Physical geography, isolation by distance and environmental variables shape genomic variation of wild barley (*Hordeum vulgare* L. ssp. *spontaneum*) in the Southern Levant

Che-Wei Chang[1], Eyal Fridman[2], Martin Mascher[3], Axel Himmelbach[3], and Karl Schmid[*][1]

[1]University of Hohenheim, Stuttgart, Germany

[2]Plant Sciences Institute, Agricultural Research Organization (ARO), The Volcani Center, Israel

[3]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

September 15, 2021

[*]Corresponding author. Tel: +49 711 459 23487 Email: karl.schmid@uni-hohenheim.de

Abstract

Determining the extent of genetic variation that reflects local adaptation in crop wild relatives is of interest to discovering useful genetic diversity for plant breeding. We investigated the association of genomic variation with geographical and environmental factors in wild barley (*Hordeum vulgare L. ssp. spontaneum*) populations of the Southern Levant using genotyping-by-sequencing (GBS) of 244 accessions of the Barley1K+ collection. Inference of population structure resulted in four genetic clusters that corresponded to eco-geographical habitats and a significant association of lower gene flow rates with geographical barriers, e.g. the Judaean Mountains and the Sea of Galilee. Redundancy analysis (RDA) revealed that spatial autocorrelation explained 45% and environmental variables explained 15% of total genomic variation. Only 4.5% of genomic variation was exclusively attributed to environmental variation if the component confounded with spatial autocorrelation was excluded. A synthetic environmental variable combining latitude, solar radiation, and accumulated precipitation explained the highest proportion of genomic variation (3.9%). After correcting for population structure, soil water capacity was the most important environmental variable explaining 1.18% of genomic variation. Genome scans with outlier analysis and genome-environment association studies were conducted to identify signatures of adaptation. RDA and outlier methods jointly detected selection signatures in the pericentromeric regions of chromosome 3H, 4H, and 5H, but they mostly disappeared after correction for population structure. In conclusion, adaptation to the highly diverse environments of the Southern Levant over short geographical ranges has a small effect on the genomic diversity of wild barley highlighting the importance of non-selective forces in genetic differentiation.

Keywords: Genotyping-by-sequencing (GBS), barley, isolation by distance, local adaptation, genetic variation

Introduction

Local adaptation is an essential strategy for plants to survive in stressful environments because they are sessile. Natural selection in heterogeneous environments leads to higher fitness of local genotypes, but gene flow can offset genetic differentiation resulting from local adaptation and reduce fitness (Kawecki and Ebert 2004). In addition, genetic drift and demographic history contribute to genetic differentiation and confound adaptive variation with neutral variation (Günther and Coop 2013; Kawecki and Ebert 2004; López-Goldar and Agrawal 2021). Therefore, the combination of selective and nonselective forces simultaneously shapes genetic variation and leads to geographic patterns of population divergence and allele frequency distribution. Determining how different population genetic processes affect the geographic distribution of genetic variation is a key component in the study of plant adaptation. Investigating the role of adaptive and non-adaptive processes on genomic variation is of particular interest for wild relatives of crop plants, as this may allow the discovery of useful genetic variation for plant breeding (Turner-Hissong et al. 2020).

Wild barley (*Hordeum vulgare L. ssp. spontaneum*) is a very suitable model species to study local adaptation of crop wild relatives, as it occurs over a wide geographical range in the Fertile Crescent (Harlan and Zohary 1966) and Eastern and Central Asia (Dai et al. 2012). Within this range, wild barley in the Levant occupies very heterogeneous environments within a short geographical distance (Hübner et al. 2009; Nevo et al. 1979; Volis et al. 2001), and its genetic diversity is much higher than in other areas of the Fertile Crescent (Jakob et al. 2014; Pankin et al. 2018; Russell et al. 2016). Wild barley populations from the southern Levant show a strong correlation of genetic and environmental distances (Hübner et al. 2009). Population structure reflects eco-geographic habitats (Hübner et al. 2012, 2009) and distinguishes between northern and southern genetic clusters correlated with latitude and precipitation gradients (Jakob et al. 2014; Russell et al. 2016). With common garden experiments, previous studies revealed that eco-geography is correlated to morphological traits (Hübner et al. 2013), phenotypic plasticity (Galkin et al. 2018), and rhizosphere microbiota (Terrazas et al. 2020).Moreover, transplantation experiments showed a correlation between the geographic origin of wild barley ecotypes and fitness in different environments, suggesting local adaptation (Volis 2011; Volis et al. 2002a,b). In addition to a broad geographic scale, environmental differences at a fine geographic scale also contribute to genetic diversification in wild barley (Bedada et al. 2014; Nevo et al. 2005; Wang et al. 2018). Overall, these results suggest a strong relationship between environmental differences, genetic divergence and phenotypic diversity of wild barley populations, supporting the hypothesis of local adaptation of wild barley in the southern Levant. However, the relative contributions of environmental and non-selective forces to genetic variation and the genetic architecture

3

of adaptive traits remain mostly unclear due to the lack of appropriate statistical approaches, fine-scale environmental data and sufficient genome-wide markers.

Wild barley serves as a valuable genetic resource for barley breeding because domestication and modern breeding have greatly reduced the genetic diversity of cultivated barley (*Hordeum vulgare L. ssp. vulgare*; Caldwell et al. 2006; Kilian et al. 2006). Since wild barley shows no reproductive barrier with cultivated barley (Nevo et al. 1979), the genetic diversity of cultivated barley can be expanded by introducing alleles from wild populations (Dawson et al. 2015). Numerous studies have shown evidence of local adaptation of wild barley to different environments (Bedada et al. 2014; Galkin et al. 2018; Hübner et al. 2013; Nevo et al. 1979; Volis 2011; Volis et al. 2002a,b, 2004; Wang et al. 2018), therefore wild barley is expected to have considerable genetic variation that contributes to adaptation to various abiotic stresses (Dawson et al. 2015). Wild barley has been repeatedly used as a source of novel alleles to improve stress tolerance in barley breeding (Baum et al. 2003; Pham et al. 2019), but widespread use of wild barley has been limited due to its large genome size and undesirable traits (Schmid et al. 2018).

Genotyping-by-sequencing (GBS; Elshire et al. 2011; Poland et al. 2012) and high-quality barley genome assembly (Jayakodi et al. 2020; Mascher et al. 2017), enable the exploration of genomic variation under environmental selection and the search for useful genetic variation in wild barley. In this study, we aimed to (1) describe the population structure of wild barley from the southern Levant and place it in the context of a worldwide sample (Milner et al. 2019), (2) examine geographic patterns of gene flow in the southern Levant, (3) characterize the relative contributions of environmental gradients and space to genomic variation and population structure, and (4) identify putative adaptive loci. Overall, our results indicate the greater importance of geography and spatial autocorrelation than selection for local adaptation in shaping genomic variation in wild barley in the southern Levant, although diverse environments, particularly water availability, show significant associations with genetic differentiation.

Materials and Methods

Plant material and genotyping by sequencing We genotyped 244 wild barley accessions collected in the southern Levant region. These accessions, hereafter named as B1K+ accessions, include 191 accessions from Barley 1K (B1K) collection (Hübner et al. 2009) and 53 accessions from an unpublished collection named as HOH, collected in 2005, 2009, and 2011 by K.S. (Figure S1; File S1). The GBS library was constructed using genomic DNA digested with the restriction enzyme *PstI* and *MspI* and the processes as described by Milner et al. (2019). In addition, published GBS data of 1,121 wild barley acces-

4

94 sions from IPK seed bank (Milner et al. 2019) was included as a part of our work. Because IPK genebank

95 includes accessions from Israel, we specify the source of accessions by mentioning accessions from Israel

96 that belong to IPK in the article to avoid confusion. Identification of single-nucleotide polymorphism (SNP)

97 was performed as Milner et al. (2019). The detailed workflow of genotypic data filtration is described in

98 Supplementary Information and summarized by Figure S2.

99 **Environmental data**   To investigate the relationship between genetic variation and environmental gra-

100 dients, we used environmental data including (1) climate data from the *WorldClim2* database (Fick and

101 Hijmans 2017) with a resolution of 30 arcseconds [∼1 km], (2) soil data from the *SoilGrids* database (Hengl

102 et al. 2017) with a resolution of 250m, (3) topographic variables based on elevation data from the *SRTM*

103 database (https://srtm.csi.cgiar.org/) with a resolution of 90m, and (4) geographic coordinates of collection

104 points (Supplementary Information; File S1). To mitigate the problem of collinearity for redundancy analysis

105 (RDA; Legendre and Legendre 2012), highly correlated environmental variables were grouped by hierar-

106 chical clustering with a customized clustering index (Supplementary Information; Figure S3), and the first

107 principal components of the grouped variables were used as synthetic variables. Next, all environmental

108 variables, including the synthetic variables, were selected based on variance inflation factors (VIFs) until all

109 VIFs were less than 5. The details of this process are described in the Supplementary Information. Finally,

110 12 environmental variables, including 7 synthetic variables and 5 nonsynthetic variables (Table S1 and S2),

111 were selected for environmental association analyses.

112 **Analysis of population structure**   The number of ancestors and ancestry coefficients were estimated

113 using the model-based method *ALStructure*. (Cabreros and Storey 2019). *ALStructure* uses a likelihood-

114 free algorithm to derive estimates through minimal model assumptions and is generally superior to existing

115 likelihood-based methods in terms of accuracy and computational speed (Cabreros and Storey 2019). The

116 method does not assume Hardy-Weinberg equilibrium within populations, but defines the number of ances-

117 tral populations (K) as the rank of a matrix consisting of individual-specific allele frequencies (Leek 2011).

118 Optimal K was calculated using the *estimate_d* function of the R package *alstructure* (Cabreros and Storey

119 2019), and ancestry coefficients were estimated using the *alstructure* function. A range of K values, from

120 2 to 8, was also used to examine stratification of population structure. In addition to *ALStructure*, princi-

121 pal component analysis (PCA) and neighbor-joining (NJ) were also performed. Missing genotypic values

122 (∼3% of the dataset) were replaced by the average number of alternative alleles at each SNP locus before

123 performing PCA.

5

124   To analyze genetic differentiation, we calculated F_{ST} and Nei genetic distance between genetic clusters

125   defined by *ALStructure*. Accessions were assigned to genetic clusters according to the highest ancestry

126   coefficient calculated by *ALStructure* with the optimal value of *K*. The F_{ST} values were calculated as the

127   ratio of the average values according to Bhatia et al. (2013) and the genetic distances of Nei were calculated

128   using the function *stamppNeisD* of the R package *StAMPP* (Pembleton et al. 2013).

129   **Investigation of gene flow pattern**   To discover gene flow barriers that may explain observed popu-

130   lation structure, analysis of estimated effective migration surfaces (*EEMS*; Petkova et al. 2016) was con-

131   ducted. To perform *EEMS*, B1K+ accessions were clustered into 58 demes corresponding to the locations

132   of collection sites. *EEMS* was conducted in three independent runs of Markov chain Monte Carlo (MCMC),

133   and the results of three runs were averaged. Each MCMC chain included 10 million burn-in iterations and

134   10 million post-burn-in iterations thinned by an interval of 5000 iterations. Outputs of *EEMS* were processed

135   with the R package *rEEMSplots*. To examine whether geographical barriers contribute to genetic isolation,

136   we classified map pixels into barriers and non-barrier pixels according to geographical elevation (details

137   in Supplementary Information; Figure S4) and conducted a Wilcoxon test to inspect the hypothesis that

138   geographical barriers are significantly associated with lower gene flow rates.

139   As a complementary method of *EEMS*, unbundled principal components (unPC; House and Hahn 2018)

140   was employed to reveal potential long-distance migration. *unPC* scores, a ratio of PCA-based genetic

141   distance on population-level to geographical distance between demes, were computed with the R package

142   *unPC*. Original *unPC* scores were transformed by Box-Cox transformation to an approximate Gaussian

143   distribution, and a Student's t-test with two-tailed significance level of 0.05 was performed to identify extreme

144   comparisons.

145   To infer asymmetric gene flows, we implemented a coalescent-based inference (CBI) method described

146   by Lundgren and Ralph (2019). We manually grouped accessions into ten geographical regions (Figure

147   S5) such that each region covered roughly equal geographical areas as suggested by Lundgren and Ralph

148   (2019) and also considered the gene flow pattern inferred by *EEMS*. Sample sizes in each region ranged

149   from 10 to 41 with an average of 24.4. Next, we created an adjacency matrix to allow gene flow between ad-

150   jacent regions (Figure S5). As the input for CBI, pairwise genetic distances were computed as the average

151   number of different alleles across SNPs. CBI was conducted by using the R package *gene.flow.inference*

152   (Lundgren and Ralph 2019) with 2 million pre-burn-in iterations, 60 million burn-in iterations, and 100 million

153   post-burn-in iterations followed by a thinning process for every 5,000 iterations to rule out serial correla-

154   tions. Medians of gene flow rates and coalescence rates were computed from posterior distributions and

155 95% credible intervals were calculated with the highest density interval method by using the R package

156 *bayestestR* (Makowski et al. 2019).

157 **Partitioning genomic variation** To partition genomic variation into components explained by different

158 factors, we conducted redundancy analysis (RDA), a multivariate method for studying a linear relationship

159 of two or several matrices (Legendre and Legendre 2012). Specifically, we used simple RDA and RDA

160 conditioned on covariates, i.e., partial RDA, to estimate the proportion of SNP variation explained by envi-

161 ronmental variables, spatial autocorrelation, and population structure. RDA was performed with the function

162 *rda* of the R package *vegan* (Oksanen et al. 2019). For all RDA models in this study, we performed 5,000

163 permutations to test the significance of explanatory variables with the R function *anova.cca*.

164 To model the effect of spatial autocorrelation on SNP variation, distance-based Moran's eigenvector

165 maps (dbMEMs) were used in RDA (Dray et al. 2006; Legendre and Legendre 2012). First, a network of

166 58 collection sites was built with the Gabriel graph, and a spatial weighting matrix of inverse geographical

167 distances (km^{-1}) was constructed accordingly by following the method of Forester et al. (2018). Next, the

168 spatial weighting matrix was decomposed to generate dbMEMs. Subsequently, forward selection was per-

169 formed to identify dbMEMs that significantly associate with spatial genetic structure by using *forward.sel*

170 function (Dray et al. 2019). The selected dbMEMs with positive and negative eigenvalues, corresponding

171 to broad-scale and fine-scale spatial structures, were both used in RDA to capture comprehensive spatial

172 autocorrelation. To represent population structure in RDA, ancestry coefficients estimated by *ALStructure*

173 with the optimal *K* values were used. We fitted RDA models by replacing SNP data with ancestry coeffi-

174 cients as response variables to quantify relative contributions of environments and spatial autocorrelation

175 to population structure.

176 To evaluate the effects of individual environmental variables on SNP variation, we sequentially fitted

177 one environmental variable as an explanatory variable at a time and treated ancestry coefficients as co-

178 variates in RDA models. Considering the correlation between environmental variables (Figure S3 C), we

179 conducted additional permutation tests for marginal effects of environmental variables in a model including

180 all the environmental variables by setting the parameter *by = 'margin'* for *anova.cca*. This method tests

181 the significance of each environmental variable while excluding the effect that confounds with the other

182 environmental variables.

183 **Linkage disequilibrium** Linkage disequilibrium (LD) was evaluated as pairwise r^2 of SNPs by using

184 *snpgdsLDMat* of the R package *SNPRelate* (Zheng et al. 2012) with a window size of 250 markers. To

7

185 evaluate genome coverage of markers, genome-wide LD decline against physical distance was fitted by

186 using local polynomial regression and the formula of Hill and Weir (1988). Local polynomial regression was

187 carried out by using the R function *loess* with a smoothing parameter of 0.005.

188 **Identification of selection signatures**   As a genome-environment association (GEA) method, RDA

189 has high detection power and a low false-positive rate in identifying adaptation signatures (Capblancq et al.

190 2018; Forester et al. 2016, 2018). Thus, we scanned SNPs by performing simple RDA and partial RDA.

191 Simple RDA was done by treating 27,147 SNPs as response variables and twelve environmental variables

192 as explanatory variables. To control false positives due to population structure, partial RDA was performed

193 by using ancestry coefficients estimated with the optimal *K* as covariates. A statistical framework proposed

194 by Capblancq et al. (2018) was used for statistical tests and controlling false discovery rates (FDR). Briefly,

195 loadings of SNPs in the first four RDA axes, selected according to the proportion of explained variation

196 (Figure S6), were converted into Mahalanobis distances that approximated to a chi-squared distribution

197 with four degrees of freedom. Next, p-values and q-values were computed accordingly, and SNPs with FDR

198 < 0.05 were considered as candidate adaptive SNPs. The statistical test was conducted with the R function

199 *rdadapt* (Capblancq et al. 2018).

200 Besides RDA, the latent factor mixed model (LFMM; Caye et al. 2019), which an univariate GEA method,

201 was performed by using the R package *lfmm* (Caye et al. 2019) with the parameter *K = 4* to correct pop-

202 ulation structure, and q-values were subsequently computed. SNPs with FDR < 0.05 were considered as

203 candidate adaptive SNPs.

204 As a complement to GEA methods, outlier SNPs with an extreme divergence between genetic clusters

205 were detected by X^TX statistics (Günther and Coop 2013). We assigned accessions to genetic clusters

206 according to the highest ancestry coefficient estimated by *ALStructure* with the optimal *K* and calculated

207 X^TX by using *BAYPASS* ver2.1 (Gautier 2015). *BAYPASS* was run by setting 25 short pilot runs, 100,000

208 burn-in iterations and 100,000 post-burn-in iterations with a thinning interval of 40 iterations. A significance

209 threshold of X^TX was determined by the 99.5% quantile of pseudo-observed X^TX (Gautier 2015) calcu-

210 lated from neutral markers simulated by *simulate.baypass* (Gautier 2015).

211 **Gene ontology enrichment**   To investigate biological functions related to putatively adaptive loci, we

212 conducted gene ontology (GO) enrichment analysis with gene annotations of the barley 'Morex v2' genome

213 (Mascher 2019). Over-representation of GO terms for genes within 500 bp adjacent intervals of candidate

214 SNPs was tested by Fisher's exact test with 10,000 runs by using *SNP2GO* (Szkiba et al. 2014), and

8

215 GO terms with FDR < 0.05 were regarded as significantly enriched. Annotations of genes within 500 bp

216 upstream and downstream of the candidate SNPs were also reported.

Results

Summary of genotypic data SNP calling and preliminary filtration resulted in 101,711 SNPs for 1,365
accessions, including 1,121 IPK accessions and 244 B1K+ accessions (Supplementary Information; Fig-
ure S2). Depending on the requirements of analyses, we selected different subsets from 101,711 SNPs
as follows. For the joint population structure analysis of IPK and B1K+ accessions, we selected 4,793
geographically diverse SNPs that are polymorphic (minor allele frequency; MAF $\geq$ 0.05) among 72 IPK
accessions originating in 13 countries (Russell et al. 2016). This joint dataset is LD-pruned and has a
missing proportion of 0.043 (Supplementary Information; Figure S2). For analyses of B1K+ accessions, we
selected 58,616 SNPs with an overall missing proportion of 0.029 and maximal individual missing propor-
tion of 0.059, and further filtration resulted in 19,601 SNPs (LD-pruned; MAF $\geq$ 0.01) and 27,147 SNPs
(unpruned; MAF $\geq$ 0.05; Details in Supplementary Information; Figure S2)

Population structure and spatial genetic pattern Inference of population structure among B1K+
accessions with *ALStructure* (Cabreros and Storey 2019) identified four clusters (Figure 1 A and B), cor-
responding to the Mediterranean northern region, semi-arid coastal region, Judaean Desert, and Negev
Desert (Figure 1C). Hereafter, we named the four B1K+ clusters as North, Coast, Eastern Desert, and
Southern Desert. With *K = 4*, 174 of 244 (71.3 %) accessions have a highest ancestry coefficient less
than 0.9. The first three principal components (PCs) represented the clusters corresponding to the *AL-
Structure* results (Figure 1 A and B). On the first PC axis, the northern cluster separated from two desert
clusters, and on the second PC axis, the coastal cluster separated from the others. On the third PC axis,
the southern desert cluster separated from the eastern desert cluster. The three PC axes explained 4.73%,
3%, and 2.83% of the variation, respectively. A hierarchical population structure was evident in the NJ tree
(Figure 2A) and in the *ALStructure* analysis with K = 2-8 (Figure 2B). To evaluate the importance of marker
density and additional samples to population structure analysis, in addition to original dataset, we performed
PCA and *ALStructure* by either including or removing the HOH accessions with random selection of 100
and 5,000 SNPs. The dataset with 100 SNPs did not allow to identify genetic clusters while datasets with
5,000 SNPs separated into four genetic clusters by using the first four PCs even with the sample of original
B1K accessions without the HOH accessions. However, *ALStructure* could only identify three ancestral

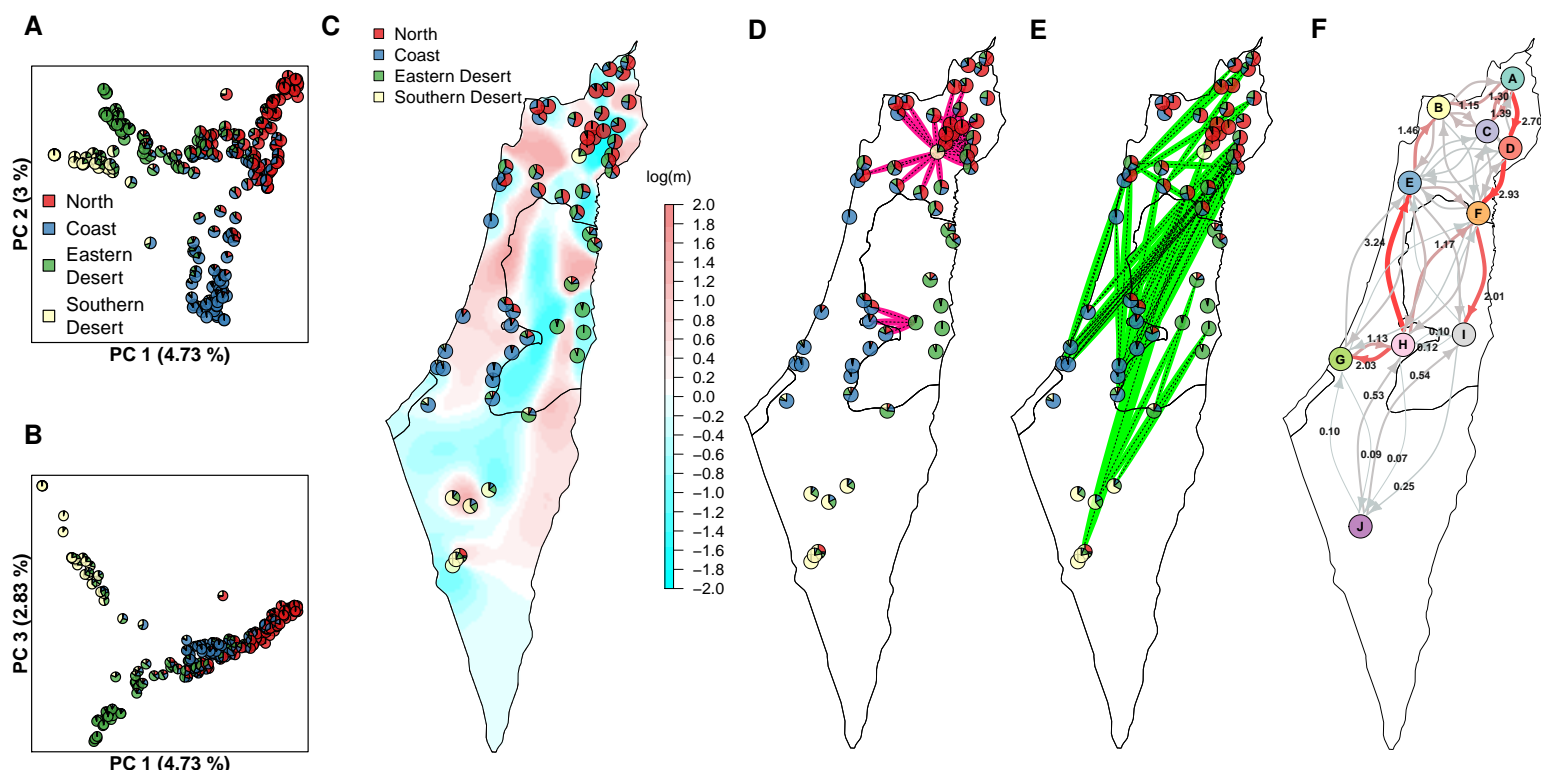**Table 1** F_{ST} and Nei's genetic distances between four genetic clusters.

		F_{ST}			
		North	Coast	Eastern Desert	Southern Desert
Nei's Distance	North	-	0.1124	0.1149	0.2593
	Coast	0.0209	-	0.1321	0.2533
	Eastern Desert	0.0216	0.0248	-	0.2125
	Southern Desert	0.0508	0.0482	0.0389	-

244 populations ($K = 3$) if the HOH accessions were excluded (File S2).

245   To evaluate genetic differentiation between four clusters, we computed pairwise F_{ST} and Nei's genetic

246 distances. The Southern Desert cluster is most strongly isolated from the other three clusters (Table 1).

247 With respect to the genomic pattern of differentiation, F_{ST} values were highest in the pericentromeric re-

248 gions of chromosome 2H, 3H, 4H, 5H, and 6H. The Southern Desert cluster differentiated from the other

249 three clusters in the most of genome except the pericentromeric regions of chromosome 3H and 4H (Figure

250 S7).

251   A joint PCA of B1K+ and IPK accessions was consistent with major clusters identified in B1K+ and

252 showed that B1K+ accessions overlapped with a large proportion of the IPK collection (Figure 3A). To

253 visualize the genetic relationship between B1K+ accessions and IPK accessions with different origins, we

254 selected 72 geographically distinct accessions used in a previous study (Russell et al. 2016). On the first PC

255 axis, most of the 72 geographically diverse accessions collected from western and central Asian countries

256 separated from B1K+ accessions but clustered more closely to the two desert clusters than to the northern

257 and coastal clusters (Figure 3A). Because an unbalanced sample size of accessions from Israel (616 of

258 1,365 accessions) may bias the PC axes, we performed another joint PCA by projecting 1,293 accessions

259 onto PC spaces of 72 geographically distinct accessions. This approach can avoid the misinterpretation

260 of sample origin and migration based on PC (McVean 2009). The PC projection showed that accessions

261 typically clustered by geographic origin, as reported in previous studies (Milner et al. 2019; Russell et al.

262 2016), and B1K+ accessions were concentrated in a small area of PC space (Figure 3B).

263 **Geographical pattern of gene flow**   To identify barriers limiting gene flow within the Levant region,

264 we performed an analysis using *EEMS* (Petkova et al. 2016) that revealed uneven gene flow across the

265 landscape. The area of low gene flow rates corresponds closely to geographic barriers, including the

266 Sea of Galilee, the Jordan Valley, and the Judea and Samaria mountain ridges (Figure 1C). A Wilcoxon
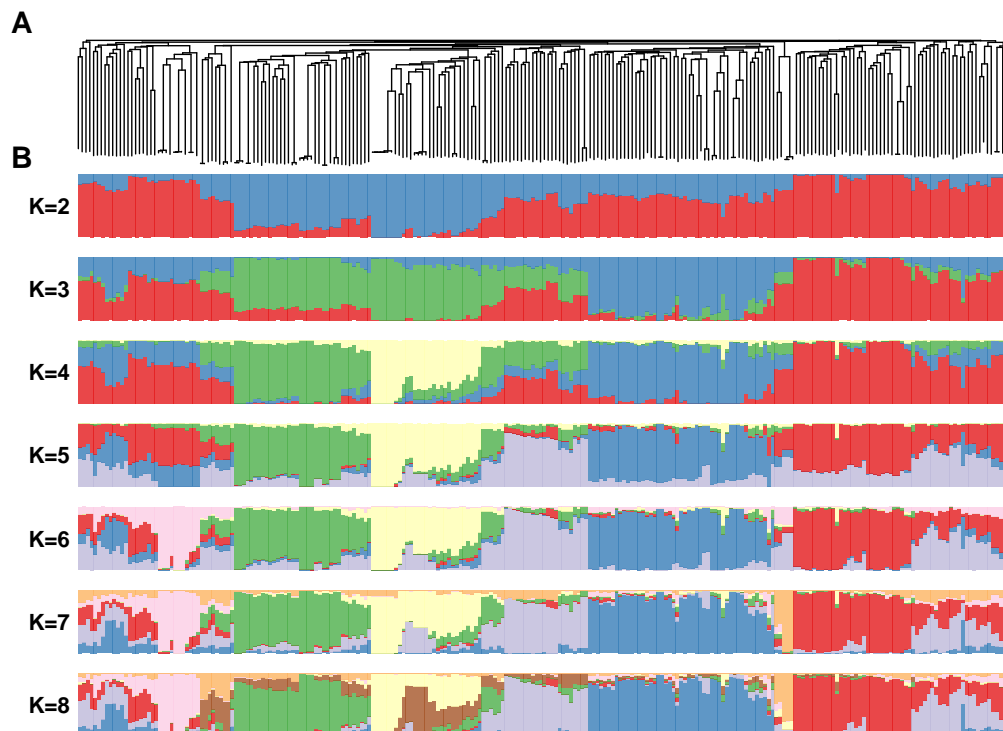
Figure 1 Spatial genetic structure of 244 B1K+ accessions and results of gene flow analysis. (A) PCA plot of the first and second PC axes. (B) PCA plot of the first and third PC axes. Pie charts in PCA plots represent ancestry coefficients of individuals estimated by *ALStructure* with *K = 4*. (C) Distribution of genetic clusters and effective migration surface. Average ancestry coefficients of individuals in collection sites are shown by pie charts. Color gradient represents gene flow rates estimated by *EEMS*. (D) Comparisons with *unPC* scores higher than the top 2.5% threshold that indicates a significantly low genetic similarity over a short geographical distance. (E) Comparisons with *unPC* scores lower than the bottom 2.5% threshold that indicates a significantly high genetic similarity over a long geographical distance. (F) Gene flow rates inferred by the coalescent-based inference method, representing the probabilities per unit of time that individuals in a region i are descended from a region j (Lundgren and Ralph 2019). The thickness of arrows and the depth of red color is proportional to gene flow rates. The complete result is shown in Table S3.

Figure 2 (A) Unrooted neighbor-joining (NJ) tree and (B) ancestry coefficients of 244 B1K+ accessions estimated by using *ALStructure* with *K = 2-8*. Accessions are sorted according to the NJ tree. With *K = 4*, red, blue, green, and yellow bars correspond to the Northern, Coastal, Eastern Desert and Southern Desert genetic clusters, respectively, as in Figure 1.

267 test supported the association between geographic barriers and lower gene flow rates ($p < 2.2 \times 10^{-22}$;

268 Figure S8 A). In addition, *EEMS* analysis showed that genetic dissimilarity between demes does not have

269 a simple linear relationship with geographic distances (Figure S8 B), indicating that isolation by distance

270 is not sufficient to explain genetic differentiation. The *EEMS* analysis also showed that effective genetic

271 diversity, which is the expected genetic dissimilarity of two individuals sampled from a site (Petkova et al.

272 2016), decreases from north to south (Figure S8 C), suggesting higher genetic diversity in the north than

273 the south. Furthermore, we performed *unPC* (House and Hahn 2018) based on the ratio of PC-based

274 genetic distances to geographic distances, which is more sensitive to long-distance migration than *EEMS*.

275 The high *unPC* score comparisons supported regions of low gene flow identified by *EEMS*, particularly the

276 majority of significant *unPC* comparisons located in the region around the Sea of Galilee in northern Israel

277 (Figure 1D). In addition, the comparisons with low *unPC* scores suggest potentially long-distance migration

278 events in the north-south direction (Figure 1E).

A

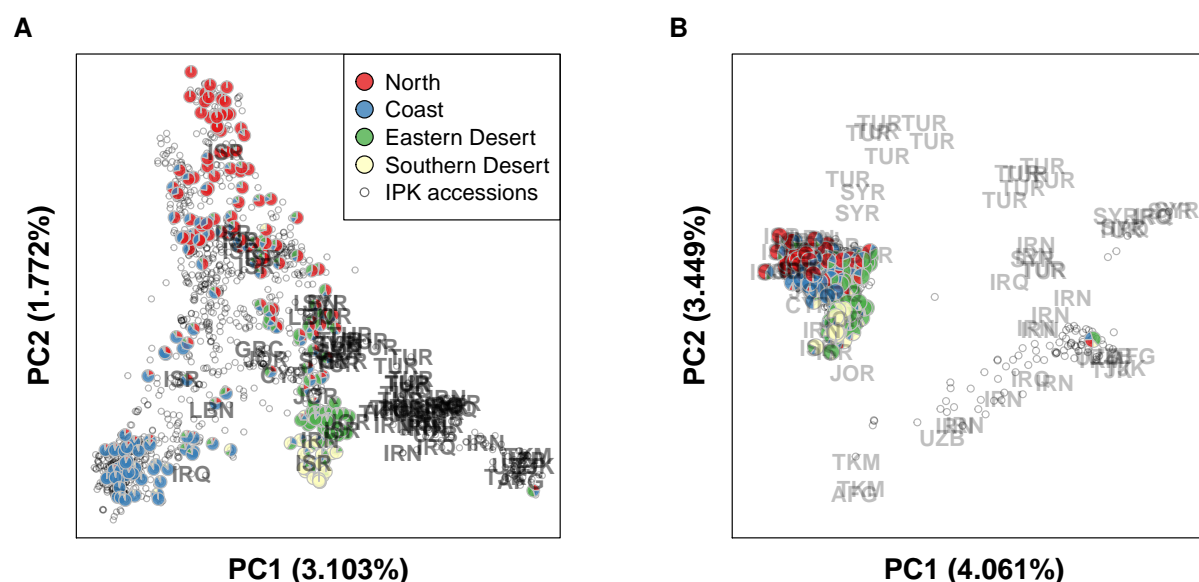**B**

PC1 (3.103%)

PC1 (4.061%)

Figure 3 PCA plots of 244 B1K+ accessions and 1,121 accessions from IPK genebank. (A) PCA performed with all of the available accessions. (B) PCA performed with 72 geographically diverse accessions and projecting the remaining accessions to PC spaces. Pie charts represent ancestry coefficients of 244 B1K+ accessions estimated by using *ALStructure* with *K = 4*. Gray open dots represent IPK accessions.

279   To evaluate asymmetric gene flows, we used CBI (Lundgren and Ralph 2019), which suggested unequal

280   gene flows in a North-South direction (Figure 1F; Table S3). There is a trend for gene flow from South

281   (region H) to North (region B) in the western region (region $H \to E \to B$; Figure 1F) and an opposite

282   trend from North (region A) to South (region I) in the eastern region (region $A \to D \to F \to I$; Figure

283   1F). The strongest gene flow (3.24 with the 95% credible interval of 0.51-6.32; Table S3) is observed from

284   populations close to Jerusalem (region H) to the surrounding areas of Mount Carmel (region E), but the

285   gene flow in the opposite direction (region $E \to H$) is much weaker (0.77 with the 95% credible interval

286   of 0-2.32; Table S3). Low gene flow rates of connections across geographical barriers, such as $C \rightleftharpoons D$

287   and $H \rightleftharpoons I$, agreed with the results of *EEMS* (Figure 1C) and *unPC* (Figure 1 D and E). Also, low gene

288   flow rates between the Negev desert (region J) and its adjacent regions indicated the isolation of Southern

289   Desert accessions, consistent with high genetic differentiation suggested by the F_{ST} values (Table 1).

290   **Genetic variation explained by environment and space**   To quantify relative contributions of envi-

291   ronment and space to genomic variation, simple RDA and partial RDA were performed. RDA showed that

292   environmental variables explained 15.12% ($R^2_{adj} = 0.107$; $p = 0.0002$) of SNP variation while spatial auto-

293   correlation captured by dbMEMs, which are eigenfunctions of a spatial network (Dray et al. 2006; Legendre

13

294 and Legendre 2012), explained 44.95% ($R^2_{adj} = 0.285$; $p = 0.0002$; Figure 4A). By using partial RDA, we

295 found 10.63% of SNP variation is jointly explained by environmental variables and spatial autocorrelation,

296 and 4.49% ($R^2_{adj} = 0.013$; $p = 0.0002$) was exclusively explained by environmental variables (Figure 4A).

297 Considering the confounding effect between environment and population structure, we treated ancestry co-

298 efficients (*K = 4*) as covariates when inspecting the effect of environmental variables on SNP variation. RDA

299 indicated that population structure explained 15.43% ($R^2_{adj} = 0.148$; $p = 0.0002$) of SNP variation and af-

300 ter controlling for population structure, environmental variables exclusively explained 8.71% ($R^2_{adj} = 0.048$;

301 $p = 0.0002$) of SNP variation (Figure 4A).
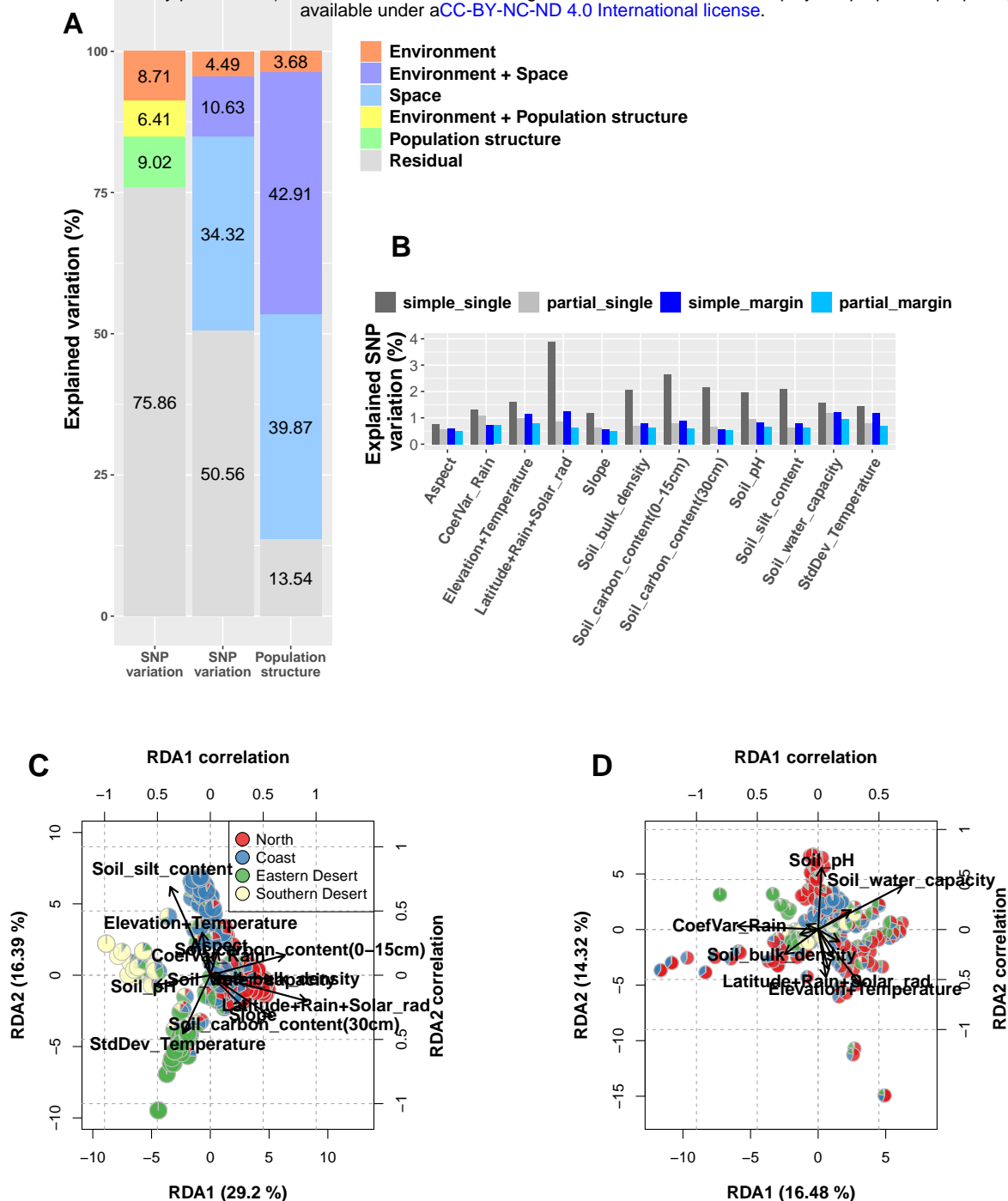
Figure 4 Results of variation partitioning and RDA biplots. (A) Variation partitioning of SNP variation and population structure. Left and middle columns: explained SNP variation estimated by RDA models using population structure and spatial autocorrelation as covariates, respectively. Right column: SNP variation explained by population structure. Environment, space and population structure are represented by twelve environmental variables, dbMEMs and ancestry coefficients ($K = 4$) in RDA models. (B) Percentage of SNP variation explained by environmental variables. The *simple_single* and *partial_single* show individual effects estimated based on RDA models fitting one environmental variable at a time. The *simple_margin* and *partial_margin* show marginal effects estimated based on RDA models fitting all environmental variables. The *partial_single* and *partial_margin* are estimated based on partial RDA conditioned on population structure. (C) Biplot of simple RDA. (D) Biplot of partial RDA conditioned on population structure. The arrows represent correlations of the environmental variables with RDA axes that are shown in Table S9 with details.

302   After confirming an association between genomic variation and the environment, we further investigated
303 the effects of individual environmental variables. In simple RDA models separately fitting each environmen-
304 tal variable, permutation tests showed all of 12 environmental variables are significantly associated with
305 SNP variation ($p < 0.005$; Table S5.1). Without constraining on population structure, the synthetic variable
306 '*Latitude+Rain+Solar_rad*' (Table S1 and S2) explained the highest proportion of SNP variation (3.89%; Fig-
307 ure 4 B; Table S5.1). In contrast, in RDA models controlling for population structure, '*Soil_water_capacity*'
308 explained the highest proportion of SNP variation (1.18%; Figure 4 B; Table S5.2) whereas the proportion of
309 SNP variation explained by '*Latitude+Rain+Solar_rad*' reduced to 0.86%. The variable '*Aspect*' showed the
310 lowest but significant association with SNP variation in both simple and partial RDA (Table S5.1 and S5.2).
311 Variables '*Soil_water_capacity*', '*CoefVar_Rain*', which refers to coefficients of variation of precipitation of
312 the growing season, and '*Aspect*' reduced explained variation less than other environmental variables af-
313 ter controlling the population structure. This indicates that they are less correlated with the population
314 structure. We also investigated marginal effects in models that incorporated all environmental variables by
315 considering correlations between environmental variables. The variables '*Latitude+Rain+Solar_rad*' and
316 '*Soil_water_capacity*' showed again the highest marginal effect in the simple RDA and partial RDA condi-
317 tioned on population structure, respectively (Figure 4 B; Table S5.3 and S5.4).

318   RDA biplots provide further information on the relative importance of environmental gradients. The biplot
319 of the simple RDA (Figure 4C) showed a population structure consistent with the four genetic clusters
320 identified by *ALStructure*. The first and second RDA axes corresponded to genetic differentiation in the
321 north-south and west-east directions, respectively (Figure 4C), and the first RDA axis is strongly (r =
322 0.911; Table S5) correlated with the variable '*Latitude+Solar_rad*' (Figure 4C). After conditioning on spatial
323 autocorrelation, the effects of all environmental variables decreased significantly (Figure S9 and Table S5),
324 indicating a strong correlation of environmental gradients with spatial autocorrelation. When conditioning on
325 population structure rather than spatial autocorrelation, two water-related variables, "*Soil_water_capacity*"
326 (r = 0.697) and "*CoefVar_Rain*" (r = -0.662), were the most influential predictors on the first RDA axis (Figure
327 4D and Table S5).

328   With reference to our hypothesis that the diverse environments in the southern Levant are an important
329 factor shaping populations, we quantified the relative contributions of environment and space to population
330 structure ($K = 4$) using RDA. As expected, a high proportion of population structure explained by envi-
331 ronmental variables (42.91 of 46.59% Figure 4A) cannot be separated from the component explained by
332 spatial autocorrelation. Only 3.68% ($R_{adj}^2 = 0.0358$; $p = 0.0002$) of population structure can be exclusively
333 explained by environments whereas spatial autocorrelation exclusively accounted for 39.87% ($R_{adj}^2 = 0.374$;

$p = 0.0002$) of population structure (Figure 4A). This result suggests that spatial autocorrelation has a larger effect on population differentiation of wild barley in the Southern Levant than environmental diversity.

Adaptive candidates and GO enrichment The association between genomic variation and environment motivated us to perform genome scans to identify putative adaptive loci. We first estimated LD decay to assess whether the marker density of the reduced sequenced library was sufficient to accurately identify adaptive genes in such scans. Fitting the *loess* model and Hill-Weir formula, we observed a rapid decay of LD because r^2 values dropped to half of the highest values of 0.377 and 0.454 after pairwise SNP distances of 213 bp and 125 bp, respectively (Figure S10). This result indicates a possible difficulty in detecting precise locations of adaptive loci except for closely linked loci with the current marker density.

Three GEA methods, simple RDA, partial RDA, and LFMM, identified 352, 364, and 307 candidate SNPs (FDR < 0.05), respectively, and the outlier method, *BAYPASS*, identified 279 candidate SNPs ($X^T X$ > 11.05). However, candidate SNPs detected by the four methods hardly overlapped, except simple RDA and *BAYPASS* with 125 common SNPs, of which 91 are located in the pericentromeric regions of the chromosome 3H, 4H, and 5H (Figure 5; File S3). By searching 500 bp adjacent intervals of candidate SNPs, the four methods jointly identified two genes on chromosome 4H. The first gene *HORVU.MOREX.r2.4HG0308420* is close to SNPs associated with the variable '*Latitude+Rain+Solar_rad*' in the LFMM analysis (File S4) and encodes an ATP-dependent RNA helicase. The second gene *HORVU.MOREX.r2.4HG0314300* is linked to SNPs associated with '*Elevation+Temperature*' and which encodes a nucleolar GTP-binding protein (Figure S11; Table S6; File S4). GO term enrichment analysis identified 2 and 10 enriched GO terms based on candidate SNPs detected by simple RDA and *BAYPASS*, respectively (Table S7). No GO term was enriched based on the results of partial RDA and LFMM.

Discussion

Our study indicated that geography and spatial autocorrelation are more predictive factors of genomic variation than environmental gradients even though the diverse environments of the Southern Levant are expected to impose strong natural selection (Hübner et al. 2009; Nevo et al. 1979). These findings imply that genomic variation of wild barley in the Southern Levant is mainly driven by neutral processes consistent with a neutralist perspective (e.g., Volis et al. 2003, 2005, 2001). However, environmental variables are still associated with a relatively small but considerable proportion of genomic variation (15.12%; Figure 4A), suggesting that natural selection and hitch-hiking may have a detectable effect on the structure of genetic
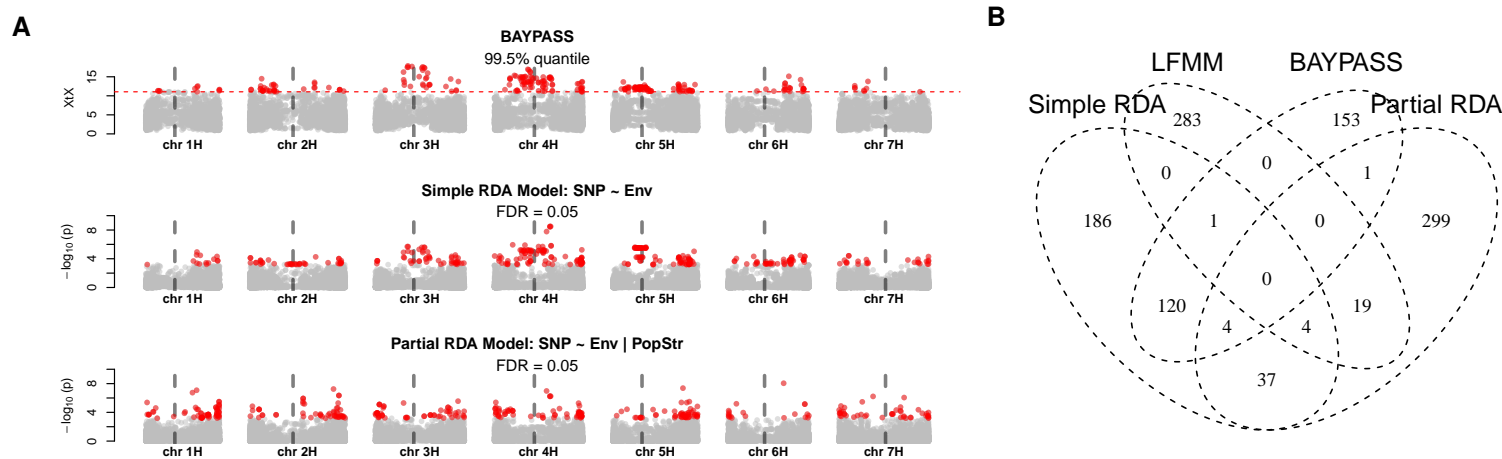
17

Figure 5 (A) Genome scans for adaptation signatures. Three manhattan plots correspond to the *BAYPASS*, simple RDA, and partial RDA conditioned on population structure. Significant SNPs are highlighted as red dots. The positions of centromeres are indicated with vertical gray dash lines. (B) Numbers of significant SNPs detected by four different methods for genome scans.

363 diversity.

364 **Population structure of B1K+ and IPK genebank collection** Three clusters corresponding to

365 eco-geographic habitats were previously characterized based on SSR markers and a SNP array developed

366 for cultivated barley (Hübner et al. 2012) and morphological traits (Hübner et al. 2013). Our results are

367 consistent with previous findings, except that the previously reported desert cluster (Hübner et al. 2012)

368 split into two clusters (Figure 1 A-C), evident in the *ALStructure* analysis with $K = 3$ and $K = 4$ (Figure 2),

369 due to not only the increased marker number but also the inclusion of additional accessions collected in the

370 Negev Desert in 2011 (File S2). Although genetic clusters are consistent with eco-geographic habitats, the

371 results of model-based methods should be interpreted cautiously. First, the number of ancestral populations

372 may be overestimated due to isolation by distance (Bradburd et al. 2018). Second, the high proportion of

373 admixed accessions (174 of 244 B1K+ accessions; 71.3%) may not result from admixture. Both spatial

374 autocorrelation (Bradburd et al. 2018) and demographic history (Lawson et al. 2018) such as bottlenecks

375 that likely occur in self-pollinating species (Hartfield et al. 2017), may lead to high admixture proportions in

376 model-based methods.

377 The joint PCA incorporating the IPK wild barley collection indicated a strong effect of an unbalanced

378 sample size of accessions from Israel (616 of 1,365 accessions) on a PCA (Figure 3 A and B). This

379 comparison highlights the importance of balanced sampling in analyzing the population structure, because

380 unequal sample sizes among groups can lead to distortion of PCs (McVean 2009). The PCA based on all

18

accessions compressed the accessions with a broad geographical origin throughout the whole distribution range of wild barley to a cluster (Figure 3A) that did not appropriately reflect their wide geographic origin. In contrast, a PCA with a more balanced sample of accessions from the whole species range revealed that the wild barley from the Southern Levant regions comprises only a small proportion of the total diversity of wild barley (Figure 3B). However, the PCA of the complete sample revealed that accessions from Greece and Cyprus clustered with accessions from the Southern Levant (Figure 3 A and S12 A) suggesting they originated from the Southern Levant or adjacent areas without a sufficiently long history of differentiation from the ancestral populations. Likewise, 579 IPK accessions with unknown origins may be closely related to the Levant region as they strongly overlapped with our B1K+ population (Figure S12).

Geographical pattern of gene flow Gene flow of wild barley is expected to be limited with a low rate of outcrossing (<2%; Abdel-Ghani et al. 2004), and seed dispersal occurs mainly within 1.2 m (Volis et al. 2010). However, self-fertilizing plants can establish a population with a single seed after long-distance dispersal (Baker 1967), and the long spiky awns attached to seeds of wild barley can facilitate dispersal by zoochory. With a sufficiently long period, gene flow across landscapes may accumulate via occasional dispersal and outcrossing. *EEMS* (Petkova et al. 2016) was previously used to identify gene flow barriers in plant populations across large geographical ranges, e.g., rice (Gutaker et al. 2020) and spruce (Tsuda et al. 2016). In our data, *EEMS* revealed fine-scale patterns of gene flow attributable to geography, particularly by the Sea of Galilee and the Jordan Valley, which were not previously identified by inferring gene flow between genetic clusters (Hübner et al. 2012). These geographic separations could promote genetic differentiation within a short geographic distance that interferes with isolation-by-distance patterns (Figure S8 B).

Coalescent-based inference (CBI; Lundgren and Ralph 2019) detected trends of gene flow with opposite directions in eastern and western regions (Figure 1F), contradicting the net gene flow from north to south identified by Hübner et al. (2012). The different conclusions regarding gene flow directions in western Israel are likely due to the mappings. While Hübner et al. (2012) assigned accessions according to genetic clustering, our assignment emphasized the geographic origin. CBI gene flow rates express the probability that a population descended from another population per unit time (Lundgren and Ralph 2019), and therefore a history of recent colonization may explain gene flow trends in our data. Incorporating historical genome recombination to infer gene flow at different time periods may provide a clearer picture (Al-Asadi et al. 2019). Errors in gene flow inference can result from sampling biases, missing and erroneous genotypic values caused by low sequencing depth, and also from uneven distributions of markers due to the nature of GBS (Elshire et al. 2011; Poland et al. 2012). However, imbalanced sampling should not bias our

19

412 results because *EEMS* and CBI are insensitive to unequal sample numbers (Lundgren and Ralph 2019;

413 Petkova et al. 2016).

Effects of environment and geographical distance on SNP variation RDA analysis indicated

415 that environmental gradients explained a substantial portion of SNP variation and population structure (Fig-

416 ure 4A). This analysis does not include all possible environmental effects because comprehensive environ-

417 mental data were not available. For example, the adaptive trait *drought stress recovery* is associated with

418 the parameter rainfall predictability in wild barley (Galkin et al. 2018), but such data are only available for

419 some collection sites. In addition, the control for collinearity and nonlinear environmental effects that RDA

420 does not account for may lead to unexplained genetic variation in our analysis.

421 Phenotypic studies suggested the importance of rainfall in the evolution of wild barley in the southern

422 Levant (Galkin et al. 2018; Hübner et al. 2013; Volis 2011; Volis et al. 2002a,b). Our RDA analysis indi-

423 cated variables related to water availability ('*Latitude+Rain+Solar_rad*' and '*Soil_water_capacity*') as most

424 important drivers of genomic variation (Figure 4 B-D; Table S5). It is not possible to specify the effects of

425 individual environmental gradients because they are highly correlated. For example, we cannot separate

426 the effect of precipitation from latitude, which is highly relevant to the flowering timing of barley (Russell

427 et al. 2016). Unlike other environmental variables, '*Aspect*' had few confounding effects with other gradi-

428 ents and population structure (Figure 4B). '*Aspect*' was the strongest predictor when conditioned on spatial

429 autocorrelation (Figure S9; Table S5). In the southern Levant, south-facing slopes might be more exposed

430 to drought and heat than north-facing slopes due to higher solar radiation, resulting in significantly stronger

431 selection within only a few hundred meters, referred to as the Evolution Canyon model (Bedada et al. 2014;

432 Nevo et al. 2005). Our results suggest that '*Aspect*' may reflect a minor but pervasive effect of microclimate

433 in the southern Levant that cannot be represented by climate data at the current resolution. In *Mimulus*

434 *guttatus*, an important locus of microgeographic adaptation was successfully identified by integrating quan-

435 titative trait loci mapping and population genomic analyses (Hendrick et al. 2016). A similar approach may

436 be used to investigate the genetic architecture of adaptation to microclimatic conditions in wild barley.

437 By using dbMEMs, which model the effects of spatial autocorrelation on SNP variation, our RDA re-

438 vealed that high proportions of SNP variation (45%) and population structure (83%) are explained by spa-

439 tial autocorrelation (Figure 4A). The lower proportion of SNP variation explained by environments (Figure

440 4A) indicates that environmental selection may be an influential but not dominant driver of genetic differ-

441 entiation. In contrast with our findings, environment had a significantly stronger effect than geographical

442 distance on diversity in *Boechera stricta* (Lee and Mitchell-Olds 2011). However, in *Arabidopsis thaliana*

443 (Lasky et al. 2012), sorghum (Lasky et al. 2015), rice (Gutaker et al. 2020) and wild tomato (Gibson and
444 Moyle 2020), the contribution of environment was comparable and highly overlapping with geographical
445 distance. This suggests that the isolation-by-distance is a robust and widespread pattern in a small ge-
446 ographic range like in our case and over large geographic scale (Gibson and Moyle 2020; Gutaker et al.
447 2020; Lasky et al. 2012, 2015). Complex spatial structure confounding with environmental gradients are a
448 pervasive challenge in the study of local adaptation (de Villemereuil et al. 2014; Excoffier et al. 2009). In
449 particular, population genetic analyses tend to be biased by spatial structure (Battey et al. 2020). For this
450 reason, phenotypic studies using crosses between accessions and common garden experiments are also
451 required to distinguish between genetic variation attributed to local adaptation and spatial autocorrelation.
452 Additionally, we noted that a high percentage of SNP variation (51%; Figure 4A) remained unexplained
453 even after incorporating dbMEMs. This could attribute to either unknown evolutionary forces independent
454 of spatial autocorrelation or the limitation of our current dataset and methodologies.

455 **Lack of strong evidence to pinpoint adaptive loci** The rapid decay of LD within a few hundred
456 base pairs (Figure S10) is consistent with similar studies of wild barley populations from the Middle East
457 and Central Asia (Morrell et al. 2005). Reduced-representation sequencing approaches tend to have lim-
458 ited power in identifying adaptive loci, especially for genomes with high levels of recombination (Tiffin and
459 Ross-Ibarra 2014). A rapid LD decay and a large genome size of ∼5.3 Gb indicate that the marker density
460 of this study may not allow precise genome scans. To account for this caveat and to control for false-positive
461 rates, we combined the results from multiple methods of genome scans (Forester et al. 2018; Rellstab et al.
462 2015). Significant polymorphisms hardly overlapped between methods (Figure 5B). This observation may
463 be explained by a (1) lack of adaptive loci with large effects, (2) strong confounding effect of population
464 structure, and (3) limitations of the dataset. Although without strong evidence to identify adaptation genes,
465 the genome scans based on $X^T X$ and simple RDA identified significant correlations with environmental
466 variables and strong genetic differentiation at the pericentromeric regions of chromosome 3H, 4H, and 5H
467 (Figure 5A). However, these associations were not observed in the partial RDA and LFMM analyses. Al-
468 though the $X^T X$ statistics accounts for covariance of allele frequencies (i.e., population structure) among
469 populations (Günther and Coop 2013), spurious signals of selection may arise if self-fertilization inflates
470 false-positive values via strong genetic drift (Hodgins and Yeaman 2019). For this reason and because of
471 a strong association of population structure with environments (Figure 4A), false positives are expected for
472 the outlier and GEA methods even with a correction for population structure. The high degree of putative
473 selection-driven differentiation is remarkable. Similar patterns of genetic differentiation in pericentromeric

474 regions were reported in previous studies of barley (Contreras-Moreira et al. 2019; Fang et al. 2014; Wang

475 et al. 2018), teosinte (Pyhäjärvi et al. 2013) and maize (Navarro et al. 2017). Theoretical studies suggested

476 that adaptation with gene flow could result in divergent linkage groups of locally beneficial alleles in low re-

477 combination regions (Akerman and Bürger 2014; Bürger and Akerman 2011; Yeaman and Whitlock 2011).

478 These conclusions were supported by simulation and empirical studies, e.g., in stickleback, sunflower, and

479 *Arabidopsis lyrata* (Berner and Roesti 2017; Hämälä and Savolainen 2019; Renaut et al. 2013; Samuk et al.

480 2017). Low-recombination pericentromeric regions of wild barley were reported to have significantly higher

481 ratios of non-synonymous to synonymous substitution (π_a/π_s) than other genomic regions (Baker et al.

482 2014), suggesting the tendency of accumulating genetic load in pericentromeric regions. Moreover, consid-

483 ering the conditional neutrality, accumulation of conditionally deleterious mutations in habitats where they

484 are neutral can lead to genotype-environment interactions of fitness if migration is weak relative to genetic

485 drift (Mee and Yeaman 2019). Taken together, given a weak gene flow, high rates of self-fertilization and

486 variable recombination rates over the genome, a long-term accumulation of conditionally deleterious muta-

487 tions may result in locally neutral linkage of alleles in low-recombination genomic regions creating a pattern

488 of polymorphism that may resemble local adaptation and explain our observations in the pericentromeric

489 regions.

490 **Conclusion and outlook** We observed a stronger effect of non-selective factors like geography and

491 isolation-by-distance on total genetic diversity in the wild barley populations of the diverse and stressful en-

492 vironments of the Southern Levant. Nevertheless, natural selection has a small but significant influence on

493 genomic variation, which is potentially valuable for barley breeding because water availability, i.e., precipita-

494 tion and soil water capacity, was the most strongly correlated environmental variable. Outlier test and simple

495 RDA identified genomic regions that may contribute to local adaptation, but these regions were not robustly

496 identified by the different tests applied. One limitation of our study is therefore that only a small proportion

497 of the wild barley genome was sequenced by the GBS approach, which is suitable for analyzing genome-

498 wide patterns of variation and mapping of causal genes (Milner et al. 2019), but not powerful enough for

499 pinpointing genomic targets of local adaptation. In the near future, whole genome sequencing of wild barley

500 accessions (Sato et al. 2021) and the development of a barley pangenome (Jayakodi et al. 2020) will greatly

501 increase the power of population genomic approaches to understand wild barley adaptation and facilitate

502 the mining of useful alleles for plant breeding. Such approaches can be combined with common garden and

503 transplantation experiments of wild barley genotypes to measure fitness effects in different environments

504 (Hübner et al. 2013; Volis 2011), gene expression studies of differentially adapted genotypes (Hübner et al.

22

505 2015) and mapping populations. Given the strong influence of isolation-by-distance on genomic varia-

506 tion, adaptive genetic variation is likely confounded with population structure. Mapping populations with

507 sufficient genome recombination evaluated in different environments allow for disentangling adaptive and

508 neutral variation, as shown in such populations developed from wild and cultivated barley (Herzig et al.

509 2018; Wiegmann et al. 2019). Whole genome resequencing followed by computational analysis can be

510 rationalized to analyze a large number of genotypes such as the complete B1K population. Therefore, we

511 believe that population genomic analysis of differentially adapted crop-wild relatives will complement other

512 approaches to understanding plant adaptation and enable use of this information for breeding.

Acknowledgements

Conflict of Interest

522 The authors declare no competing interests.

Data archiving

524 (The FASTQ files will be archived at European Nucleotide Archive. The geographic coordinates and envi-

525 ronmental data is available in the File S1. The R codes will be archived at https://gitlab.com/kjschmidlab.)

23

References

Abdel-Ghani, A. H., Parzies, H. K., Omary, A. and Geiger, H. H. (2004). Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan, *Theoretical and Applied Genetics* **109**(3): 588–595.

Akerman, A. and Bürger, R. (2014). The consequences of gene flow for local adaptation and differentiation: a two-locus two-deme model, *Journal of mathematical biology* **68**(5): 1135–1198.

Al-Asadi, H., Petkova, D., Stephens, M. and Novembre, J. (2019). Estimating recent migration and population-size surfaces, *PLoS genetics* **15**(1): e1007908.

Baker, H. G. (1967). Support for Baker's law—as a rule, *Evolution* **21**(4): 853–856.

Baker, K., Bayer, M., Cook, N., Dreißig, S., Dhillon, T., Russell, J., Hedley, P. E., Morris, J., Ramsay, L., Colas, I. et al. (2014). The low-recombining pericentromeric region of barley restricts gene diversity and evolution but not gene expression, *The Plant Journal* **79**(6): 981–992.

Battey, C., Ralph, P. L. and Kern, A. D. (2020). Space is the place: Effects of continuous spatial structure on analysis of population genetic data, *Genetics* **215**(1): 193–214.

Baum, M., Grando, S., Backes, G., Jahoor, A., Sabbagh, A. and Ceccarelli, S. (2003). QTLs for agronomic traits in the mediterranean environment identified in recombinant inbred lines of the cross' Arta'$\times$ *H. spontaneum* 41-1, *Theoretical and Applied Genetics* **107**(7): 1215–1225.

Bedada, G., Westerbergh, A., Nevo, E., Korol, A. and Schmid, K. J. (2014). DNA sequence variation of wild barley *Hordeum spontaneum* (L.) across environmental gradients in israel, *Heredity* **112**(6): 646–655.

Berner, D. and Roesti, M. (2017). Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate, *Molecular Ecology* **26**(22): 6351–6369.

Bhatia, G., Patterson, N., Sankararaman, S. and Price, A. L. (2013). Estimating and interpreting FST: the impact of rare variants, *Genome research* **23**(9): 1514–1521.

Bradburd, G. S., Coop, G. M. and Ralph, P. L. (2018). Inferring continuous and discrete population genetic structure across space, *Genetics* **210**(1): 33–52.

Bürger, R. and Akerman, A. (2011). The effects of linkage and gene flow on local adaptation: A two-locus continent–island model, *Theoretical population biology* **80**(4): 272–288.

24

553 Cabreros, I. and Storey, J. D. (2019). A likelihood-free estimator of population structure bridging admixture
554   models and principal components analysis, *Genetics* **212**(4): 1009–1029.

555 Caldwell, K. S., Russell, J., Langridge, P. and Powell, W. (2006). Extreme population-dependent linkage
556   disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*, *Genetics* **172**(1): 557–567.

557 Capblancq, T., Luu, K., Blum, M. G. and Bazin, E. (2018). Evaluation of redundancy analysis to identify
558   signatures of local adaptation, *Molecular ecology resources* **18**(6): 1223–1233.

559 Caye, K., Jumentier, B., Lepeule, J. and François, O. (2019). LFMM 2: fast and accurate inference of gene-
560   environment associations in genome-wide studies, *Molecular biology and evolution* **36**(4): 852–860.

561 Contreras-Moreira, B., Serrano-Notivoli, R., Mohammed, N. E., Cantalapiedra, C. P., Beguería, S., Casas,
562   A. M. and Igartua, E. (2019). Genetic association with high-resolution climate data reveals selection foot-
563   prints in the genomes of barley landraces across the Iberian Peninsula, *Molecular ecology* **28**(8): 1994–
564   2012.

565 Dai, F., Nevo, E., Wu, D., Comadran, J., Zhou, M., Qiu, L., Chen, Z., Beiles, A., Chen, G. and Zhang, G.
566   (2012). Tibet is one of the centers of domestication of cultivated barley, *Proceedings of the National*
567   *Academy of Sciences* **109**(42): 16969–16973.

568 Dawson, I. K., Russell, J., Powell, W., Steffenson, B., Thomas, W. T. and Waugh, R. (2015). Barley: a
569   translational model for adaptation to climate change, *New Phytologist* **206**(3): 913–931.

570 de Villemereuil, P., Frichot, É., Bazin, É., François, O. and Gaggiotti, O. E. (2014). Genome scan meth-
571   ods against more complex models: when and how much should we trust them?, *Molecular Ecology*
572   **23**(8): 2006–2019.

573 Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G.,
574   Legendre, P., Madi, N. and Wagner, H. H. (2019). *adespatial: Multivariate Multiscale Spatial Analysis*. R
575   package version 0.3-7.
576   **URL:** *https://CRAN.R-project.org/package=adespatial*

577 Dray, S., Legendre, P. and Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for
578   principal coordinate analysis of neighbour matrices (PCNM), *ecological modelling* **196**(3-4): 483–493.

579 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. (2011).
580   A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, *PloS one* **6**(5).

Excoffier, L., Hofer, T. and Foll, M. (2009). Detecting loci under selection in a hierarchically structured population, *Heredity* **103**(4): 285–298.

Fang, Z., Gonzales, A. M., Clegg, M. T., Smith, K. P., Muehlbauer, G. J., Steffenson, B. J. and Morrell, P. L. (2014). Two genomic regions contribute disproportionately to geographic differentiation in wild barley, *G3: Genes, Genomes, Genetics* **4**(7): 1193–1203.

Fick, S. E. and Hijmans, R. J. (2017). Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas, *International journal of climatology* **37**(12): 4302–4315.

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L. and Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes, *Molecular ecology* **25**(1): 104–120.

Forester, B. R., Lasky, J. R., Wagner, H. H. and Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations, *Molecular Ecology* **27**(9): 2215–2233.

Galkin, E., Dalal, A., Evenko, A., Fridman, E., Kan, I., Wallach, R. and Moshelion, M. (2018). Risk-management strategies and transpiration rates of wild barley in uncertain environments, *Physiologia plantarum* **164**(4): 412–428.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates, *Genetics* **201**(4): 1555–1579.

Gibson, M. J. S. and Moyle, L. C. (2020). Regional differences in the abiotic environment contribute to genomic divergence within a wild tomato species, *Molecular Ecology* .
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15477*

Günther, T. and Coop, G. (2013). Robust identification of local adaptation from allele frequencies, *Genetics* **195**(1): 205–220.

Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., Slayton, E. R., Wilkins, O., Castillo, C. C., Negrão, S. et al. (2020). Genomic history and ecology of the geographic spread of rice, *Nature Plants* **6**(5): 492–502.

Hämälä, T. and Savolainen, O. (2019). Genomic patterns of local adaptation under gene flow in *arabidopsis lyrata*, *Molecular Biology and Evolution* **36**(11): 2557–2571.

Harlan, J. R. and Zohary, D. (1966). Distribution of wild wheats and barley, *Science* **153**(3740): 1074–1080.

Hartfield, M., Bataillon, T. and Glémin, S. (2017). The evolutionary interplay between adaptation and self-fertilization, *Trends in Genetics* **33**(6): 420–431.

Hendrick, M. F., Finseth, F. R., Mathiasson, M. E., Palmer, K. A., Broder, E. M., Breigenzer, P. and Fishman, L. (2016). The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in yellowstone mimulus guttatus, *Molecular Ecology* **25**(22): 5647–5662.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B. et al. (2017). Soilgrids250m: Global gridded soil information based on machine learning, *PLoS one* **12**(2): e0169748.

Herzig, P., Maurer, A., Draba, V., Sharma, R., Draicchio, F., Bull, H., Milne, L., Thomas, W. T. B., Flavell, A. J. and Pillen, K. (2018). Contrasting genetic regulation of plant development in wild barley grown in two European environments revealed by nested association mapping, *Journal of Experimental Botany* **69**(7): 1517–1531. Publisher: Oxford Academic.
URL: *https://academic.oup.com/jxb/article/69/7/1517/4816162*

Hill, W. and Weir, B. (1988). Variances and covariances of squared linkage disequilibria in finite populations, *Theoretical population biology* **33**(1): 54–78.

Hodgins, K. A. and Yeaman, S. (2019). Mating system impacts the genetic architecture of adaptation to heterogeneous environments, *New Phytologist* .

House, G. L. and Hahn, M. W. (2018). Evaluating methods to visualize patterns of genetic differentiation on a landscape, *Molecular ecology resources* **18**(3): 448–460.

Hübner, S., Bdolach, E., Ein-Gedy, S., Schmid, K., Korol, A. and Fridman, E. (2013). Phenotypic landscapes: phenological patterns in wild and cultivated barley, *Journal of Evolutionary Biology* **26**(1): 163–174.

Hübner, S., Günther, T., Flavell, A., Fridman, E., Graner, A., Korol, A. and Schmid, K. J. (2012). Islands and streams: clusters and gene flow in wild barley populations from the Levant, *Molecular ecology* **21**(5): 1115–1129.

635 Hübner, S., Höffken, M., Oren, E., Haseneyer, G., Stein, N., Graner, A., Schmid, K. and Fridman, E.
636 (2009). Strong correlation of wild barley (*Hordeum spontaneum*) population structure with temperature
637 and precipitation variation, *Molecular Ecology* **18**(7): 1523–1536.

638 Hübner, S., Korol, A. B. and Schmid, K. J. (2015). Rna-seq analysis identifies genes associated with
639 differential reproductive success under drought-stress in accessions of wild barley hordeum spontaneum,
640 *BMC plant biology* **15**(1): 1–14.

641 Jakob, S. S., Rödder, D., Engler, J. O., Shaaf, S., Özkan, H., Blattner, F. R. and Kilian, B. (2014). Evolu-
642 tionary history of wild barley (*Hordeum vulgare* subsp. *spontaneum*) analyzed using multilocus sequence
643 data and paleodistribution modeling, *Genome biology and evolution* **6**(3): 685–702.

644 Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., Lux, T., Kamal, N.,
645 Lang, D., Himmelbach, A., Ens, J., Zhang, X.-Q., Angessa, T. T., Zhou, G., Tan, C., Hill, C., Wang, P.,
646 Schreiber, M., Boston, L. B., Plott, C., Jenkins, J., Guo, Y., Fiebig, A., Budak, H., Xu, D., Zhang, J., Wang,
647 C., Grimwood, J., Schmutz, J., Guo, G., Zhang, G., Mochida, K., Hirayama, T., Sato, K., Chalmers, K. J.,
648 Langridge, P., Waugh, R., Pozniak, C. J., Scholz, U., Mayer, K. F. X., Spannagl, M., Li, C., Mascher, M.
649 and Stein, N. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding, *Nature*
650 pp. 1–6. Publisher: Nature Publishing Group.
651 **URL:** *https://www.nature.com/articles/s41586-020-2947-8*

652 Kawecki, T. J. and Ebert, D. (2004). Conceptual issues in local adaptation, *Ecology letters* **7**(12): 1225–
653 1241.

654 Kilian, B., Özkan, H., Kohl, J., von Haeseler, A., Barale, F., Deusch, O., Brandolini, A., Yucel, C., Martin, W.
655 and Salamini, F. (2006). Haplotype structure at seven barley genes: relevance to gene pool bottlenecks,
656 phylogeny of ear type and site of barley domestication, *Molecular Genetics and Genomics* **276**(3): 230–
657 241.

658 Lasky, J. R., Des Marais, D. L., McKAY, J. K., Richards, J. H., Juenger, T. E. and Keitt, T. H. (2012).
659 Characterizing genomic variation of *arabidopsis thaliana*: the roles of geography and climate, *Molecular*
660 *Ecology* **21**(22): 5512–5529.

661 Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., Juenger, T. E., Hyma,
662 K., Acharya, C., Mitchell, S. E. et al. (2015). Genome-environment associations in sorghum landraces
663 predict adaptive traits, *Science advances* **1**(6): e1400218.

28

Lawson, D. J., Van Dorp, L. and Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots, *Nature Communications* **9**(1): 1–11.

Lee, C.-R. and Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation, *Molecular ecology* **20**(22): 4631–4642.

Leek, J. T. (2011). Asymptotic conditional singular value decomposition for high-dimensional genomic data, *Biometrics* **67**(2): 344–352.

Legendre, P. and Legendre, L. (2012). Canonical analysis, *Numerical ecology, 3rd English Edition*, The Netherlands: Elsevier Science BV, chapter 11, pp. 625–710.

López-Goldar, X. and Agrawal, A. A. (2021). Ecological interactions, environmental gradients, and gene flow in local adaptation, *Trends in Plant Science* **0**(0). Publisher: Elsevier.
URL: *https://www.cell.com/trends/plant-science/abstract/S1360-1385(21)00065-0*

Lundgren, E. and Ralph, P. L. (2019). Are populations like a circuit? Comparing isolation by resistance to a new coalescent-based method, *Molecular ecology resources* **19**(6): 1388–1406.

Makowski, D., Ben-Shachar, M. and Lüdecke, D. (2019). bayestestR: Describing effects and their un-certainty, existence and significance within the Bayesian framework, *Journal of Open Source Software* **4**(40): 1541.

Mascher, M. (2019). Pseudomolecules and annotation of the second version of the reference genome sequence assembly of barley cv. morex [morex v2].
URL: *https://doi.ipk-gatersleben.de:443/DOI/83e8e186-dc4b-47f7-a820-28ad37cb176b/d1067eba-1d08-42e2-85ec-66bfd5112cd8/2*

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P. E., Russell, J. et al. (2017). A chromosome conformation capture ordered sequence of the barley genome, *Nature* **544**(7651): 427–433.

McVean, G. (2009). A genealogical interpretation of principal components analysis, *PLoS genetics* **5**(10).

Mee, J. A. and Yeaman, S. (2019). Unpacking conditional neutrality: genomic signatures of selection on conditionally beneficial and conditionally deleterious mutations, *The American Naturalist* **194**(4): 529–540.

Milner, S. G., Jost, M., Taketa, S., Mazón, E. R., Himmelbach, A., Oppermann, M., Weise, S., Knüpffer, H., Basterrechea, M., König, P. et al. (2019). Genebank genomics highlights the diversity of a global barley collection, *Nature genetics* **51**(2): 319–326.

Morrell, P. L., Toleno, D. M., Lundy, K. E. and Clegg, M. T. (2005). Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization, *Proceedings of the National Academy of Sciences* **102**(7): 2442–2447.

Navarro, J. A. R., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., Preciado, E., Terron, A., Delgado, H. V., Vidal, V. et al. (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces, *Nature genetics* **49**(3): 476.

Nevo, E., Beharav, A., Meyer, R., Hackett, C., Forster, B., Russell, J. and Powell, W. (2005). Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution Canyon', Israel, *Biological Journal of the Linnean Society* **84**(2): 205–224.

Nevo, E., Zohary, D., Brown, A. and Haber, M. (1979). Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel, *Evolution* pp. 815–833.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E. and Wagner, H. (2019). *vegan: Community Ecology Package*. R package version 2.5-6.
URL: *https://CRAN.R-project.org/package=vegan*

Pankin, A., Altmüller, J., Becker, C. and von Korff, M. (2018). Targeted resequencing reveals genomic signatures of barley domestication, *New Phytologist* **218**(3): 1247–1259.

Pembleton, L., Cogan, N. and Forster, J. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations, *Molecular Ecology Resources* **13**: 946–952.

Petkova, D., Novembre, J. and Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces, *Nature genetics* **48**(1): 94.

Pham, A.-T., Maurer, A., Pillen, K., Brien, C., Dowling, K., Berger, B., Eglinton, J. K. and March, T. J. (2019). Genome-wide association of barley plant growth under drought stress using a nested association mapping population, *BMC plant biology* **19**(1): 134.

718 Poland, J. A., Brown, P. J., Sorrells, M. E. and Jannink, J.-L. (2012). Development of high-density genetic
719   maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach, *PloS one*
720   **7**(2): e32253.

721 Pyhäjärvi, T., Hufford, M. B., Mezmouk, S. and Ross-Ibarra, J. (2013). Complex patterns of local adaptation
722   in teosinte, *Genome biology and evolution* **5**(9): 1594–1609.

723 Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. and Holderegger, R. (2015). A practical guide to
724   environmental association analysis in landscape genomics, *Molecular Ecology* **24**(17): 4348–4370.

725 Renaut, S., Grassa, C., Yeaman, S., Moyers, B., Lai, Z., Kane, N., Bowers, J., Burke, J. and Rieseberg, L.
726   (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers, *Nature*
727   *communications* **4**(1): 1–8.

728 Russell, J., Mascher, M., Dawson, I. K., Kyriakidis, S., Calixto, C., Freund, F., Bayer, M., Milne, I., Marshall-
729   Griffiths, T., Heinen, S. et al. (2016). Exome sequencing of geographically diverse barley landraces and
730   wild relatives gives insights into environmental adaptation, *Nature Genetics* **48**(9): 1024.

731 Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J. and Schluter, D. (2017). Gene flow
732   and selection interact to promote adaptive divergence in regions of low recombination, *Molecular Ecology*
733   **26**(17): 4378–4390.

734 Sato, K., Mascher, M., Himmelbach, A., Haberer, G., Spannagl, M. and Stein, N. (2021). Chromosome-
735   scale assembly of wild barley accession 'OUH602', *G3 Genes|Genomes|Genetics* (jkab244).
736   **URL:** *https://doi.org/10.1093/g3journal/jkab244*

737 Schmid, K., Kilian, B. and Russell, J. (2018). Barley domestication, adaptation and population genomics,
738   *The Barley Genome*, Springer, pp. 317–336.

739 Szkiba, D., Kapun, M., von Haeseler, A. and Gallach, M. (2014). SNP2GO: functional analysis of genome-
740   wide association studies, *Genetics* **197**(1): 285–289.

741 Terrazas, R. A., Balbirnie-Cumming, K., Morris, J., Hedley, P. E., Russell, J., Paterson, E., Baggs, E. M.,
742   Fridman, E. and Bulgarelli, D. (2020). A footprint of plant eco-geographic adaptation on the composition
743   of the barley rhizosphere bacterial microbiota, *Scientific reports* **10**(1): 1–13.

744 Tiffin, P. and Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local
745   adaptation, *Trends in ecology & evolution* **29**(12): 673–680.

31

746 Tsuda, Y., Chen, J., Stocks, M., Källman, T., Sønstebø, J. H., Parducci, L., Semerikov, V., Sperisen, C., Poli-
747 tov, D., Ronkainen, T. et al. (2016). The extent and meaning of hybridization and introgression between
748 siberian spruce (*picea obovata*) and norway spruce (*picea abies*): cryptic refugia as stepping stones to
749 the west?, *Molecular Ecology* **25**(12): 2773–2789.

750 Turner-Hissong, S. D., Mabry, M. E., Beissinger, T. M., Ross-Ibarra, J. and Pires, J. C. (2020). Evolutionary
751 insights into plant breeding, *Current Opinion in Plant Biology* **54**: 93–100.
752 **URL:** *http://www.sciencedirect.com/science/article/pii/S1369526620300339*

753 Volis, S. (2011). Adaptive genetic differentiation in a predominantly self-pollinating species analyzed by
754 transplanting into natural environment, crossbreeding and QST-FST test, *New Phytologist* **192**(1): 237–
755 248.

756 Volis, S., Mendlinger, S. and Ward, D. (2002a). Adaptive traits of wild barley plants of Mediterranean and
757 desert origin, *Oecologia* **133**(2): 131–138.

758 Volis, S., Mendlinger, S. and Ward, D. (2002b). Differentiation in populations of *Hordeum spontaneum* along
759 a gradient of environmental productivity and predictability: life history and local adaptation, *Biological
760 Journal of the Linnean Society* **77**(4): 479–490.

761 Volis, S., Shulgina, I., Ward, D. and Mendlinger, S. (2003). Regional subdivision in wild barley allozyme
762 variation: adaptive or neutral?, *Journal of Heredity* **94**(4): 341–351.

763 Volis, S., Verhoeven, K., Mendlinger, S. and Ward, D. (2004). Phenotypic selection and regulation of repro-
764 duction in different environments in wild barley, *Journal of evolutionary biology* **17**(5): 1121–1131.

765 Volis, S., Yakubov, B., Shulgina, I., Ward, D. and Mendlinger, S. (2005). Distinguishing adaptive from
766 nonadaptive genetic differentiation: comparison of q st and f st at two spatial scales, *Heredity* **95**(6): 466–
767 475.

768 Volis, S., Yakubov, B., Shulgina, I., Ward, D., Zur, V. and Mendlinger, S. (2001). Tests for adaptive RAPD
769 variation in population genetic structure of wild barley, *Hordeum spontaneum* Koch, *Biological Journal of
770 the Linnean Society* **74**(3): 289–303.

771 Volis, S., Zaretsky, M. and Shulgina, I. (2010). Fine-scale spatial genetic structure in a predominantly selfing
772 plant: role of seed and pollen dispersal, *Heredity* **105**(4): 384–393.

Wang, X., Chen, Z.-H., Yang, C., Zhang, X., Jin, G., Chen, G., Wang, Y., Holford, P., Nevo, E., Zhang, G. et al. (2018). Genomic adaptation to drought in wild barley is driven by edaphic natural selection at the Tabigha Evolution Slope, *Proceedings of the National Academy of Sciences* **115**(20): 5223–5228.

Wiegmann, M., Maurer, A., Pham, A., March, T. J., Al-Abdallat, A., Thomas, W. T. B., Bull, H. J., Shahid, M., Eglinton, J., Baum, M., Flavell, A. J., Tester, M. and Pillen, K. (2019). Barley yield formation under abiotic stress depends on the interplay between flowering time genes and environmental cues, *Scientific Reports* **9**(1): 6397.

URL: *https://www.nature.com/articles/s41598-019-42673-1*

Yeaman, S. and Whitlock, M. C. (2011). The genetic architecture of adaptation under migration–selection balance, *Evolution: International Journal of Organic Evolution* **65**(7): 1897–1911.

Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C. and Weir, B. (2012). A high-performance computing toolset for relatedness and principal component analysis of snp data, *Bioinformatics* **28**(24): 3326–3328.