New resources for environmental genomics in the wild Mediterranean grass *B. distachyon*

Nikolaos Minadakis[1], Hefin Williams[2], Robert Horvath[1], Danka Caković[3], Christoph Stritt[4], Michael Thieme[1], Yann Bourgeois[5,6], Anne C. Roulin[1]

[1] Department of Plant and Microbial Biology, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland.

[2] Department of Life Sciences, Aberystwyth University, Aberystwyth, Wales.

[3] Department of Biology, University of Montenegro, Džordža Vašingtona bb, 81000 Podgorica, Montenegro.

[4] Swiss Tropical Institute, Kreuzstrasse 2, 4123 Allschwil, Switzerland.

[5] DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France.

[6] University House, Winston Churchill Ave, Portsmouth PO1 2UP, United Kingdom.

Corresponding authors: nikolaos.minadakis@botinst.uzh.ch; anne.roulin@botinst.uzh.ch

ABSTRACT

Closely related to economically important crops, the grass *Brachypodium distachyon* has been originally established as a pivotal species for grass genomics but more recently flourished as a model for developmental biology. Grasses encompass more than 10,000 species and cover more than 40% of the world land area from tropical to temperate regions. Unlocking the sources of phenotypic variation in *B. distachyon* is hence of prime interest in fundamental and applied research in agronomy, ecology and evolution. We present here the *B. distachyon* diversity panel which encompasses 332 fully sequenced accessions covering the whole species distribution from Spain to Iraq. By combining population genetics, niche modelling and landscape genomics, we show that *B. distachyon* recolonized Europe and the Middle East following the last glacial maximum. As a consequence, the species faced new environmental conditions which led to clear association between bioclimatic variables and genetic factors as well as footprints of positive selection in the genome. Altogether, this genomic resource offers a powerful alternative to *Arabidopsis thaliana* to investigate the genetic bases of adaptation in plants and more specifically in monocots.

Keywords: genetic diversity, local adaptation, genomic landscape, grasses, *Brachypodium distachyon*

INTRODUCTION

In the face of accelerating climate change, understanding how plant populations adapt to new environmental conditions has sparked great interest (Lasky et al. 2023). Environmental constraints varying in space and time alter the local frequencies of adapted genotypes, leaving detectable signatures at selected genes. Identifying these signatures is valuable for understanding the mechanistic processes underlying adaptation, and providing candidate genes for functional studies. In this context, advances in genomics and DNA sequencing technology have revolutionized the field of landscape genomics by enabling high-resolution genome-wide analyses (for review Bourgeois and Warren 2021). For instance, genotype-environment association analyses (GEA) are powerful tools to identify alleles associated with ecologically relevant factors (Rellstab et al. 2015; Lasky et al. 2023). On the other hand, genome-wide scans of selection (Nielsen et al. 2005; Tang et al. 2007; Gautier 2015), have been largely used to identify genetic factors under selection without a prior neither on the phenotype under selection nor on the selective constraints acting on it. Those approaches, however, all rely on one fundamental aspect: the availability of genomic resources for many accessions occurring in contrasting habitats.

Despite the pressing need for new model species (Marks et al. 2023), research on local adaptation in plants is still largely dominated by *A. thaliana* (Provart et al. 2016; Woodward and Bartel 2018). The *A. thaliana* 1001 genome project, combined with decades of functional studies, has indeed produced an unmatched wealth of knowledge about the processes of adaptation and evolution in plants (e.g. Horton et al. 2012; Hancock et al. 2011; Durvasula et al. 2017; Fulgione et al. 2018; Takou et al. 2019; Exposito-Alonso 2020; Monroe et al. 2022; Lee et al. 2017; Wieters et al. 2021; Exposito-Alonso et al. 2019; Wu et al. 2017). Even though the development of pan-genomes for crops (Montenegro et al. 2017; Wang et al. 2018; Haberer et al. 2020; Jayakodi et al. 2020; Walkowiak et al. 2020; Lovell et al. 2021) will undoubtedly help to understand stress-resilience in plants of agronomical interest, they nonetheless remain of limited value to tackle the diversity of paths to adaptation found in natural systems. Developing alternative wild systems, especially in grasses which play a key role in ecosystem functioning (Groves 2000), is thus very timely.

Initially established as a model for crop genomics (International Brachypodium Initiative 2010), the grass species *Brachypodium distachyon* is now a pivotal system for developmental

biology (Hasterok et al. 2022; Raissig and Woods 2022) including the study of flowering time (Ream et al. 2014; Woods et al. 2014; Sharma et al. 2017; Woods et al. 2020; Bouché et al. 2022), stomata traits (McKown and Bergmann 2020; Nunes et al. 2020; Zhang et al. 2022; Slawinska et al. 2023) or cell-wall development (Coomey et al. 2020). In addition to harbouring a large mutant collection (Dalmais et al. 2013) and modern tools for mutagenesis (Hus et al. 2020), *B. distachyon* is a wild and widespread species with a small diploid genome (280Mb). As such, it features functional and population genomics resources that are not combined in any other grass system for which a germplasm diversity panel has been established, such as rice or switchgrass (Wang et al. 2018; Lovell et al. 2021).

B. distachyon occurs naturally in oligotrophic habitats around the Mediterranean rim (López-Alvarez et al. 2015; Catalán et al. 2016). The combination of a geographic mosaic and climatic stability allowed Mediterranean species to diversify at regional and local scales (Nieto Feliner 2014). In this context *B. distachyon* constitutes an excellent system to investigate the genetic bases of local adaptation in a species adapted to arid climates. Earlier works on this topic mostly focused on a small set of accessions originating from Turkey and Spain (Del'Acqua et al. 2014; Des Marais et al. 2017; Skalska et al. 2020; Stritt et al. 2018; Bourgeois et al. 2018, Wyler et al. 2018) and the genetic basis of trait variation is still largely characterized through QTL mapping and mutant screening in this species (e.g. Barbieri et al. 2012; Woods et al. 2020; 2017; Jiang et al. 2017). Yet, genome-wide sequencing data have been produced for about 260 accessions originating from Spain, France, Italy, Turkey, Lesser Caucasus and Iraq (Gordon et al. 2017; 2020; Skalska et al. 2020; Stritt et al. 2022). Those resources, however, were never analyzed at once (but see Stritt et al. 2022 for the analysis of 196 accessions) despite their great potential to unlock the source of genetic and phenotypic diversity in this species. In the current study, we filled-up a last geographical gap by collecting and sequencing an additional set of *B. distachyon* accessions from Greece and Montenegro. We combined all available genomic data into a diversity panel encompassing 332 accessions and asked: i) What is the recent demographic history of *B. distachyon* populations ii) To what extent have environmental factors shaped genetic diversity in this species and, iii) Which genes have been selected by the environment. This resource is made freely available to the community to stimulate research and facilitate comparisons across established and emerging model species for plant adaptation.

RESULTS AND DISCUSSION

Population structure and demographic history of *B. distachyon*

We used publicly available sequencing data from 196 natural *B. distachyon* accessions originating from Spain, France, Italy, Turkey, and Iraq (Gordon et al. 2017; Skalska et al. 2020; Stritt et al. 2022) previously analyzed in Stritt et al. (2022). We also added 65 additional sequencing data produced by Gordon et al. (2020) for accessions from Spain, Turkey and lesser Caucasus. Finally, we collected and sequenced an additional set of 71 accessions from Montenegro, Greece and France. Altogether, we assembled a diversity panel that encompasses 332 accessions (Fig. 1a, Table S1) for which whole-genome sequencing data are available with a minimum coverage of 20X.

B. distachyon belongs to the Brachypodium species complex. The three species in this complex, *B. distachyon*, *B. stacei* and the allopolyploid *B. hybridum* can be difficult to distinguish morphologically but are straightforward to identify through genotyping (Giraldo et al. 2012; Catalán et al. 2016). Following extensive fieldwork, all accessions collected in the past in Morocco, Afghanistan, Iran, Pakistan, Australia and USA have been genotyped as *B. hybridum* (Stritt et al. 2022; Wilson et al. 2019; Fig. S1 for the geographical distribution of the 2420 accessions genotyped). In contrast to *B. hybridum* and *B. stacei, B. distachyon* is extremely rare in Israel (Wilson et al. 2019) and in France only recorded in the South of the country (https://www.tela-botanica.org/bdtfx-nn-10075-synthese). The 332 *B. distachyon* accessions currently at hand (Fig. 1a, Table S1) are thus likely to cover comprehensively the species range.

We identified 10,227,760 high-confidence single nucleotide polymorphisms (SNPs) in the diversity panel and applied different filtering criteria according to the requirement of each analysis in the following sections, as described in Materials and Methods. While we merged datasets from different studies, a hierarchical clustering based on 75,000 random SNPs shows that samples do not cluster according to the study of origin (Fig. S2), indicating no large technical biases. Based on 196 samples, Stritt et al. (2022) found that *B. distachyon* accessions cluster in three main genetic lineages (A, B, C) that further split into five clades (A_East, A_Italia, B_East, B_West and C). Our current analysis based on 332 accessions did not reveal any additional discrete genetic clusters (Fig. 1a, b, c), even after expanding the sampling range. Indeed, the principal component analysis (PCA) (Fig. 1b), sNMF (Fig. 1c and Fig. S3a,b)

and the Evanno method (Evanno et al. 2005; Fig. S3c) all suggest that our accessions group into five main genetic clades.

Accessions from Spain and France share common ancestry and cluster together in the B_West clade (n = 72). One accession from France (MonSV5), however, shares common ancestry with the main Italian clade (A_Italia) while four accessions from Spain have a mixed ancestry and cannot be assigned to any of the clades (Spa-Sou-GR6_4, Spa-Nor-S6D, Mon3, Arn1). Accessions from Eastern regions belong to two clades, one comprising Balkan and

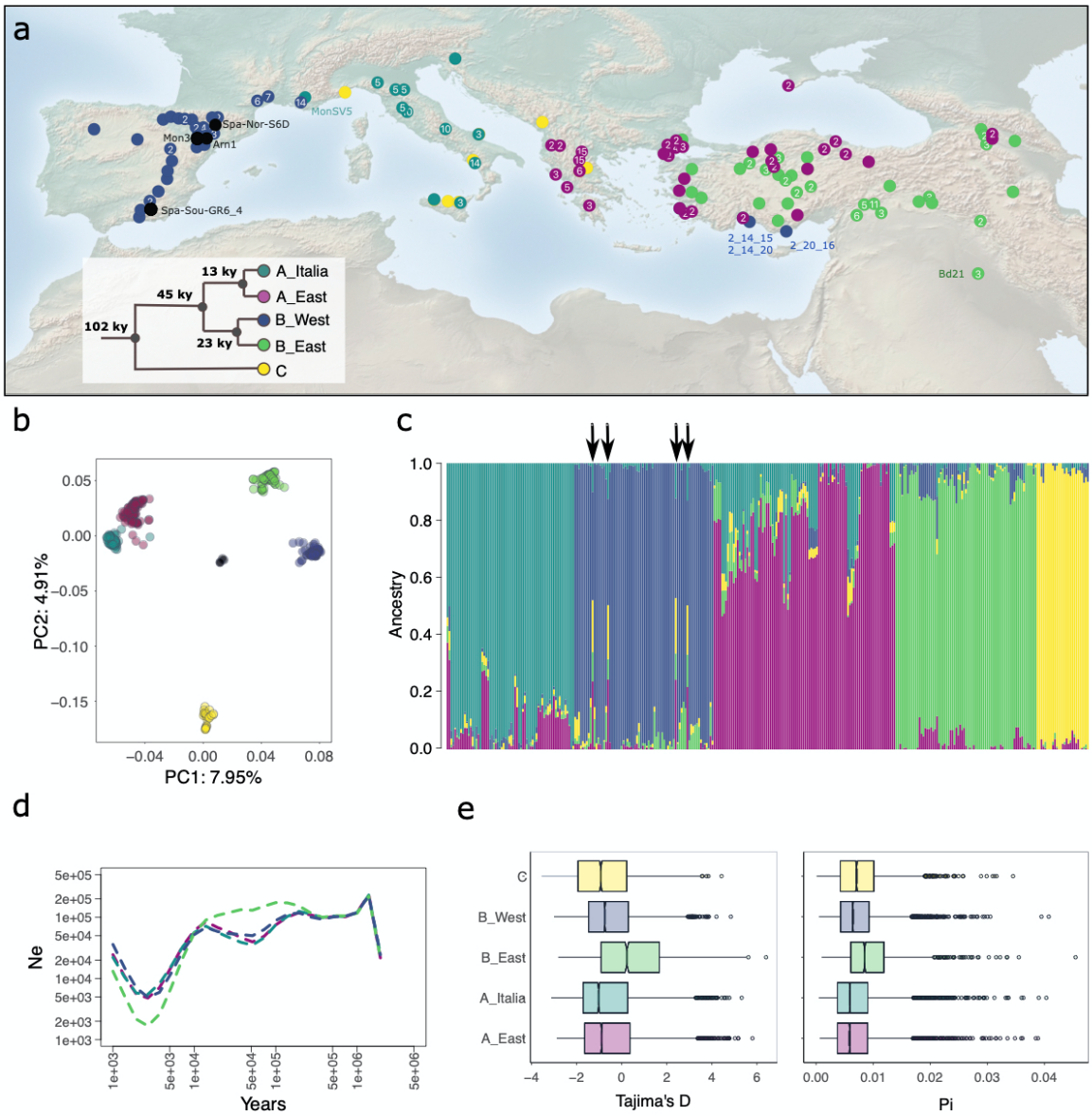

Figure 1 - Sample distribution and population structure a) the map displays the origin of the 332 accessions used in the study. The tree represents the phylogeny of the five genetic clades and indicates the divergence estimates. The color code will apply to the rest of the study b) PCA based on 16.381 independent SNPs c) Structure plot displaying the shared ancestry among accessions. The black arrow points at the four admixed accessions found in Spain d) Population size evolution over time computed for the four derived genetic clades e) Tajima's D and pi per genetic clade computed in 5kb windows over the entire genome.

131 coastal Turkish individuals (A_East, n = 94), and the other including mainland Turkish and
132 Anatolian individuals (B_East, n = 73). In Italy, two genetic clades, the A_Italia (n = 66) and C
133 (n = 27) co-occur. We also observed that all the accessions from Montenegro share ancestry
134 with the rest of the C clade, as does one individual from Greece (Meg7; Fig. 1a-c). Eleven
135 accessions from the lesser Caucasus have been identified as belonging to two different
136 genetic clusters, with three accessions belonging to the A_East cluster and eight accessions
137 belonging to the B_East cluster. Finally, three accessions from coastal Turkey (2_14_15,
138 2_14_20, 2_20_16) clustered with the B_West, as already reported in previous studies
139 (Skalska et al. 2020; Stritt et al. 2022).

140 Using a multispecies coalescent approach, we estimated that the split between the
141 ancestral C lineage and the A/B lineages occurred 102 thousand years ago (kya, 90% highest
142 posterior density interval [HPDI] 50–170 ky), the split between the A and B lineages 45 kya
143 (90% HPDI 21–76 ky) and the split within the A and B lineages 13 kya (90% HPDI 5-23 ky) and
144 23 kya (90% HPDI 11-39ky) respectively (Fig. 1a). These estimates are in the same order of
145 magnitude as the ones we obtained with Relate, a method used for estimating genealogies
146 genome-wide. For this analysis, the ancestral C clade was used to polarize SNPs. Relative
147 cross-coalescence rates (ccr) between the remaining four genetic clades indicate that the split
148 (ccr < 0.5) between the A and B lineages occurred within the last 100,000 years ago while the
149 splits (ccr < 0.5) within the A and B lineages occurred within the last 5 kya (Fig. S4).

150 We also used Relate to compute population size (Ne) evolution over time. This analysis
151 revealed an abrupt decline in effective population size (Ne) across all clades around 30 kya,
152 all followed by population expansion in a very recent past (Fig. 1d). Congruent with the recent
153 population decline and expansion observed for each genetic clade, we found overall negative
154 Tajimas'D values but still substantial remaining genetic diversity (pi) for all genetic clades
155 (Fig. 1e). Nonetheless, the B_East clade displays higher Tajima's D and genetic diversity levels
156 than the other genetic clades (Kruskall-Wallis test, all p-values < 5.2e-05) which could be
157 explained by its more stable effective population size before the most recent drop/expansion
158 ca. 5 kya. Note that all our dating methods rely on the use of a molecular clock and must thus
159 be interpreted with some caution. Our results nonetheless indicate major changes in the
160 recent demographic history of *B. distachyon.* Because these demographic changes
161 accelerated within the last 10,000 years, we speculated that the species experienced a shift

162 of its distribution following the last glacial maximum (LGM, 22 kya) and more specifically
163 during the Holocene period (11.7 kya) which marks the beginning of deglaciation in Europe.

164

165 **Ecological Niche Modeling and distribution of *B. distachyon* during the last glacial period**

166 The climate, vegetation and landscape of Northern Eurasia (north of ca. 40°N and from 10°W
167 to 180°E) underwent massive changes during the last glacial period (Binney et al. 2017; Davis
168 et al. 2022). Under the cooler and dryer conditions faced during the LGM specifically, forests
169 retreated to glacial refugees in Spain, Italy or Balkans (for review Feliner 2011; Nieto Feliner
170 2014) and land remained mostly covered by steppe and tundra (Binney, Edwards, Macias-
171 fauria, et al. 2017). Following the LGM, deglaciation lead to the recolonization of Eurasia by
172 woody plants and forests (Binney, Edwards, Macias-fauria, et al. 2017) and hence to
173 substantial changes in dominant biomes. These recent climatic events have shaped the
174 biogeography and genetic diversity of plants at large (Feliner 2011). In *A. thaliana* for
175 instance, relict populations occupied post-glacial Eurasia first and were later replaced by non-
176 relict populations whose range expansions, accompanied by admixture, largely shaped
177 modern populations (Lee et al. 2017).

178 To test for a potential shift in the distribution of *B. distachyon* in the recent past, we
179 performed niche modelling analyses using the derivation of Maxent climate niche models
180 (Phillips et al. 2006). Due to the high correlation among bioclimatic variables for current time
181 (Fig. S5a), we proceeded with a strategy based on selecting variables using an *a priori*
182 understanding (Burnham and Anderson, 2004) of the *B. distachyon* life cycle: minimum
183 temperature averaged from November to February (hereafter *tmin_Nov-Feb*) was chosen as
184 potentially important for the vernalization process; precipitation levels averaged from March
185 to June (hereafter *prec_March-June*), solar radiation levels averaged from March to June
186 (*srad_March-June*) and elevation were chosen as relevant for plant phenology during the
187 growing season.

188 We first fitted consecutive models for the whole species distribution with these four
189 variables under current environmental conditions (see Materials and Methods), before
190 projecting them over past conditions. Models with combinations of two variables had the
191 lowest AIC values generally, however, only the model using *prec_March-June* and *tmin_Nov-*

192 *Feb* showed Area under the Curve (AUC) values and predictive ability better than null models
193 (p-value < 0.05) and was consequently chosen as our final model (see Fig. S5b for the
194 distribution of these variables across genetic clades).

195 Based on this model and using paleoclimatic datasets, we eventually projected the
196 potential whole-species distribution in Europe, North Africa, and Middle East since the Last
197 Glacial Maximum (ca. 22 kya), as well as for the five genetic clades (Fig. S6). In contrast to the
198 niche modelling performed with the geographically restricted current bioclimatic variables,
199 we extended the analysis to a much wider geographic area in order to be able to detect
200 putative changes in the species distribution. Our results suggest that climatic conditions
201 similar to currently suitable conditions for *B. distachyon* extended southwards under LGM
202 conditions with Northern Africa, Levant countries as well as the Iberic and Arabian Peninsula
203 providing the most suitable habitats for the species (Fig. S6). During LGM, the sea level was
204 about 120–150 meters lower than at present (Yokoyama et al. 2000) which may have
205 facilitated migration between Southern Europe and North Africa (Ortiz et al. 2007).
206 Furthermore, Levant countries were mostly dry to subhumid habitats (Jennings et al. 2015).
207 Hence, their geographical proximity with modern accessions and habitat suitability makes
208 Northern Africa, Levant countries and the Iberic peninsula good candidates for *B. distachyon*
209 glacial refuges. While the high suitability in the Arabian Peninsula might be seen as an overfit
210 of our models, it is worth noting that annual rainfall levels under LGM conditions were much
211 more important than nowadays in this region, making part of the Arabian Peninsula a dry to
212 subhumid habitat (Jennings et al. 2015) likely suitable for *B. distachyon*.

213 We did not include information about soil as those are not projected for past conditions.
214 For an oligotroph species as *B. distachyon*, the fundamental niche we computed might
215 therefore differ from the realized one and without modern samples available for all these four
216 putative glacial refugees, one can only speculate about a migration scenario (Lee et al. 2017).
217 Yet, a natural conclusion is that *B. distachyon*, like most plants, operated a shift during the
218 last glaciation and recolonized Europe and the Middle East following deglaciation. Our results
219 contrast to some extend with those of López-Alvarez et al. (2015), who found that the
220 distribution of *B. distachyon* extended southwards without a complete shift. This early study
221 was based on the only samples collected at that time in Spain and Turkey. As it failed to
222 recover the presence of the species in Italy and Balkans under current environmental

223 conditions, we argue that the projection to past conditions might be less reliable than the one
224 we presented here based on a more comprehensive sampling of the species.

225 The split among the three main genetic lineages of *B. distachyon* predates LGM (Fig. 1a,
226 Fig. S4) and it is therefore likely that accessions from the A_East/A_Italia, B_East/B_West and
227 C genetic clades experienced different demographic histories. The geographic distribution of
228 the five genetic clades is more parsimoniously explained by independent expansions and is
229 sustained by the rapid increase of the effective population size for each clade within the last
230 5ky (Fig. 1d). It has already been proposed that lineages and genetic clades may have used
231 different migration corridors to recolonize Europe and the Middle East (Stritt et al. 2022). An
232 East–West phylogeographical break, as often reported in Mediterranean species (Nieto
233 Feliner 2014), could explain that accessions from the B_West and B_East genetic clades group
234 together phylogenetically despite the current geographic gap (Fig. 1a). This scenario further
235 supports a North African corridor for accessions from the B lineage as already speculated in
236 Stritt et al. (2022). Such a demographic scenario has been demonstrated in the herbaceous
237 perennial species *Erophaca baetica* for instance, where plants from the Iberic peninsula are
238 clearly derived from North-African populations while geographically disconnected from the
239 Greek and Turkish ones (Casimiro-Soriguer et al. 2010; Nieto Feliner 2014). In contrast, we
240 previously observed a South-to-North gradient of declining N_e, genetic diversity, and shared
241 ancestry for A_Italia clade, which suggests a northwards expansion of this clade (Stritt et al.
242 2022).

243

244 **Genetic clades occupy different ecological niches and display adaptive loci**

245 The Mediterranean basin comprises a mosaic of habitats (Nieto Feliner 2014) and consistent
246 with recolonizations through different routes, the five genetic clades of *B. distachyon* occupy
247 nowadays different geographical areas and ecological niches (Fig. 2). To detect significant
248 differences among the ecological niche models (ENMs) of the five clades under current
249 conditions, we created pseudoreplicate niche comparisons, and compared them with the
250 niche similarity score produced using the empirical data of the realized distribution (Fig. 2a-
251 e). Note that the three accessions from Turkey clustering with the B_West clade (2_14_15,
252 2_14_20, 2_20_16, Fig. 1a) were excluded from the analysis. According to our models, there
253 was no overlap between the niche suitability among the five clades. For the A_Italia vs C,

254 B_West vs A_East, B_West vs A_Italia comparisons, p-values ranged from 0.01 and 0.05. For
255 all the other comparisons, p-values were <0.01. We however acknowledge that the ENM of
256 the C clade must be interpreted cautiously due to the small sample size.

257 More specifically, the higher predicted niche suitability for the A_East clade was at the
258 coasts of Turkey, in mainland Greece, Italy, south France, and parts of the Iberian Peninsula
259 (Fig. 2a). The A_Italia clade displays a high suitability in Italy, Iraq, northwest Iberian Peninsula
260 as well as on the east coast of the Adriatic Sea (Fig. 2b). For B_East clade, the areas with the
261 highest suitability scores were mainland and east Turkey, northern Greece, coastal Bulgaria,
262 as well as a part of central Spain (Fig. 2c), while the B_West clade displayed higher suitability
263 in northern Spain, southern France, parts of Italy and Greece, as well as northwest Turkey
264 (Fig. 2d). Finally, the C clade shows high niche suitability in south Italy, south Iberian Peninsula,
265 the south Greek islands, and Iraq (Fig. 2e).

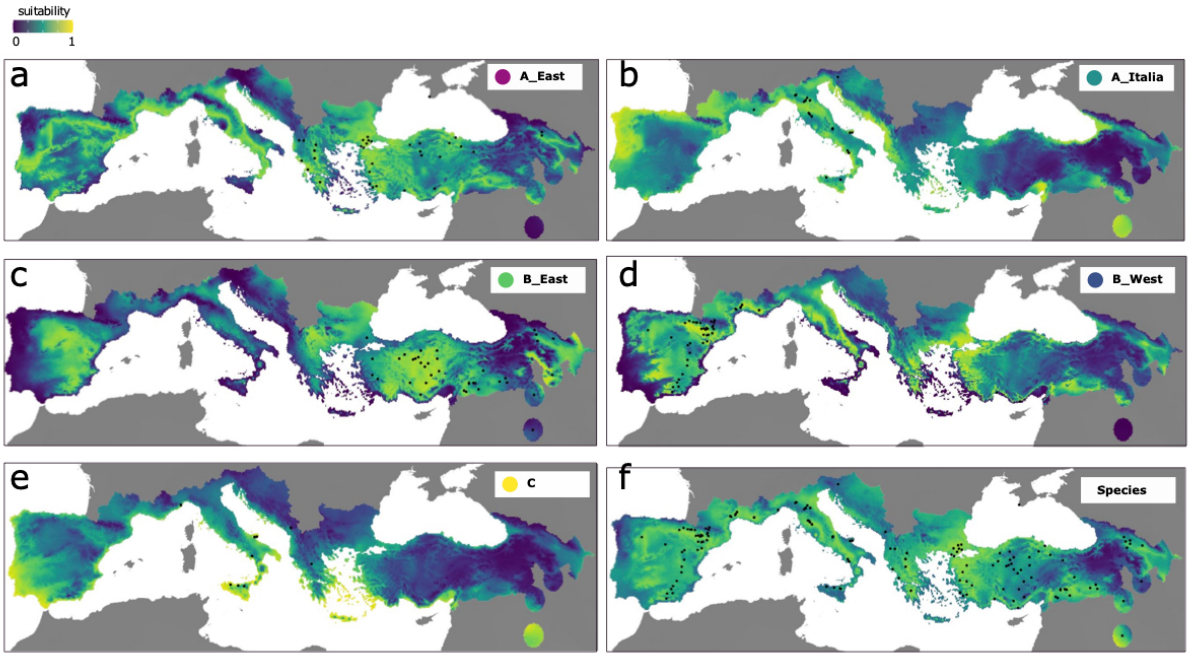

Figure 2 - Environmental Niche Modeling for current conditions using the masking region. The maps display environmental suitability for the five genetic clades a) A_East b) A_Italia c) B_East d) B_West e) C and f) at the species level.

266 We have already shown that adaptation at a regional scale lead to specific footprints of
267 positive selection in the genome of accessions from the B_East and B_West genetic clades
268 (Bourgeois et al. 2018) but this previous analysis was somewhat limited by the number of
269 samples then sequenced (27 and 17 accessions respectively). As we found that *B. distachyon*
270 genetic clades now occupy different niches, we extended this initial study by testing for

271 genome wide footprints of positive selection at a regional level, using the genetic clades as
272 focal populations. To do so, we calculated X^TX statistic, a measure comparable to single SNP
273 Fst that accounts for the neutral genetic covariance across populations (Günther and Coop
274 2013) and detects highly differentiated alleles across several populations at once without SNP
275 polarization into ancestral or derived alleles (Gautier 2015b). By selecting the top 0.1% X^TX
276 outliers (Fig. 3a), we identified 1477 highly differentiated genes putatively under positive
277 selection (Fig. 3a and c). The GO annotation of the X^TX gene set (Table S3) revealed a
278 significant over-representation of genes involved in response to chemical (p-value = 2.01E-4)
279 and metabolic process (p-value = 1.72E-04) including glutathione metabolic process (3.99E-
280 06), a metabolic process involved in the control of reactive oxygen species (ROS) and hence
281 stress response (Mittler et al. 2022) .

282 However, local adaptation is likely to occur at a much finer geographical scale (Gloss et al.
283 2022) and populations may each carry adaptations to their local climates (Lasky et al.
284 2023). Under the hypothesis that distinct lineages adapt to environmental gradients using the
285 same traits (Lasky et al. 2023) through balancing selection, GEAs are powerful bottom-up
286 tools to characterize the genetic basis of adaptation and identify which selective constraints
287 might have shaped genetic diversity in a species. After excluding alleles displaying a minor
288 allele frequency < 0.05, we tested for significant associations between the 2,867,335
289 remaining SNPs and 32 environmental variables related to precipitation levels, temperature,
290 aridity index, solar radiation or elevation (Fig. 3b, Fig. S5a) with GEMMA while correcting for
291 population structure (Zhou and Stephens 2012). We considered 5kb windows significantly
292 associated with a given environmental variable when they display at least two SNPs above
293 the False discovery rate threshold (See Materials and Methods).

294 For each bioclimatic variable, we extracted genes located in significantly associated
295 regions. Out of the 32,432 total genes already annotated in *B. distachyon* genome, 2,379 are
296 significantly associated to at least one of the environmental associations (Fig. 3c, Fig. S7, Table
297 S2). Annual mean precipitation (bio12, Fig. 3a), together with bioclimatic variables associated
298 to conditions from March through June (precipitation or aridity) yielded the highest number
299 of genes (Fig. 3c). On the other hand, only 32 and 21 genes are located in regions associated
300 with elevation or annual mean temperature (bio1) respectively. Interestingly, even though
301 the 32 bioclimatic variables we chose show some levels of correlation (Fig. 3b, Fig. S5a), we

only observed a partial overlap among the associated gene sets. For instance, while we identified 313 genes associated with precipitation levels in spring (*prec_March_June*) and 145 genes associated with precipitation levels in winter (*prec_Nov_Feb*), those two gene sets only share 113 genes indicating that genes playing a role in adaptation to climate may also be season specific. The Gene Ontology (GO) annotation performed with the 2,379 genes associated with at least one bioclimatic variable (Table S3), however, did not reveal any significant process or molecular function.

We found little overlap between the GEAs and the scan of positive selection (Fig. 3c). This was expected given that we use the five genetic clades as focal populations for the X^TX analysis

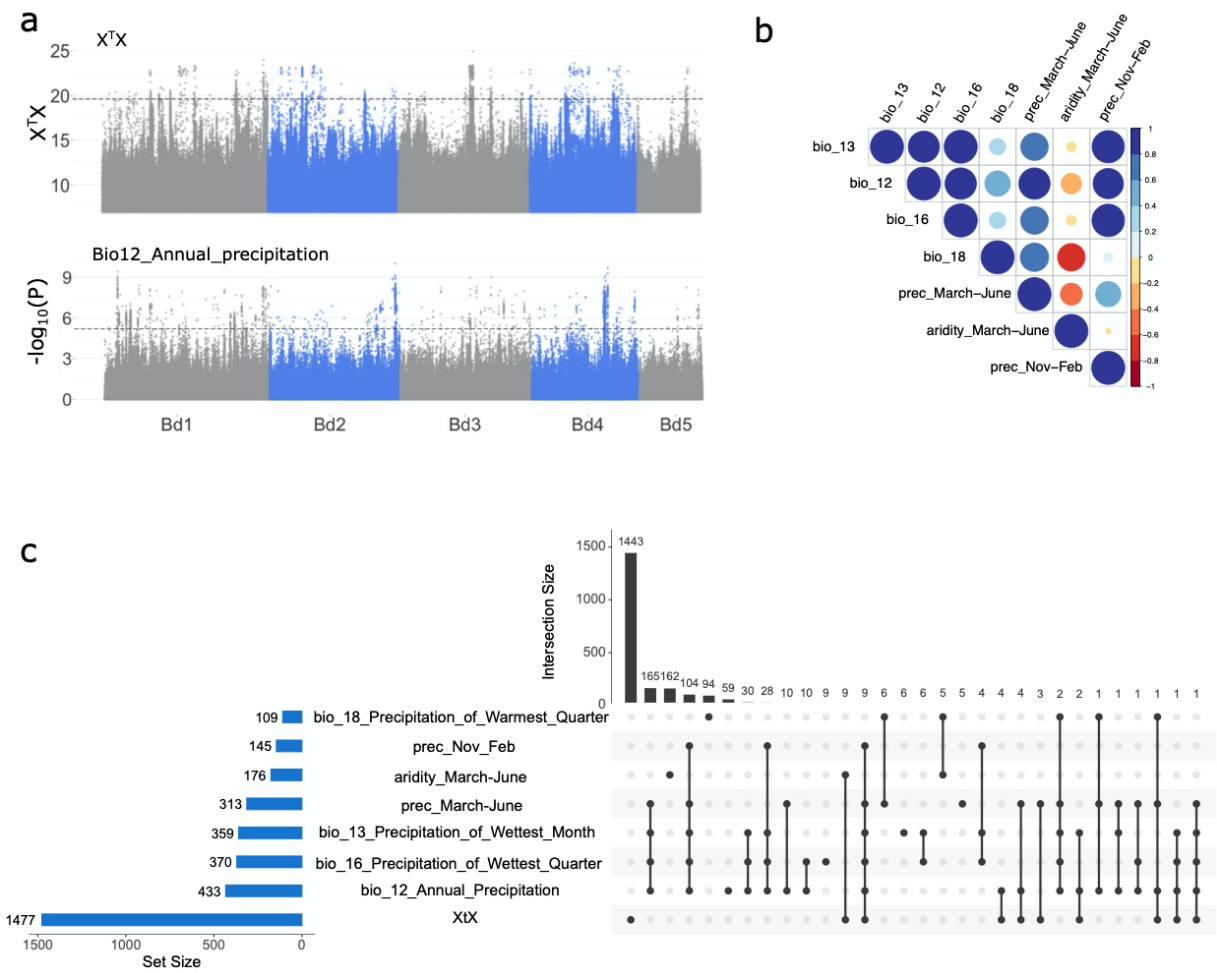

Figure 3 - Genotype-environment association and X^TX analyses a) Manhattan plots displaying regions under positive selections (the dotted line indicates the 0.1% outlier threshold) and the association between genomic regions and annual precipitation levels bio12 (the dotted line indicates the FDR threshold) b) Correlations among top 7 bioclimatic variables c) Upset plot displaying the overlaps among gene sets associated with bioclimatic variables or positive selection. The top 11 variables (associated with more than 100 genes) are presented. Set size shows the number of genes in significant regions of each specific variable. Intersection size shows the number of candidate genes associated with a variable (single dot) or shared among variables (multiple dots linked).

311 and therefore putatively detect loci influencing adaptation at a larger geographical scale than
312 with the GEAs. We also estimated the average age of alleles per gene with GEVA (Albers and
313 McVean 2020). To avoid spurious age estimates, we only kept the 29,761 genes which harbor
314 at least five SNPs (referred hereafter as genome-wide level). We used the gene lists produced
315 by the GEA and the X^TX analyses only if they contained at least fifteen genes. Even though
316 statistical differences were observed among GEA and X^TX gene sets, age estimates indicate
317 that the large majority of alleles we detected as potentially involved in adaptation emerged
318 40 to 20 kya (Fig. S7b) and are slightly older than the recolonization of Europe and the Midle-
319 East by *B. distachyon*. In contrast to positive selection on *de-novo* mutations (Barrett &
320 Schluter, 2008), selection from standing variation is predicted to promote faster evolution
321 (Hermisson and Pennings 2005) and this latter result suggests that the recycling of older
322 alleles may have played an important role in local adaptation in our system.

323

324 CURRENT LIMITATIONS AND PERSPECTIVES

325 We have not considered here that plants may also adapt along gradients via distinct strategies
326 and traits, as shown in *A. thaliana*, where genetic factors causing local adaptation to elevation
327 vary across regions and populations (Yan et al. 2021; Gamba et al. 2022). Because we filtered
328 SNPs with more than two alleles, we may have generally excluded such genetic factors from
329 our analyses and the subsequent GEAs might thus largely underestimate the number of
330 adaptive genes in our system. This might explain why we found so few genes associated with
331 elevation, while altitude is well known to influence key traits such as plant size (Moles et al.
332 2009), flowering time (Kooyers et al. 2015; Vidigal et al. 2016; Wadgymar et al. 2018) or
333 freezing tolerance (Zhen and Ungerer 2008). As we also observed substructure within genetic
334 clades (Fig. S3), within-genetic clades GEAs combined with common garden experiments
335 might be more pertinent to identify population-specific loci. Furthermore, most traits exhibit
336 a polygenic architecture and detecting a large number of variants with subtle effects (for
337 review Yeaman 2022) can be challenging with classical GEA or genome-wide association
338 analyses (de Miguel et al. 2022). Hence, many crucial questions remain open to understand
339 the genetic architecture as well as the geographical scale of adaptation in *B. distachyon.* With
340 this diversity panel, the plant community is provided with a new mean to unlock the source

341 of natural variation in a monocot adapted to arid climate, beyond what previous pan-genome
342 and QTL mapping studies permitted in this system.
343

344 MATERIALS AND METHODS

345 Sampling and Genotyping

346 Seeds from 299 *Brachypodium* plants were collected in 2018 and 2019 from south France,
347 Greece, and Montenegro. A representative subset of 110 plants was grown from seeds in
348 greenhouse conditions, and DNA was extracted from their leaves with a DNeasy plant kit from
349 Qiagen, following the manufacturer instructions. To confirm that only *B. distachyon*
350 individuals were selected for whole genome sequencing, plants were genotyped using the
351 microsatellite marker ALB165 (Giraldo et al. 2012) and by Sanger sequencing the GIGANTEA
352 gene (López-Alvarez et al. 2012). Out of the 88 accessions identified as *B. distachyon*, 71 were
353 selected for sequencing, while the remaining accessions were excluded from the rest of the
354 analysis. The libraries were prepared and sequenced using Illumina HiSeq2500 (150 bp PE) by
355 Novogene. Information about additional accessions collected and genotyped by genotyping-
356 by-sequencing (n = 1897) were obtained from Wilson et al. (2019) and displayed on a map
357 with QGIS (version 3.4.13) together with all the *B. distachyon* used in this study.
358

359 Sequencing, SNP calling, and Filtering

360 Paired-end reads from the 71 newly sequenced accessions, together with the 65 *B. distachyon*
361 accessions sequenced by Gordon et al. (2020) were aligned to version 3.0 of the *B. distachyon*
362 Bd21 reference genome (https://phytozome-next.jgi.doe.gov) with bwa mem version 0.7.17-
363 r1188 (Li 2013). Aligned reads were converted to bam files and sorted with samtools version
364 1.7 (Li et al. 2009), while accessions with multiple lanes were merged using the same program.
365 Read duplicates were then removed with PICARD MarkDuplicates version 2.23.3, and variants
366 were called with GATK v. 4.1.2.0 (McKenna et al. 2010). The resulted gvcf files were combined
367 with the 196 accessions used in the study of Stritt et al. (2022), leading to a dataset of 332
368 accessions. Subsequently, only SNPs with quality-by-depth higher than 8 and Phred-score
369 more than 20 were kept using VCFTOOLS version 0.1.15 (Danecek et al. 2011).
370
371

372 **Population Genetic Structure and demographic history**

373 The genetic structure of the samples was estimated by using principal component analysis
374 and admixture analysis. To obtain independent SNPs, they were filtered for minor allele
375 frequency > 0.05 and linkage disequilibrium < 0.4 using the function snpgdsLDpruning of the
376 R package SNPRelate (Zheng et al. 2012). The principal components were calculated using the
377 resulting 15,960 independent SNPs, and the function snpgdsPCA of SNPRelate. On the same
378 dataset, ancestry coefficients were estimated using the function sNMF of the R package LEA
379 (Frichot and François 2015). Results for 2-10 K values for 20 replicates per K were retained,
380 and the optimal K value was determined by calculating ΔK as described by (Evanno et al.
381 2005). The hierarchical cluster analysis was eventually performed using a set of 75,000
382 random SNPs with the R package pvclust version 2.2-0 (Suzuki and Shimodaira 2006).

383 We obtained population size history using Relate v1.1.7 (Speidel et al. 2019) while filtering
384 for FS > 60, SOR > 3, MQ > 40, -5.0 < MQRankSum < 5.0, QD < 2, ReadPosRankSum < -4.0,
385 INFO/DP > 16,000. We also excluded genotypes with a genotype quality GQ > 20, an individual
386 depth DP < 8x, and excluded sites with more than 20% missing genotypes, but importantly
387 did not filter for a maf. This resulted in 6,728,435 SNPs. Relate also requires alleles to be
388 polarized into ancestral and derived alleles. To do so, we used the genotypes from ancestral
389 C clade (Stritt et al. 2022) assuming the most frequent allele in that group to be ancestral. We
390 then assigned each allele in the four other clades as ancestral or derived. We assume that
391 individuals are strongly inbred and each own two copies of the same haplotype. We therefore
392 keep only one haplotype per individual. We used the Relate package (Speidel et al. 2019) to
393 infer genome-wide genealogies and coalescence rates across all individuals. To take into
394 account local variation in recombination rates, we used a recombination map for *B.*
395 *distachyon* obtained from Huo et al. (2011). To obtain demographic estimates of times and
396 population sizes, we used a mutation rate of $7x10^{-9}$ substitution/generation (Lynch et al. 2016)
397 and a generation time of one year as *B. distachyon* is annual. We set the prior for the effective
398 haploid population size to 75,000 individuals based on previous estimates Stritt et al. (2022).
399 We subsequently fitted a time-varying population size history and inferred genome-wide
400 topologies using the EstimatePopulationSize.sh script provided with the Relate package.

401 For each annotated gene (https://phytozome-next.jgi.doe.gov), the nucleotide diversity
402 (pi) was calculated using pixy version 1.2.7.beta1 in 5Kb windows (Korunes and Samuk 2021).

Tajima's D were derived from these pi values in R using theta(pi) and the number of segregating sites.

Estimating divergence times

We used the multispecies coalescent approach implemented in bpp v.4.2.9 (Rannala and Yang 2003; Flouri et al. 2018) to estimate the age of the splits among genetic lineages and clades as also described in Stritt et al. (2022). 28 accessions (Cm4, Ren22, Msa27, Lb13, BdTR7a, 2_14_13, 4_52_6, 1c_25_14, Luc1, ABR6, Tso18, Bd30-1, BdTR2B, BdTR3C, 1a_32_12, Bd21, Mca12, Cro24, Cb23, San12, Arm-Arm-2B, Geo-G30i2, Geo-G31i4, Alb-AL1A, Ko2, MonSV13, Myt1, Vyt1) were selected to represent the five genetic clades. 200 random genomic regions of 1 kb length and at least 100 kb apart were chosen. For each of the 208 accessions, the 200 sequences were obtained by calling consensus sequences from the respective bam file. 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. The rooted species tree was defined as (((A_East, A_Italia), (B_East, B_West)), C_Italia), as inferred by Stritt et al. (2022). The MCMC was run four times independently, each time with 408,000 iterations, including a burnin of 8,000 iterations. Highest posterior density intervals (HPDI) were estimated with the R package HDInterval version 0.2.4 (Meredith and Kruschke 2022). Relative cross-coalescence extracted from the output of Relate were also used to estimate divergence time, with relative cross-coalescence < 0.5 considered as populations being fully separated.

Ecological Niche Modeling

Raster maps were downloaded for all bioclimatic variables from WorldClim 2.1 (Fick and Hijmans 2017) at a 30 sec resolution (about one kilometer). Global aridity index raster maps were obtained from https://cgiarcsi.community/data/global-aridity-and-pet-database/. We used the geolocations of the 332 accessions to extract bioclimatic variables at each locality with the R package raster 3.5-29. The 19 classical WorldClim variables (bio1 to bio19), and elevation were used unmodified. New variables were created from precipitation, solar radiation, average temperature, maximum temperature, minimum temperature, and aridity

434 monthly values, by calculating the averages of four months of putative biological interest
435 (March to June, and November to February) (Table S1). Elevation data were directly
436 downloaded using Google Earth. Correlations among bioclimatic variables were plotted as a
437 bubble plot with the R package corrplot Version 0.92 (Wei and Simko 2021).

438 We used ecological niche models (ENMs) to estimate the climatic niches of the five
439 genomic clades of *B. distachyon* and to investigate the degree of climatic niche overlap in
440 both current and past conditions. As inputs for the models, we used the geographical
441 coordinates of sequenced individuals and a set of environmental variables, that were selected
442 based both on biological relevance, and on minimum collinearity between them: minimum
443 temperature averaged from November to February, precipitation levels averaged from March
444 to June, solar radiation levels averaged from March to June and elevation. The ENMs were
445 produced with the MaxEnt software version 3.4.4 (Phillips et al. 2006), which uses machine-
446 learning to predict potential geographic distributions of species across different landscapes
447 (Elith et al. 2006). The strength of association between the current distribution of different
448 genomic clades and present climate variables was tested following the methods used by
449 Williams et al. (2015) and Skalska et al. (2020). In brief, ENMs were considered to have
450 identified relationships between the distribution of *B. distachyon* and the environmental
451 predictors that were stronger than expected by chance given the spatial patterns in the data
452 if the Area under the Curve (AUC) values of the real model were in the top five of the 100
453 models (99 null and 1 real) as described by Beale et al. (2008).

454 The ENMs were visualized using QGIS version 3.4.13-Madeira and maps were created to
455 show the predicted suitable habitats for each clade. ENM modelling was restricted to the
456 relevant study area by using a prepared mask (seen in Fig. 2) which was informed by very
457 broad estimates of the extent of the sampling efforts (Fig. S1) and field observations. This
458 included mainland Spain, Portugal, Turkey, Greece, Slovenia, Croatia, Bosnia, Montenegro,
459 Albania, North Macedonia, Bulgaria, Georgia, Armenia and Azerbaijan. In France, *B.*
460 *distachyon* is only found in the South (https://www.tela-botanica.org/bdtfx-nn-10075-
461 synthese) and we only included the Occitanie and Provence-Alpes-Côte d'Azur administrative
462 regions within the mask, rather than the whole country. For the same reason, we only
463 included the Italian administrative regions to the south of Torino, Milan and the west of the
464 Po valley. Sardinia, Corsica and the Balearic Islands were excluded. The dataset also included

465 three sample points in Northern Iran and Iraq. No suitable administrative region could be
466 found to effectively delineate these areas. We therefore created a buffer (1 decimal degree
467 radians) around the sampling points.

468 To study past clade distributions, we projected current climatic niches onto paleoclimate
469 maps using variables with a resolution of 2.5 arc-min for the Last Glacial Maximum (LGM)
470 obtained from WorldClim1.4 (Fick and Hijmans 2017). The atmospheric general circulation
471 models used in this study are the Community Climate System Model version 4 (CCSM4) (Gent
472 et al. 2011). Past Maxent model projections were 'clamped' which restricts them to the range
473 of climatic conditions that are seen in the current climate conditions (where the model was
474 trained). This is recommended to avoid spurious model extrapolations when projecting into
475 novel environments (Elith et al. 2010). Clamped areas can be seen coloured grey in Fig. S6.

476 The program ENMTools version 1.3 (Warren et al. 2010; Warren et al. 2021) was used to
477 test whether there are statistically significant ecological differences between the clades,
478 based on the ENMs produced by Maxent. Niche similarity was first calculated between each
479 clade combination for the empirical data. Ninety-nine pseudoreplicate niche comparisons
480 between each clade were then produced by pooled and randomized occurrence points of the
481 empirical data. The overlap between ENMs generated from the empirical data for each clade
482 comparison was then compared to the null distribution obtained using the pseudoreplicates,
483 using Schoener's D and the I statistic (Warren et al. 2010). We concluded that ENMs produced
484 by two clades are significantly different if the empirical data had values smaller than the 95
485 of the pseudoreplicate values.

486

487 **Genotype-environment association studies**

488 We used GEMMA 0.98.5 (Zhou and Stephens 2012) to fit a linear mixed model and test for
489 association between SNPs and the 32 bioclimatic variables described above, while correcting
490 for population structure. A centered relatedness matrix was first produced with the option -
491 gk 1. Association tests were performed using the option -maf 0.05 to exclude SNPs with minor
492 allele frequency with values less than 0.05, as power is lacking for detecting associations using
493 rare alleles (Marees et al. 2018).

494 We applied a False Discovery Rate (FDR, Benjamin and Hochberg 1995) threshold to control
495 for the expected false positives rates among the rejected null hypotheses. To further reduce

496 false positives, a sliding window approach was implemented across the whole genome, with
497 window size of 5 kb and 2.5 kb overlap, using the R package rehh version 3.2.2 (Gautier et al.
498 2012). The selected windows contained at least two SNPs above the FDR threshold. We then
499 used BEDTOOLS intersect version 2.26.0 (Quinlan and Hall 2010) to extract genes located in
500 or overlapping with significant regions based on the v.3.2 annotation of the *B. distachyon*
501 (https://phytozome-next.jgi.doe.gov).
502

503 **Scans of selection**

504 The X^TX analysis was performed with BayPass using the five genetic clades as focal populations
505 (Gautier 2015). We generated the input file by using vcftools --count to calculated the allele
506 frequency of each SNP present in our vcf (no filtering on minimum allele frequency). We then
507 ran Baypass on our actual dataset with the following parameters: -pilotlength 500 -npilot 15
508 -burnin 2500 -nthreads 6. Top 0.1% outliers were chosen as a threshold of significance, as
509 more conservative than the threshold calculated with a pseudo-observed dataset simulated
510 with the co-variance matrix (Gautier 2015). We also used a sliding window approach, with
511 window size of 5 kb and 2.5 kb overlap, using the R package rehh (Gautier et al. 2012).
512

513 **Age estimates and Gene Ontology Enrichment**

514 The age of each single SNP was computed with GEVA (Albers and McVean 2020) to estimate
515 the average SNP age for each annotated gene (https://phytozome-next.jgi.doe.gov). All
516 private SNPs to the combined A and B lineages were polarized using the ancestral C lineage
517 using custom R scripts. GEVA was run on the five main scaffolds (corresponding to the five
518 chromosomes) using the genetic map produced by (Huo et al. 2011) and the polarized SNP
519 dataset. The average age of genes associated with the 32 bioclimatic variables described
520 above or outliers in the X^TX analysis was compared to the average age of genes at the genome
521 wide level with Wilcoxon test. We only used the gene lists produced by the GEA and the X^TX
522 analyses if they contained at least fifteen genes. Gene Ontology search was performed with
523 PANTHER v17.0 (http://www.pantherdb.org) using the built-in database for *B. distachyon.*

DATA AVAILABILITY

Seeds will be distributed through GRIN but can be in the meanwhile obtained in small quantities upon request. Illumina paired-end sequencing data generated for this project are archived on the European Nucleotide Archive, project number PRJEB61986. Archive numbers for the reads produced by Gordon et al. (2018, 2020), Skalska et al. (2020) and Stritt et al. (2022) are available in the respective publications.

AUTHOR CONTRIBUTION

NM conceived the study, performed the niche modelling, population genetics and GEA analyses, wrote the manuscript. RH and YB performed age estimates, population size evolution and statistics. CS, DC and MT collected samples. HW performed the niche modelling analysis. ACR conceived the study and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDMENTS

We are extremely grateful to the Swiss National Science foundation (project 31003A_182785) and the Research Priority Program Evolution in action from the University of Zurich for their generous funding. We would like to thank Chris Thomas and Léa Frachon for valuable discussions on niche modelling and GEA respectively.

SUPPLEMENTARY INFORMATION

Figure S1 - Geographic distribution of *B. distachyon*, *B. hybridum* and *B. stacei*. The distribution is based on the 1897 samples genotyped by Wilson et al (2019) and the 332 samples described in this study.

Figure S2 - Clustering analysis performed with 75,000 random SNPs for the 332 *B. distachyon* accessions described in the study. The colors indicate the study of origin.

Figure S3 - Admixture analysis a) sNMF analysis with K from 2 to 10 b) Cross-entropy plot with K from 2 to 20 c) Evanno-like evaluation of the most conspicuous level of population structure. The same calculations are performed as described in Evanno et al. 2005, with the difference that likelihoods are replaced by cross entropy (CE), the measure of model fit used by sNMF. Distributions for single values of K represent the 20 repetitions performed for each K**.**

Figure S4 - Relative cross-coalescence rate between pairs of genetic clades. The C clade was used to polarize the data and is not depicted here.

Figure S5 - 32 Bioclimatic variable plots. a) Correlation plots among 32 bioclimatic variables. Variables highlighted in red are the ones tested for niche modelling analyses b) Variation of prec_march_June and tmin_Nov_Feb across the four genetic clades.

Figure S6 - Niche modelling for last glacial maximum (LGM). The maps display environmental suitability for the five genetic clades a) A_East b) A_Italia c) B_East d) B_West e) C and f) at the species level. The grey areas on the maps do not have any model projection due to clamping.

Figure S7 - Genotype-environment associations analyses with all 32 bioclimatic variables. a) The upset plot displays the gene set overlap among the 32 GEA and XtX analyses. b) Age estimates of genes associated with bioclimatic variables or under positive selection compared to the 29761 remaining annotated genes in the genome (total gene). The numbers at the top represent the number of genes analyzed per variable. Y-axis the variable ordered based on mean. * indicate whether the age of specific gene sets is significantly different than the age of the total gene set. ns = no significant, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$; Wilcoxon signed-rank test corrected for multiple testing.

Table S1 - Origin of the 332 accessions. The table further provides information about the admixture levels as well as the bioclimatic variable data.

Table S2 - genes identified by GEA and XtX analyses

Table S3 - Gene Ontology annotation (biological process) for the genes identified by XtX analysis.

REFERENCES

Albers PK, McVean G. 2020. Dating genomic variants and shared ancestry in population scale sequencing data. PLoS Biol. 18:1–26.

Anderson D, Burnham K. 2004. Model selection and multi-model inference. Second. NY: Springer-Verlag, 63(2020), p.10.

Barbieri M, Marcel TC, Niks RE, Francia E, Pasquariello M, Mazzamurro V, Garvin DF. 2012. QTLs for resistance to the false brome rust *Puccinia brachypodii* in the model grass *Brachypodium distachyon*. 55:152–163.

Beale CM, Lennon JJ, Gimona, A, 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. Proceedings of the National Academy of Sciences 105(39), pp.14908-14912.

Benjamin Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Stat. Methodol. 57:289–300.

Binney H, Edwards M, Macias-fauria M, Lozhkin A, Anderson P, Kaplan JO, Andreev A, Bezrukova E, Blyakharchuk T, Jankovska V, et al. 2017. Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. Quat. Sci. Rev. 157:80–97.

Bouché F, Woods DP, Linden J, Li W, Mayer KS, Amasino RM, Périlleux C. 2022. EARLY FLOWERING 3 and Photoperiod Sensing in *Brachypodium distachyon*. Front. Plant Sci. 12:1–14.

Bourgeois Y, Stritt C, Walser JC, Gordon SP, Vogel JP, Roulin AC. 2018. Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass *Brachypodium distachyon*. Plant J. 96:438–451.

Casimiro-Soriguer R, Talavera M, Balao F, Terrab A, Herrera J, Talavera S. 2010. Phylogeny and genetic structure of *Erophaca* (Leguminosae), a East-West Mediterranean disjunct genus from the Tertiary. Mol. Phylogenet. Evol.56:441–450.

Catalán P, López-Álvarez D, Bellosta C, Villar L. (2016). Updated taxonomic descriptions, iconography, and habitat preferences of Brachypodium distachyon, B. stacei, and B. hybridum (Poaceae). Anales Del Jardín Botánico De Madrid, 73(1), e028.

Coomey JH, Sibout R, Hazen SP. 2020. Grass secondary cell walls*, Brachypodium distachyon* as a model for discovery. New Phytol. 227:1649–1667.

Dalmais M, Antelme S, Ho-Yue-Kuang S, Wang Y, Darracq O, d'Yvoire MB, Cézard L, Légée F, Blondet E, Oria N, et al. 2013. A TILLING Platform for Functional Genomics in *Brachypodium distachyon*. PLoS One 8.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Davis BAS, Fasel M, Kaplan JO, Russo E, Burke A. 2022. The climate and vegetation of Europe, North Africa and the Middle East during the Last Glacial Maximum (21, 000 years BP) based on pollen data. Preprint: https://doi.org/10.5194/cp-2022-59.

Del'Acqua MD, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME. 2014. Targeting environmental

621 adaptation in the monocot model *Brachypodium distachyon*: a multi-faceted approach.
622 BMC Genomics 15:801.

623 Durvasula A, Fulgione A, Gutaker RM, Alacakaptan SI, Flood PJ, Neto C, Tsuchimatsu T,
624 Burbano HA, Picó FX, Alonso-Blanco C, et al. 2017. African genomes illuminate the early
625 history and transition to selfing in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. 114:5213–
626 5218.

627 Elith J, Graham C, P. Anderson R, Dudík M, Ferrier S, Guisan A, J. Hijmans R, Huettmann F, R.
628 Leathwick J, Lehmann A, et al. 2006. Novel methods improve prediction of species'
629 distributions from occurrence data. Ecography. 29:129–151.

630 Elith J., Phillips SJ, Hastie T, Dudík M, Chee YE and Yates CJ. 2011. A statistical explanation of
631 MaxEnt for ecologists. Diversity and distributions, 17:43-57.

632 Exposito-Alonso M. 2020. Seasonal timing adaptation across the geographic range of
633 *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. U. S. A. 117:9665–9667.

634 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using
635 the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.

636 Exposito-Alonso M, Exposito-Alonso M, Gómez Rodríguez R, Barragán C, Capovilla G, Chae E,
637 Devos J, Dogan ES, Friedemann C, Gross C, et al. 2019. Natural selection on the
638 *Arabidopsis thaliana* genome in present and future climates. Nature 573:126–129.

639 Feliner GN. 2011. Southern European glacial refugia: A tale of tales. Taxon 60:365–372.

640 Feliner GN. 2014. Patterns and processes in plant phylogeography in the Mediterranean
641 Basin. A review. Perspect. Plant Ecol. Evol. Syst.16:265–278.

642 Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global
643 land areas. Int. J. Climatol. 37:4302–4315.

644 Flouri T, Jiao X, Rannala B, Yang Z. 2018. Species tree inference with *BPP* using genomic
645 sequences and the multispecies coalescent. Mol. Biol. Evol. 35:2585–2593.

646 Frichot E, François O. 2015. LEA: An *R* package for landscape and ecological association
647 studies. Methods Ecol. Evol. 6:925–929.

648 Fulgione A, Koornneef M, Roux F, Hermisson J, Hancock AM. 2018. Madeiran *Arabidopsis*
649 *thaliana* reveals ancient long-range colonization and clarifies demography in Eurasia.
650 Mol. Biol. Evol. 35:564–574.

651 Gamba D, Lorts C, Haile A, Sahay S, Lopez L, Xia T, Kulesza E, Elango D, Kerby J, Yifru M, et al.
652 2022. The genomics and physiology of abiotic stressors associated with global elevation
653 gradients in *Arabidopsis thaliana*. bioRxiv https://doi.org/10.1101/2022.03.22.485410.

654 Gautier M. 2015. Genome-Wide Scan for Adaptive Divergence and Association with
655 Population-Specific Covariates. Genetics 201:1555–1579.

656 Gautier M, Vitalis R, Baillarguet D, Cedex F-M. 2012. rehh : an R package to detect footprints
657 of selection in genome-wide SNP data from haplotype structure. Bioinformatics
658 28:1176–1177.

659 Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale
660 RB, Rasch PJ, Vertenstein M, et al. 2011. The community climate system model version

661 4. J. Clim. 24:4973–4991.

662 Giraldo P, Rodríguez-Quijano M, Vázquez JF, Carrillo JM, Benavente E. 2012. Validation of
663 microsatellite markers for cytotype discrimination in the model grass *Brachypodium*
664 *distachyon*. Genome 55:523–527.

665 Gloss AD, Vergnol A, Morton TC, Laurin PJ, Roux F, Bergelson J. 2022. Genome-wide
666 association mapping within a local *Arabidopsis thaliana* population more fully reveals
667 the genetic architecture for defensive metabolite diversity. Philos. Trans. R. Soc. B Biol.
668 Sci. 377.

669 Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session
670 A, Martin J, Cartwright A, Katz A, et al. 2020. Gradual polyploid genome evolution
671 revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors.
672 Nat. Commun.11:1–16.

673 Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin
674 AC, Schackwitz W, Tyler L, et al. 2017. Extensive gene content variation in the
675 *Brachypodium distachyon* pan-genome correlates with population structure. Nat.
676 Commun.8.

677 Groves RH. 2000. Temperate grasslands of the southern hemisphere. In Grasses Systematics
678 and Evolution. Jacobs, S. W. L. and Everett, J., eds. Melbourne: CSIRO Publishing, pp. 3781–
679 3791.

680 Günther, T. & Coop, G. Robust Identification of Local Adaptation from Allele Frequencies.
681 Genetics 195, 205–220 (2013).

682 Haberer G, Kamal N, Bauer E, Gundlach H, Fischer I, Seidel MA, Spannagl M, Marcon C, Ruban
683 A, Urbany C, et al. 2020. European maize genomes highlight intraspecies variation in
684 repeat and gene content. Nat. Genet.52:950–957.

685 Hancock AM, Brachi B, Faure N, Horton MW, Jarymowycz LB, Sperone FG, Toomajian C, Roux
686 F, Bergelson J. 2011. Adaptation to climate across the *Arabidopsis thaliana* genome.
687 Science 334:83–86.

688 Hasterok R, Catalan P, Hazen SP, Roulin AC, Vogel JP, Wang K, Mur LAJ. 2022. Brachypodium:
689 20 years as a grass biology model system; the way forward? Trends Plant Sci.27:1002–
690 1016.

691 Hermisson J, Pennings PS. 2005. Soft Sweeps: Molecular Population Genetics of Adaptation
692 From Standing Genetic Variation. 169:2335–2352.

693 Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A,
694 Sperone FG, Vilhjálmsson BJ, et al. 2012. Genome-wide patterns of genetic variation in
695 worldwide *Arabidopsis thaliana* accessions from the RegMap panel. Nat. Genet.44:212–
696 216.

697 Huo N, Garvin DF, You FM, Luo SMM, Gu YQ, Lazo GR, Philip J. 2011. Comparison of a high-
698 density genetic linkage map to genome features in the model *grass Brachypodium*
699 *distachyon*. Theor. Appl. Genet. 123:455–464.

700 Hus K, Betekhtin A, Pinski A, Rojek-Jelonek M, Grzebelus E, Nibau C, Gao M, Jaeger KE, Jenkins
701 G, Doonan JH, et al. 2020. A CRISPR/Cas9-Based Mutagenesis Protocol for *Brachypodium*

distachyon and Its Allopolyploid Relative*, Brachypodium hybridum*. Front. Plant Sci. 11:1–21.

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

Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C, Lux T, Kamal N, Lang D, Himmelbach A, et al. 2020. The barley pan-genome reveals the hidden legacy of mutation breeding. Nature 588:284–289.

Jennings RP, Singarayer J, Stone EJ, Krebs-Kanzow U, Khon V, Nisancioglu KH, Pfeiffer M, Zhang X, Parker A, Parton A, et al. 2015. The greening of Arabia: Multiple opportunities for human occupation of the Arabian Peninsula during the Late Pleistocene inferred from an ensemble of climate model simulations. Quat. Int.382:181–199.

Jiang Y, Wang X, Yu X, Zhao X, Luo N, Pei Z. 2017. Quantitative Trait Loci Associated with Drought Tolerance in *Brachypodium distachyon*. Front. Plant Sci. 8:1–11.

Kooyers NJ, Greenlee AB, Colicchio JM, Oh M, Blackman BK. 2015. Replicate altitudinal clines reveal that evolutionary flexibility underlies adaptation to drought stress in annual *Mimulus guttatus*. New Phytol. 206:152–165.

Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. Mol. Ecol. Resour. 21:1359–1368.

Lasky JR, Josephs EB, Morris GP. 2023. Genotype-environment associations to reveal the molecular basis of environmental adaptation. Plant Cell 35:125–138.

Lee CR, Svardal H, Farlow A, Exposito-Alonso M, Ding W, Novikova P, Alonso-Blanco C, Weigel D, Nordborg M. 2017. On the post-glacial spread of human commensal *Arabidopsis thaliana*. Nat. Commun. 8:1–12.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

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

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

Lovell JT, MacQueen AH, Mamidi S, Bonnette J, Jenkins J, Napier JD, Sreedasyam A, Healey A, Session A, Shu S, et al. 2021. Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. Nature 590:438–444.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. 17:704–714.

Des Marais DL, Lasky JR, Verslues PE, Chang TZ, Juenger TE. 2017. Interactive effects of water

743 limitation and elevated temperature on the physiology, development and fitness of
744 diverse accessions of *Brachypodium distachyon*. New Phytol. 214:132–144.

745 Marees, AT, de Kluiver, H, Stringer, S, et al. A tutorial on conducting genome-wide association
746 studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;
747 27:e1608. https://doi.org/10.1002/mpr.1608

748 Marks RA, Amézquita EJ, Percival S, Rougon-Cardoso A, Chibici-Revneanu C, Tebele SM,
749 Farrant JM, Chitwood DH, VanBuren R. 2023. A critical analysis of plant science literature
750 reveals ongoing inequities. Proc. Natl. Acad. Sci.120:10 e2217564120.

751 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler
752 D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce
753 framework for analyzing next-generation DNA sequencing data**.** Genome Res*,* 20:1297-
754 1303

755 McKown KH, Bergmann DC. 2020. Stomatal development in the grasses: lessons from models
756 and crops (and crop models). New Phytol. 227:1636–1648.

757 Meredith M, Kruschke J. 2022. HDInterval. HDInterval: Highest (Posterior) Density Intervals.
758 R package. https://CRAN.R-project.org/package=HDInterval

759 de Miguel M, Rodríguez-Quilón I, Heuertz M, Hurel A, Grivet D, Jaramillo-Correa JP,
760 Vendramin GG, Plomion C, Majada J, Alía R, et al. 2022. Polygenic adaptation and
761 negative selection across traits, years and environments in a long-lived plant species
762 (*Pinus pinaster* Ait., Pinaceae). Mol. Ecol. 31:2089–2105.

763 Mittler R, Zandalinas SI, Fichman Y, Van Breusegem F. 2022. Reactive oxygen species signaling
764 in plant stress responses. Nat. Rev. Mol. Cell Biol. 23:663–679.

765 Moles AT, Warton DI, Warman L, Swenson NG, Laffan SW, Zanne AE, Pitman A, Hemmings FA,
766 Leishman MR. 2009. Global patterns in plant height. J. Ecol. 97:923–932.

767 Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M,
768 Hildebrandt J, Neumann M, Kliebenstein D, et al. 2022. Mutation bias reflects natural
769 selection in *Arabidopsis thaliana*. Nature 602:101–105.

770 Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Batley J, Edwards D. 2017. The
771 pangenome of hexaploid bread wheat. Plant J. 90:1007–1013.

772 Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for
773 selective sweeps using SNP data. Genome Research 15:1566–1575.

774 Nunes TDG, Zhang D, Raissig MT. 2020. Form, development and function of grass stomata.
775 Plant J. 101:780–799.

776 Nunes TDG, Berg LS, Slawinska MW, Zhang D, Redt L, Sibout R, Vogel J, Laudencia-Chingcuano
777 D, Jesenofsky B, Lindner H, Raissig MT. 2023. Report Regulation of hair cell and stomatal
778 size by a hair cell-specific peroxidase in the *grass Brachypodium distachyon*. Curr.
779 Biol.33:1–11.

780 Ortiz MÁ, Tremetsberger K, Talavera S, Stuessy T, García-Castaño JL. 2007. Population
781 structure of *Hypochaeris salzmanniana* DC. (Asteraceae), an endemic species to the
782 Atlantic coast on both sides of the Strait of Gibraltar, in relation to Quaternary sea level
783 changes. Mol. Ecol. 16:541–552.

784 Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species
785 geographic distributions. Ecol. Modell. 190:231–259.

786 Provart NJ, Alonso J, Assmann SM, Bergmann D, Brady SM, Brkljacic J, Browse J, Chapple C,
787 Colot V, Cutler S, et al. 2016. 50 years of Arabidopsis research: Highlights and future
788 directions. New Phytol. 209:921–944.

789 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
790 features. Bioinformatics 26:841–842.

791 Raissig MT, Woods DP. 2022. The wild grass *Brachypodium distachyon* as a developmental
792 model system. Current Topics in Developmental Biology 147: 33-71

793 Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral
794 population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

795 Ream TS, Woods DP, Schwartz CJ, Sanabria CP, Mahoy JA, Walters EM, Kaeppler HF, Amasino
796 RM. 2014. Interaction of photoperiod and vernalization determines flowering time of
797 *Brachypodium distachyon*. Plant Physiol. 164:694–709.

798 Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R. 2015. A practical guide to
799 environmental association analysis in landscape genomics. Mol. Ecol. 24:4348–4370.

800 Sharma N, Ruelens P, Mariëlla D, Maggen T, Dochy N, Torfs S, Kaufmann K, Rohde A, Geuten
801 K, Leuven KU, et al. 2017. A Flowering Locus C Homolog Is a Vernalization-Regulated
802 Repressor *in Brachypodium* and Is Cold Regulated in Wheat. Plant Phys. 173:1301–1315.

803 Skalska A, Stritt C, Wyler M, Williams Hefin W, Vickers M, Han J, Tuna M, Savas Tuna G, Susek
804 K, Swain M, et al. 2020. Genetic and Methylome Variation in Turkish *Brachypodium*
805 *distachyon* Accessions Differentiate Two Geographically Distinct Subpopulations. Int. J.
806 Mol. Sci. 21:6700.

807 Speidel L, Forest M, Shi S, Myers SR. 2019. A method for genome-wide genealogy estimation
808 for thousands of samples. Nat. Genet. 51:1321–1329.

809 Stritt C, Gimmi EL, Wyler M, Bakali AH, Skalska A, Hasterok R, Mur LAJ, Pecchioni N, Roulin AC.
810 2022. Migration without interbreeding: Evolutionary history of a highly selfing
811 Mediterranean grass inferred from whole genomes. Mol. Ecol. 31:70–85.

812 Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. 2018. Recent activity in expanding
813 populations and purifying selection have shaped transposable element landscapes
814 across natural accessions of themediterranean grass *Brachypodium distachyon*. Genome
815 Biol. Evol. 10:304–318.

816 Suzuki R, Shimodaira H. 2006. Pvclust: An R package for assessing the uncertainty in
817 hierarchical clustering. Bioinformatics 22:1540–1542.

818 Takou M, Wieters B, Kopriva S, Coupland G, Linstädter A, De Meaux J. 2019. Linking genes
819 with ecological strategies *in Arabidopsis thaliana*. J. Exp. Bot. 70:1141–1151.

820 Tang K, Thornton KR, Stoneking M. 2007. A New Approach for Using Genome Scans to Detect
821 Recent Positive Selection in the Human Genome. PLoS Biol. 5.

822 Vidigal DS, Marques ACSS, Willems LAJ, Buijs G, Méndez-Vigo B, Hilhorst HWM, Bentsink L,
823 Picó FX, Alonso-Blanco C. 2016. Altitudinal and climatic associations of seed dormancy
824 and flowering traits evidence adaptation of annual life cycle timing in *Arabidopsis*

825 *thaliana*. Plant Cell Environ. 39:1737–1748.

826 Wadgymar SM, Ogilvie JE, Inouye DW, Weis AE, Anderson JT. 2018. Phenological responses
827 to multiple environmental drivers under climate change: insights from a long-term
828 observational study and a manipulative field experiment. New Phytol. 218:517–529.

829 Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, Brinton J, Ramirez-Gonzalez RH, Kolodziej
830 MC, Delorean E, Thambugala D, et al. 2020. Multiple wheat genomes reveal global
831 variation in modern breeding. Nature 588:277–283.

832 Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, et
833 al. 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature
834 557:43–49.

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

837 Warren DL, Matzke NJ, Cardillo M, Baumgartner JB, Beaumont LJ, Turelli M, Glor RE, Huron
838 NA, Simões M, Iglesias TL, et al. 2021. ENMTools 1.0: an R package for comparative
839 ecological biogeography. Ecography 44:504–511.

840 Wei T, Simko V. 2021. R package 'corrplot': Visualization of a Correlation Matrix. (Version
841 0.92), https://github.com/taiyun/corrplot.

842 Wieters B, Steige KA, He F, Koch EM, Ramos-Onsins SE, Gu H, Guo YL, Sunyaev S, de Meaux J.
843 2021. Polygenic adaptation of rosette growth in Arabidopsis thaliana. PLoS Genet. 17:1–
844 27.

845 Wyler M, Stritt C, Walser JC, Baroux C, Roulin AC. 2020. Impact of transposable elements on
846 methylation and gene expression across natural accessions of *Brachypodium distachyon.*
847 Genome Biol. Evol. 12:1994–2001.

848 Williams HW, Cross DE, Crump HL, Drost CJ, Thomas CJ. 2015. Climate suitability for European
849 ticks: Assessing species distribution models against null models and projection under
850 AR5 climate. Parasites and Vectors 8:1–15.

851 Wilson PB, Streich JC, Murray KD, Eichten SR, Cheng R, Aitken NC, Spokas K, Warthmann N,
852 Gordon SP, Vogel JP, et al. 2019. Global diversity of the Brachypodium species complex
853 as a resource for genome-wide association studies demonstrated for agronomic traits in
854 response to climate. Genetics 211:317–331.

855 Woods DP, Bednarek R, Bouché F, Gordon SP, Vogel JP, Garvin DF, Amasino RM. 2017. Genetic
856 architecture of flowering-time variation in *Brachypodium distachyon*. Plant Physiol.
857 173:269–279.

858 Woods DP, Dong Y, Bouché F, Mayer K, Varner L, Ream TS, Thrower N, Wilkerson C, Cartwright
859 A, Sibout R, et al. 2020. Mutations in the predicted DNA polymerase subunit POLD3
860 result in more rapid flowering of *Brachypodium distachyon*. New Phytol.227:1725-1735.

861 Woods DP, Ream TS, Minevich G, Hobert O, Amasino RM. 2014. PHYTOCHROME C is an
862 essential light receptor for photoperiodic flowering in the temperate grass,
863 *Brachypodium distachyon*. Genetics 198:397–408.

864 Woodward AW, Bartel B. 2018. Biology in bloom: A primer on the *Arabidopsis thaliana* model
865 system. Genetics 208:1337–1349.

866 Wu Q, Han T, Chen X, Chen J, Zou Y, Li Z, Xu Y. 2017. Long-term balancing selection contributes
867 to adaptation in *Arabidopsis* and its relatives. Genome Biol. 18:217.

868 Yan W, Wang B, Chan E, Mitchell-Olds T. 2021. Genetic architecture and adaptation of
869 flowering time among environments. New Phytol. 230:1214–1227.

870 Yeaman S. 2022. Evolution of polygenic traits under global vs local adaptation. Genetics
871 220:iyab134

872 Yokoyama Y, Lambeck K, De Deckker P, Johnston P, Fifield LK. 2000. Timing of the Last Glacial
873 Maximum from observed sea-level minima. Nature 406:713-716.

874 Zhang D, Abrash EB, Nunes TDG, Prados IH, Gil MXA, Jesenofsky B, Lindner H, Bergmann DC,
875 Raissig MT. 2022. Opposite polarity programs regulate asymmetric subsidiary cell
876 divisions in grasses. Elife11:e79913:2022.04.24.489281.

877 Zhen Y, Ungerer MC. 2008. Clinal variation in freezing tolerance among natural accessions of
878 *Arabidopsis thaliana*. New Phytol. 177:419–427.

879 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance
880 computing toolset for relatedness and principal component analysis of SNP data.
881 Bioinformatics 28:3326–3328.

882 Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association
883 studies. Nat. Genet. 44:821–824.

884

885

886

887