ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Migration without interbreeding: Evolutionary history of a highly selfing Mediterranean grass inferred from whole genomes

Christoph Stritt[1] | **Elena L. Gimmi**[1] | **Michele Wyler**[1] | **Abdelmonaim H. Bakali**[2] | **Aleksandra Skalska**[3] | **Robert Hasterok**[3] | **Luis A. J. Mur**[4] | **Nicola Pecchioni**[5] | **Anne C. Roulin**[1]

[1]Institute for Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

[2]National Institute of Agronomy, Regional Center of Errachidia, Errachidia, Morocco

[3]Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, Katowice, Poland

[4]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Wales, UK

[5]Research Centre for Cereal and Industrial Crops, CREA – Council for Agricultural Research and Economics, Foggia, Italy

Correspondence
Anne C. Roulin, Institute for Plant and Microbial Biology, University of Zurich, Zurich, Switzerland.
Email: anne.roulin@botinst.uzh.ch

Present address
Christoph Stritt, Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland

Gimmi Gimmi, Department Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology Eawag, Dubendorf, Switzerland

Funding information
Swiss National Science Foundation; University of Zurich Research priority program "Evolution in Action"

Abstract

Wild plant populations show extensive genetic subdivision and are far from the ideal of panmixia which permeates population genetic theory. Understanding the spatial and temporal scale of population structure is therefore fundamental for empirical population genetics – and of interest in itself, as it yields insights into the history and biology of a species. In this study we extend the genomic resources for the wild Mediterranean grass *Brachypodium distachyon* to investigate the scale of population structure and its underlying history at whole-genome resolution. A total of 86 accessions were sampled at local and regional scales in Italy and France, which closes a conspicuous gap in the collection for this model organism. The analysis of 196 accessions, spanning the Mediterranean from Spain to Iraq, suggests that the interplay of high selfing and seed dispersal rates has shaped genetic structure in *B. distachyon*. At the continental scale, the evolution in *B. distachyon* is characterized by the independent expansion of three lineages during the Upper Pleistocene. Today, these lineages may occur on the same meadow yet do not interbreed. At the regional scale, dispersal and selfing interact and maintain high genotypic diversity, thus challenging the textbook notion that selfing in finite populations implies reduced diversity. Our study extends the population genomic resources for *B. distachyon* and suggests that an important use of this wild plant model is to investigate how selfing and dispersal, two processes typically studied separately, interact in colonizing plant species.

KEYWORDS
flowering time, heterozygosity, population structure, seed dispersal, selfing

1 | INTRODUCTION

In population genetic theory, a population is a matter of definition. In the natural world, populations are part of complex geographic mosaics; they can be discrete, continuous, or somewhere in-between; and they can expand and contract, appear and disappear. In short, answering the question "What is a population?" is not straightforward for most species (Waples & Gaggiotti, 2006). Yet answering it, implicitly or explicitly, establishes the link between theory and data and is the first step for many studies in ecology, evolution, and conservation.

The concept of population structure or subdivision subsumes many of the challenges of empirical population genetics (Hey & Machado, 2003; Wright, 1931). Population structure results from the phylogeographic and demographic history of the species (Charlesworth, 2009; Nieto Feliner, 2014), life history traits such as dispersal mode and breeding system (Loveless & Hamrick, 1984), and ecological interactions and adaptation (Linhart & Grant, 1996). Disentangling how these different processes contribute to genetic diversity and differentiation is a major goal of evolutionary biology. As processes are inferred from patterns, for any species the disentangling ideally begins with a thorough analysis of population subdivision, that is, of the spatial and temporal scale of genetic differentiation (e.g., Platt et al., 2010).

In the past years, the resolution at which population structure can be analysed has increased greatly due to the availability of genomic data (Pannell & Fields, 2014). Sampling genealogies along genomes allows to overcome problems due to the high stochasticity of single gene trees (Hey & Machado, 2003). Furthermore, measures of absolute genetic divergence such as d_{XY} can replace or complement more assumption-rich statistics such as F_{ST}, which rely on *a priori* definitions of populations and can be confounded by within-population structure or the sampling of close relatives (Putman & Carbone, 2014).

In this study we investigate the scale of population structure in the Mediterranean grass *Brachypodium distachyon* and infer its underlying history of population divergence. *Brachypodium distachyon* naturally occurs around the Mediterranean Basin and is a specialist of arid, oligotrophic habitats characterized by recurrent disturbance (Catalán et al., 2016; López-Alvarez et al., 2015). It has a fully assembled diploid 272 Mb genome (International Brachypodium Initiative, 2010) and is an established model to study various aspects of grass biology (Scholthof et al., 2018; Vogel, 2016). Its value as a model for plant evolution lies in its Mediterranean distribution, providing a window into an ecologically precious yet understudied region (Thompson, 2005); its belonging to the Poaceae, a plant family dominating terrestrial ecosystems around the globe (Gibson, 2009); and, finally, its genome of intermediate complexity, where numerous gene copy number variants (Gordon et al., 2017) and transposable element polymorphisms (Stritt et al., 2018, 2020) contribute to molecular evolution.

A key question which has emerged from previous population genetic studies with *B. distachyon* concerns the role of seed dispersal in shaping genetic structure. A large-scale microsatellite survey in Turkey reported high genotypic diversity within and little differentiation between populations (Vogel et al., 2009). This was explained through a combination of long-distance seed dispersal and extreme selfing rates, the former shuffling genotypes across the landscape, the latter preventing gene flow between diverged genotypes. Other studies, however, found little diversity within and high differentiation between populations in Turkey (Dell'Acqua et al., 2014) and Spain (Marques et al., 2017), in agreement with the increased genetic drift expected in selfing organisms (Charlesworth, 2003). However, as these studies considered different populations and used different sampling schemes, it remains difficult to tell whether their disagreement reflects different biological realities or study designs.

A second pattern highlighted by previous studies is the sympatry of highly diverged lineages differing in flowering phenology (Gordon et al., 2017; Sancho et al., 2018; Tyler et al., 2016). As in other widespread annual plants (e.g., Engelmann & Purugganan, 2006), genetic variation for the control of flowering is present in *B. distachyon*, in particular in the vernalization pathway preventing the development of flowers in the cold season (Ream et al., 2014). It has been suggested that differences in flowering phenology could explain the lack of admixture between sympatric lineages (Gordon et al., 2017). This hypothesis, however, is difficult to evaluate as long as the role of phenotypic plasticity in flowering and the phylogeographic history underlying the large-scale distribution of lineages remain unknown.

Here we extend the genomic resources for *B. distachyon* with whole-genome sequences for 75 Italian and 11 French accessions sampled at local and regional scales. In addition to filling a conspicuous geographical gap in the *B. distachyon* collection, our aim was to sample across different spatial scales in order to better understand the scale of population structure in this species. Combining these resources with publicly available data, we describe the scale of population structure in 196 accessions spanning the Mediterranean from Spain to Iraq, infer the underlying history of population divergence, and discuss how the interplay of selfing and dispersal has shaped genetic structure across spatial and temporal scales. Furthermore, we conducted flowering time experiments illustrating the phenotypic plasticity of flowering phenology and suggesting caution when extrapolating from greenhouse experiments to natural populations.

2 | MATERIALS AND METHODS

2.1 | Sampling and sequencing

719 dry *Brachypodium* plants were collected in 2016 across Italy, in the Languedoc region in Southern France, and in 2018 across the Atlas Mountains of Morocco (Table S1). A representative set of 276 accessions were grown in the greenhouse from single seeds. Their DNA was extracted from leaves with a CTAB-based method (Stein et al., 2001), quality-checked with Nanodrop (NanoDrop Technologies Inc), and quantified with Qubit (Invitrogen).

As *B. distachyon* may be confounded with its morphologically similar sister species *B. stacei* and *B. hybridum* (Catalán et al., 2016), we first identified the species of the 276 accessions using the microsatellite marker ALB165 (Giraldo et al., 2012) and by sequencing the GIGANTEA gene (López-Alvarez et al., 2012) with ABI SOLiD. 143 individuals were identified as *B. distachyon*, 130 as *B. hybridum*, and three as *B. stacei* (Table S1). All 70 accessions collected in Morocco proved to be *B. hybridum* and were not included in this study.

86 out of the 143 genotyped *B. distachyon* accessions were selected for sequencing. In order to reduce the chance of sequencing close relatives, only accessions collected at least 50 m apart were included. In addition, we also sequenced the reference accession Bd21 as a control and the *B. stacei* accession Cef2 as an outgroup for phylogenetic analyses. Paired-end libraries were prepared with the PCR-free KAPA Hyper Prep Kit (KAPA Biosystems). The 150 bp paired-end reads were sequenced with Illumina HiSeq2500 at the Genomics Facility Basel to a mean coverage of 23 (range 18–38, Table S1). The resulting sequencing data were joined with 57 recently sequenced Turkish accessions (Skalska et al., 2020; mean coverage 36) and with data for 52 accessions originating mainly from Spain and Turkey (Gordon et al., 2017). To speed up computation and avoid bias due to unequal coverage, the latter reads were downsampled to a mean coverage of 30 using PICARD tools v.1.97 (http://broadinstitute.github.io/picard). The total number of *B. distachyon* accessions analysed in this study is thus 86 + Bd21 + 57 + 52 = 196.

2.2 | Alignment, variant calling & filtering

Paired-end reads were aligned to version 3.0 of the *B. distachyon* Bd21 reference genome, downloaded from Phytozome 12, with BWA MEM v. 0.7.12-r1039 (Li, 2013). Aligned reads were converted to the bam format and sorted with SAMTOOLS v. 1.9 (Li et al., 2009), and read duplicates were removed with sambamba (Tarasov et al., 2015). Variants were then called with GATK v. 4.0.2.1. The following hard-filtering thresholds were applied to the combined variants: only SNPs were retained with a quality-by-depth >20, a Phred score <20 for Fisher's exact test for strand bias, and a Z-score between –1 and 1 for Wilcoxon's rank sum test of Alt versus Ref read mapping qualities. Additionally we removed SNPs in low complexity regions identified with dustmasker (Morgulis et al., 2006), resulting in a total of 8,426,015 hard-filtered SNPs.

2.3 | Dealing with artificial heterozygosity due to structural variants

Identifying heterozygous SNPs in *B. distachyon* is complicated by a high number of gene copy number variants (Gordon et al., 2017). If a stretch of DNA is duplicated in a sequenced accession but present as a single copy in the reference, sequencing reads derived from the duplicates are collapsed onto a single region when aligning them to the reference genome. Depending on the divergence of the duplicates, this may give rise to islands of extended heterozygosity (IEH): regions of up to few thousand base pairs with extreme numbers of heterozygous sites (Figure S1). As wrongly identified heterozygosity might severely bias downstream analyses, in particular the estimation of selfing rates, we first quantified the issue by scanning each accession for IEH. An IEH was recorded if there were ten or more heterozygous SNPs along a distance of at least 300 bp. The program implementing this approach (detectIEH.py) is available on http://github.com/cstritt/popgen.

Because hard filtering proved insufficient to deal with the artificial heterozygosity identified in the previous step, we used a masking approach to obtain high-confidence SNPs, including heterozygous sites. For each accession a mask was created indicating the regions in the genome within 1.5 standard deviations of the accession-specific mean read coverage. In this way, an average of 85% of the genome was included per accession (range 76% [Bd1-1] to 88% [Bd21]). In a next step, the individual masks were combined with BEDTOOLS (Quinlan & Hall, 2010) to define high-confidence regions as regions with "normal" coverage in at least 90% of the accessions. Removing SNPs outside these regions resulted in a set of 5,774,928 SNPs across the 196 accessions. These variants are referred to as high-confidence SNPs in the following.

2.4 | Population structure and historical gene flow

Genetic structure was analysed using principal component analysis and ancestry coefficient estimation. For the principal component analysis, the 5,774,928 SNPs in high-confidence regions were used and additionally filtered with the snpgdsLDpruning function of the R package SNPRELATE (Zheng et al., 2012). Only SNPs were retained with a minor allele frequency >0.05 and a composite linkage disequilibrium <0.4, resulting in 14,485 informative SNPs. The snpgdsPCA function of the same package was used to calculate the components. Ancestry coefficients were estimated with sNMF, as implemented in the R package LEA (Frichot & François, 2015), a method appropriate for high levels of inbreeding as it does not assume Hardy-Weinberg equilibrium. Using the same SNPs as for the PCA, sNMF was run for K values from 2 to 20, with 10 repetitions per K.

A rooted phylogeny was estimated with SNPhylo (Lee et al., 2014). For this analysis we included the *B. stacei* accession Cef2 as an outgroup and regenerated a high-confidence SNP set as discribed above. Synonymous sites were used and additionally filtered for a minor allele frequency of 0.05, a maximum missing rate of 0.1, and a maximum linkage disequilibrium (r^2) of 0.4, resulting in 5,432 LD-pruned synonymous sites. One hundred bootstrap replicates were performed to assess the robustness of the tree.

FINESTRUCTURE v. 4.0.1 (Lawson et al., 2012) was used to obtain a more detailed picture of population subdivision in *B. distachyon*. This method leverages haplotype information to infer shared ancestry between individuals. More specifically, it estimates the number and lengths of chunks along the genome of individual *i* coalescing with individual *j* before they coalesce with any other individual. The

program output we present is a similarity matrix which indicates the recombination map distance a "recipient" genotype (rows) receives from the "donor" genotypes (columns). As FINESTRUCTURE models linkage disequilibrium and performs best with dense haplotype data, the 8,426,015 hard-filtered SNPs were used after removing heterozygous sites and missing data, resulting in 3,467,964 SNPs. By removing heterozygous sites the accessions are treated as haploid and no phasing is thus required. Recombination distances between loci were obtained from the linkage map of Huo et al. (2011). FINESTRUCTURE was run in its haploid mode, using default settings.

To test whether shared ancestry is better explained by common descent or historical gene flow, TreeMix (Pickrell & Pritchard, 2012) was used to model migration. The high-confidence SNP set was used, with 4,473,162 variable sites after removing missing data. To account for linkage disequilibrium, a block size k of 2,000 SNPs was defined, which corresponds to an average distance of 120 kb in the genome. Models with zero to four migration edges *m* were tested, with 100 independent runs for each *m*. For each model, the run achieving the highest likelihoods was used for model comparison.

2.5 | Linkage disequilibrium

Selfing has important consequences for the extent of linkage disequilibrium (LD). The effective recombination rate is 1 – s times lower with selfing because meiotic recombination is mainly reshuffling homozygous alleles (Nordborg, 2000), resulting in extensive linkage disequilibrium. On the other hand, population substructure in the absence of gene flow decreases LD as mutations accumulate in independently evolving lineages (Hartfield et al., 2018). We estimated linkage disequilibrium at different scales of population structure with POPLDDECAY (Zhang et al., 2019). LD is reported as r^2 averaged over pairwise comparisons. The high-confidence SNP set was used with default program settings (-MAF 0.05, -Miss 0.25). To assess the effect of population subdivision on LD, we used the sum of terminal branch lengths in subtrees of the SNPhylo phylogeny as a measure of how much independent evolution has occurred in different subtrees.

2.6 | Estimation of divergence times

To put a timescale on the divergence in *B. distachyon*, we used the multispecies coalescent approach implemented in BPP v.4.2.9 (Flouri et al., 2018; Rannala & Yang, 2003). 20 diverse accessions were selected to represent the *B. distachyon* phylogeny. 200 random genomic regions of 1 kb length and at least 100 kb apart were chosen, as BPP assumes no recombination within and free recombination between loci. For each of the 20 accessions the 200 sequences were obtained by calling consensus sequences from the respective bam file. Potential problems of paralogy were avoided by considering only reads with a mapping quality >30, which excludes reads mapping to multiple places in the genome. Inverse gamma priors were set to (3, 0.014) for the root age τ and to (3, 0.002) for the population

size parameter θ, which corresponds to a mean theta of 0.001 and a mean root age of one million years, assuming a constant mutation rate of 7×10^{-9} substitutions per site per generation (Ossowski et al., 2010). The rooted species tree was defined as (((A_East, A_Italia), (B_East, B_West)), C_Italia), as inferred with the SNPhylo analysis described above. The MCMC was run four times independently, each time with 408,000 iterations, including a burnin of 8,000 iterations and samples taken every second iteration. Convergence of the four runs was assessed with Tracer (Rambaut et al., 2018).

2.7 | Genetic diversity and population differentiation

To estimate genetic diversity and differentiation, we estimated absolute genetic differentiation d_{XY} at synonymous sites. The --includeNonVariantSites option of GATK v.3.4 was used to obtain variable and nonvariable synonymous sites. For each pairwise comparison between the 196 accessions, we then calculated the number of variable and the total number of sites. Heterozygous sites were ignored and only sites were considered for which both accessions in the comparison had a read coverage within 1.5 standard deviations of the accession-specific mean coverage. The script implementing this procedure (pairwise_nucdiff.py) is provided on http://github.com/cstritt/popgen. Pairwise geographic distances were calculated as Great Circle (WGS84 ellipsoid) distances between each pair, using the spDistsN1 function of the R package SP. (Pebesma & Bivand, 2005). Mantel tests were used to test for isolation by distance, using the mantel.rtest function of the ADE4 package with 100 permutations to assess the significance of the correlations (Dray & Dufour, 2007). To test if genetic diversity differs between regions, a Kruskal-Wallis test was performed followed by Dunn's test for multiple comparisons with a Benjamini-Hochberg correction.

2.8 | Flowering time experiments

Two experiments were conducted to investigate how flowering phenology relates to genetic structure in *B. distachyon*. Firstly, we tested how the newly sampled accessions respond to vernalization treatments ranging from 0 to 12 weeks at 4°C under short-day conditions in the greenhouse. Twelve Italian accessions with different geographic origins were analysed (Cro24, Cm2, Leo2, Mca23, Mca18, Msa27, Ren22, San12, Sap49, Spm16, Sul5, Tar1). We further used four control accessions from Turkey and Iraq (Bd1-1, Bd3-1, BdTR3C, BdTR7a) previously investigated (Ream et al., 2014; Sharma et al., 2017) in order to confirm the reproducibility of the flowering behaviour and vernalization requirements. Plants from all 16 accessions were germinated for 3 weeks in the greenhouse. They were subsequently put in a cooling chamber for 0, 4, 6, 8, 10 or 12 weeks, respectively, and transferred back to the greenhouse. Three to five replicates were used per accession per time point. Time until flowering was counted from the day after
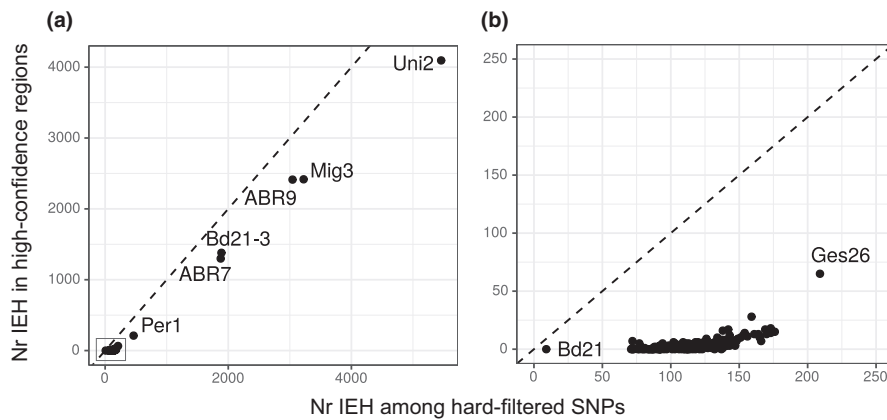
(a)

(b)

FIGURE 1 Islands of extended heterozygosity (IEH) among hard-filtered SNPs (x-axis) and among SNPs in high confidence regions (y-axis), defined as regions where at least 90 percent of the accessions have a read coverage within 1.5 standard deviations of their mean coverage. (a) Overview of all 196 accessions, highlighting the six outlier accessions. (b) Zoom-in on the small square in panel a

the return to warm conditions (directly after the first 3 weeks of germination for the plants with 0 weeks vernalization). Flowering was defined as the emergence of the first spike. The cooling chamber was set to 4°C, 8 h light and 16 h dark. The greenhouse conditions were 16 h day at 20–21°C and 8 h dark at 18°C with a light intensity of 200 umol PAR and 60% relative humidity over the entire experiment.

In the wild, plants all go through a prolonged vernalization process during winter which might produce different outcomes than the conditions tested in the greenhouse. To test for this possibility, we planted seeds of 105 accessions outdoors in November 2017, in the Botanical Garden of Zurich, Switzerland, with six replicates per accession. Only six plants did not germinate and 624 plants were used for phenotyping. To ensure no bias due to position effect, replicates were distributed randomly across trays. Mild heating was applied at the bottom of the compartment to prevent freezing when outdoor temperature fell below –10°C. From April 2018 on, plants were checked every three to four days for flowering. The first plant flowered 21 April 2018, which was used as a reference (day 1). About 14% of the replicates from the A lineage desiccated before the end of the experiment and did not reach flowering. To test for differences in flowering time between clades a linear mixed effects model was used, as implemented in the R package MCMCGLMM (Hadfield, 2010), and individual accessions were treated as random effects to control for the six biological replicates.

3 | RESULTS

3.1 | Occurrence of *Brachypodium distachyon*

Brachypodium distachyon, *B. stacei*, and their allotetraploid hybrid *B. hybridum* were described as different species only in 2012 (Catalán et al., 2012), while before they were considered different cytotypes of *B. distachyon*. As a consequence, the relative distribution of the three species is still somewhat vague in many regions, including Italy. In agreement with general habitat descriptions (Catalán et al., 2016), we found that *B. hybridum* was more frequent in the drier South of Italy, while *B. distachyon* was

predominant in Northern Italy and France (Table S1). Only three plants were genotyped as *B. stacei*, in agreement with this species being the rarest of the three (Table S1). Furthermore, we found only *B. hybridum* in Morocco, suggesting that *B. distachyon* might be rare on the African continent.

3.2 | Single nucleotide polymorphisms: Beware of artificial heterozygosity

As mentioned above, a significant proportion of the initial 8,426,015 hard-filtered SNPs were observed to aggregate in islands of extended heterozygosity (IEH, Figure S1). IEH can be artefacts which arise when reads from paralogous stretches of DNA are collapsed onto a single region in the reference genome; but they can also represent real heterozygosity in the case of recent admixture between diverged lineages. To distinguish the two possibilities, we compared IEH among the initial hard-filtered SNPs and 5,774,928 SNPs in high-confidence regions, defined as regions where at least 90 percent of the accessions have a read coverage within 1.5 standard deviations of their mean coverage. If a region has a coverage within this "normal" range, this indicates that it was not deleted or duplicated in the sequenced accession, and that heterozygous sites are thus not due to collapsed paralogs.

A median of 116 IEH were identified per accession in the hard-filtered set as compared to only two in the high-confidence set (Figure 1), indicating that most IEH are artefacts caused by structural variants. Six accessions (Uni2, Mig3, ABR9, Bd21-3, ABR7, Per1) deviate from the general pattern and have hundreds of IEH even in high confidence regions (Figure 1a). As the sequencing depth at these IEH is within the normal range and does not suggest the presence of structural variation (Figure S2), we propose that this is real heterozygosity left over from recent interbreeding between diverged genotypes. On average 46% of the IEH identified in any accession overlap with annotated genes, with a median of 50 genes per accession showing evidence of extended heterozygosity. As expected from the composition of the pan-genome (Gordon et al., 2017), the total set of 4,211 genes with IEH is enriched for the GO terms "defence response", "response to stress" and other functions associated with highly dynamic gene families.

3.3 | Three major lineages and five large geographic clades of *Brachypodium distachyon*

The different clustering methods capture a clear subdivision of the 196 accessions into three major lineages and five large geographic clades, except for two accessions from the Pyrenees which cannot be assigned (Arn1 and Mon3; Figure 2a). This subdivision is particularly clear in the sNMF analysis, where model fit improves in large steps until *K* = 5, after which it declines in smaller steps as regional and local populations are recovered (Figure 2b; Figure S3). At the base of the rooted *B. distachyon* phylogeny (Figure 2c,e; Figure S4; an interactive tree/map is provided on http://github.com/cstritt/bdis-phylogeo), 15 Italian accessions form a clade (hereafter called C_Italia) consisting of nine plants from the Gargano peninsula in south-eastern Italy and six collected across Italy, namely in Sicily, southern Campania and Liguria (Figure 2a). After the divergence of this basal lineage, the tree splits into two large lineages: one comprising all the remaining plants from Italy (A_Italia) and plants primarily from the coastal regions of Turkey (A_East), the other comprising plants from the western Mediterranean (B_West) and plants primarily from the Anatolian plateau and south-eastern Turkey/Iraq (B_East). The same subdivision is evident in the PCA (Figure 2d). A geographic outlier are three accessions from southern Turkey (2_20_16, 2_14_15 and 2_14_20; Figure 2a) which cluster with accessions from Spain and France. As these Turkish accessions were collected recently (Skalska et al., 2020) and not grown in contact with Spanish or French accessions, this finding can hardly be explained by mislabelling or seed contamination.

Using the multispecies coalescent approach implemented in BPP (Figure 2c), the split between the basal C lineage and the others is dated to 107 thousand years ago (kya, 95% highest posterior density interval [HPDI] 86–128 ky) and the split between the A and B lineages to 52 kya (95% HPDI 35–72 ky). Within the A lineage, the separation between A_East and A_Italia is dated to 18 kya (95% HPDI 8–27 ky), while within the B lineage the separation between B_East and B_West is estimated at 22 kya (95% HPDI 15–30 ky).

3.4 | Little evidence for historical gene flow between diverged lineages

At this coarse level of genetic structure, we thus observe an incongruence between genetic structure and geographic occurrence. Roughly half of the plants collected in Turkey and Iraq are more closely related to plants from the western Mediterranean than to other plants from the east. Similarly, in Italy the majority of the collected plants are more closely related to plants from the east than to the 15 Italian plants at the base of the phylogeny. In the east, the A and the B lineage might occur in the same locality, yet there is a clear overall geographic division (Figure 2a; http://github.com/cstritt/bdis-phylogeo; see also Skalska et al., 2020): lineage A is predominant in the coastal region, lineage B on the inland Anatolian plateau and in the Fertile Crescent. In Italy the occurrence of the A and the

C lineage is less clear-cut: in two of the four localities where lineage C was found, plants of the A and the C lineage were present on the same meadow and thus sympatric in a narrow sense (Figure S5).

Estimates of shared ancestry show little evidence for recent gene flow between co-occurring lineages in Turkey and Italy (Figure 2a,e). Historical gene flow between geographic clades appears possible, yet the evidence is not particularly strong: a tree-like model without migration edges explains 99.7% of the covariance of allele frequencies between geographic clades (Figure S6), suggesting that shared ancestry apparent in Figure 2 is due to common descent and shared genetic drift rather than admixture. Adding a migration edge leading from A_East to B_East leads to an increase to 99.98%, suggesting that, if historical gene flow has occurred, it has not left strong signatures.

3.5 | High selfing rates and a strong effect of subdivision on linkage disequilibrium

Selfing rates often vary within species (Barrett, 2003; Griffin & Willi, 2014), yet studies with *B. distachyon* have unanimously reported a near absence of heterozygosity (Marques et al., 2017; Vogel et al., 2009). To test whether this pattern holds across geographic regions, we estimated inbreeding coefficients and selfing rates for the 196 accessions. Using SNPs in high-confidence regions, as defined above, we estimated a median inbreeding coefficient *F* of 0.996 among the 196 accessions (mad = 0.003, Table S1), which translates to a median selfing rate of 0.998 using $s = 2F/1 + F$. When population structure is taken into account and F is estimated within geographic clades, median *F* is 0.991 and thus only marginally lower than at the species level. As expected, lower *F* values are estimated for the six accessions whose genomes contain numerous IEH. Overall, our results show that extreme levels of homozygosity and high selfing rates predominate throughout the species' range and vary little between subpopulations.

LD is low at the species level (Figure 3a) and successively increases as one considers finer scales of subdivision (Figure 3b–d). A comparison of LD decay and the topology of the *B. distachyon* phylogeny support the expectation that LD decay reflects substructure rather than recombination. Using the sum of terminal branch lengths in subtrees as a measure of how much independent accumulation of mutations has occurred at different levels, we observe a strong negative relation between mean r^2 at a distance of 100 kb and the sum of terminal branch lengths (Figure S7): LD decay is slower in subtrees containing either few leaves, for example, in the C lineage, or containing short terminal branches, as in B_East, where there is a surprising number of genetically highly similar accessions.

3.6 | Strong isolation by distance within the three major lineages

To understand how genetic diversity is partitioned across space and the different levels of genetic structure in *B. distachyon*, we
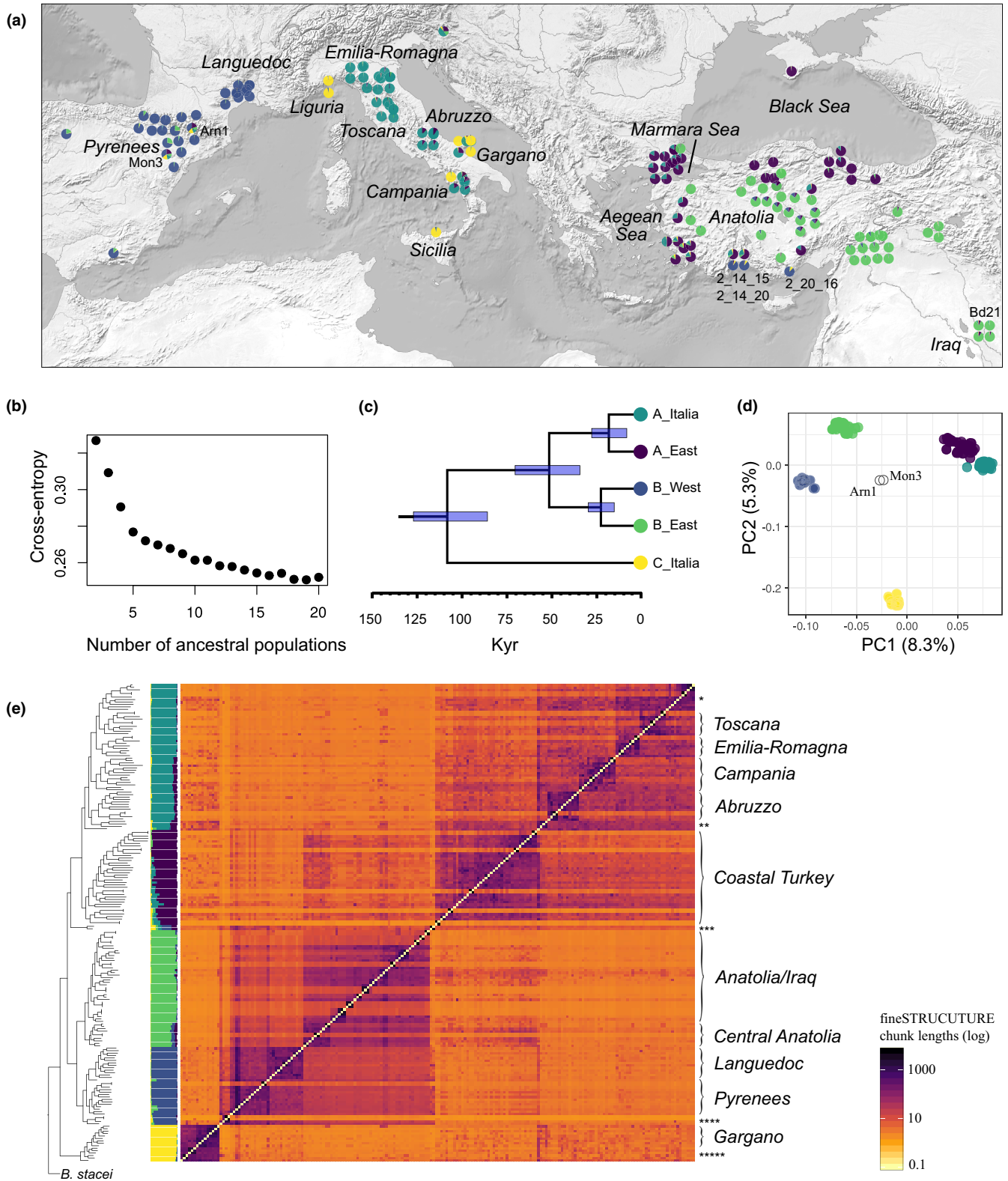
FIGURE 2 Geographic origin and genetic structure. (a) Map and pie charts showing shared ancestry. Only a subset of accessions is displayed for the dense local samples in Italy and France. (b) Cross-entropy for increasing K in the ancestry analysis with sNMF. (c) Rooted phylogeny with divergence times, estimated under a multispecies coalescent model. (d) Principal component analysis. (e) Rooted phylogeny and coancestry matrix estimated with FINESTRUCTURE, as well as sNMF barplot for *K* = 5. The FINESTRUCTURE chunk length is the recombination map distance donated to individuals in rows from individuals in columns (log scale). Names in italics highlight regional populations, asterisks indicate small clusters: * Accessions from northern Italy, including a cluster of five accessions from Taro valley in Emilia-Romagna. ** ABR9 collected in Albania and three accessions from the Gargano peninsula. *** Arn1 and Mon3. **** Three Turkish accessions clustering with Bd30-1 from southern Spain. ***** Accessions from Sicily, Campania and Liguria. An interactive phylogenetic tree/map can be explored on http://github.com/cstritt/bdis-phylogeo

FIGURE 3 Decay of linkage disequilibrium at different scales of population structure: (a) species-wide, (b) in the three major genetic lineages, (c) in four geographic clades (C_Italia is not included as it is the sole representative of lineage C and thus already included at the lineage level), (d) four regional populations in Italy
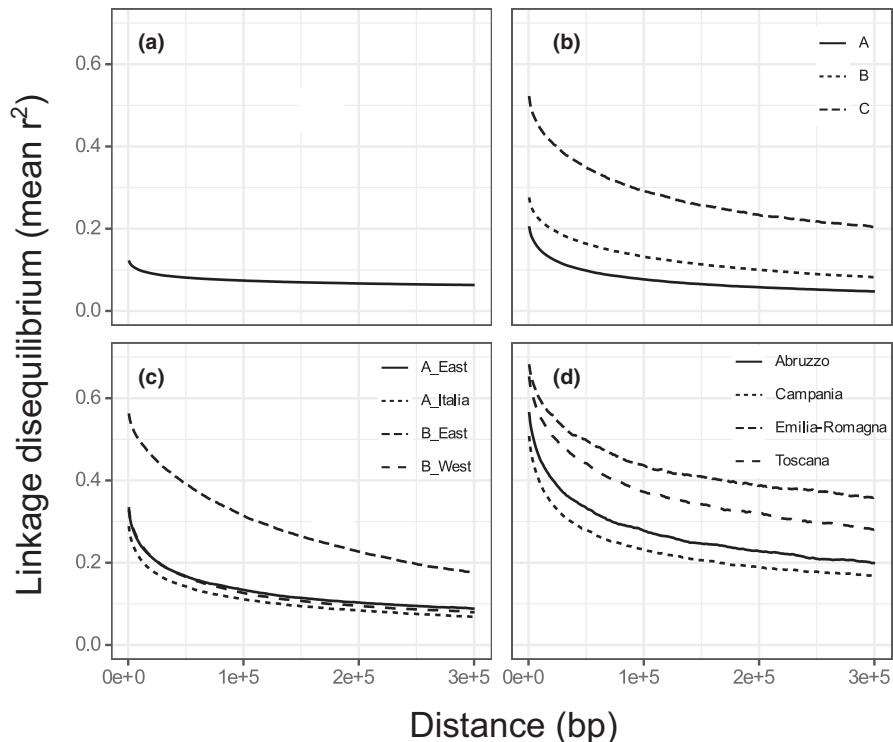

FIGURE 4 Genetic diversity and differentiation. (a) Pairwise genetic differences (d_{XY}) within and between the five geographic clades. One panel is shown for each clade, the colours within the panel showing with which clade pairwise comparisons are made. (b) Correlation between geographic and genetic distance within the three major genetic lineages A, B and C. Colours within the panels indicate between which geographic regions pairwise comparisons are made. Correlation coefficients from Mantel test and their significance are displayed
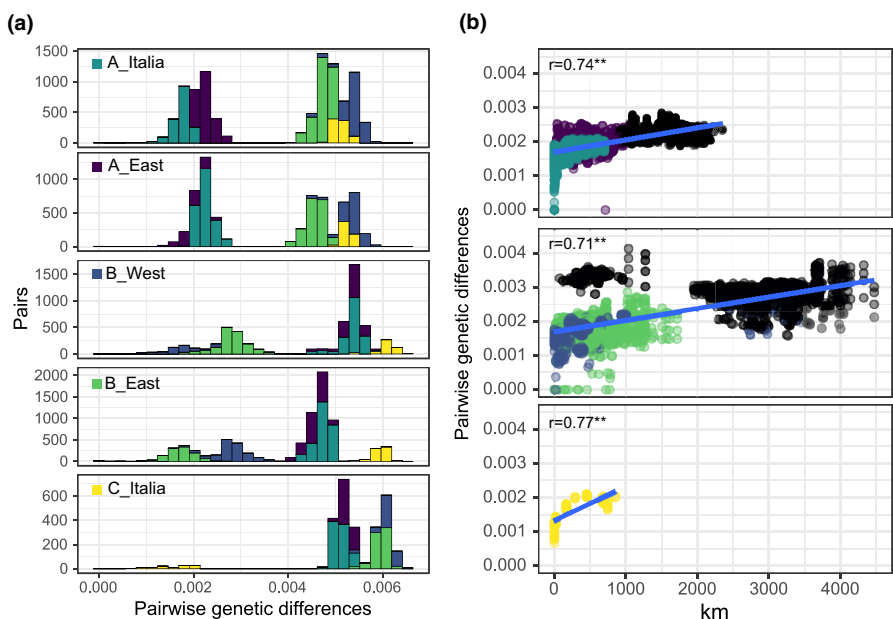
estimated pairwise genetic differences (d_{XY}) at synonymous sites. Within the five geographic clades, median d_{XY} is 0.0018, that is, 1.8 different sites per 1,000 bp (Figure 4a). Values differ somewhat among clades: d_{XY} is higher in A_East (0.0019) than in the other lineages ($p < 2 \times 10^{-16}$, Kruskal-Wallis), and lower in C_Italia (0.0017) than in the others ($p < .03$). Levels of diversity in the B_East and B_West clades are similar (0.0017, $p = .37$). Between clades, d_{XY} reaches up to 0.0065 in comparisons between C_Italia and B_West.

Within all three major lineages, d_{XY} increases with geographic distance, as shown by significant Mantel correlations (Figure 4b). While

correlation coefficients are similar between lineages ($r_A = 0.75$, $r_B = 0.71$, $r_C = 0.77$), some deviations are observed in the B lineage. On the one hand, there are genetically similar accessions with distant geographic origins, particularly in the B_East clade, in agreement with the previously observed poor mapping between geography and phylogeny in this clade. On the other hand, genetically distant individuals can be geographically close (upper left of the middle panel in Figure 4b). These are comparisons of Turkish accessions with the three accessions collected in Turkey clustering genetically with B_West accessions (Figure 2a).

3.7 | The hierarchical nature of population subdivision

Population subdivision in *B. distachyon* is hierarchical and does not end at level of the five geographic clades. Substructure is evident for increasing values of K in the sNMF analysis (Figure S8), in within-lineage PCAs (Figure S9), as subtrees in the phylogeny (Figure S4), and most clearly in the FINESTRUCTURE analysis (Figure 2e). The clustering algorithm implemented in FINESTRUCTURE identifies 59 populations with a median size of two (range 1–13). This inflated number of populations most likely reflects the limitations of formal approaches to identify the "most realistic" number of populations in a highly selfing organism, where the homogenizing force of gene flow is limited and populations are better understood as phylogenetic clades than as panmictic units.

In *B. distachyon*, many subtrees cluster geographically into regional populations (Figure 2e; http://github.com/cstritt/bdis-phylogeo). In the A_Italia clade, populations from Campania, Abruzzo and Gargano in the South and from Toscana and Emilia-Romagna in the North can be distinguished. Also in the other lineage present in Italy, C_Italia, the 15 samples cluster according to their geographic origin. In the western Mediterranean, the Pyrenees separate a Spanish from a French population comprising all the newly collected accessions from the Languedoc region. Substructure is less obvious in the east. A_East is a diffuse clade occupying primarily coastal areas in Turkey. In the B_East clade, ten accessions from central Anatolia form a well-supported group at the base of the clade. The rest of the samples cover a large area from Baghdad to the Marmara Sea region and do not show an obvious mapping between phylogeny and geography.

3.8 | Local and regional patterns of genetic diversity in Italy

The local and regional sampling in Italy allows us to take a closer look at how genetic diversity is distributed across small spatial scales. In Italy, the Apennines separate *B. distachyon* into genetically differentiated regional populations (Figure 5a). We here focus on four regional populations belonging to the A_Italia clade for which the largest number of samples is available (7–14). A TreeMix analysis with these populations suggests that no gene flow has occurred across the Appenines, as a tree-like model without migration edges explains 99.99% of the covariance of allele frequencies between populations (Figure S10).

Figure 5b shows pairwise genetic differences in the four populations, for which samples were collected few dozen meters to few kilometres apart. Within populations, some more closely related individuals stand out ($d_{XY} < 0.001$); these were typically collected within 1 km of each other. Most plants, however, even if growing close-by, are genetically quite distinct and can be more different from each other than from accessions kilometres away. Isolation by distance is weak at this scale and, while still significant in the two southern populations, is not significant in the Toscana and Emilia-Romagna populations. Genetic diversity declines step-wise from south to north, with a median d_{XY} of 0.0016 in Campania, 0.0015 in Abruzzo, 0.0014 in Toscana, and 0.0013 in Emilia-Romagna (Figure 5b).

3.9 | Flowering phenology in the greenhouse and outdoors

Previous studies described the co-occurrence of the A_East and B_East clades in Turkey and interpreted the lack of gene flow between these lineages as a result of variation in flowering time: accessions from the A_East clade require long vernalization and show a delay in flowering under greenhouse conditions (Gordon et al., 2017; Tyler et al., 2016).

Our greenhouse experiment replicated the previously identified flowering behaviour of the two B_East accessions Bd3-1 and BdTR3C and the two A_East accessions Bd1-1 and BdTR7a, which in this order required four, six, eight, and more than 12 weeks to satisfy the vernalization response (Figure S11a, Table S2). With the exception of two C_Italia accessions which required 6 weeks of cold, all other 10 newly tested Italian accessions required long periods (>10 weeks) of cold followed by up to 40 days in warm conditions to flower (Figure S11a, Table S2).

Flowering behaviour in the outdoor experiment was quite different (Figure S11b, Table S2). All plants emerged in autumn and flowered next spring within 2 weeks. However, clades still differ in their flowering behaviour. C_Italia and B_East accessions tended to produce flowers more rapidly than A_East accessions, though this difference is not statistically significant at the 5% level. A_Italia and B_West, on the other hand, take longer to flower than the other clades (LME, $p < .006$ for all comparisons between these two clades and the others), even though overall only 2 weeks separate the early from late flowering accessions.

4 | DISCUSSION

Brachypodium distachyon was initially developed as a model to study gene function in grasses, with an emphasis on applications in crop breeding. With this study we extend the genomic resources for this Mediterranean grass and prepare the ground for further evolutionary studies by describing the spatial and temporal scale of population subdivision in the species. Population structure in *B. distachyon* is strongly hierarchical. It originated in the Upper Pleistocene (c. 129,000 to 11,700 years ago) and reflects, as we will argue in the following, the combined effect of high selfing rates and effective seed dispersal. At the continental scale, dispersal and selfing help to explain why highly diverged lineages might be present in the same locality without signs of admixture. At the local scale, selfing and dispersal interact and maintain high local genetic diversity, with important implications for how natural selection is expected to act in this species.
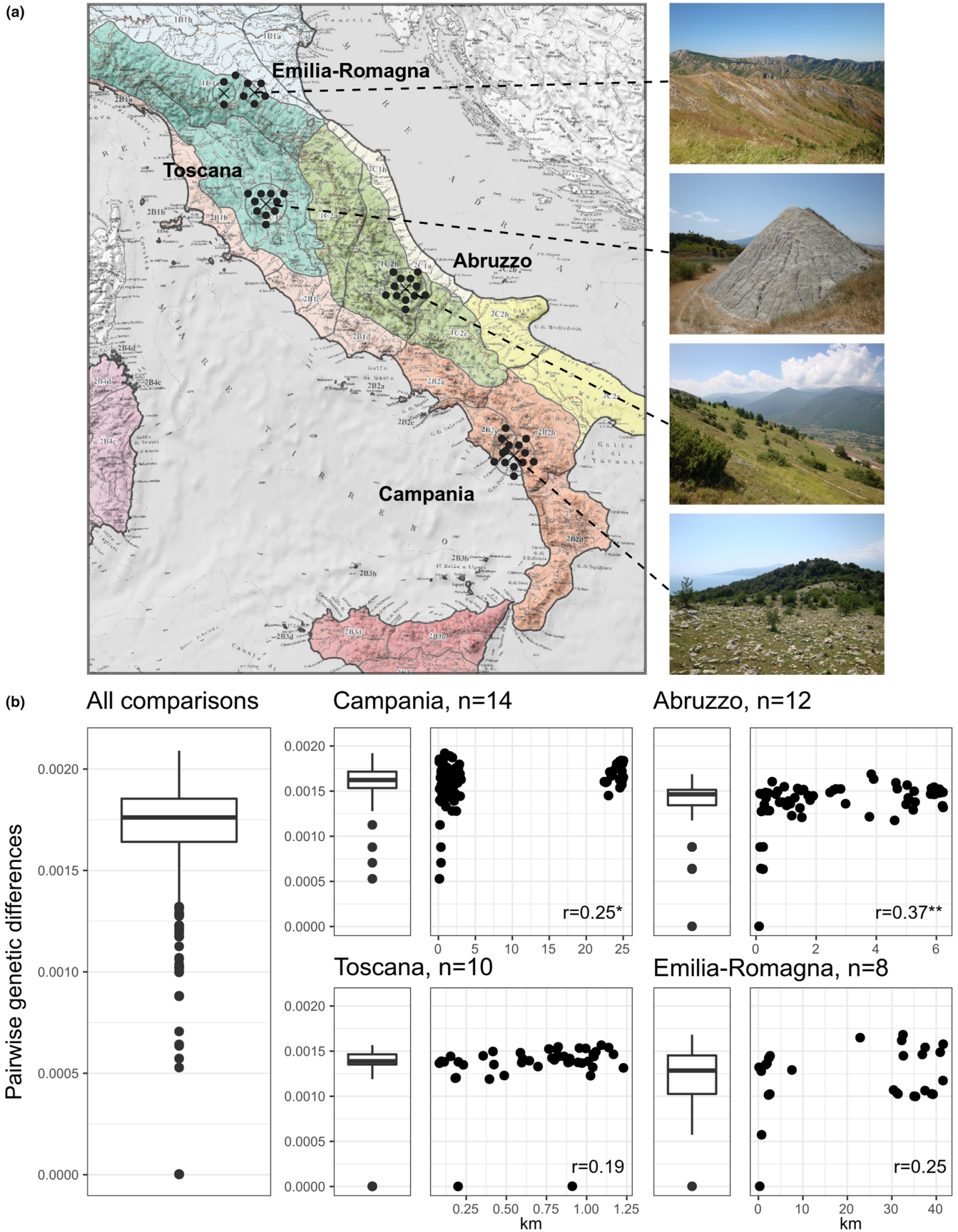
FIGURE 5 Genetic differentiation across ecoregions in Italy. (a) Map showing the ecoregion classification of Blasi et al. (2014) and *Brachypodium distachyon* habitats in the four sampled regional populations. (b) Genetic diversity within the A_Italia clade, for which the most extensive sampling is available. Boxplots show the distribution of d_{XY} across the whole clade and within regional populations, while dotplots show how d_{XY} relates to geographic distance

4.1 | Extreme selfing and extended heterozygosity as a result of recent outcrossing

In plants, the wide occurrence of sexual reproduction and the complex adaptations to ensure it suggest that self-fertilization is disadvantageous under a wide range of conditions (Barrett, 2010). The best known exception are colonizing plants (Baker, 1955; Stebbins, 1957): here, the advantage of reproductive assurance during colonization can outweigh the costs of inbreeding. *Brachypodium distachyon* clearly is a colonizing plant: it thrives in disturbed areas like pastures or along paths and declines when stable conditions allow the establishment of perennial competitors (San Miguel, 2008).

Considering this life style, as well as the highly reduced cleistogamous flowers (Thi et al., 2016; Khan et al., 1999), the highly selfing nature of *B. distachyon* comes at no surprise. Nevertheless, the degree of homozygosity revealed by molecular markers is astonishing. In a large-scale survey of 187 accessions from Turkey with 43 microsatellite markers, only four heterozygous sites in four individuals were detected (Vogel et al., 2009). Similarly, Marques et al. (2017), looking at 137 plants in 14 populations with 10 markers, estimated inbreeding coefficients of one in all except one population.

Our study agrees with this finding and suggests that high selfing rates are prevalent throughout the species' range. It also corrects previous estimates of heterozygosity based on genome-wide SNPs, where we reported a median of 7.8 and 6.3% of heterozygous SNPs in the B_East and B_West clades (Bourgeois et al., 2018). These estimates were most likely too high as we did not take into account artificial heterozygosity due to structural variation. Indeed, distinguishing real from artificial heterozygosity proved to be challenging. As we show here, a method based on the clustering of heterozygous sites and sequencing depth can be used to identify artificial heterozygosity arising from deletions and duplications. As plant population genomics is extending to more complex and dynamic genomes, searching for islands of extended heterozygosity can be a simple first step to identify problems with heterozygous sites and to limit biases in downstream analyses.

While self-fertilization appears to be the norm in *B. distachyon*, we also find evidence for recent outcrossing. A large part of the heterozygosity we observed occurred in only six accessions, in the form of extended islands of heterozygosity. Extended heterozygosity can result from the purging of deleterious homozygosity, as observed in inbred maize (Brandenburg et al., 2017), or from balancing selection (Nordborg et al., 1996). These processes, however, imply a selective advantage of heterozygosity and would hardly create hundreds of IEH in single accessions, as observed here, but rather few IEH present at higher frequencies. Our results thus provide evidence that the IEH in the six accession are due to the admixture of diverged genotypes. As heterozygosity disappears within a few generations of selfing (Hedrick, 2011), the outcrossing events leading to the six heterozygous accessions must be recent and might have occurred during propagation of the accessions for research and seed distribution.

4.2 | Shared ancestry reflects common descent rather than admixture

Historical gene flow can be inferred when the covariance of allele frequencies between populations is poorly explained by tree-like evolution and genetic drift (Pickrell & Pritchard, 2012). Our tests for gene flow between the five geographic clades and between regional populations in Italy suggest that patterns of shared ancestry can largely be explained by tree-like descent rather than historical admixture. These results contrast with previous studies which have interpreted shared ancestry as admixture (Gordon et al., 2017; Stritt et al., 2018) and inferred frequent introgressions from incongruences between plastome and nuclear phylogenies (Sancho et al., 2018). In the latter study, several B_East accessions are shown to have B_West plastome genotypes, but no admixture is evident at the nuclear level, which the authors explain through repeated backcrossing and loss of the B_West nuclear ancestry. Here too, however, common descent and incomplete lineage sorting of plastome haplotypes are a plausible alternative explanation, as for example described for *Hordeum* (Jakob & Blattner, 2006).

While outcrossing and historical gene flow remain a possibility, our results suggest that alternative explanations for shared ancestry, more compatible with the extreme homozygosity and high selfing rates observed throughout *B. distachyon*, should be considered in future studies.

4.3 | A scenario of separate migration corridors

To our current knowledge, *B. distachyon* in the Mediterranean region comprises three highly diverged genetic lineages which can be further divided into geographic clades and regional populations. This basic tripartition was described previously in a genotyping-by-sequencing study (Tyler et al., 2016), yet a phylogeographic scenario to account for the distribution of the three lineages is now only starting to emerge as samples from the entire species range are being collected (Gordon et al., 2020; Skalska et al., 2020). Here we show that the A lineage is present throughout Italy, and we describe a so far poorly known lineage of *B. distachyon*, C_Italia.

The C lineage diverged from the others 86–128 kya and forms the basal branch in the phylogeny. Previous studies included single samples of this lineage, all originating from Sicily (Gordon et al., 2020; Tyler et al., 2016). We found this lineage to be present yet rare throughout Italy, in two instances growing on the very same meadow as plants belonging to lineage A. Recently a similar situation has been described for *B. hybridum*, where individuals with a common ancestor more than one million years ago coexist in the same place (Gordon et al., 2020; Shiposha et al., 2020). *Brachypodium* thus provides striking examples of cryptic grass diversity in the Mediterranean region which would be difficult to appreciate without the help of molecular markers.

The geographic co-occurrence of genetically diverged lineages in *B. distachyon* suggests that three lineages of *B. distachyon* have

expanded independently in the past, and that today, where their ranges overlap, extreme selfing and possibly other reproductive barriers prevent interbreeding. While the biogeography of the C lineage remains vague because of the low number of samples, an intriguing scenario which could explain the disjunct genetic structure in *B. distachyon* is that lineages A and B used different migration corridors: a North African corridor for lineage B, and a European corridor for lineage A. To our knowledge, all samples collected in the Balkans so far belong to the A or the C lineage (Tyler et al., 2016; Gordon et al., 2020), suggesting that the B lineage is absent or rare between the Ligurian Alps and inland Turkey. With additional samples from the Balkans, the Middle East, southern Spain and Africa, some of which have already been sequenced, this scenario should be testable in the near future.

Our molecular dating analysis and the strong isolation by distance observed at the lineage level suggest that the three lineage expansions in *B. distachyon* are not recent and might predate the end of the Last Glacial Period c.11,700 years ago. The timescale we infer in this study, with a time to the most recent common ancestor of 86–128 thousand years ago, hinges on an assumed mutation rate of 7×10^{-9}. This estimate was obtained in an experimental evolution study with *A. thaliana* (Ossowski et al., 2010); it is at the fast end of the spectrum as it reflects the accumulation of mutations before purifying selection had a chance to remove them (Ho et al., 2011). Our estimates are therefore more likely too recent than too old. A similar estimate, 162 ky to the common ancestor of the three lineages of *B. distachyon*, was recently obtained by time-calibrating an internal node rather than assuming a mutation rate (Gordon et al., 2020), providing further support for the timescale inferred here.

4.4 | Genetic structure and flowering time differences

Previous studies have demoted geography to "a secondary factor in shaping population structure" (Gordon et al., 2017) and instead identified flowering time differences as "a main factor driving rapid intraspecific divergence" (Sancho et al., 2018; see also Tyler et al., 2016). This view is problematic as it is based on flowering phenotypes observed in the greenhouse. Flowering is a plastic trait and depends on environmental conditions (Gaudinier & Blackman, 2020) and the timing of the other major life history event in annual plants, germination (Martínez-Berdeja et al., 2020). In our outdoors experiment, all accessions produced flowers within 2 weeks when they went through prolonged vernalization during winter. This experiment, performed in a place outside the native distribution of *B. distachyon* although with climatic conditions similar to those of cold climate native sites of *B. distachyon* ecotypes, does obviously not reflect all the environmental variation present in natural habitats. Yet it suggests that differences observed in the greenhouse might be exaggerated through artificially short vernalization times.

While the role of flowering in reproductive isolation is less clear than suggested by previous studies, flowering remains an intriguing trait in *B. distachyon*. A geographic pattern of flowering time differences is observed in Turkey but not in Italy. In Turkey, the A lineage, which requires vernalization to flower, is present along the coast, while the B lineage, which requires little to no vernalization, is mainly found in inland Anatolia (Figure 2; Skalska et al., 2020). Skalska et al. (2020) recently showed that the two lineages also differ in drought tolerance, indicating that the B_East accessions are adapted to the arid climate of central Anatolia. In Italy, no such geographic differences in vernalization requirement and flowering are observed, neither among lineages nor among regional populations. Possibly this reflects the stronger ecological contrasts in Turkey, where rain is scarce on the inland Anatolian plateau even in winter, while the coastal regions have the typical Mediterranean winter rainfalls (Sensoy, 2008). Investigating the molecular evolution of flowering genes in *B. distachyon* will allow to further characterize the putative adaptive role of flowering time in this species.

4.5 | Local consequences of selfing and dispersal

Within regional populations in Italy, isolation by distance is weak, and sampling within few hundred meters can yield individuals as different from each other as individuals kilometres apart (Figure 5). This pattern is discernible in the distribution of pairwise genetic differences and as long terminal branch lengths in the phylogeny, especially in the A lineage. Clearly the plants collected locally did not descend from single recent founder events, which would result in high local similarity. Our study thus supports the hypothesis that the interplay of high selfing and seed dispersal rates shapes local and regional genetic diversity, as proposed by Vogel et al. (2009) for populations in Turkey.

In population genetics textbooks, migration, gene flow, and interbreeding are often used synonymously (e.g., Charlesworth & Charlesworth, 2012), reflecting the dominance of animal models in the field. In plants, however, migration can occur through pollen and seed, and the migration of seeds need not imply the exchange of genetic material. With a selfing rate close to one, lineages can evolve independently of each other for a considerable period without the homogenizing effect of interbreeding, and lineage-specific mutations may accumulate. Seed dispersal could then break up neighbourhoods of close relatives and shuffle genotypes across the landscape.

Central to the selfing-dispersal hypothesis is that *B. distachyon* can disperse its seeds easily and over long distances. Morphological adaptations to do so are present: the species has adherent seeds which may facilitate zoochory. Surprisingly, however, so far no study has mentioned the two dispersal vectors which could accomplish genotype shuffling at a large scale and frequency: sheep and goats. From our fieldwork and previous reports (Vogel et al., 2009) it is clear that a habitat in which *B. distachyon* thrives are arid, oligotrophic pastures grazed by sheep and goats. Not only do these animals keep away more competitive perennial plants, but they are

also known to carry plant seeds over long distances and to be an important vector of long-distance seed dispersal (Manzano & Malo, 2006). In the Mediterranean landscape, which has been shaped by human activity for thousands of years (Grove & Rackham, 2001), the movement of domestic animals might thus be a crucial factor in the population dynamics of *B. distachyon* and similar plants.

Given the context dependence and stochasticity of seed dispersal (Beckman et al., 2020), the extent of genotype shuffling is likely to vary in space and time. The poor mapping between geography and phylogeny in central Anatolia suggests elevated dispersal rates in this region. On the other hand, the highly homogenous populations in Spain described by Marques et al. (2017) document local neighbourhoods largely unaffected by the immigration of diverged genotypes. In this study a minimum sampling distance of ten meters was employed; here genetic structure might reflect the movement of small-scale dispersal vectors like ants (Catalán et al., 2016b). Formal studies on the seed dispersal ecology of *B. distachyon* would be valuable to better understand variation in dispersal patterns and their implications for sampling design.

The possibility that genetic diversity is strongly influenced by long-distance seed dispersal has important implications for the study of natural selection in *B. distachyon*. While selfing itself decreases the fixation time of adaptive mutations and might favour local adaptation (Glémin & Ronfort, 2013; Hartfield et al., 2017), selfing combined with high rates of seed migration is expected to favour phenotypic plasticity over local adaptation (Sultan & Spencer, 2002). It is notable that no instance of a single lineage rising to high local frequencies is observed. Instead, it appears that seed dispersal maintains high local genetic diversity and might compensate for what is considered a main disadvantage of selfing: low diversity and failure to adapt to spatially and temporally heterogeneous environments (Wright et al., 2013).

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

DATA AVAILABILITY STATEMENT

Illumina paired-end sequencing data generated for this project are available on the European Nucleotide Archive, project number PRJEB40344. A nexus file with the rooted phylogeny and a vcf containing the 5,774,928 hard-filtered, high-confidence SNPs are available on Dryad, https://doi.org/10.5061/dryad.0cfxpnw15.

ORCID

Christoph Stritt https://orcid.org/0000-0002-3167-6658
Michele Wyler https://orcid.org/0000-0003-1097-5322
Luis A. J. Mur https://orcid.org/0000-0002-0961-9817
Anne C. Roulin https://orcid.org/0000-0002-6668-3321

REFERENCES

Baker, H. G. (1955). Self-compatibility and establishment after "long-distance" dispersal. *Evolution*, *9*(3), 347–349. https://doi.org/10.2307/2405656

Barrett, S. C. H. (2003). Mating strategies in flowering plants: the outcrossing–selfing paradigm and beyond. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1434), 991–1004. https://doi.org/10.1098/rstb.2003.1301

Barrett, S. C. H. (2010). Darwin's legacy: The forms, function and sexual diversity of flowers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1539), 351–368. https://doi.org/10.1098/rstb.2009.0212

Beckman, N. G., Aslan, C. E., Rogers, H. S., Kogan, O., Bronstein, J. L., Bullock, J. M., Hartig, F., HilleRisLambers, J., Zhou, Y., Zurell, D., Brodie, J. F., Bruna, E. M., Cantrell, R. S., Decker, R. R., Efiom, E., Fricke, E. C., Gurski, K., Hastings, A., Johnson, J. S., ... Zambrano, J. (2020). Advancing an interdisciplinary framework to study seed dispersal ecology. *AoB PLANTS*, *12*(2), 1–18. https://doi.org/10.1093/aobpla/plz048

Blasi, C., Capotorti, G., Copiz, R., Guida, D., Mollo, B., Smiraglia, D., & Zavattero, L. (2014). Classification and mapping of the ecoregions of Italy. *Plant Biosystems*, *148*(6), 1255–1345. https://doi.org/10.1080/11263504.2014.985756

Bourgeois, Y., Stritt, C., Walser, J.-C., Gordon, S. P., Vogel, J. P., & Roulin, A. C. (2018). Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass *Brachypodium distachyon*. *The Plant Journal*, *96*(2), 438–451. https://doi.org/10.1111/tpj.14042

Brandenburg, J.-T., Mary-Huard, T., Rigaill, G., Hearne, S. J., Corti, H., Joets, J., Vitte, C., Charcosset, A., Nicolas, S. D., & Tenaillon, M. I. (2017). Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS Genetics*, *13*(3), 1–30. https://doi.org/10.1371/journal.pgen.1006666

Catalán, P., López-Álvarez, D., Bellosta, C., & Villar, L. (2016). Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei*, and *B. hybridum* (Poaceae). *Anales Del Jardin Botanico De Madrid*, 73(1), 1–14. https://doi.org/10.3989/ajbm.2428

Catalán, P., López-Álvarez, D., Díaz-Pérez, A., Sancho, R., & López-Herránz, M. L. (2016b). Phylogeny and evolution of the genus *Brachypodium*. In *Genetics and genomics of* Brachypodium (pp. 9–38). Springer.

Catalán, P., Müller, J., Hasterok, R., Jenkins, G., Mur, L. A. J., Langdon, T., Betekhtin, A., Siwinska, D., Pimentel, M., & López-Alvarez, D. (2012). Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany*, 109(2), 385–405. https://doi.org/10.1093/aob/mcr294

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205. https://doi.org/10.1038/nrg2526

Charlesworth, B., & Charlesworth, D. (2012). *Elements of evolutionary genetics*. Roberts and Company Publishers.

Charlesworth, D. (2003). Effects of inbreeding on the genetic diversity of populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1434), 1051–1070. https://doi.org/10.1098/rstb.2003.1296

Dell'Acqua, M., Zuccolo, A., Tuna, M., Gianfranceschi, L., & Pè, M. (2014). Targeting environmental adaptation in the monocot model *Brachypodium distachyon*: A multi-faceted approach. *BMC Genomics*, 15(1), 801. https://doi.org/10.1186/1471-2164-15-801

Dinh Thi, V. H., Coriton, O., Le Clainche, I., Arnaud, D., Gordon, S. P., Linc, G., Catalan, P., Hasterok, R., Vogel, J. P., Jahier, J., & Chalhoub, B. (2016). Recreating stable *Brachypodium hybridum* allotetraploids by uniting the divergent genomes of *B. distachyon* and *B. stacei*. *PLoS One*, 11(12), 1–27. https://doi.org/10.1371/journal.pone.0167171

Dray, S., & Dufour, A.-B.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20. https://doi.org/10.18637/jss.v022.i04

Engelmann, K., & Purugganan, M. (2006). The molecular evolutionary ecology of plant development: flowering time in *Arabidopsis thaliana*. *Advances in Botanical Research*, 44(06), 507–526. https://doi.org/10.1016/S0065-2296(06)44013-1

Flouri, T., Jiao, X., Rannala, B., & Yang, Z. (2018). Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Molecular Biology and Evolution*, 35(10), 2585–2593. https://doi.org/10.1093/molbev/msy147

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Gaudinier, A., & Blackman, B. K. (2020). Evolutionary processes from the perspective of flowering time diversity. *New Phytologist*, 225(5), 1883–1898. https://doi.org/10.1111/nph.16205

Gibson, D. J. (2009). *Grasses & grassland ecology*. Oxford University Press.

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

Glémin, S., & Ronfort, J. (2013). Adaptation and maladaptation in selfing and outcrossing species: New mutations versus standing variation. *Evolution*, 67(1), 225–240. https://doi.org/10.1111/j.1558-5646.2012.01778.x

Gordon, S. P., Contreras-Moreira, B., Levy, J. J., Djamei, A., Czedik-Eysenberg, A., Tartaglio, V. S., Session, A., Martin, J., Cartwright, A., Katz, A., Singan, V. R., Goltsman, E., Barry, K., Dinh-Thi, V. H., Chalhoub, B., Diaz-Perez, A., Sancho, R., Lusinska, J., Wolny, E., … Vogel, J. P. (2020). Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nature Communications*, 11(1), 1–16. https://doi.org/10.1038/s41467-020-17302-5

Gordon, S. P., Contreras-Moreira, B., Woods, D. P., Des Marais, D. L., Burgess, D., Shu, S., Stritt, C., Roulin, A. C., Schackwitz, W., Tyler, L., Martin, J., Lipzen, A., Dochy, N., Phillips, J., Barry, K., Geuten, K., Budak, H., Juenger, T. E., Amasino, R., … Vogel, J. P. (2017). Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure. *Nature Communications*, 8(1). https://doi.org/10.1038/s41467-017-02292-8

Griffin, P. C., & Willi, Y. (2014). Evolutionary shifts to self-fertilisation restricted to geographic range margins in North American *Arabidopsis lyrata*. *Ecology Letters*, 17(4), 484–490. https://doi.org/10.1111/ele.12248 Eahead of print.

Grove, A., & Rackham, O. (2001). *The nature of mediterranean Europe - An ecological history*. Yale University Press.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1–22. https://doi.org/10.18637/jss.v033.i02

Hartfield, M., Bataillon, T., & Glémin, S. (2017). The evolutionary interplay between adaptation and self-fertilization. *Trends in Genetics*, 33(6), 420–431. https://doi.org/10.1016/j.tig.2017.04.002

Hartfield, M., Wright, S. I., & Agrawal, A. F. (2018). Coalescence and linkage disequilibrium in facultatively sexual diploids. *Genetics*, 210(2), 683–701. https://doi.org/10.1534/genetics.118.301244

Hedrick, P. (2011). *Genetics of populations*. Jones & Bartlett Publishers.

Hey, J., & Machado, C. A. (2003). The study of structured populations–new hope for a difficult and divided science. *Nature Reviews Genetics*, 4(7), 535–543. https://doi.org/10.1038/nrg1112

Ho, S. Y. W. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., & Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology*, 20(15), 3087–3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x

Huo, N., Garvin, D. F., You, F. M., McMahon, S., Luo, M.-C., Gu, Y. Q., Lazo, G. R., & Vogel, J. P. (2011). Comparison of a high-density genetic linkage map to genome features in the model grass *Brachypodium distachyon*. *Theoretical and Applied Genetics*, 123(3), 455–464. https://doi.org/10.1007/s00122-011-1598-4

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

Jakob, S. S., & Blattner, F. R. (2006). A chloroplast genealogy of Hordeum (Poaceae): Long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Molecular Biology and Evolution*, 23(8), 1602–1612. https://doi.org/10.1093/molbev/msl018

Khan, M. A., & Stace, C. A. (1999). Breeding relationships in the genus *Brachypodium* (Poaceae: Pooideae). *Nordic Journal of Botany*, 19(3), 257–269. https://doi.org/10.1111/j.1756-1051.1999.tb01108.x

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics*, 8(1), 11–17. https://doi.org/10.1371/journal.pgen.1002453

Lee, T.-H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, 15(1), 162. https://doi.org/10.1186/1471-2164-15-162

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv*. https://doi.org/10.1093/bioinformatics/btw748

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., & Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Research*, 19(6), 1124–1132. https://doi.org/10.1101/gr.088013.108

Linhart, Y. B., & Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, 27(1), 237–277. https://doi.org/10.1146/annurev.ecolsys.27.1.237

López-Alvarez, D., López-Herranz, M. L., Betekhtin, A., & Catalán, P. (2012). A DNA barcoding method to discriminate between the

model plant *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). *PLoS One*, *7*(12), e51058. https://doi.org/10.1371/journal.pone.0051058

Lopez-Alvarez, D., Manzaneda, A. J., Rey, P. J., Giraldo, P., Benavente, E., Allainguillaume, J., Mur, L., Caicedo, A. L., Hazen, S. P., Breiman, A., Ezrati, S., & Catalan, P. (2015). Environmental niche variation and evolutionary diversification of the Brachypodium distachyon grass complex species in their native circum-Mediterranean range. *American Journal of Botany*, *102*(7), 1073–1088. https://doi.org/10.3732/ajb.1500128

Loveless, M. D., & Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, *15*, 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433

Manzano, P., & Malo, J. E. (2006). Extreme long-distance seed dispersal via sheep. *Frontiers in Ecology and the Environment*, *4*(5), 244–248. https://doi.org/10.1890/1540-9295(2006)004[0244:ELSDVS]2.0.CO;2

Marques, I., Shiposha, V., López-Alvarez, D., Manzaneda, A. J., Hernandez, P., Olonova, M., & Catalán, P. (2017). Environmental isolation explains Iberian genetic diversity in the highly homozygous model grass *Brachypodium distachyon*. *BMC Evolutionary Biology*, *17*(1), 139. https://doi.org/10.1186/s12862-017-0996-x

Martínez-Berdeja, A., Stitzer, M. C., Taylor, M. A., Okada, M., Ezcurra, E., Runcie, D. E., & Schmitt, J. (2020). Functional variants of DOG1 control seed chilling responses and variation in seasonal life-history strategies in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(5), 2526–2534. https://doi.org/10.1073/pnas.1912451117

Morgulis, A., Gertz, E. M., Schäffer, A. A., & Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *Journal of Computational Biology*, *13*(5), 1028–1040. https://doi.org/10.1089/cmb.2006.13.1028

Nieto Feliner, G. (2014). Patterns and processes in plant phylogeography in the Mediterranean Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics*, *16*(5), 265–278. https://doi.org/10.1016/j.ppees.2014.07.002

Nordborg, M. (2000). Linkage disequilibrium, gene trees and selfing: An ancestral recombination graph with partial self-fertilization. *Genetics*, *154*(2), 923–929. https://doi.org/10.1093/genetics/154.2.923

Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. *Proceedings of the Royal Society B: Biological Sciences*, *263*(1373), 1033–1039. https://doi.org/10.1098/rspb.1996.0152

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., & Lynch, M. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, *327*(5961), 92–94. https://doi.org/10.1126/science.1180677

Pannell, J. R., & Fields, P. D. (2014). Evolution in subdivided plant populations: Concepts, recent advances and future directions. *New Phytologist*, *201*(2), 417–432. https://doi.org/10.1111/nph.12495

Pebesma, E. J., & Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13. Retrieved from http://cran.r-project.org/doc/Rnews/

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1002967

Platt, A., Horton, M., Huang, Y. S., Li, Y., Anastasio, A. E., Mulyati, N. W., Ågren, J., Bossdorf, O., Byers, D., Donohue, K., Dunning, M., Holub, E. B., Hudson, A., Le Corre, V., Loudet, O., Roux, F., Warthmann, N., Weigel, D., Rivero, L., … Borevitz, J. O. (2010). The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics*, *6*(2), e1000843. https://doi.org/10.1371/journal.pgen.1000843

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies.

Ecology and Evolution, *4*(22), 4399–4428. https://doi.org/10.1002/ece3.1305

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rannala, B., & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *1656*, 1645–1656. https://doi.org/10.1093/genetics/164.4.1645

Ream, T. S., Woods, D. P., Schwartz, C. J., Sanabria, C. P., Mahoy, J. A., Walters, E. M., & Amasino, R. M. (2014). Interaction of photoperiod and vernalization determines flowering time of *Brachypodium distachyon*. *Plant Physiology*, *164*(2), 694–709. https://doi.org/10.1104/pp.113.232678

San Miguel, A. (2008). Management of Natura 2000 habitats. 6220: Pseudo-steppe with grasses and annuals of the Thero-Brachypodietea. European Commission. https://ec.europa.eu/environment/nature/natura2000/management/habitats/models_en.html

Sancho, R., Cantalapiedra, C. P., López-Alvarez, D., Gordon, S. P., Vogel, J. P., Catalán, P., & Contreras-Moreira, B. (2018). Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytologist*, *218*(4), 1631–1644. https://doi.org/10.1111/nph.14926

Scholthof, K. B. G., Irigoyen, S., Catalan, P., & Mandadi, K. K. (2018). *Brachypodium*: A monocot grass model genus for plant biology. *The Plant Cell*, *30*(8), 1673–1694. https://doi.org/10.1105/tpc.18.00083

Sensoy, S. (2008). *Climate of Turkey.* Turkish State Meteorological Service.

Sharma, N., Ruelens, P., D'Hauw, M., Maggen, T., Dochy, N., Torfs, S., & Geuten, K. (2017). A flowering locus C homolog is a vernalization-regulated repressor in *Brachypodium* and is cold regulated in wheat. *Plant Physiology*, *173*(2), 1301–1315. https://doi.org/10.1104/pp.16.01161

Shiposha, V., Marques, I., López-Alvarez, D., Manzaneda, A. J., Hernandez, P., Olonova, M., & Catalán, P. (2020). Multiple founder events explain the genetic diversity and structure of the model allopolyploid grass *Brachypodium hybridum* in the Iberian Peninsula hotspot. *Annals of Botany*, *125*(4), 625–638. https://doi.org/10.1093/aob/mcz169

Skalska, A., Stritt, C., Wyler, M., Williams, H. W., Vickers, M., Han, J., Tuna, M., Savas Tuna, G., Susek, K., Swain, M., Wóycicki, R. K., Chaudhary, S., Corke, F., Doonan, J. H., Roulin, A. C., Hasterok, R., & Mur, L. A. J. (2020). Genetic and methylome variation in Turkish *Brachypodium distachyon* accessions differentiate two geographically distinct subpopulations. *International Journal of Molecular Sciences*, *21*(18), 1–17. https://doi.org/10.3390/ijms21186700

Stebbins, G. L. (1957). Self fertilization and population variability in the higher plants. *The American Naturalist*, *91*(861), 337–354. https://doi.org/10.1086/281999

Stein, N., Herren, G., & Keller, B. (2001). A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breeding*, *120*(4), 354–356. https://doi.org/10.1046/j.1439-0523.2001.00615.x

Stritt, C., Gordon, S. P., Wicker, T., Vogel, J. P., & Roulin, A. C. (2018). Recent activity in expanding populations and purifying selection have shaped transposable element landscapes across natural accessions of the Mediterranean grass *Brachypodium distachyon*. *Genome Biology and Evolution*, *10*(1), 304–318. https://doi.org/10.1093/gbe/evx276

Stritt, C., Wyler, M., Gimmi, E. L., Pippel, M., & Roulin, A. C. (2020). Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. *New Phytologist*, *227*(6), 1736–1748. https://doi.org/10.1111/nph.16308

Sultan, S. E., & Spencer, H. G. (2002). Metapopulation structure favors plasticity over local adaptation. *The American Naturalist*, *160*(2), 271–283. https://doi.org/10.1086/341015

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Thompson, J. D. (2005). *Plant Evolution in the Mediterranean*. Oxford University Press.

Tyler, A. L., Lee, S. J., Young, N. D., Deiulio, G. A., Benavente, E., Reagon, M., & Caicedo, A. L. (2016). Population structure in the model grass *Brachypodium distachyon* is highly correlated with flowering differences across broad geographic areas. *The Plant Genome*, 1–55. https://doi.org/10.3835/plantgenome2015.08.0074

Vogel, J. P. (2016). *Genetics and genomics of Brachypodium*. Springer International Publishing.

Vogel, J. P., Tuna, M., Budak, H., Huo, N., Gu, Y. Q., & Steinwand, M. A. (2009). Development of SSR markers and analysis of diversity in Turkish populations of *Brachypodium distachyon*. *BMC Plant Biology*, *9*, 88. https://doi.org/10.1186/1471-2229-9-88

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, *15*(6), 1419–1439. https://doi.org/10.1111/j.1365-294X.2006.02890.x

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(97), https://doi.org/10.4161/hv.21408

Wright, S. I., Kalisz, S., & Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proceedings Biological Sciences/The Royal Socity*, *280*(1760), 20130133. https://doi.org/10.1098/rspb.2013.0133

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*, *35*(10), 1786–1788. https://doi.org/10.1093/bioinformatics/bty875

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.