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 gene flow 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 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) 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). In contrast to other model plants, the annual *B. distachyon* has a rich combination of desirable 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 0.631 pg/2C 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 the diploid *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 Mbp) are also available (Scholthof *et al.*, 2018). Despite having twice the number of chromosomes, the genome size of *B. stacei* is roughly similar to that of *B. distachyon* due to the small sizes of the *B. stacei* chromosomes compared to those of *B. distachyon*, whereas the genome size of *B. hybridum* corresponds to the sum of the two progenitor genomes (Catalán et al. 2012). 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, have 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. We have followed two strategies to analyse both the genetic and phylogeographic patterns of the Iberian *B. hybridum* populations and the multiple origins of their individuals. For population genetics and phylogeography we have considered each individual as a single evolutionary and dispersal unit, and have thus analysed together its decoupled S and D subgenomic allelic loci. However, for the investigation of their multiple origins we have analysed separately their decoupled S and D phenotypes, together with those of Iberian populations of the *B. stacei* and *B. distachyon* progenitor species. 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 Supplementary data Table S1. Because *B. hybridum* can be morphologically confused with the parental species *B. distachyon* (and less frequently *B. stacei*), the identities of the samples were 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, deleting extraneous peaks and discarding low sizing quality peaks across all the samples. 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* (Supplementary data 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 (Supplementary data Table S3).

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 chose 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 relationships 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 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 322 individuals of *B. hybridum* (34 populations, current study), and single diploid genotypes from the 181 *B. stacei* and 148 *B. distachyon* individuals (Supplementary data 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 showing unexpected allelic patterns of unclear origin (Supplementary data 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 Bayesian 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. This unrooted tree was artificially rooted according to the

strongly supported *B. stacei* + *B. hybridum*S (*stacei*-like) and *B. distachyon* + *B. hybridum*D (*distachyon*-like) divergences observed in the outgroup-rooted plastid trnLF-ndhF tree of the *B. stacei-B. distachyon-B. hybridum* complex (see below) and in nuclear and plastid phylogenies of the three species of the complex (López-Alvarez *et al.*, 2012). 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 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, Supplementary data 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 (accessions xxxx to xxxx) (Supplementary data 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 (Supplementary data 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 although 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) (Supplementary data 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* (Supplementary data 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 (Supplementary data 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 Hardy-Weinberg Equilibrium (HWE), we found that only one locus showed a significant deviation after the Bonferroni correction (ALB006D: $P<0.05$; Supplementary data Table S5).

Within populations, observed heterozygosity varied between 0 and 0.214 in the NE population (Supplementary data 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 Supplementary data Table S1). The rate of self-fertilization estimated for *B. hybridum* varied considerably between populations in that some plants, but not all, were highly selfing (Supplementary data 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% (Supplementary data 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 with the 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 (Supplementary data Fig. 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 (Supplementary data 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 (Supplementary data Fig. 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 (Supplementary data Table S6). A hierarchical AMOVA performed between the four geographical regions found the highest percentage of variance among populations within groups (61.3%). The hierarchical AMOVA between the *K*=2 populations defined by STRUCTURE showed the highest partition of variance among populations within groups ($P<0.00001$, Supplementary data 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 (Supplementary data Table S3) showed evidence of close genetic relatedness between the *B. hybridum* S and D subgenomic phenotypes and some phenotypes from the sampled Iberian progenitor species. The bidimensional 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 partially 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 sampled Iberian *B. stacei* phenotypes, except for a few SE Iberia *B. hybridum* S phenotypes (Fig. 3A). A tridimensional PCO plot showed the predominant distribution of *B. stacei* plus *B. hybridum* S phenotypes and of *B. distachyon* plus *B. hybridum* D phenotypes in the respective right and left sides of the SSR space defined by axes C1, C2 and C3 (10.82% of variance) (Supplementary data Fig. S2).

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 ML SSR 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, NE and SW diverged after *B. stacei* (Fig. 4). By contrast, within the *B. distachyon*+D clade the *B. hybridum* D lineages from the NW diverged earlier than the *B. distachyon* NW lineages, whereas 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 the *B. distachyon* S lineages (Fig. 4). Surprisingly, both *B. hybridum* D and S phenotypes 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 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, Supplementary data Fig. S3). 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 (84% BS): 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 (99% BS): 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; Supplementary data Fig. S3). By contrast, only one clear type of *B. hybridum* S plastotypes could be identified within the S clade (100% BS), 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, Supplementary data Fig. S3).

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 evaluation 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 (López-Alvarez *et al*. 2012; Catalán *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 (Supplementary data 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). Taken altogether, this lends support to our 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* (Supplementary data 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), detected considerable genetic variation across the 36 searched populations although most populations are highly homogeneous and exhibit few distinct genotypes (Fig. 1, Supplementary data Table S1). The level of heterozygosis is very low 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 (Supplementary data Table S1). Mixed, but not genetically admixed, *B. hybridum* populations occur in southern Iberia, probably caused by multiple founder events or multiple introductions coupled with a low outcrossing rate due to selfing. These findings are corroborated by our multiple origins analysis of *B. hybridum* that shows a high genetic diversity of S and D phenotypic profiles in *B. hybridum* individuals from the SE and SW compared to the rest (Supplementary data Figs. 3A, B). It is also confirmed by our plastid tree where *B.*

hybridum individuals from these regions show any of the two (SW: S, D1) or three (SE: S, D1, D2) plastotypes compared to the single plastotype found in all individuals from the other regions (NE, NW: S) (Fig. 5). 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, Supplementary data 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 (Fig. 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 (Supplementary data Fig. S1B). The strong geographically structured genetic differentiation observed for *B. hybridum* in the Iberian Peninsula is comparable to 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 of long distance seed dispersal and the 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 the 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 lead 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 together with our previous findings 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 (Fig. 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 (Fig. 5), corroborating earlier findings that the western Mediterranean region has one of the highest accumulations of plant genetic variation in the northern hemisphere, throughout recent Cenozoic evolutionary time (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 analysis of the *B. hybridum* S and D lineages and progenitor *B. stacei* and *B. distachyon* SSR profiles suggests 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 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 corroborated by their respective separations in the PCoA and the Structure K=5 and K=7 (Fig. 3) plots. Our data suggest that at least two alternative hybridizations may have occurred regarding the contribution of the progenitor *B. stacei* genome. The first, more ancestral hybridization resulted in the SE Iberian S subgenomes, and a second more recent hybridization originated 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). This has been confirmed for *B. hybridum*, a successful colonizer of 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 scenario 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 *Aegilops* (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. Our large nuclear and plastid barcoding analysis of wild circumMediterranean *B. stacei*, *B. distachyon* and *B. hybridum* accessions demonstrate that most of the studied *B. hybridum* individuals were derived from the cross of a maternal *B. stacei* and a paternal *B. distachyon* parent, although some were the result of a reciprocal cross (López-Alvarez et al. 2012). Moreover, comparative phenotypic analyses of the same set of morphological traits of *B. hybridum* and its progenitor *B. distachyon* and *B. stacei* parents, spanning their respective native Mediterranean regions, reveals key phenotypic differences within each of the diploids, but not within the allotetraploid (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. Despite the geographic limitations of our study, our analysis has demonstrated the occurrence of at least three natural multiple origins of the allopolyploid *B. hybridum*. The three 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 germoplasm for future analysis of allopolyploid formation and genomic expression in the *B. hybridum* – *B. distachyon* – *B. stacei* grass model complex.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. **Table S1**: Sampled populations of *Brachypodium hybridum* sorted by geographical areas. **Table S2**: Genotypes of the 342 studied *B. hybridum* individuals from 36 Iberian populations encoded for 17 SSR loci. **Table S3**: SSR phenotypes of 973 individuals of *B. hybridum*, *B. stacei*, and *B. distachyon* encoded for 98 binary alleles. **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. **Table S5**: Characteristics of the SSRs markers used in the population genetic study of Iberian Peninsula *Brachypodium hybridum* populations. **Table S6:** Analysis of molecular variance for 36 populations of *Brachypodium hybridum*. **Figure S1**: Population structure of *Brachypodium hybridum* in the Iberian Peninsula at *K*=14, following the best assignments retrieved by STRUCTURE. **Figure S2:** Tridimensional Principal Coordinate Analysis (PCoA) plot of the *Brachypodium hybridum-B. stacei-B. distachyon* complex. **Figure S3**: Maximum Likelihood plastid phylogenetic tree of the *Brachypodium hybridum-B. stacei-B. distachyon* complex based on trnLF and ndhF DNA sequences.

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 Science and Higher Education fellowship and grant RFBR № 18-34-00901. 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:** 213-218.

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, Catalán 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, Catalán 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, Catalán P, Vogel JP. 2017. Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with phenotypic variation. *Nature Communications* **8:** 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* **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:** 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:** 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 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:** 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 AJ, Hernández P, Olonova M, Catalán 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, Catalán P, Contreras-Moreira B. 2018. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist* **218:** 1631-1644.

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

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

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.

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.

Figure legends:

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 and squares indicate population samples used in the SSR and the plastid DNA sequence analyses, respectively, as shown in the chart (see also Supplementary data Tables S1, S4). Six individuals analysed for SSR data were also used in the plastid analysis [ALFR, CALA, CIMB, LEPE (Bhyb30_1; Bhyb30_2), LLIG].

Figure 2A: Bidimensional Principal Coordinate analysis (PCoA) plot 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 hybridum* 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 Supplementary data Table S1.

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 Supplementary data Table S3. Major geographic areas are labeled below the graph.

Figure 4: 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 Supplementary data Table S3. p.p., pro parte.

Figure 5: Maximum Likelihood plastid phylogenetic tree of the *Brachypodium hybridum-B. stacei-B. distachyon* complex based on trnLF and ndhF DNA sequences. Phylogram showing the main S (*B. stacei*-type) and D (*B. distachyon*-type) *Brachypodium* clades and 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 Supplementary data Table S4 and their respective geographical suffixes to regions of origin: NE, NW, SE and SW (Iberian Peninsula NE, NW, SE and SW areas), AFG (Afghanistan), BAL (Balearic Islands), CAN (Canary Islands), COR (Corsica), FRA (France, Pyrenees), IRA (Iraq), MOR (Morocco), SLO (Slovenia), TUR (Turkey).

Figure 1

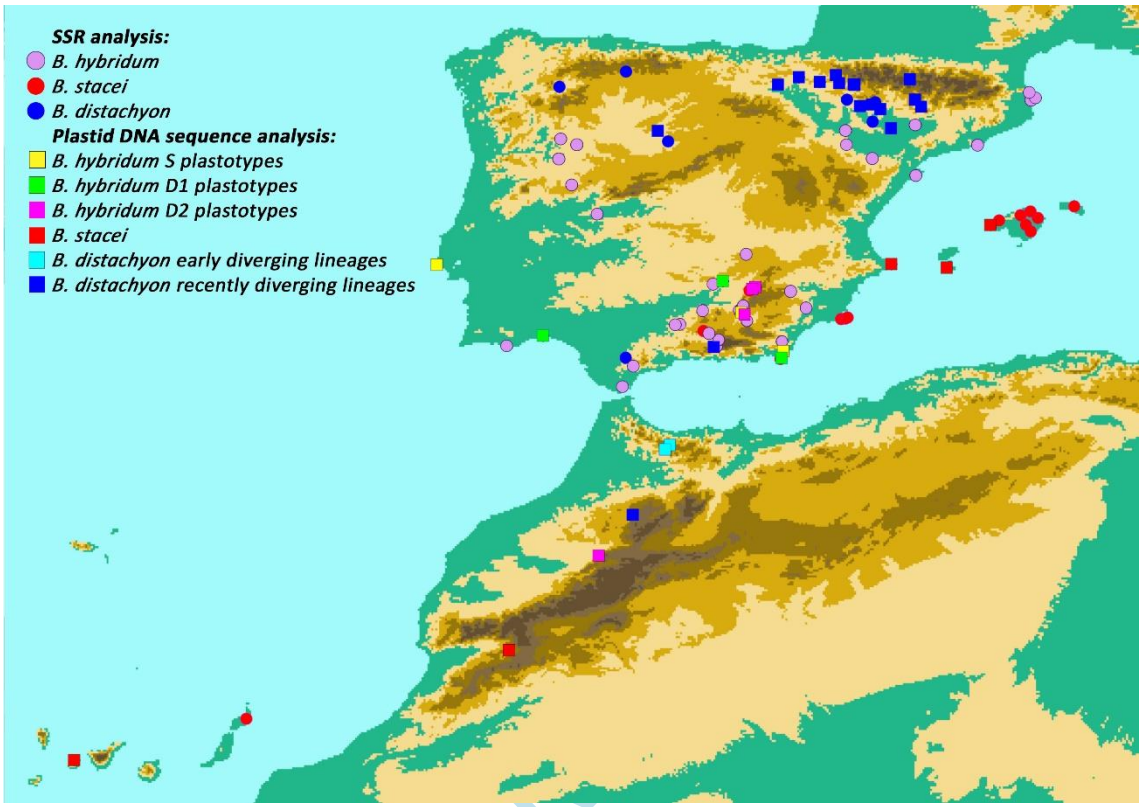

Figure 2A

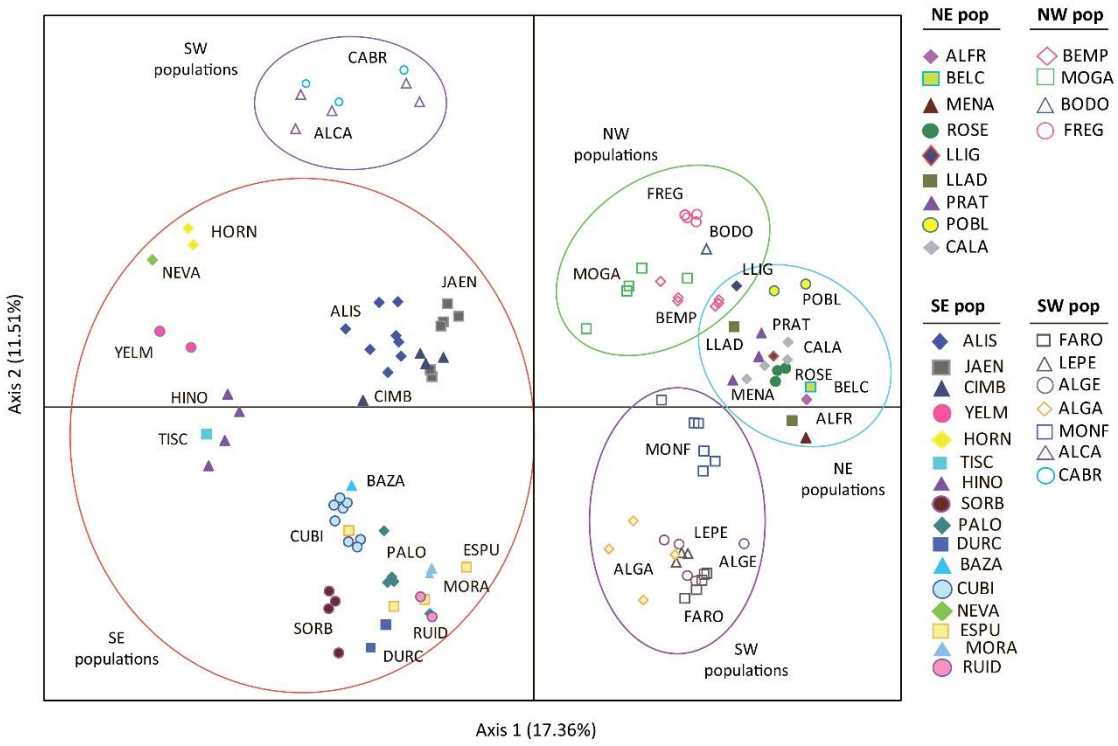

Figure 2B

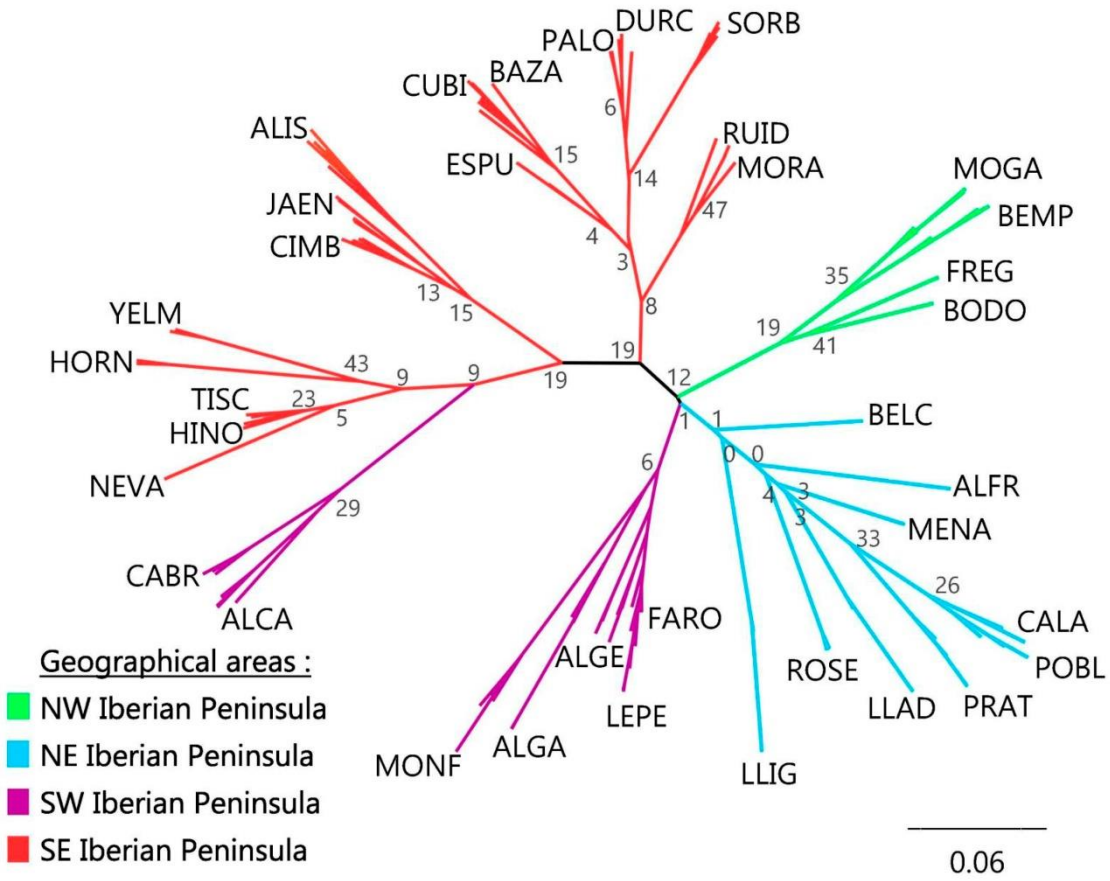

Figure 2C

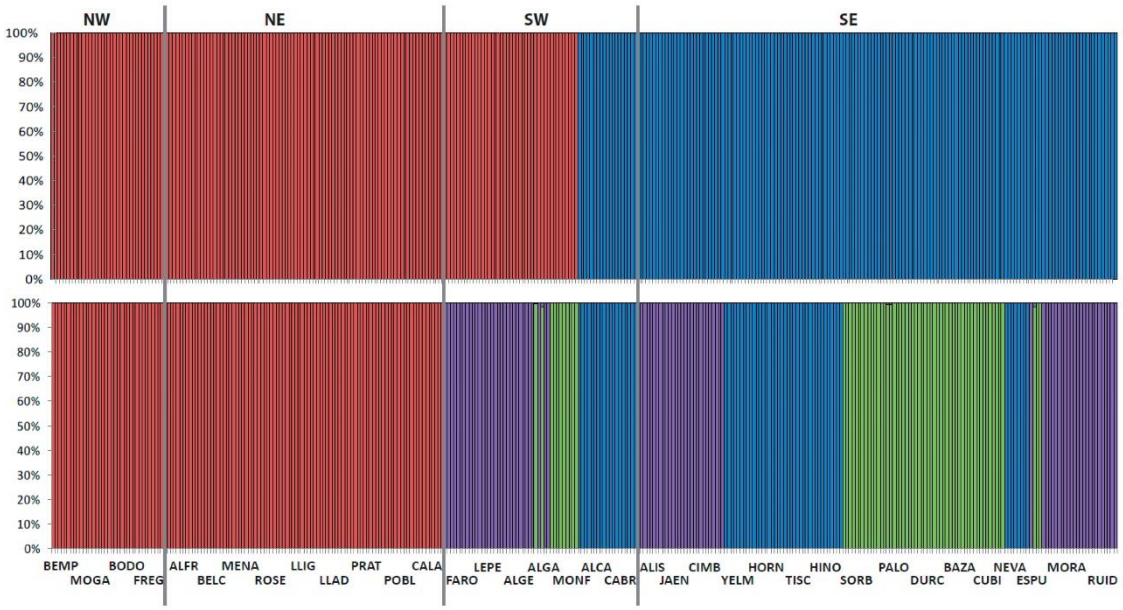

Figure 3A

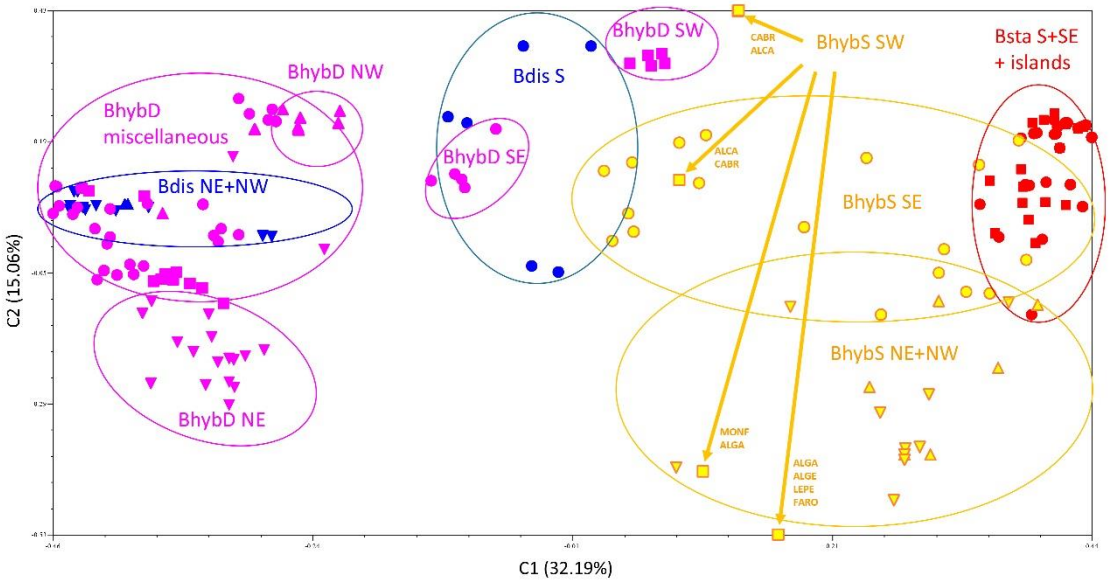

Figure 3B

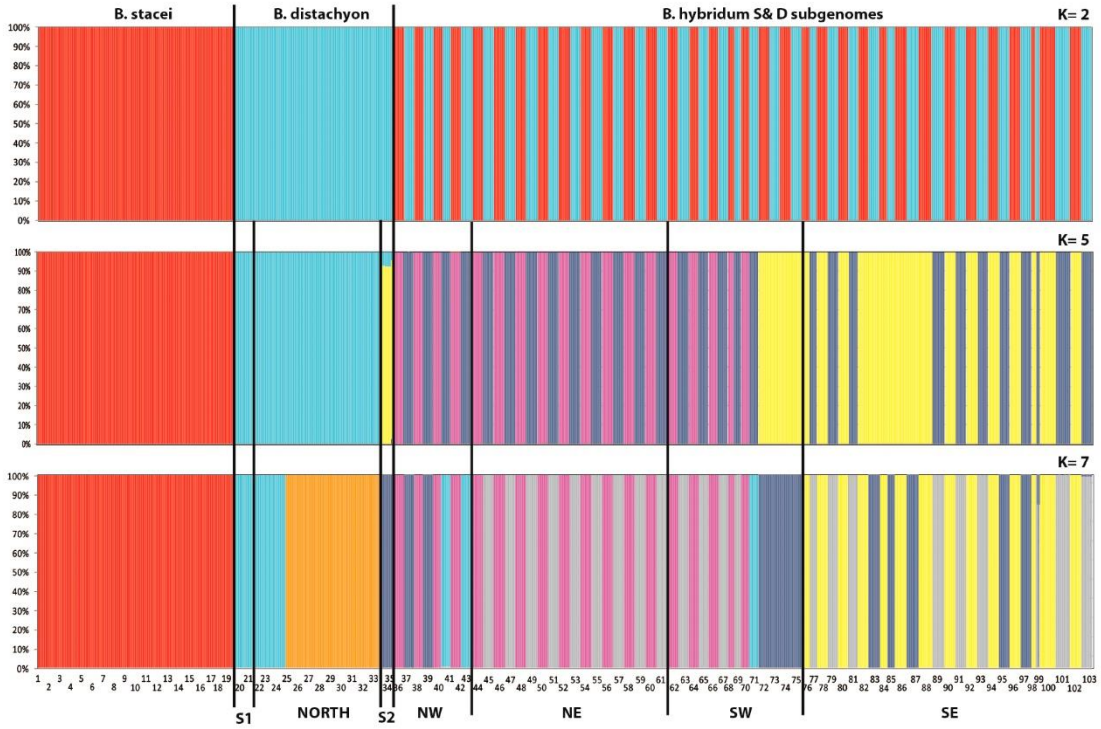

Figure 4

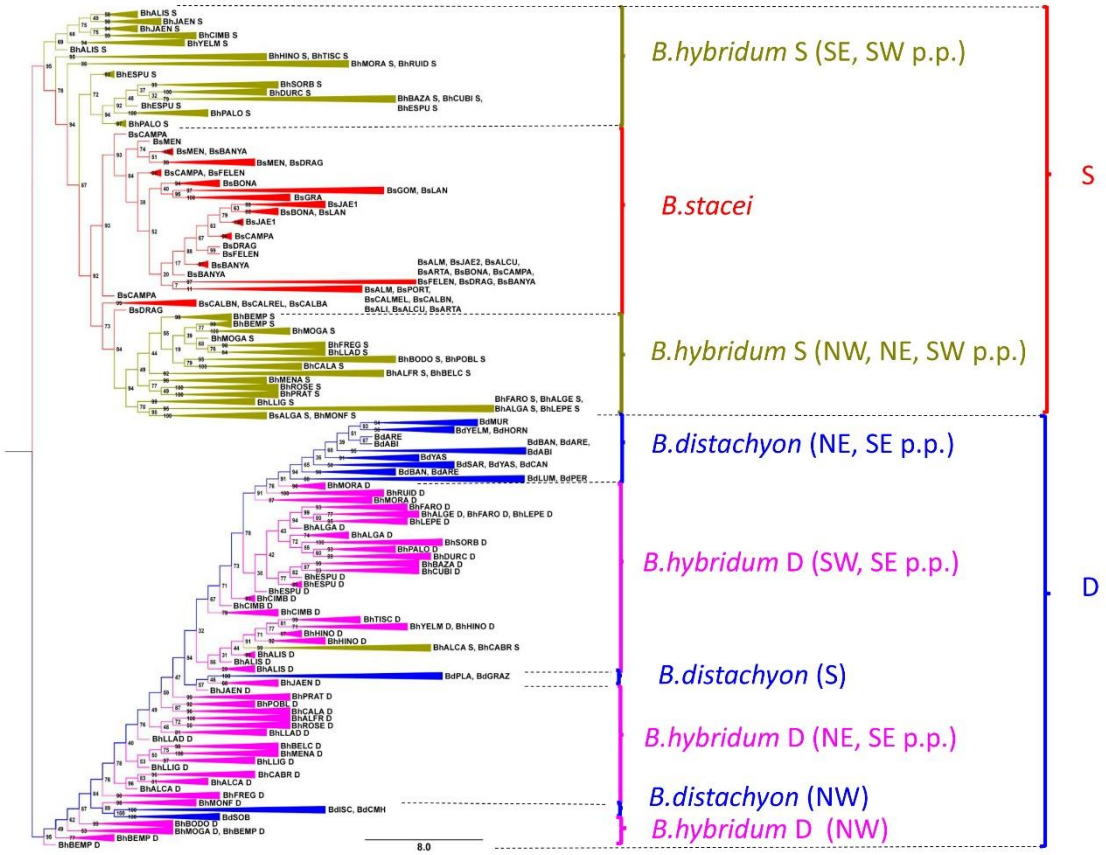

Figure 5