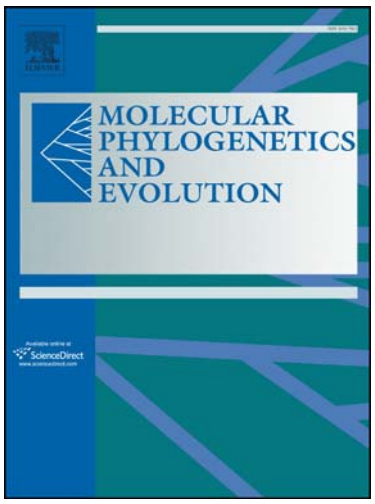


Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches

Antonio Díaz-Pérez, Diana López-Álvarez, Rubén Sancho, Pilar Catalán

PII: S1055-7903(18)30091-5
DOI: https://doi.org/10.1016/j.ympev.2018.06.003
Reference: YMPEV 6186

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 13 February 2018
Revised Date: 25 May 2018
Accepted Date: 2 June 2018

Please cite this article as: Díaz-Pérez, A., López-Álvarez, D., Sancho, R., Catalán, P., Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches, *Molecular Phylogenetics and Evolution* (2018), doi: https://doi.org/10.1016/j.ympev.2018.06.003

Regular article

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

Antonio Díaz-Pérez[a,1], Diana López-Álvarez[a], Rubén Sancho[a], Pilar Catalán[a,*]

[a] Escuela Politécnica Superior de Huesca, Universidad de Zaragoza, Carretera de Cuarte km 1. 22071 Huesca, Spain

[1]Present address: Agricultural Biotechnology Research Center (CIBA) and Institute of Genetics, School of Agronomy, Universidad Central de Venezuela

[*] Corresponding author: Pilar Catalán. Departamento de Ciencias Agrarias y del Medio Natural, Escuela Politecnica Superior-Huesca, Universidad de Zaragoza, Ctra. Cuarte km 1, 22071-Huesca, Spain. Email: pcatalan@unizar.es

Abstract

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

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

Keywords: divergence times (lineages) and coalescence ages (allopolyploids), genomic biogeography, grafted species tree and species network, homeologous genomes, model *Brachypodium* grasses, reticulate historical scenarios.

1. Introduction

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

Brachypodium has received considerable attention since the selection of the annual *B. distachyon* as model functional plant for temperate cereals and biofuel grasses (Mur et al., 2011; Vogel et al., 2010) and of its three annual species as a model group for allopolyploid speciation (Catalán et al., 2014; Gordon et al., 2016). This genus, characterized by its small-size and compact genomes (Betekhtin et al., 2014), is an ideal

model for comparative genomics of monocots (Kellogg, 2015b). *Brachypodium* belongs to the monotypic tribe Brachypodieae and contains between 18 and 20 taxa (Catalán et al., 2016) (Fig. 1). Dated phylogenies of plastid and nuclear rDNA genes support a rapid and relatively recent radiation of the genus since the Mid-Miocene, showing the early divergences of annual and short-rhizomatose lineages and the recent split of the strong-rhizomatose core perennial lineages (Catalán et al., 2012). Phylogenetic trees reconstructed from single-copy nuclear genes supported this hypothesis, but also showed homeologous copies in all of the polyploid lineages studied to date (Catalán et al., 2012, 2016; Wolny et al., 2011).

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

A preliminary evolutionary analysis of the *Brachypodium* taxa was performed in our previous work (Catalán et al., 2016). We grafted the polyploid alleles into a diploid species tree using a minimum evolution criterion aiming to draft a general scenario explaining the putative origins of the polyploid species. We observed four main placements of polyploid allelic copies in basal, *stacei*, *distachyon* and core perennial

clade branches, with some putative recent polyploids sharing also basal allelic copies. Nevertheless, statistical refinements are necessary to correct the excess of allelic copies grafted to different branches of the skeleton diploid species tree in order to properly infer the origins and the hybridization patterns of the homeologous genomes present in the allopolyploids.

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

Given these considerations, the objectives of our research were i) to incorporate statistical support for the allele grafting method to identify specific *Brachypodium* homoeologous genomes; ii) to reconstruct a robust explicit phylogenetic framework using a multigenic Species Network to disentangle the complex reticulate history of diploid and allopolyploid taxa, including all the identified genomic copies; iii) to build a dated chronogram for the multigenic allelic copies of *Brachypodium*; iv) to reconstruct

the historical biogeography of its genomes using parametric dispersion-extinction-cladogenesis models, inferring the paleo-scenarios for the dispersals and merging of genomes; and v) to estimate the coalescence ages of polyploid genomes from their closest diploid relatives, identifying and dating the hybridization events that gave rise to the allopolyploid species and cytotypes.

2. Materials and methods

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

2.1. Sampling, DNA sequence data processing and haplotype networks

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

DNA sequences from three nuclear [rDNA ETS and ITS, and a single-copy GIGANTEA (GI)] and two plastid (*ndh*F, *trn*LF) loci were used to reconstruct the phylogeny of *Brachypodium*. The protocols used for DNA isolation, amplification, cloning and sequencing are described in Appendix B.2. Five clones per sample were used for each nuclear locus in both diploid and polyploid taxa, aiming to detect all potential copies. A total of 973 *Brachypodium* sequences were aligned with sequences retrieved from GenBank (Table A.1 and Appendix B.2). The final data sets consisted of 431 sequences/682 aligned positions for ETS, 368/645 for ITS, 280/831 for GI, 95/564 for *ndh*F, and 100/941 for *trn*LF. The non-recombinant *ndh*F + *trn*LF plastid (cpDNA) sequences were concatenated into a combined 105/1505 data set. In order to discard spurious variation generated from PCR or cloning artifacts, intraspecific consensus (type) sequences were generated following Díaz-Pérez et al. (2014). Closely related sequences of the same species that showed a p-distance lower than 0.01 base differences per site were collapsed into a consensus type sequence using MEGA v. 5 (Tamura et al., 2011) and BIOEDIT v. 7.0.9.0 (Hall, 1999) (Table A.3 and Appendix B.2). The consensus types that were represented by a single clone were discarded. The haplotype networks were constructed using statistical parsimony (Clement et al., 2002) using POPART (Leigh and Bryant, 2015), with a 95% cut-off for the maximum number of mutational connections between pairs of sequences.

2.2. Diploid species tree reconstruction

A Bayesian diploid backbone species tree was constructed from consensus sequences (types) from each separate locus (Table A.3) with *BEAST v.2.1.3 (Bouckaert et al., 2014), using *Festuca pratensis* to root the tree. All parameters were unlinked across loci to allocate different evolutionary models in the species tree

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

2.3. Grafting polyploid alleles into the diploid species tree

A modified procedure of Cai et al. (2012) was used to graft individual alleles of polyploid species to specific branches of the diploid species tree using the Minimum Evolution criterion. In this analysis, all polyploid and skeleton diploid alleles (used to generate the species tree) per locus were analyzed to construct a gene tree. Different pruned gene trees were generated by pruning all polyploid alleles except one, per analysis. This excluded allele was treated as missing in the remaining gene trees of the other three loci, which were solely composed of skeleton diploid alleles. Several integrated distance matrices were constructed by averaging distances between diploid

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

2.4. Dating analysis

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

The selection of tree priors were based on Bayes factors (BF) where Marginal Likelihood Estimators (MLE) were generated according to the Path-Sampling (PS) and Stepping-Stone (SS) methods as implemented in BEAST. The Uncorrelated Relaxed Clock (UCLD)-Birth-Death model was chosen over the UCLD-YULE with a PS and SS MLE of -13211.5 vs -13236.9 and -13211.6 vs -13237.0, respectively, yielding a decisive BF of 22.5 with both estimators. The Strict Clock tree prior did not reach convergence so we could not estimate BF to test them against UCLD models. MLE are highly influenced by prior distributions, but we did not detect any mismatch between simulated and theoretical prior distributions for multiple calibrated internal nodes (see

below), as suggested by Heled and Drummond (2012). Moreover, "the ucld.sdev" estimate obtained from UCLD models was clearly different from zero, indicating variability of branch rates, giving an indirect support to UCLD over the Strict prior. Because there are no described fossils of *Brachypodium*, we dated the more inclusive data sets. For this, we calibrated the crown node of the BOP clade imposing a secondary calibration of 54.9 ± 5.7 Ma (normal prior distribution) according to the family-wide analysis of Bouchenak-Khelladi et al. (2010). A pooid epidermal phytolith fossil from the Middle Eocene (Strömberg, 2011) provided a minimum age for the crown node of Pooideae of 48.4 Ma [log-normal prior distribution mean=3.88, stdev=0.05, 95% highest posterior density (HPD) interval 44.6 to 52.58 Ma].

2.5. Divergence times of homeologous genomes and plausible ages of hybridization events

We assumed that a homeologous *Brachypodium* genome diverged from an ancestral diploid parental lineage, represented by the current diploid closest relative(s) identified in the Minimum Evolution tree. Pairwise divergence times were computed using an "Isolation-with-Migration" model according to the Bayesian method of Hey and Nielsen (2004) implemented in the program IM v.3.5. The bidirectional migration rates and population size parameters were enforced to be the same in all cases. These parameters were used to simplify the model and to maintain agreement with the recent radiation observed for the *Brachypodium* clade lineages (Catalán et al., 2016, 2012). Population parameters were scaled by μ (the neutral mutation rate), the effective number of gene copies (*Ne*), the migration rate (*M*) and the divergence time (*T*). These parameters were estimated from the model parameters $\theta = 4Ne\mu$, $m = M/\mu$ and $t = T\mu$. The estimated IM coalescent diverging times should not be confused with the estimated

*BEAST lineage diverging times; *BEAST estimates the relative divergence times of diploid genome lineages, whereas IM estimates the demographic divergence time of each homeologous genome from its diploid relative. Three simulations per pairwise divergence estimation between a homeologous genome and its counterpart diploid genome were performed with $2x10^6$ burn-in and $3x10^6$ iterations to check for convergence, in addition to ESS > 300. A total of 22x3=66 pairwise runs were performed (Table 1). Wide uniform priors were assigned in the first run to set appropriate limits for the priors of the two subsequent independent runs. There were a variable number of loci available for pairwise comparisons, ranging from one to four loci depending on the genome (Table 1). In this case, we suggest that most estimates should be taken as approximate values, despite the fact that convergence was achieved and the replicated runs generated similar values. Considering that homeologous genomes could never have originated before than their more recent genome donors, we equated the time of the putative hybridization event with the time of the origin of the most recent counterpart diploid genome.

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

2.6. Species Network reconstruction

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

2.7. Biogeographic reconstruction of *Brachypodium* genomes

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

We defined 10 operational areas (OAs) for reconstructing the biogeography of the *Brachypodium* genomes (Fig. 1; Table A.4). The OAs were selected according to the current distribution of taxa in their respective native ranges (excluding recent anthropogenic introductions of *B. hybridum* and *B. sylvaticum* in non-native ranges), but also reflected the geological history of the study area: A) western Mediterranean; B) eastern Mediterranean + SW Asia; C) western Eurasia (from the Atlantic to the Urals); D) eastern Eurasia (from the Urals to the Pacific and eastern Asia); E) Canary Islands; F) America (from Mexico to Peru-Bolivia); G) Africa (Tropical Africa and South Africa); H) Madagascar; I) Taiwan; and J) Malesia (including Papua-New Guinea). Given the relatively disjunct and isolated distribution of most current *Brachypodium* taxa, the DEC analyses were constrained to a maximum number of two areas at ancestral nodes, assuming that ancestors (and genomes) were not more widespread than their extant descendants.

Two alternative DEC models were used to infer the biogeographical events along the branches of the *Brachypodium* chronogram, an unconstrained model (M0), where dispersal rates between all biogeographic areas were constant through time, and a constrained stratified model (M1), where the topology was divided into three temporal windows, each with a specific matrix of dispersal rates set according to paleogeographic connectivity (Table A.4). Three time slices were defined: TSI, Mid-Miocene (Langhian) to Messinian (16.2-7.2 Ma); TSII, Messinian to Pleistocene (7.21-2.6 Ma); and TSIII, Quaternary (2.61-0 Ma). These time slices were used to reflect the foremost paleogeographic events of both hemispheres that could have affected the divergence of the current *Brachypodium* lineages.

3. Results

3.1. The *Brachypodium* species tree and inference of allopolyploid homeologous genome lineages

Single-locus haplotypic networks and phylogenetic trees of *Brachypodium* based on plastid, ITS, ETS and GI data were in agreement with in the earliest divergences of *B. stacei*, *B. mexicanum* and *B. distachyon* lineages, and of a more recent split of the core perennial group (Figs. 3A-D, C.1-C.4; Appendix B.4). The ETS and ITS data also detected the early divergence of the African *B. bolusii*/*B. flexum*, the Canarian *B. arbuscula* and the Mediterranean *B. retusum* lineages within the core perennials clade, and the clustering of endemic East Asia-Madagascar [*B. sylvaticum* (China)/*B. kawakamii*, *B. madagascariense*] and East Asia-New Guinea (*B. kawakamii*/*B. sylvaticum* var. *pseudodistachyon*) haplotypes, respectively. The three nuclear genes (ETS, ITS, and GI) identified co-inherited *B. stacei*-type and *B. distachyon*-type parental copies in *B. hybridum*, and a number of co-inherited ancestral and recently

evolved homeologous copies among the perennial allopolyploid species (Figs. 3B-D, C.2-C.4 and Appendix B.4).

Our diploid tree, which included only *Brachypodium* species of confirmed diploid nature (Fig. 4), showed the earliest divergence for the annual *B. stacei* lineage, then the annual *B. distachyon* and lastly the clade of core perennial taxa, which successively split into the *B. arbuscula*, *B. genuense*, *B. sylvaticum*, *B. glaucovirens*, and *B. pinnatum* 2x (2n=18)/*B. rupestre* 2x (2n=18) lineages. The grafting of *Brachypodium* polyploid alleles, inferred from the minimum evolution approach along the branches of the species tree, suggested there were six homeologous genomes that could have participated in allopolyploidization events within *Brachypodium*, spanning several levels of phylogenetic depth (Figs. 4 and 5). We have named core genomes all recently evolved genomes falling within the core perennial clade, and out-core genomes those showing more ancestral divergences. We also traced the sources of one of the most ancestral out-core type genomes (*ANCESTRAL*), two more recently diverged out-core diploid genomes [*STACEI* (stacei-like)] and *DISTACHYON* (distachyon-like)], one ancestral core-type genome (*ARBUSCULA*), and two recently diverged core-type diploid genomes [*SYLVATICUM* (sylvaticum-genuense-like) and *PINNATUM* (pinnatum-rupestre-like)] (Figs. 4, 5). Both *SYLVATICUM* and *PINNATUM* were represented by polyploid alleles grafted to *B. sylvaticum* + *B. genuense* and *B. pinnatum* + *B.rupestre* terminal branches, respectively. However, we considered each of them as constituting a single genome, because they were grafted to both branches with similar though moderate-to-low bootstrap support. In addition, the *GLAUCOVIRENS* genome was represented by alleles grafted to *B. glaucovirens* + *B. sylvaticum* branches; although, in this case, strong bootstrap support was also observed for alleles grafted to the *B. glaucovirens* terminal branch.

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

3.2. Divergence times and biogeography of the *Brachypodium* lineages

The consensus maximum clade credibility chronogram indicated that the *Brachypodium* lineage branched off from its stem node (S) in the Late Eocene (38.8 Ma) and the split of the crown node (CR) occurred in the Mid-Miocene (12.6 Ma) (Fig. 6). Our analyses also showed successive Late-Miocene and Early-Pliocene divergences for the basalmost currently extant *Brachypodium* genome lineages (*B. stacei*, 6.8 Ma; *B. distachyon*, 5.1 Ma). This was followed by a rapid radiation of the core perennial

genome lineages from the end of the Pliocene (2.4 Ma) through the Quaternary, showing the sequential divergence of *B. arbuscula* (1.5 Ma), *B. genuense* (0.7 Ma), *B. sylvaticum* (0.6 Ma), *B. glaucovirens* (0.5 Ma), and *B. rupestre*/*B. pinnatum* lineages (0.3 Ma).

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

The stratified DEC model (M1) of *Brachypodium* showed a better fit for the data than the unconstrained (M0) model (-ln likelihood 196.7 *vs*. 206.3, respectively; Likelihood Ratio Test (LRT)=19.2, p =0.001), and we will refer to this model hereafter (Fig. 7). The global estimated dispersal rate (*dis*: 0.8314) was five times higher than the

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

The western and eastern Mediterranean ranges hosted the most complex hybridization and genome doubling processes, which generated the high ploidy level *Brachypodium* allopolyploids (*B. boissieri*, *B. retusum*) (Table A.1). The genomes of several recent lineages from western Eurasia (*SYLVATICUM*, *PINNATUM*) have

converged with the ancestral local core lineage (*ARBUSCULA*) in *B. retusum* or with local out-core western Mediterranean genomes (*DISTACHYON+ANCESTRAL*) in *B. boissieri* (Figs. 7, 8). Similar patterns of genomic colonization, but involving long distance transoceanic dispersal, mostly from eastern to western Mediterranean regions (N_{AR}, N_{SG}, N_{ρ}), but also from eastern Eurasia (N_{z}, N_{μ}) to Africa and Madagascar, could have contributed to the presumed allopolyploids *B. bolusii*, *B. flexum* and *B. madagascariense*. In Taiwan, the putative allopolyploid *B. kawakamii* likely resulted from the merging of colonizing genomes from eastern Eurasia (N_{γ}, N_{ζ}) and the western Mediterranean region (N_{φ}, N_{ς}) (Figs. 7, 8).

4. Discussion

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

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

allopolyploids (Brassac and Blattner, 2015; Brysting et al., 2007; Marcussen et al., 2015)— and accounting for plausible gene copy losses and lineage sorting events (Brassac and Blattner, 2015; Kamneva et al., 2017) that could confound the recovery of all homeologous genomes.

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

(10.1 Ma; Sancho et al., 2017). Evolutionary relationships have been corroborated for six poorly studied taxa (*B. bolusii*, *B. flexum*, *B. genuense*, *B. kawakamii*, *B. madagascariense*, *B. sylvaticum* var. *pseudodistachyon*), all falling within the core perennial clade (Figs. 3A-D, 4, 5, Figs. C.1-C.4). Approximately half of the species in the genus are diploids (8) and most of the remaining taxa (10) are likely allopolyploids (Figs. 3A-D, 4, 5; Figs. C.1-C.4), as determined for other model grasses, such as *Oryza* (Zhou et al., 2015).

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

potential origins (*ANCESTRAL*, *STACEI*) for the alleles of *B. mexicanum* (Figs. 4, 5). This connection to the *STACEI* genome could explain the shared biological, morphological and genomic features of *B. mexicanum* and *B. stacei* (Catalán et al., 2016).

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

Our study has revealed the evolutionary origins of *B. bolusii*, *B. flexum*, *B. kawakamii* and *B. madagascariense* (Figs. 4, 5). These lineages show homeologous *ARBUSCULA* allelic copies grafted to the core perennial clade, indicating a putative

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

doubling, or with x=7, a chromosome base number not confirmed yet in the cytogenetically studied diploid *Brachypodium* species (Betekhtin et al., 2014; Wolny and Hasterok, 2009; Catalán et al., 2016), but indicated by some authors (Robertson, 1981).

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

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

Our study, using the species and cytotypes of the grass genus *Brachypodium* as models, represents the first attempt to reconstruct the biogeography of ancestral genomes inherited by current diploid and allopolyploid taxa. The proposed biogeographical scenarios for the *Brachypodium* genomes and taxa fit the conceptual requirements for appropriate ancestral range reconstruction, and show i) that the splits of the allopolyploids' homeologous (sub)genomes from those of their diploid counterparts occurred in the same ancestral areas, although they could have dispersed

independently (Fig. 7), and ii) that following the genome mergings, the homeologous genomes participating in the new allopolyploids had the same biogeographical patterns (Figs. 7, 8). The inferred existence of parallel evolution of homeologous genomes within the allopolyploid *Brachypodium* species might have artificially increased the global rate of dispersion estimated by LAGRANGE (*dis*: 0.8314). This is predicated on our approach that considered a dispersal event of an allopolyploid as two or three independent events, each related to a single subgenome. We contend that this was not important in *Brachypodium* because all homeologous genomes of *B. hybridum*, *B. boissieri*, *B. bolusii*, *B. retusum*, *B. mexicanum* and *B. phoenicoides* originated in the same geographic location (Table 1, Fig. 8), thus precluding these species acting as genetic sources for additional dispersions. For the remaining allopolyploids (*B. madagascariense*, *B. flexum* and *B. kawakamii*), some dispersion events were observed (Table 1, Fig. 8), but they were limited to a single genome at a time.

Our DEC M1 model has provided a biogeographical scenario for the *Brachypodium* genomes and taxa that supports the origin of their MRCA in the Holarctic region, followed by successive dispersals to Northern and Southern Hemisphere ranges from the Miocene to the present (Figs. 7, 8). This parallels similar cases with other temperate grasses and angiosperms (e. g., Cardueae, Barres et al., 2013; *Hordeum*, Blattner, 2006; Loliinae, Minaya et al., 2017). Of 32 total inferred dispersals, 25 occurred in the Quaternary (TSIII), 5 in the the Pliocene (TSII) and two in the Miocene (TSI), (Fig. 7). This indicates that most *Brachypodium* genomes and species, especially those of the core perennial clade, emerged very recently. The western Mediterranean and American ranges were reconstructed as the ancestral areas with the highest marginal probabilities for the MRCA of *Brachypodium* (CR, 12.6 Ma). In the Mid-Miocene the areas were probably connected through Asia and the Bering Land

Bridge, favoring the migrations of these and other xerophytic ancestors (Sanmartin et al., 2001). A Mid-Miocene vicariance (CR; A/F), coincident with a major temperature drop in the global climate (Meijer and Krijgsman, 2005), would explain the distribution of an isolated W Mediterranean genome (N_a), later inherited by the local polyploid *B. boissieri* and by the American *B. mexicanum* (Figs. 7, 8). Several connections between America and Asia through Beringia enabled genomic exchanges between the two areas (e. g. *Rosa*, Fougere-Danezan et al., 2015). A Mid-Late Miocene range expansion from America to Asia (N_e, 9.1 Ma; DF), followed by peripheral isolations, probably originated the *ANCESTRAL* genome of *B. mexicanum*, whereas a Late Miocene American/Asian vicariance (N_g, 5.4 Ma; F/D), followed by dispersal of the Old World lineage to the Mediterranean region in the Pliocene (N_j, 3.0 Ma; AB), likely separated the *STACEI* genome of *B. mexicanum* from that of *B. stacei* (Figs. 7, 8, Table 1).

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

in the region at different Pleistocene and Holocene times (Catalán et al., 2012) that could have facilitated the recurrent origin of the species (López-Álvarez et al., 2015).

Multiple colonizations of Eurasia and other continents by ancestral perennial *Brachypodium* genomes (x=9, 8) were inferred to have occurred profusely in the Pliocene-Pleistocene (Fig. 7). These genomes merged with more ancestral annual-type genomes (x=10, 5), giving rise to a dysploid series of strongly-rhizomatose core perennial allopolyploid taxa (Fig. 8) (Betekhtin et al., 2014; Catalán et al., 2016). In addition, a Late Miocene-Pliocene range expansion from the eastern Mediterranean region to Africa would explain the widespread distribution of ancestral genomes of the core perennial clade (N_{DS}-N_{AR}, 5.1-2.4 Ma; BG). This migration likely occurred through the southwest Asian and Arabian platform corridor, a main migratory pathway of temperate Holarctic elements into East Africa and South Africa (Gehrke and Linder, 2009). Subsequent peripheral isolations and colonization of Asia, Madagascar and Taiwan, concomitantly with the Quaternary climatic oscillations (Hewitt, 2000) and the recent uplifts of the high African and Central and East Asian mountains, were inferred to explain the origins of the oldest core-type *ARBUSCULA* genome. This genome was inherited from a putative polyploid African (*B. bolusii*, *B. flexum*), Malagasy (*B. madagascariense*) and Taiwanese (*B. kawakamii*) species (Figs. 7, 8). A Mid-Quaternary LDD of a perennial genome from the eastern Mediterranean region to Macaronesia (Canary Islands), followed by vicariance (N_{γ}-N_{α}, 1.47-0.14 Ma), would explain the origin of the Canarian endemic *B. arbuscula*, following the emergence of these volcanic islands. New range expansions from the E Mediterranean region to Africa, and separate migrations from Africa to Asia (N_{SG}-N_{γ}, 1.17-0.92 Ma; DG) and from the Mediterranean region to Europe (N_{SG}-N_{θ}, 1.17-0.92 Ma; BC), were inferred to have caused the disjunct distributions of the ancestral genomes of the East and West

Palaearctic perennial lineages (Figs. 7, 8). In the East, Late Quaternary LDDs of genomes from Africa to Madagascar (N_{ε}-N_{η}, 0.74-0.23 Ma), and from Asia to Taiwan (N_{ε}-N_{ζ}, 0.74-0.48 Ma), over the respective straits, would explain the origins of newly recruited genomes, inherited by the local polyploids. The diploid *B. sylvaticum* var. *pseudodistachyon* could have originated following transoceanic colonization of an African genome in Malesia (N_{γ}-N_{δ}, 0.92-0.21 Ma), possibly facilitated by the mountain chains in New Guinea (Heads, 2006) (Figs. 7, 8). In the West, Upper Pleistocene range expansion from Europe to the Mediterranean region (N_{θ}-N_{β}, 0.92-0.73 Ma AC), and their respective Ionian-Holocene dispersals to Asia, were inferred to have been the origin of the most recent genomes of Mediterranean diploids *B. genuense* and *B. glaucovirens* and local polyploids, and of Eurosiberian *B. sylvaticum*, *B. rupestre* and *B. pinnatum* diploids. Some of the recent *SYLVATICUM* and *PINNATUM* genomes were also inferred to have migrated to Africa, Madagascar and Taiwan, contributing to the genomic dosage of the local polyploids (Figs. 7, 8). The current widespread Palearctic distribution of *B. sylvaticum* and *B. pinnatum* (Figs. 1, 8) probably resulted from recent Holocene postglacial colonizations from different Eurasian refugia, as indicated for other temperate grass lineages (Inda et al., 2014).

Our reconstruction is the first implementation of a spatio-temporal scenario for the successive splitting and merging of genomes in different ancestral areas that is in agreement with the formation and distribution of the extant diploid and polyploid *Brachypodium* taxa (Figs. 7, 8). Additionally, our DEC analysis infers that long distance dispersal is associated only with diploid genomes, since all of the known hybridizations and genome doublings have occurred within ancestral ranges, without further expansion to other areas (Table 1; Figs. 7, 8). These results are consistent with previous findings showing that allopolyploids evolved mostly *in situ* and did not disperse or dispersed

only to geographically close areas (e. g., *Hordeum*; Brassac and Blattner, 2015). In contrast, some authors have concluded that polyploids were more successful than diploids for long distance dispersal, although they used only a single terminal lineage per polyploid taxon for their analysis (Linder & Barker 2014). Our genome-wide *Brachypodium* biogeographical scenario supports more frequent past dispersals of diploid (and counterpart homeologous) genomes than allopolyploid lineages (Figs. 7, 8). This suggests that the dispersal capabilities of diploid and polyploid lineages should be revisited within a dissected genomic scenario.

6. Acknowledgements

We thank Karen Beth G. Scholthof for her helpful comments on a revised version of this manuscript, and Clive Stace, Maria Luisa López-Herranz and Emily Lemonds for taxonomic, laboratory and linguistic assistance, respectively, several National Parks for permission for field collections, and several colleagues, herbaria and germplasm banks for providing us with some *Brachypodium* accessions and DNA samples (see Appendix B.5).

7. Funding

The work was supported by two consecutive Spanish Ministry of Economy and Competitiveness research project grants CGL2012-39953-C02-01 and CGL2016-79790-P. DL and RS were both funded by Spanish Ministry of Science and Innovation PhD FPI fellowships. PC, ADP, DL and RS were partially funded by a Bioflora research team grant co-funded by the Spanish Aragón Government and the European Social Fund.

8. References

Barres, L., Sanmartín, I., Anderson, C.L., Susanna, A., Buerki, S., Galbany-Casals, M., Vilatersana, R., 2013. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). Am. J. Bot. 100, 867–882.

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

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

Bouchenak-Khelladi, Y., Verboom, G.A., Savolainen, V., Hodkinson, T.R., 2010. Biogeography of the grasses (Poaceae): A phylogenetic approach to reveal evolutionary history in geographical space and geological time. Bot. J. Linn. Soc. 162, 543–557.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Brassac J., Blattner F.R., 2015. Species-level phylogeny and polyploid relationships in *Hordeum* (Poaceae) inferred by next-generation sequencing and in silico cloning of multiple nuclear loci. Syst. Biol. 64, 792–808.

Brysting, A.K., Oxelman, B., Huber, K.T., Moulton, V., Brochmann, C., 2007. Untangling complex histories of genome mergings in high polyploids. Syst. Biol. 56, 467–476.

Cai, D., Rodríguez, F., Teng, Y., Ané, C., Bonierbale, M., Mueller, L.A., Spooner, D.M., 2012. Single copy nuclear gene analysis of polyploidy in wild potatoes

(*Solanum* section *Petota*). BMC Evol. Biol. 12, 70.

Carlsen, T., Bleeker, W., Hurka, H., Elven, R., Brochmann, C., 2009. Biogeography and phylogeny of *Cardamine* (Brassicaceae). Ann. Mo. Bot Gard. 96, 215–236.

Catalán, P., Chalhoub, B., Chochois, V., Garvin, D.F., Hasterok, R., Manzaneda, A.J., Mur, L.A.J., Pecchioni, N., Rasmussen, S.K., Vogel, J.P., Voxeur, A., 2014. Update on the genomics and basic biology of *Brachypodium*. International Brachypodium Initiative (IBI). Trends Plant Sci. 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*. In: Vogel, J.P. (Ed.), Genetics and Genomics of *Brachypodium*. Plant Genetics and Genomics: Crops Models. Switzerland, Springer. pp. 9–38.

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

Clement, M., Snell, Q., Walker, P., Posada, D., Crandall, K., 2002. TCS: estimating gene genealogies. Proceedings of the 16th International Parallel Distributed Processing Symposium, p. 184.

Desper, R., Gascuel, O., 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. J. Comput. Biol. 9, 687–705.

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

Estep, M.C., Mckain, M.R., Vela Díaz, D., Zhong, J., Hodge, J.G., Hodkinson, T.R., Layton, D.J., Malcomber, S.T., Pasquet, R., Kellogg, E.A., 2014. Allopolyploidy,

diversification, and the Miocene grassland expansion. Proc. Natl. Acad. Sci. U.S.A. 111, 15149–15154.

Fougère-Danezan, M., Joly, S., Bruneau, A., Gao, X.F., Zhang, L.B., 2015. Phylogeny and biogeography of wild roses with specific attention to polyploids. Ann. Bot. 115, 275–291.

Gehrke, B., Linder, H.P., 2009. The scramble for Africa: pan-temperate elements on the African high mountains. Proc. Biol. Sci. 276, 2657–2665.

Gordon, S.P., Liu, L., Vogel, J., 2016. The genus *Brachypodium* as a model for perenniality and polyploidy. In: Vogel, J. (Ed.). Genetics and Genomics of Brachypodium. Plant Genetics and Genomics: Crops Models. Switzerland, Springer. pp. 313–326.

Gruenstaeudl, M., Carstens, B.C., Santos-Guerra, A., Jansen, R.K., 2017. Statistical hybrid detection and the inference of ancestral distribution areas in *Tolpis* (Asteraceae). Biol. J. Linn. Soc. Lond. 121, 133–149.

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

Hall, T., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.

Heads, M., 2006. Biogeography, ecology and tectonics in New Guinea. J. Biogeogr. 33, 957–958.

Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogeneticsand divergence time estimation. Syst. Biol. 61, 138–149.

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

Hey J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747–760.

Huber, K.T., Oxelman, B., Lott, M., Moulton, V., 2006. Reconstructing the evolutionary history of polyploids from multilabeled trees. Mol. Biol. Evol. 23, 1784–1791.

Huson, D.H., Scornavacca, C., 2012. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. Syst. Biol. 61, 1061–1067.

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

Jones, G., Sagitov, S., Oxelman, B., 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. Syst. Biol. 62, 467–478.

Kamneva, O.K., Syring, J., Liston, A., Rosenberg, N.A., 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. BMC Evol. Biol. 17, 180.

Kellogg, E.A., 2015a. The Families and Genera of Vascular Plants. Volume XIII. Flowering Plants. Monocots. Poaceae. Springer, New York. Pp. 1–408.

Kellogg, E.A., 2015b. *Brachypodium distachyon* as a genetic model system. Annu. Rev. Genet. 49, 1–20.

Krijgsman, W., 2002. The Mediterranean: Mare Nostrum of Earth sciences. Earth Planet. Sci. Lett. 205, 1–12.

Leigh J., Bryant, D., 2015. POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116.

Linder, H.P., Barker, N.P., 2014. Does polyploidy facilitate long-distance dispersal? Ann. Bot. 113, 1175–1183.

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

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

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). Syst. Biol. 64, 84–101.

Meijer, P.T., Krijgsman, W., 2005. A quantitative analysis of the desiccation and re-filling of the Mediterranean during the Messinian Salinity Crisis. Earth Planet. Sci. Lett. 240, 510–520.

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

Meulenkamp, J.E., Sissingh, W., 2003. Tertiary palaeogeography and tectonostratigraphic evolution of the Northern and Southern Peri-Tethys platforms and the intermediate domains of the African-Eurasian convergent plate boundary zone. Palaeogeogr. Palaeoclimatol. Palaeoecol. 196, 209–228.

Meyer, S., von Haeseler, A., 2003. Identifying site-specific substitution rates. Mol. Biol. Evol. 20, 182–189.

Minaya, M., Hackel, J., Namaganda, M., Brochmann, C., Vorontsova, M.S., Besnard, G., Catalán, P., 2017. Contrasting dispersal histories of broad- and fine-leaved temperate Loliinae grasses: range expansion, founder events, and the roles of distance and barriers. J. Biogeogr. 44, 1980–1993.

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

Nieto-Feliner G., Rosselló J.A., 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. Mol. Phylogenet. Evol. 44, 911–919.

Paradis, E., Claude, J., & Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/

Ree, R.H., Smith, S.A., 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst. Biol. 57, 4–4.

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

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. Plant Cell 20, 11–24.

Sancho, R., Cantalapiedra, C.P., López-Álvarez, D., Gordon, S.P., Vogel, J.P., Catalán,

P., Contreras-Moreira, B., 2018. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. New Phytol. 218, 1631–1644.

Sanmartin, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linnean Soc. 73, 345–390.

Schippmann, U., 1991. Revision der europäischen Arten der Gattung *Brachypodium* Palisot de Beauvois (Poaceae). Boissiera 45, 1–249.

Soltis, P.S., Soltis, D.E., 2016. Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. 30, 159–165.

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

Stamatakis, A., 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

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. Ann. Mo. Bot. Gard. 72, 824–832.

Strömberg, C.A.E., 2011. Evolution of grasses and grassland ecosystems. Annu. Rev. Earth Planet. Sci. 39, 517–544.

Swofford, D.L., 2003. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4b10.* Sinauer Associates, Sunderland, Massachusetts, USA.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Masatoshi, N., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood,

evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Vogel, J.P., Garvin, D.F., Mockler, T.C., *et al*., 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463, 763–768.

Wolny, E., Lesniewska, K., Hasterok, R., Langdon, T., 2011. Compact genomes and complex evolution in the genus *Brachypodium*. Chromosoma 120, 199–212.

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

Xiang, Q.P., Wei, R., Shao, Y.Z., Yang, Z.Y., Wang, X.Q., Zhang, X.C., 2015. Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. Mol. Phylogenet. Evol. 82, 1–14.

Zou, X.-H., Du, Y.-S., Tang, L., Xu X.-W., Doyle, J.J., Sang, T., Ge, S., 2015. Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). Sci. Rep. 5, 14876.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix A Supplementary Tables. Lists of taxa, number of sequences and haplotypes analysed and dispersal rate matrices used in the stratified biogeographical model.

Appendix B Expanded materials and methods and results.

Appendix C Supplementary Figures. MrBayes trees obtained from plastid and nuclear genes.

Figure 1. The worldwide geographic distribution of the 18 *Brachypodium* taxa and the boundaries of the 10 operational areas used in the biogeographic study [A) western Mediterranean; B) eastern Mediterranean + SW Asia; C) western Eurasia (from Atlantic to Urals); D) eastern Eurasia (from Urals to Pacific and eastern Asia); E) Canary Isles; F) America (from Mexico to Peru-Bolivia); G) Africa (Tropical Africa and South Africa); H) Madagascar; I) Taiwan; J) Malesia (including Papua-New Guinea)]. The species ranges colors and marks are indicated in the chart. This image was modified from Catalan et al. (2016) with permission of the authors and the publisher.

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

Figure 3. The statistical parsimony networks constructed with POPART for (A) the chloroplast (*ndh*F + *trn*LF), (B) the nuclear ITS, (C) the nuclear ETS, and (D) the nuclear GIGANTEA (GI) haplotypic data sets (Table A.2). The species colors are indicated in the charts. The size of the circles is correlated with the number of samples showing the haplotype.

Figure 4. Minimum Evolution grafting of single-locus polyploid alleles into the *BEAST diploid species tree. The polyploid alleles of each species are grafted (in color) along the branches, according to the bootstrap pseudoreplications. The thick, medium, and thin lines represent allele placement with >75, 51-75, and <51 bootstrap support, respectively The different colors differentiate the groups of alleles associated with

several homoeologous genomes (dark green, *SYLVATICUM*; light green, *PINNATUM*; purple, *ARBUSCULA*; dark blue, *DISTACHYON*; red, *STACEI*; brown, *ANCESTRAL*; and, light blue, *GLAUCOVIRENS*). The polyploid alleles grafted to the same branches are considered copies of the same homeologous genome. *Festuca pratensis* (Poeae) was used to root the tree. The color codes for the *Brachypodium* species are indicated in the chart.

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

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

Pooideae nodal ancestors, respectively (see text). The vertical red lines are used to separate the three time slices (TSI-TSIII) used in the LAGRANGE analysis (see Fig. 7). The time scale bars below each panel represent million years ago (Ma).

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

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

glaucovirens; hyb, *B. hybridum*; kaw, *B. kawakamii*; mad, *B. madagascariense*; mex, *B. mexicanum*; pho, *B. phoenicoides*; pin, *B. pinnatum*; pse, *B. sylvaticum* var. *pseudodistachyon*; ret, *B. retusum*; rup, *B. rupestre*; sta, *B. stacei*; and, syl, *B. sylvaticum*.

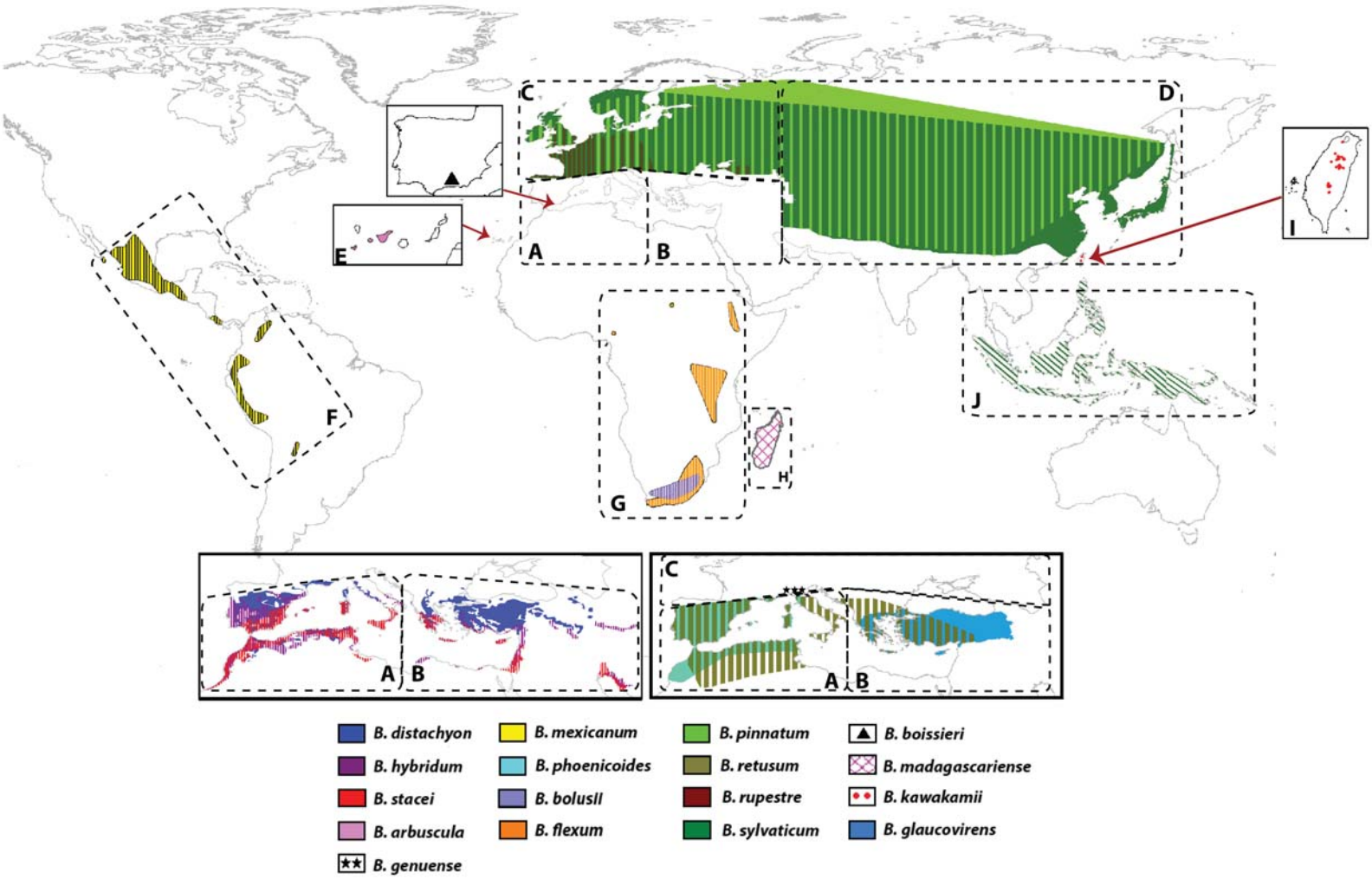

A
B
C
D
E
F
G
H
I
J
A
B
C
A
B
B. distachyon
B. hybridum
B. stacei
B. arbuscula
B. genuense
B. mexicanum
B. phoenicoides
B. bolusII
B. flexum
B. pinnatum
B. retusum
B. rupestre
B. sylvaticum
B. boissieri
B. madagascariense
B. kawakamii
B. glaucovirens

Fig.2

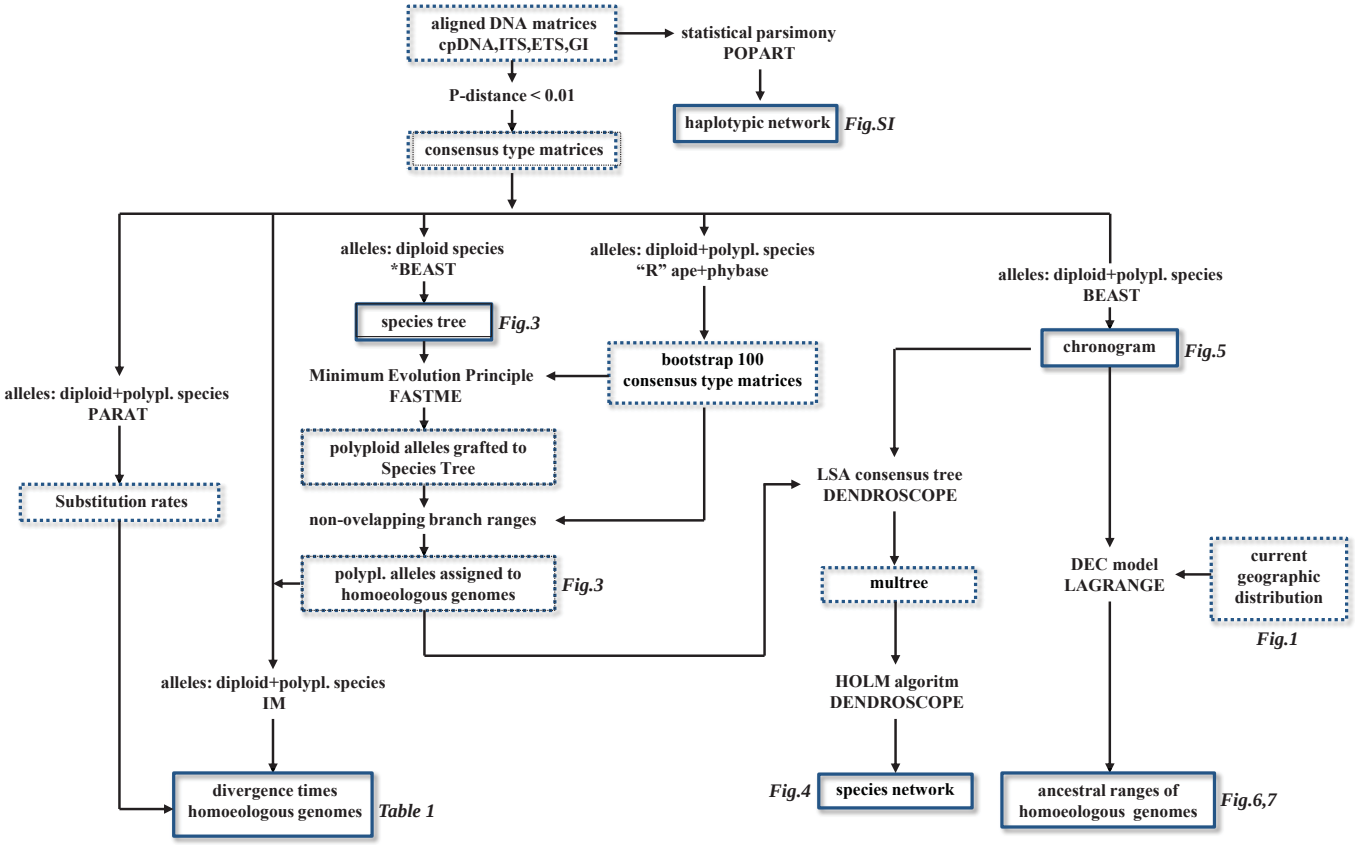

aligned DNA matrices cpDNA,ITS,ETS,GI
statistical parsimony POPART
P-distance < 0.01
haplotypic network
Fig.SI
consensus type matrices
alleles: diploid species *BEAST
alleles: diploid+polypl. species "R" ape+phybase
alleles: diploid+polypl. species BEAST
species tree
Fig.3
chronogram
Fig.5
alleles: diploid+polypl. species PARAT
Minimum Evolution Principle FASTME
bootstrap 100 consensus type matrices
polyploid alleles grafted to Species Tree
LSA consensus tree DENDROSCOPE
Substitution rates
non-ovelapping branch ranges
polypl. alleles assigned to homoeologous genomes
Fig.3
multree
DEC model LAGRANGE
current geographic distribution
Fig.1
alleles: diploid+polypl. species IM
HOLM algoritm DENDROSCOPE
divergence times homoeologous genomes
Table 1
Fig.4
species network
ancestral ranges of homoeologous genomes
Fig.6,7

Fig.3A

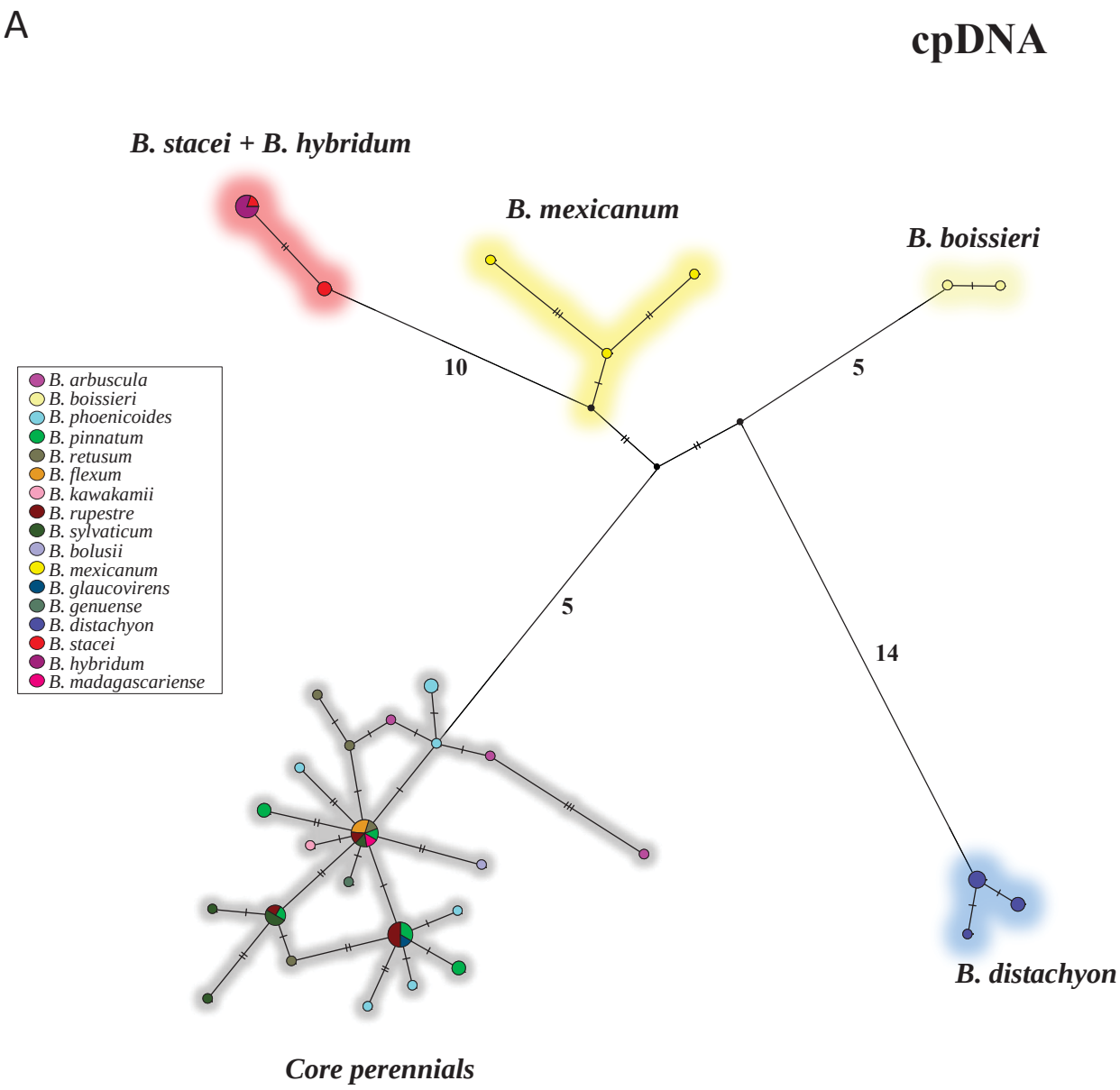

A
cpDNA
B. stacei + B. hybridum
B. mexicanum
B. boissieri
B. distachyon
Core perennials
10
5
5
14
B. arbuscula
B. boissieri
B. phoenicoides
B. pinnatum
B. retusum
B. flexum
B. kawakamii
B. rupestre
B. sylvaticum
B. bolusii
B. mexicanum
B. glaucovirens
B. genuense
B. distachyon
B. stacei
B. hybridum
B. madagascariense

Fig.3B

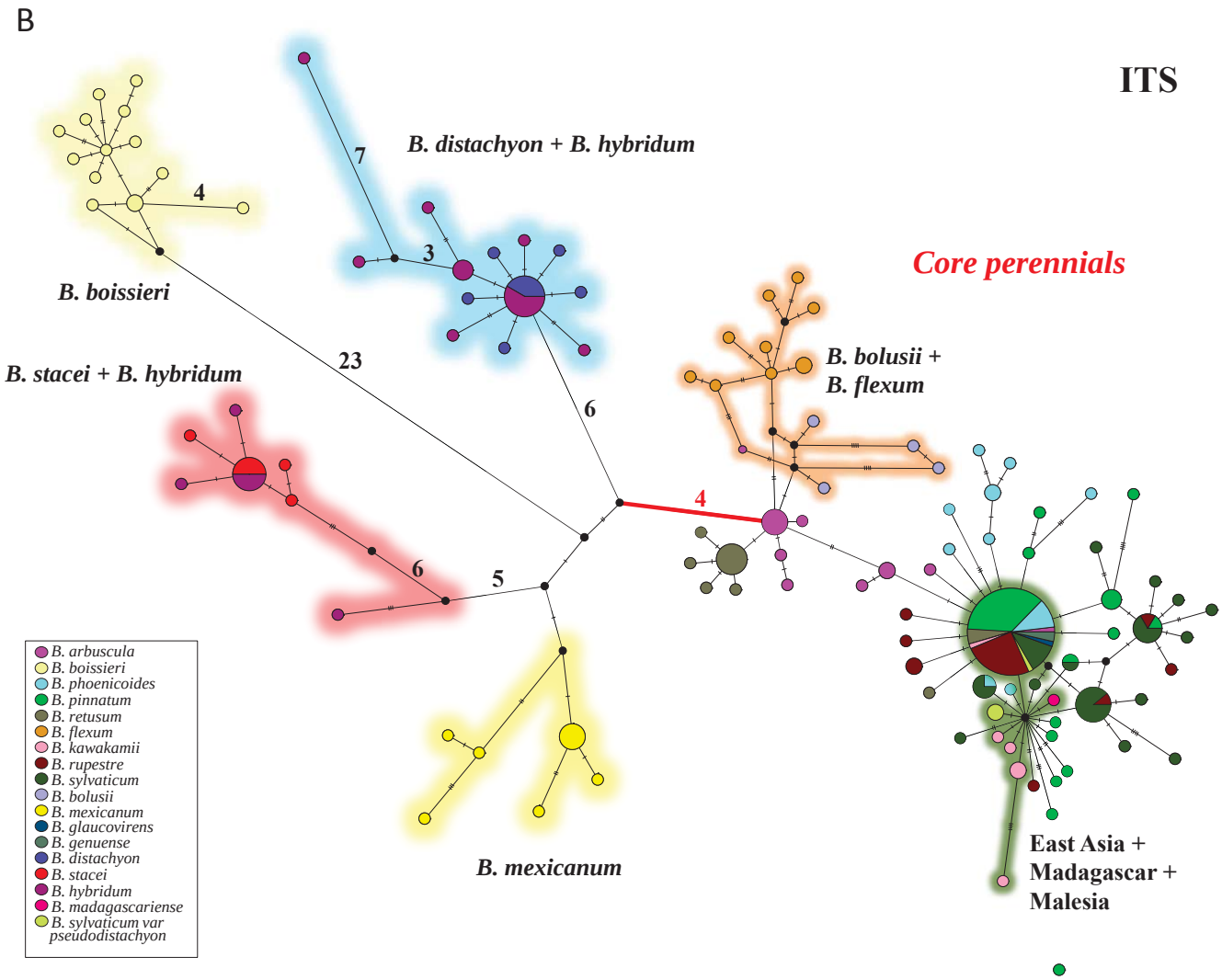

B
ITS
B. distachyon + B. hybridum
7
4
3
B. boissieri
23
B. stacei + B. hybridum
6
Core perennials
B. bolusii +
B. flexum
4
6
5
B. mexicanum
East Asia +
Madagascar +
Malesia
B. arbuscula
B. boissieri
B. phoenicoides
B. pinnatum
B. retusum
B. flexum
B. kawakamii
B. rupestre
B. sylvaticum
B. bolusii
B.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

Fig.3C

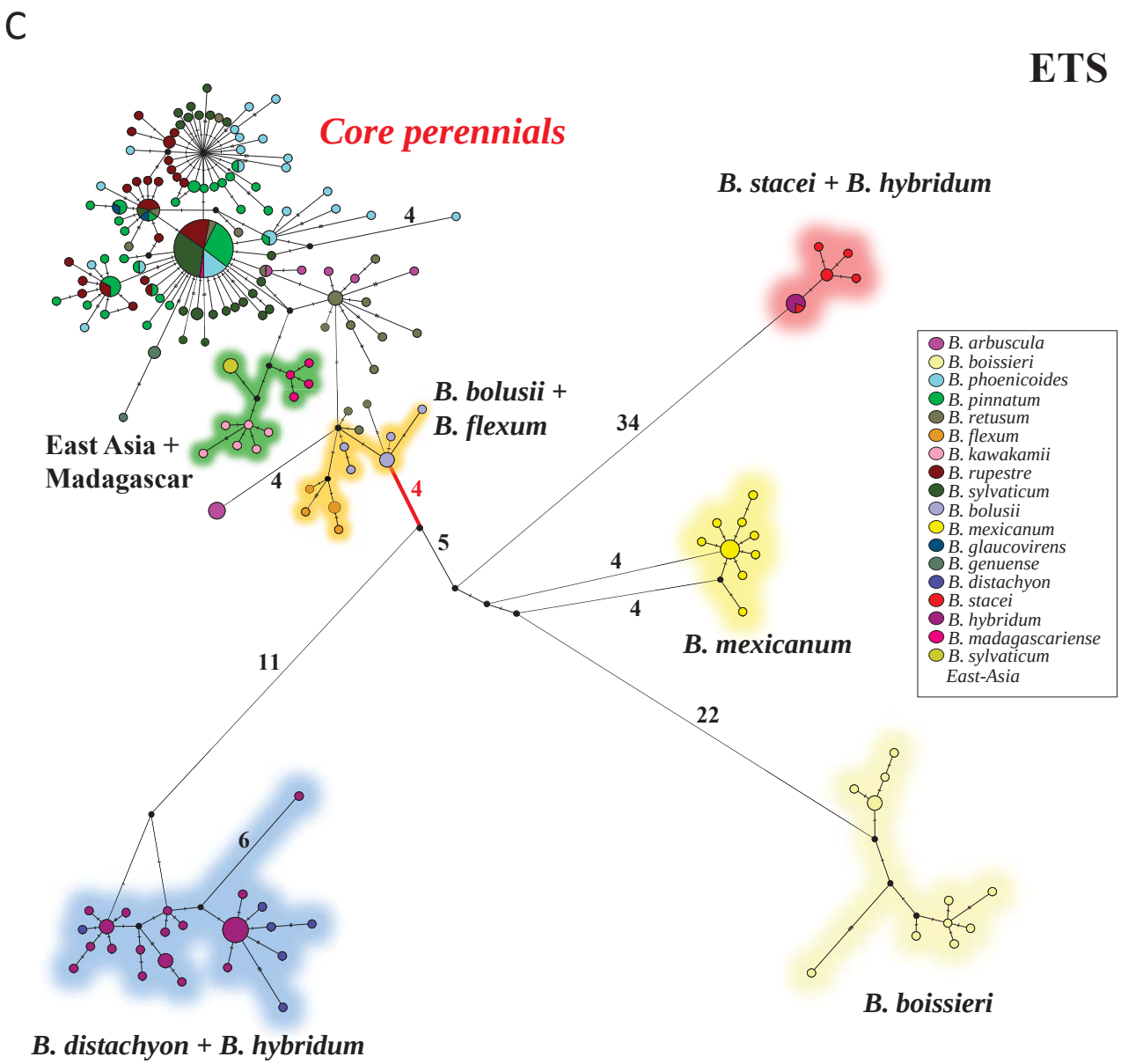

C
ETS
Core perennials
B. stacei + B. hybridum
B. bolusii +
B. flexum
East Asia +
Madagascar
B. mexicanum
B. boissieri
B. distachyon + B. hybridum
4
4
4
5
34
4
4
11
22
6
B. arbuscula
B. boissieri
B. phoenicoides
B. pinnatum
B. retusum
B. flexum
B. kawakamii
B. rupestre
B. sylvaticum
B. bolusii
B. mexicanum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

Fig.3D

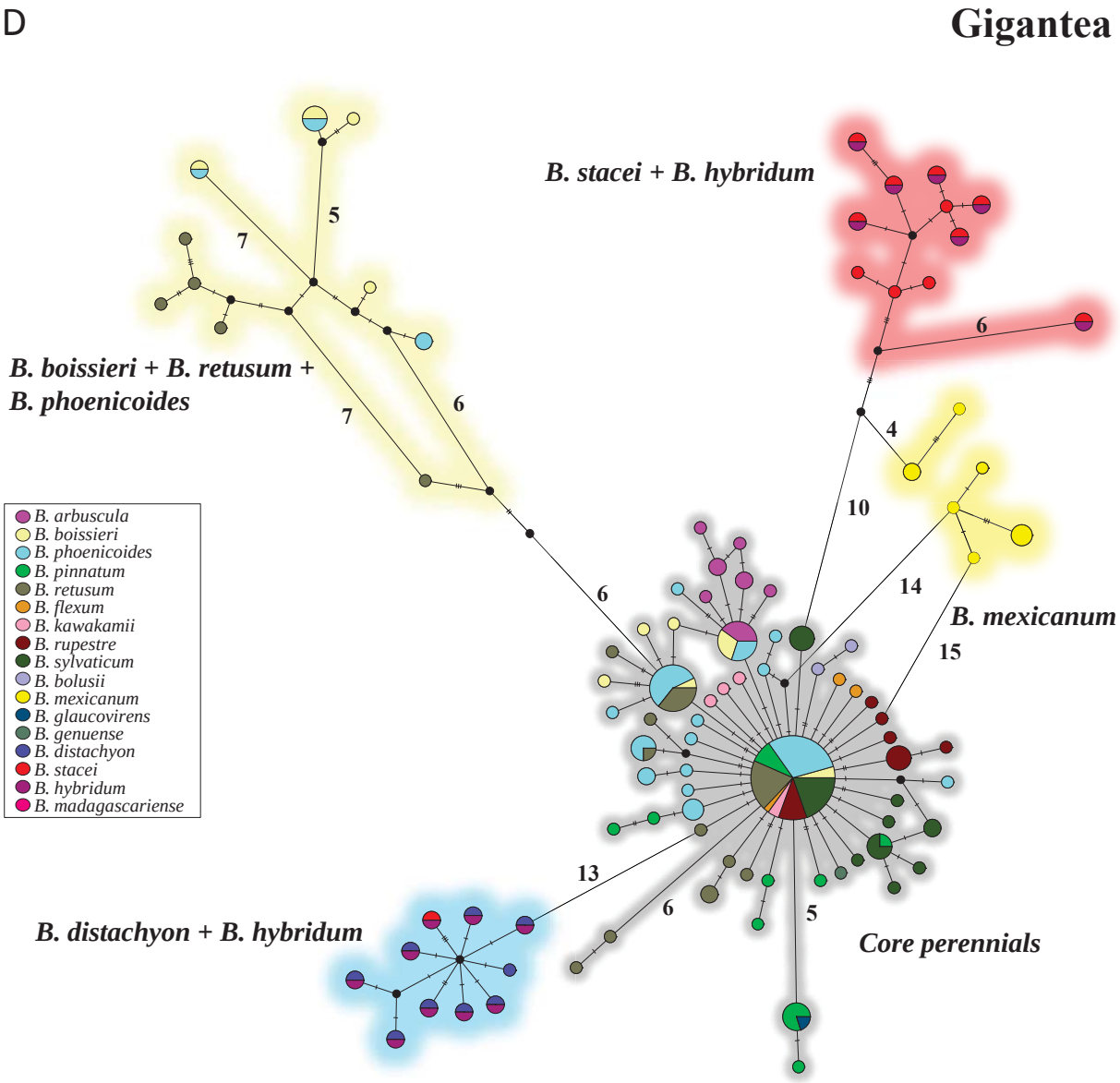

D
Gigantea
B. stacei + B. hybridum
B. boissieri + B. retusum +
B. phoenicoides
B. distachyon + B. hybridum
B. mexicanum
Core perennials
5
7
7
6
6
4
10
14
15
13
6
5
B. arbuscula
B. boissieri
B. phoenicoides
B. pinnatum
B. retusum
B. flexum
B. kawakamii
B. rupestre
B. sylvaticum
B. bolusii
B. mexicanum
B. glaucovirens
B. genuense
B. distachyon
B. stacei
B. hybridum
B. madagascariense

Fig.4

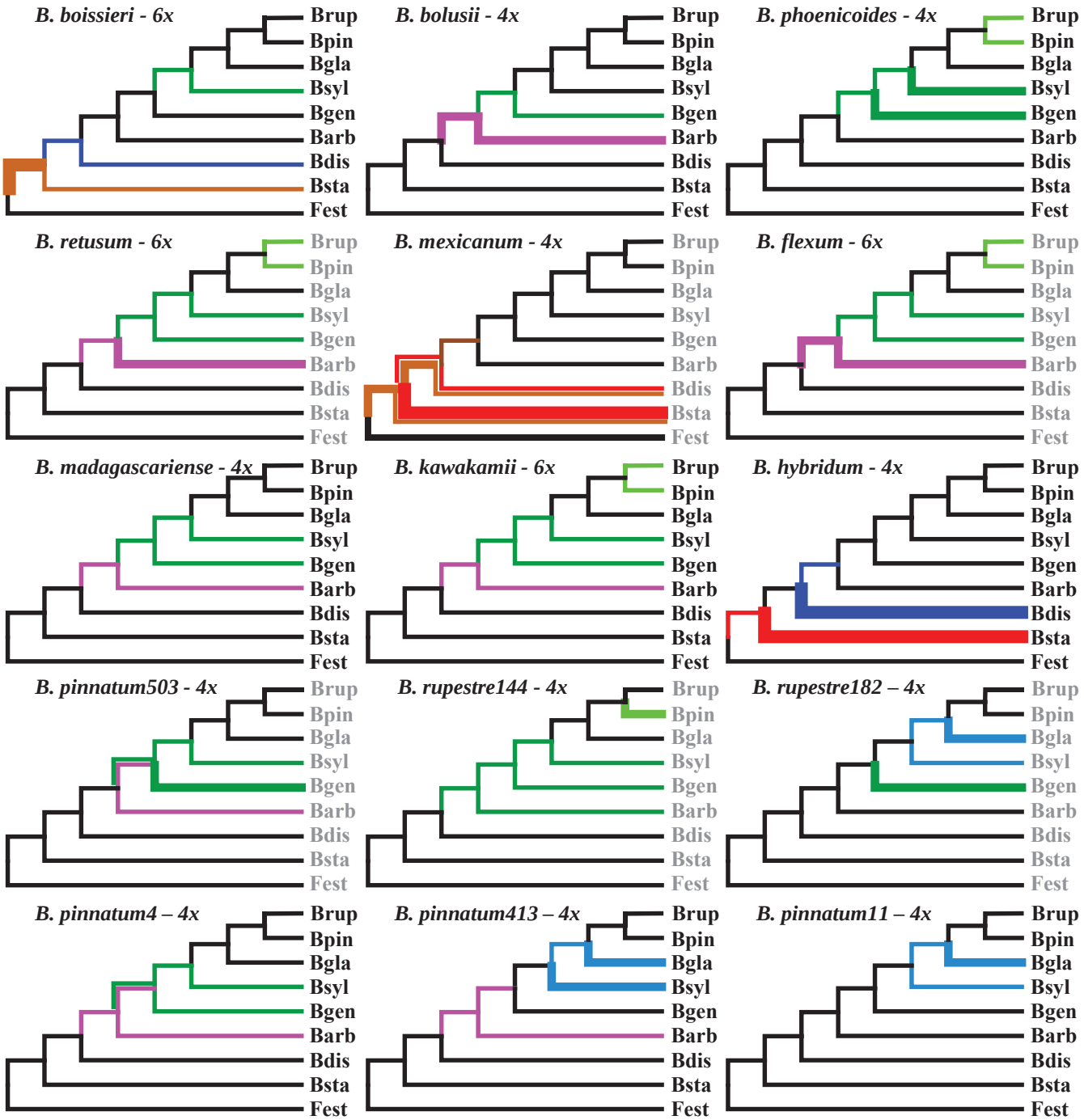

B. boissieri - 6x
B. bolusii - 4x
B. phoenicoides - 4x
B. retusum - 6x
B. mexicanum - 4x
B. flexum - 6x
B. madagascariense - 4x
B. kawakamii - 6x
B. hybridum - 4x
B. pinnatum503 - 4x
B. rupestre144 - 4x
B. rupestre182 – 4x
B. pinnatum4 – 4x
B. pinnatum413 – 4x
B. pinnatum11 – 4x
Brup
Bpin
Bgla
Bsyl
Bgen
Barb
Bdis
Bsta
Fest

Fig.5

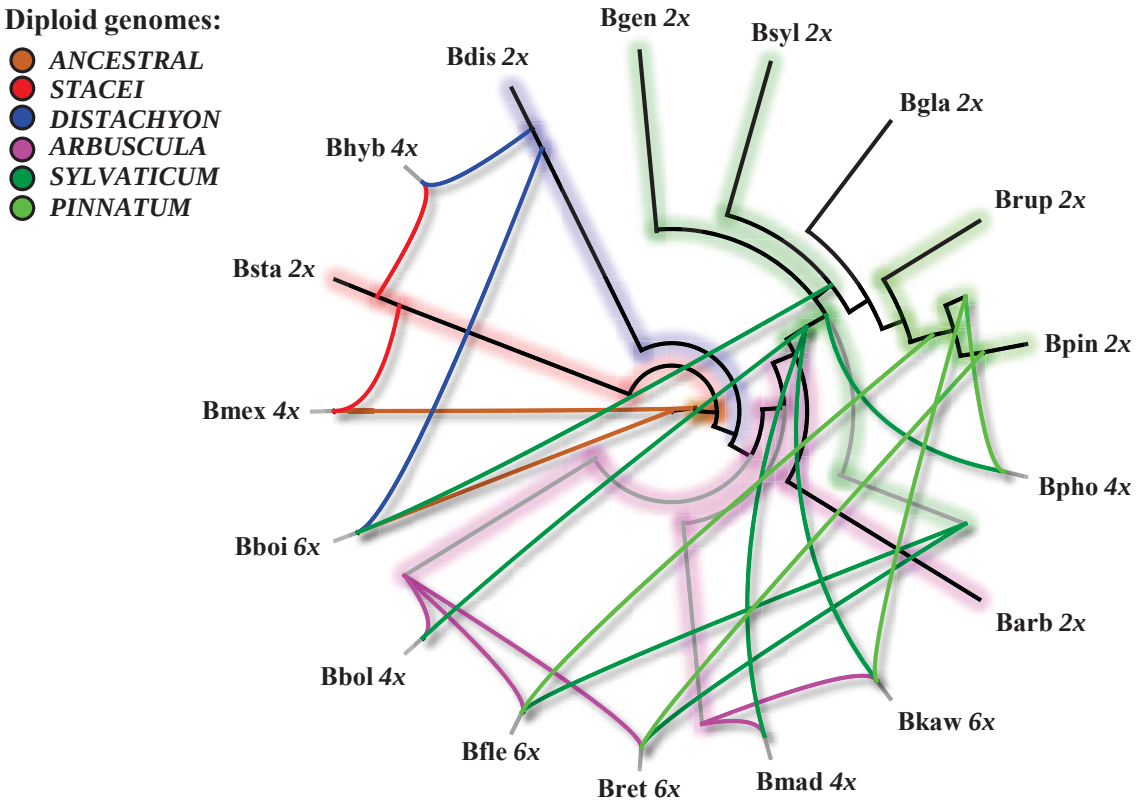

Diploid genomes:
ANCESTRAL
STACEI
DISTACHYON
ARBUSCULA
SYLVATICUM
PINNATUM
Bgen 2x
Bsyl 2x
Bdis 2x
Bgla 2x
Bhyb 4x
Brup 2x
Bsta 2x
Bpin 2x
Bmex 4x
Bpho 4x
Bboi 6x
Barb 2x
Bbol 4x
Bkaw 6x
Bfle 6x
Bmad 4x
Bret 6x

Fig.6

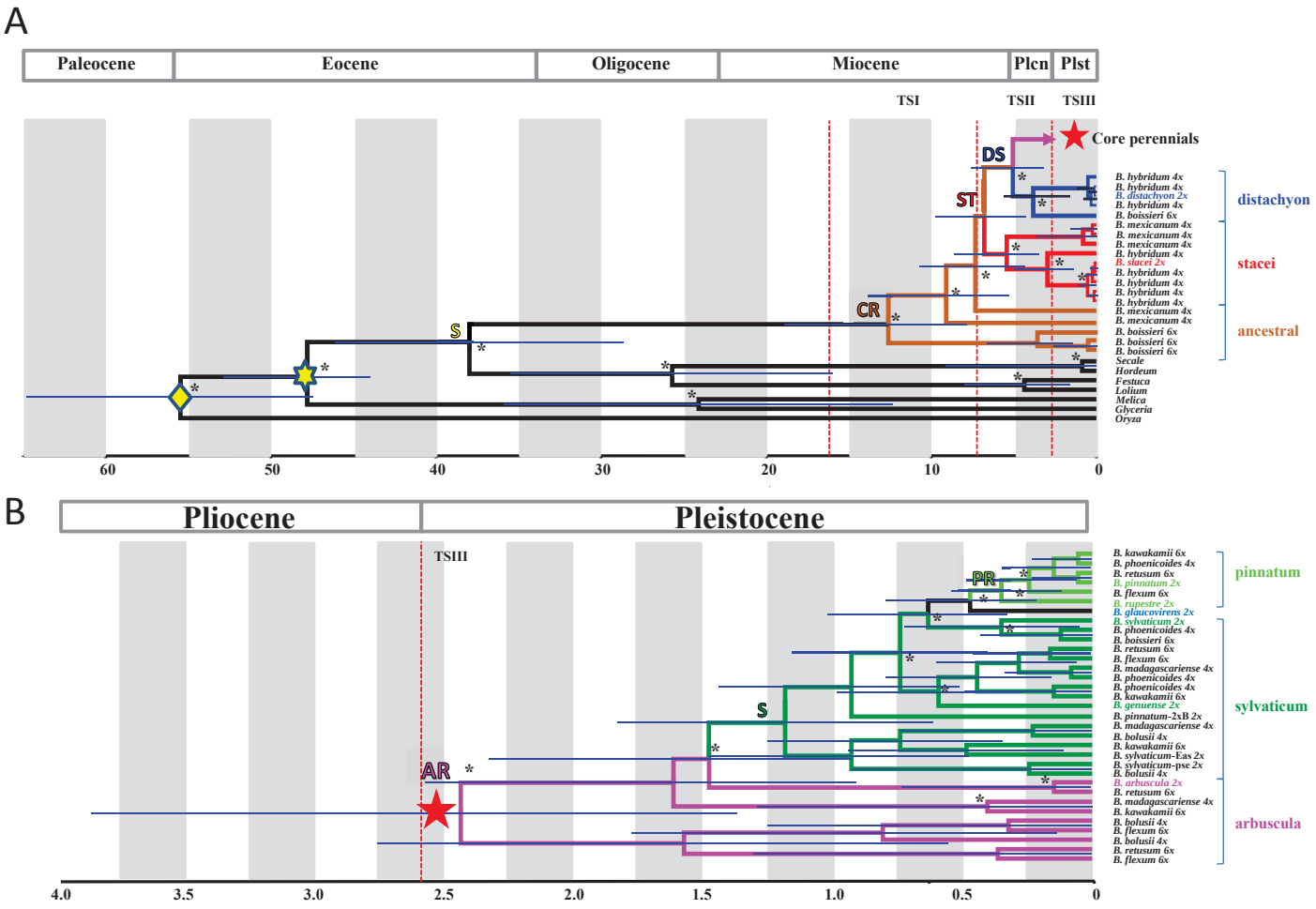

A
Paleocene
Eocene
Oligocene
Miocene
Plcn
Plst
TSI
TSII
TSIII
Core perennials
DS
ST
CR
S
B. hybridum 4x
B. hybridum 4x
B. distachyon 2x
B. hybridum 4x
B. boissieri 6x
B. mexicanum 4x
B. mexicanum 4x
B. mexicanum 4x
B. hybridum 4x
B. stacei 2x
B. hybridum 4x
B. hybridum 4x
B. hybridum 4x
B. hybridum 4x
B. mexicanum 4x
B. mexicanum 4x
B. boissieri 6x
B. boissieri 6x
B. boissieri 6x
Secale
Hordeum
Festuca
Lolium
Melica
Glyceria
Oryza
distachyon
stacei
ancestral
60
50
40
30
20
10
0
B
Pliocene
Pleistocene
TSIII
PR
S
AR
B. kawakamii 6x
B. phoenicoides 4x
B. retusum 6x
B. pinnatum 2x
B. flexum 6x
B. rupestre 2x
B. glaucovirens 2x
B. sylvaticum 2x
B. phoenicoides 4x
B. boissieri 6x
B. retusum 6x
B. flexum 6x
B. madagascariense 4x
B. phoenicoides 4x
B. phoenicoides 4x
B. kawakamii 6x
B. genuense 2x
B. pinnatum-2xB 2x
B. madagascariense 4x
B. bolusii 4x
B. kawakamii 6x
B. sylvaticum-Eas 2x
B. sylvaticum-pse 2x
B. bolusii 4x
B. arbuscula 2x
B. retusum 6x
B. madagascariense 4x
B. kawakamii 6x
B. bolusii 4x
B. flexum 6x
B. bolusii 4x
B. retusum 6x
B. flexum 6x
pinnatum
sylvaticum
arbuscula
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0

Fig.7

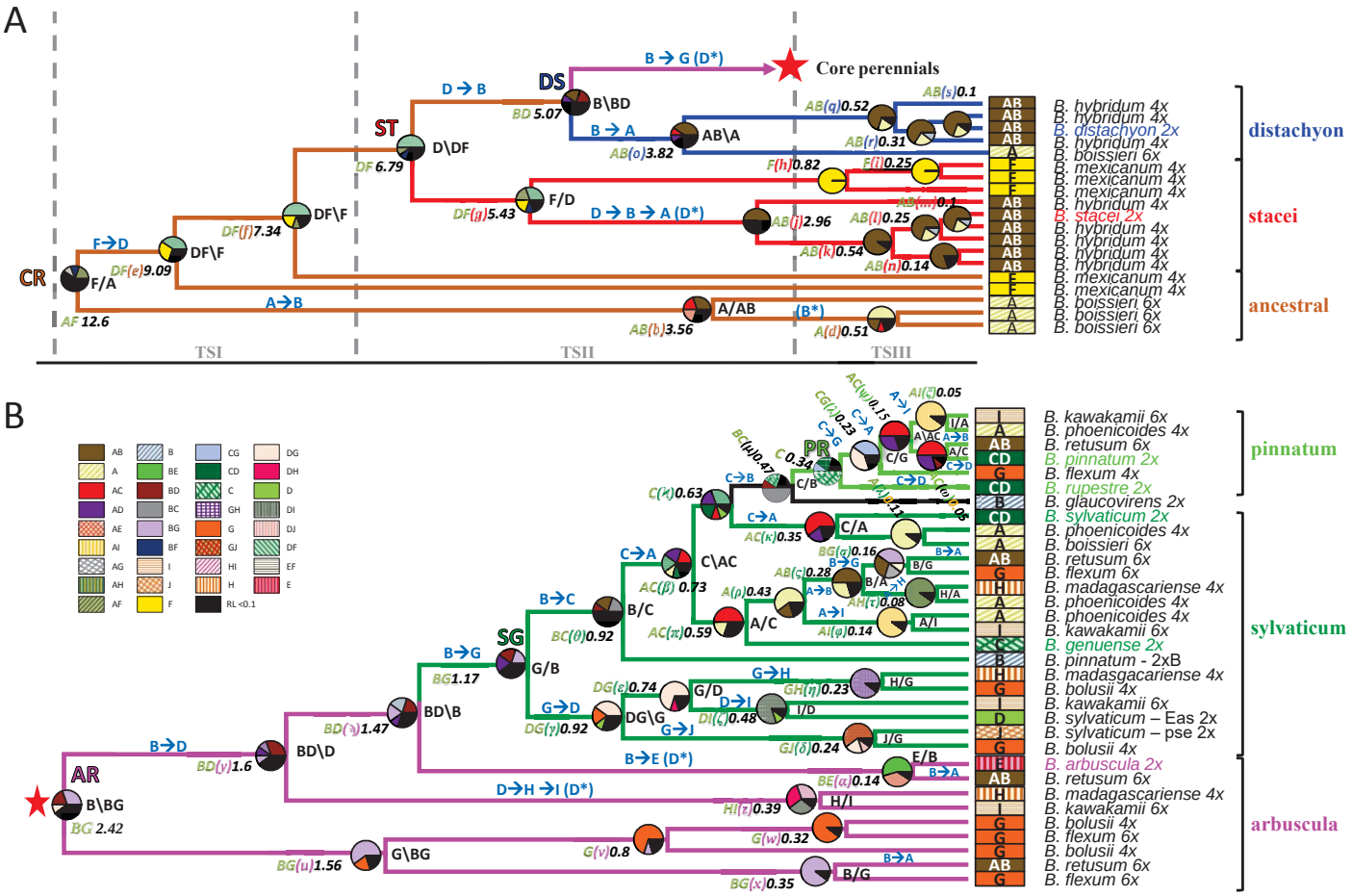

A
CR
F/A
F→D
DF\F
DF(e)9.09
DF(f)7.34
DF\F
ST
D\DF
DF 6.79
D→B
DS
B\BD
BD 5.07
B→A
AB\A
AB(o)3.82
B→G (D*)
Core perennials
AB(q)0.52
AB(s)0.1
AB(r)0.31
F(h)0.82
F(i)0.25
F/D
DF(g)5.43
D→B→A (D*)
AB(j)2.96
AB(l)0.25
AB(k)0.54
AB(n)0.14
A→B
A/AB
AB(b)3.56
(B*)
A(d)0.51
AF 12.6
TSI
TSII
TSIII
B. hybridum 4x
B. hybridum 4x
B. distachyon 2x
B. hybridum 4x
B. boissieri 6x
B. mexicanum 4x
B. mexicanum 4x
B. mexicanum 4x
B. hybridum 4x
B. stacei 2x
B. hybridum 4x
B. hybridum 4x
B. hybridum 4x
B. hybridum 4x
B. mexicanum 4x
B. mexicanum 4x
B. boissieri 6x
B. boissieri 6x
B. boissieri 6x
distachyon
stacei
ancestral
B
AR
B\BG
BG 2.42
B→D
BD\D
BD(y)1.6
BD(x)1.47
BD\B
B→G
SG
G/B
BG1.17
G→D
DG\G
DG(γ)0.92
DG(ε)0.74
G/D
G→J
G→H
B→E (D*)
D→H→I (D*)
BG(u)1.56
G\BG
G(v)0.8
G(w)0.32
BG(x)0.35
B/G
B→C
B/C
BC(θ)0.92
C→A
C\AC
AC(β) 0.73
AC(π)0.59
A/C
C→B
C/B
C(κ)0.63
C→G
PR
C 0.34
BC(μ)0.47
C→A
C/A
AC(κ)0.35
C→D
A(ρ)0.43
AB(ς)0.28
BG(σ)0.16
B/A
AH(τ)0.08
H/A
A/I
AI(φ)0.14
AC(ψ)0.15
CG(ι)0.23
AI(ξ)0.05
I/A
A/C
A(λ)0.11
C(ω)0.0
DG(ζ)0.48
I/D
GH(η)0.23
H/G
GJ(δ)0.24
J/G
F/B
BE(α)0.14
H(ζ)0.39
H/I
B. kawakamii 6x
B. phoenicoides 4x
B. retusum 6x
B. pinnatum 2x
B. flexum 4x
B. rupestre 2x
B. glaucovirens 2x
B. sylvaticum 2x
B. phoenicoides 4x
B. boissieri 6x
B. retusum 6x
B. flexum 6x
B. madagascariense 4x
B. phoenicoides 4x
B. phoenicoides 4x
B. kawakamii 6x
B. genuense 2x
B. pinnatum - 2xB
B. madagascariense 4x
B. bolusii 4x
B. kawakamii 6x
B. sylvaticum – Eas 2x
B. sylvaticum – pse 2x
B. bolusii 4x
B. arbuscula 2x
B. retusum 6x
B. madagascariense 4x
B. kawakamii 6x
B. bolusii 4x
B. flexum 6x
B. bolusii 4x
B. retusum 6x
B. flexum 6x
pinnatum
sylvaticum
arbuscula

Fig.8

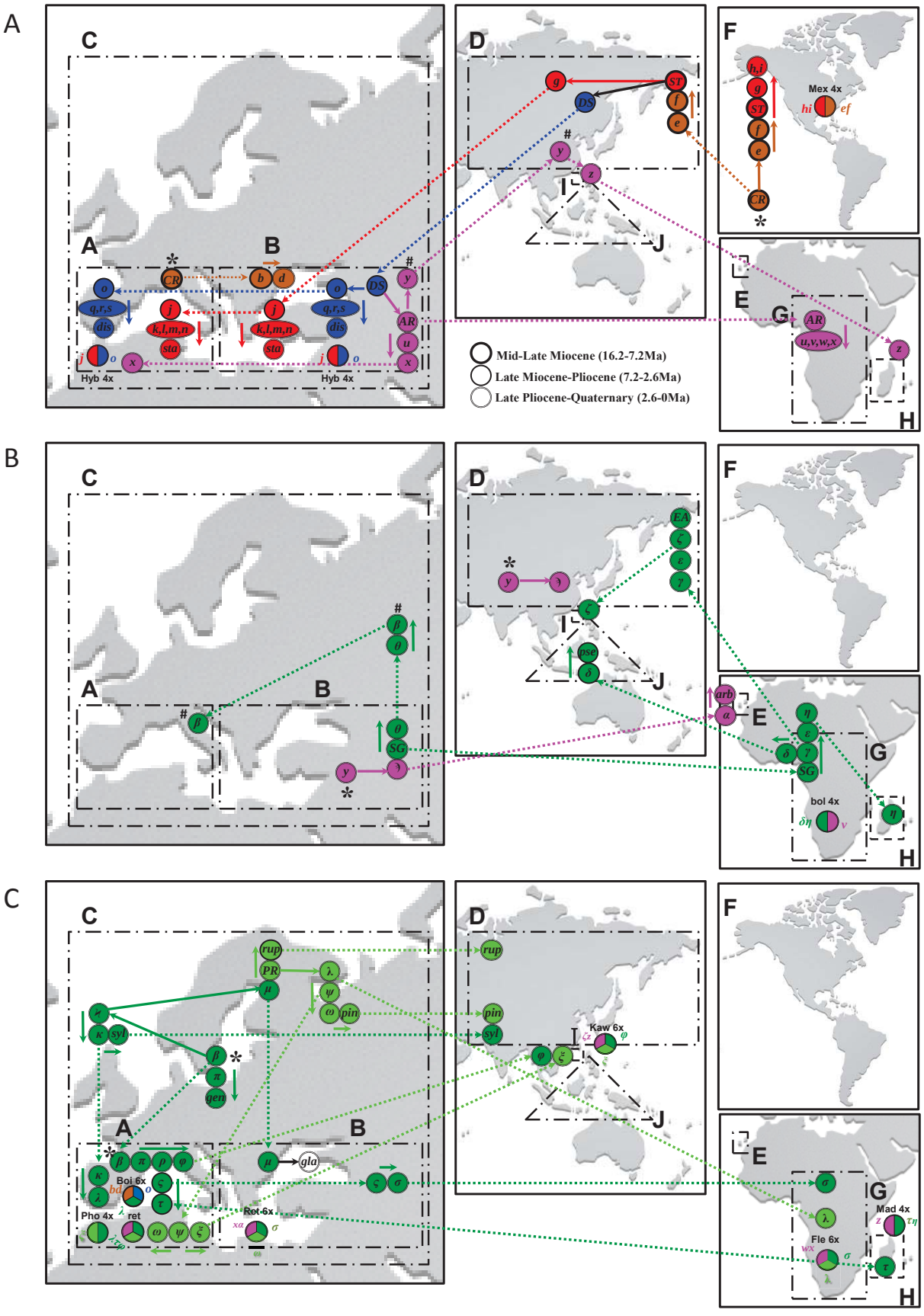
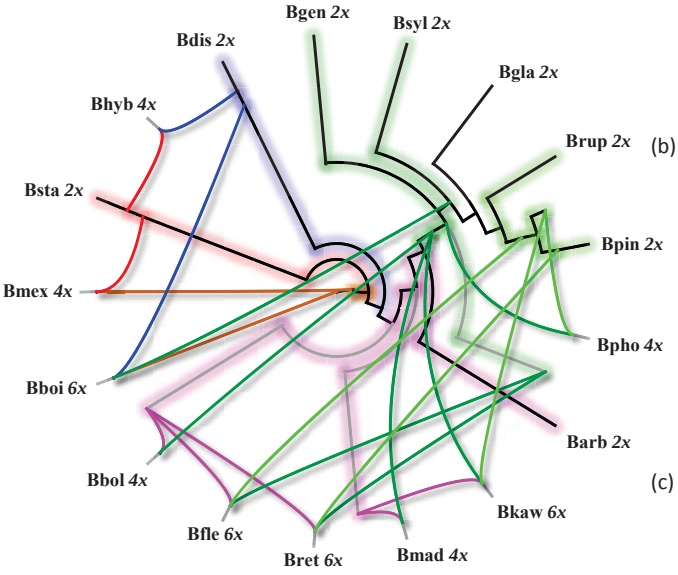
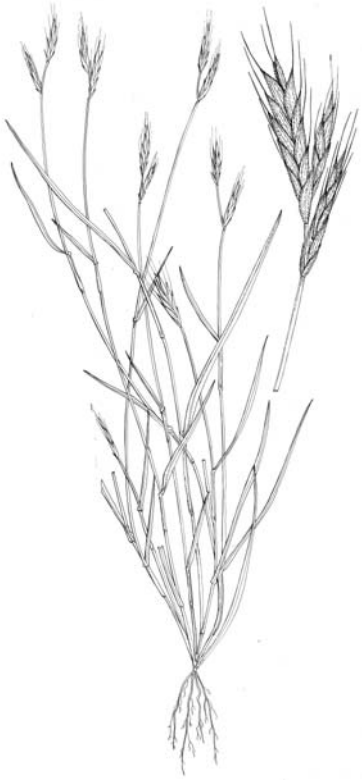

A
B
C
D
E
F
G
H
I
J
Mid-Late Miocene (16.2-7.2Ma)
Late Miocene-Pliocene (7.2-2.6Ma)
Late Pliocene-Quaternary (2.6-0Ma)
Hyb 4x
Hyb 4x
Mex 4x
Boi 6x
Pho 4x
ret
Ret-6x
Kaw 6x
bol 4x
Mad 4x
Fle 6x

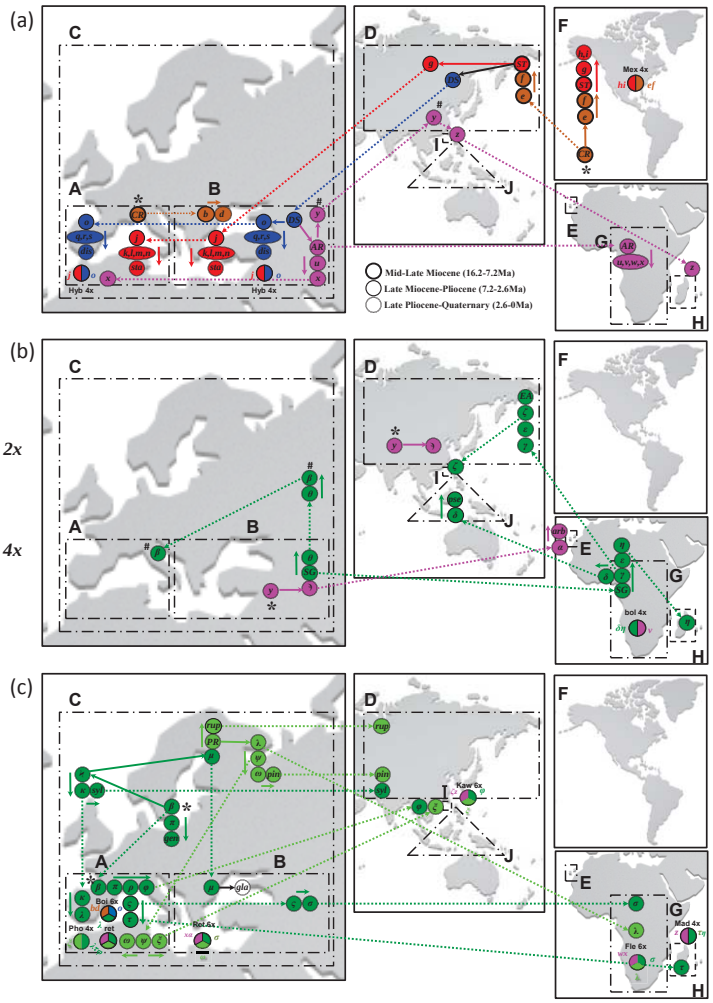
Table 1. The estimated age (Ma) of homeologous genomes present in the allopolyploid *Brachypodium* species. This is inferred from the coalescent splits from their respective closest counterpart diploid genome lineages, computed through the Isolation-Migration model implemented in IM. A square box represents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

Polyploid species	time**	AA	
		(2)	(1)
***B. hybridum* (4x)**			
(1) *STACEI* [4]	0.035		AB
(2) *DISTACHYON* [2]	0.060	AB	AB
***B. bolusii* (unknown)**			
(1) *ARBUSCULA* [2]	0.027		G
(2) *SYLVATICUM* [2]	0.379	G	G
***B. madagascariense* (unknown)**			
(1) *ARBUSCULA* [1]	0.390		H
(2) *SYLVATICUM* [2]	0.441	AG	AG
***B. mexicanum* (4x)**			
(1) *STACEI* [2]	3.377		F
(2) *ANCESTRAL* [2]	11.070	F	F
***B. phoenicoides* (4x)**			
(1) *PINNATUM* [1]	0.048		A
(2) *SYLVATICUM* [3]	0.052	A	A

Polyploid species	time	AA		
		(3)	(2)	(1)
***B. boissieri* (cf. 8x)**				
(1) *SYLVATICUM* [1]	0.030			A
(2) *DISTACHYON* [1]	3.750		A	A
(3) *ANCESTRAL* [3]	16.915*	A	A	A
***B. flexum* (unknown)**				
(1) *PINNATUM* [1]	0.024			G
(2) *SYLVATICUM* [1]	0.197		B	G
(3) *ARBUSCULA* [2]	0.609	BG	G	G
***B. retusum* (6x)**				
(1) *PINNATUM* [1]	0.036			A
(2) *ARBUSCULA* [1]	0.037		B	B
(3) *SYLVATICUM* [2]	0.466	A	B	B
***B. kawakamii* (unknown)**				
(1) *PINNATUM* [1]	0.067			A
(2) *ARBUSCULA* [1]	0.309		I	I
(3) *SYLVATICUM* [2]	0.476	AI	AI	I


Bgen 2x
Bsyl 2x
Bdis 2x
Bgla 2x
Bhyb 4x
Brup 2x
Bsta 2x
Bpin 2x
Bmex 4x
Bpho 4x
Bboi 6x
Barb 2x
Bbol 4x
Bkaw 6x
Bfle 6x
Bret 6x
Bmad 4x




(a)
(b)
(c)
A
B
C
D
E
F
G
H
I
J
Mid-Late Miocene (16.2-7.2Ma)
Late Miocene-Pliocene (7.2-2.6Ma)
Late Pliocene-Quaternary (2.6-0Ma)

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

Antonio Díaz-Pérez, Diana López-Álvarez, Rubén Sancho, Pilar Catalán

Highlights:

- A comprehensive 5-gene phylogeny of the model grass genus *Brachypodium* was built
- Minimum-Evolution and Species-Network approaches identified the concurring genomes
- Detected homeologous genomes matched the expected ploidy levels of allopolyploids
- Splittings and mergings of genomes occurred in different spatio-temporal scenarios
- Our biogeographical study infers dispersals of diploids but not of allopolyploids