Title: Evaluating the persistence and utility of five wild *Vitis* species in the context of climate change

Running title: *Vitis'* evolutionary response to climate change

Jonas A. Aguirre-Liguori*[1], Abraham Morales-Cruz*[1] and Brandon S. Gaut[1,2]

* joint first authors

[1]Department of Ecology and Evolutionary Biology

University of California, Irvine

[2] Corresponding author:

bgaut@uci.edu

321 Steinhaus Hall

UC Irvine,

Irvine, CA 92697

ABSTRACT

Crop wild relatives (CWRs) have the capacity to contribute novel traits to agriculture. Given climate change, these contributions may be especially vital for the persistence of perennial crops, because perennials are often clonally propagated and consequently do not evolve rapidly. By studying the landscape genomics of samples from five *Vitis* CWRs (*V. arizonica, V. mustangensis, V. riparia, V. berlandieri* and *V. girdiana*) in the context of projected climate change, we addressed two goals. The first was to assess the relative potential of different CWR accessions to persist in the face of climate change. By integrating species distribution models with adaptive genetic variation, additional genetic features such as genomic load and a phenotype (resistance to Pierce's Disease), we predicted that accessions from one species (*V. mustangensis*) are particularly well-suited to persist in future climates. The second goal was to identify which CWR accessions may contribute to bioclimatic adaptation for grapevine (*V. vinifera*) cultivation. To do so, we evaluated whether CWR accessions have the allelic capacity to persist if moved to locations where grapevines (*V. vinifera*) are cultivated in the United States. We identified six candidates from *V. mustangensis* and hypothesized that they may prove useful for contributing alleles that may mitigate climate impacts on viticulture. By identifying candidate germplasm, this work takes a conceptual step toward assessing the genomic and bioclimatic characteristics of CWRs.

KEYWORDS

Climate change, Crop Wild Relatives, Genetic offset, Genomic Load, Local adaptation, Migration Load, Species Distribution Models

38

39 **INTRODUCTION**

40 Natural populations respond to climate change by adapting to new conditions, by
41 tolerating wider environmental ranges (phenotypic plasticity), by migrating to suitable
42 areas, or by going extinct (Feeley, Rehm, & Machovina, 2012). There is a pressing need to
43 better understand each of these mechanisms to manage and conserve species (Des
44 Roches, Pendleton, Shapiro, & Palkovacs, 2021; Frankham, 2005; Waldvogel et al., 2019).
45 This need is even more pressing for the wild relatives of crop plants (or crop wild relatives,
46 CWRs), because CWRs often contain novel genetic diversity that can contribute to crop
47 improvement (Janzen, Wang, & Hufford, 2019). The potential loss of CWRs has
48 repercussions beyond biodiversity: they are crucial for breeding crops that meet the
49 climate challenge (McCouch, 2013).

50 Given their importance, numerous studies have focused on the geographic
51 distributions, *ex situ* germplasm collections and *in situ* conservation of CWRs (e.g.,
52 (Castañeda-Álvarez et al., 2016)). Surprisingly fewer studies have assessed the potential
53 effects of climate change on CWRs (Jarvis, Lane, & Hijmans, 2008), but recent notable
54 exceptions include studies of CWRs in Europe (Aguirre-Gutiérrez, van Treuren, Hoekstra, &
55 van Hintum, 2017) and the United States (US) (Khoury et al., 2020). These and similar
56 studies typically employ species distribution models (SDMs) to infer the climatic niche of
57 CWRs based on the locations where the species is (or is not) found. This niche is then
58 projected into the future, based on the predicted climate. The outcome of SDMs is an
59 estimate of the predicted geographic niche under climate change, which helps inform the

60 fate of species. SDMs are limited, however, because they usually ignore biotic
61 interactions (Lawler, White, Neilson, & Blaustein, 2006), phenotypic plasticity, and the
62 potential for evolutionary adaptation. As a consequence, SDMs likely yield biased
63 predictions about species persistence (Exposito-Alonso et al., 2018; Hällfors et al., 2016;
64 Ikeda et al., 2017; Razgour et al., 2018)

65 Although the number of climate-based studies of CWRs is growing, fewer studies
66 have assessed the fate of CWRs by integrating climate predictions with landscape genomic
67 data (Aguirre-Liguori, Ramírez-Barahona, & Gaut, 2021; Aguirre-Liguori, Ramírez-
68 Barahona, Tiffin, & Eguiarte, 2019). Genomic data can be used to model how adaptive
69 genetic variation will change in the context of climate predictions. In this way, one can
70 identify species or populations that appear genetically poised to adapt to climatic change
71 and, conversely, populations that may require especially dramatic genetic changes and are
72 thus genetically endangered. This integration of genomic data with climate predictions is
73 relatively new (Capblancq, Fitzpatrick, Bay, Exposito-Alonso, & Keller, 2020; Fitzpatrick &
74 Keller, 2015), and hence it has been applied to only a handful of taxa (see Capblancq et al.,
75 2020). Studies of poplars, pearl millet and *Arabidopsis* have been especially notable,
76 because they have shown that predicted shifts in allelic variants correlate with
77 components of fitness (Exposito-Alonso, Burbano, Bossdorf, Nielsen, & Weigel, 2019;
78 Fitzpatrick, Chhatre, Soolanayakanahally, & Keller, 2021; Rhoné et al., 2020), thus arguing
79 for the relevance of the approach.

80 While genomic data include potential information about the adaptive process, the
81 data also contain information about non-adaptive processes that can provide insights into

82 the potential evolutionary fate of species (Aguirre-Liguori et al., 2021; Waldvogel et al.,
83 2019). Such information includes the history of genetic migration among populations,
84 insights into the history and potential effects of genetic drift, and the magnitude of
85 genomic load, which correlates inversely with fitness (Frankham, 2005). To incorporate
86 additional genetic and ecological information into species predictions, Aguirre-Liguori et
87 al. (2021) proposed the FOLDS model, a conceptual framework to evaluate potential
88 responses to climate change. The goal of the FOLDS model is to predict which populations
89 are most likely to respond adequately to climate change, based on multiple layers of
90 information. Those layers can include SDMs, information about putative climate-adaptive
91 alleles, phenotypic data and other population genetic information, such as population size
92 and genomic load.

93 In this study, we employ the FOLDS model to assess potential responses to climate
94 change for samples from five North American (NA) *Vitis* CWRs. We focus on the CWRs of
95 *Vitis* because they are climatically diverse (Callen, Klein, & Miller, 2016) and agronomically
96 crucial. *Vitis* CWRs are employed for hybrid scion breeding and also as rootstocks, to the
97 extent that ~80% of viticulture worldwide utilizes rootstocks from NA *Vitis* species (Ollat et
98 al., 2016). Among ~25 NA *Vitis* species (Wan et al., 2013), those from the American
99 Southwest are of particular interest, because some are resistant to important diseases and
100 others grow under abiotic stresses like extreme conditions that may 'preadapt' them to
101 some aspects of climate change (Heinitz, Uretsky, Peterson, Acosta, & Walker, 2019). As
102 concrete examples, *V. berlandieri* is used as a rootstock especially in limestone soils and
103 hot, dry environments (Heinitz et al., 2019), and *V. arizonica* has been used to create

104 hybrid grapevine scions that are resistant to Pierce's Disease (Summaira Riaz et al., 2009),
105 an economically devastating disease caused by a bacterium that infects several
106 economically important crops (Rapicavoli, Ingel, Blanco-Ulate, Cantu, & Roper, 2018).
107 We also focus on *Vitis* because climate change is predicted to disrupt the
108 phenology and production of domesticated grapevines (*V. vinifera*) (Morales-castilla et al.,
109 2020), representing substantial impact to a ~$5 billion per annum industry (USDA, 2013).
110 In addition, despite their ubiquitous use, rootstocks have a remarkably narrow genetic
111 foundation. Currently, a collection of ~10 rootstocks are used for 90% of grafted
112 grapevines. The most common rootstocks include hybrid or solo contributions from seven
113 NA *Vitis* species. However, the contribution from each species is often (and, in fact,
114 usually) a single accession (Marín et al., 2021). These observations underscore the
115 pressing need to identify additional germplasm for rootstock and scion breeding (Heinitz
116 et al., 2019; Summaira Riaz et al., 2019), given the pressures of climate change.
117 With genomic data from samples representing five *Vitis* species, this study has two
118 goals. The first is to employ the FOLDs model to evaluate which sets of individuals in the
119 sample are most likely to persist through climate change at their sampled locations in
120 nature. Our evaluation considers climate predictions, the complement of putatively
121 adaptive alleles, resistance to Pierce's Disease and aspects of population history. The
122 second goal is to assess whether accessions have adaptive alleles that may allow them to
123 persist where *V. vinifera* is currently grown in the United States in predicted future
124 climates. The motivation for this second goal is to evaluate whether CWRs have
125 combinations of alleles that may make them potentially useful, pending extensive future

126 functional validation, for scion or rootstock breeding. Overall, our work takes a
127 conceptual step toward estimating the climate impacts on CWRs and toward identifying
128 candidate species and accessions that may prove useful for the *in situ* adaptation of an
129 important crop.
130

131 **METHODS**

132 **Plant material, resequencing and variant identification**

133 The plant material used in this study consisted of 105 individuals from five
134 American *Vitis* species (*V. arizonica, n* = 22; *V. mustangensis, n* = 24; *V. berlandieri, n* = 22;
135 *V. girdiana, n* = 18 and *V. riparia; n* = 19), which were described previously in Morales-Cruz
136 et al. (2021) (Table S1). The 105 accessions were also assayed previously for resistance to
137 *Xylella fastidiosa* (Morales-Cruz et al., 2021; S. Riaz, Huerta-Acosta, Tenscher, & Walker,
138 2018), the causative agent of PD. Briefly, PD resistance was evaluated using greenhouse
139 experiments in which *X. fastidiosa* was inoculated in different individuals and
140 concentration was evaluated using ELISA tests 10 to 14 weeks after inoculation.
141 Individuals were considered to be resistant to PD if they had concentrations of *X.*
142 *fastidiosa* < 13 Least Square Means of colony forming units (CFUs) per mL (Summaira Riaz,
143 Tenscher, Heinitz, Huerta-Acosta, & Andrew Walker, 2020).

144 We used the sequencing data generated by Morales-Cruz *et al.* (2021), which is in
145 the Short Read Archive at NCBI under BioProject ID: PRJNA731597, and their SNP calls.
146 Briefly, they filtered Illumina paired-end reads of 150 base pairs (bp) and mapped them to
147 the *V. arizonica* b40-14 v1.0 genome (Morales-Cruz et al., 2021)

148 (https://doi.org/10.5281/zenodo.4977234 and www.grapegenomics.com/pages/Vari/).
149 Joint SNP calling was conducted using the HaplotypeCaller in the GATK v.4.0 pipeline
150 following (Zhou et al. (2017). The VCF files were split by species, and the raw SNPs were
151 filtered with bcftools v1.9 (https://samtools.github.io/bcftools/) and vcftools v0.1.15
152 (https://vcftools.github.io/). SNPs sites were kept for downstream analyses if they were
153 biallelic, had no missing data, had quality higher than 30, had a depth of coverage higher
154 than five reads, and also had no more than three times the median coverage depth.
155 Additionally, the following expression was applied under the exclusion argument of the
156 filter function in bcftools: "QD < 2.0 | FS > 60.0 | MQ < 40.0 | MQRankSum < -12.5 |
157 ReadPosRankSum < -8.0 | SOR > 3.0". These steps resulted in a range of high-quality
158 filtered SNPs from 3.6 million in *V. girdiana* to 5.6 million in *V. riparia* (Table S2).
159

160 **Species Distribution Models**

161 For each CWR *Vitis* species and for *V. vinifera* we downloaded occurrences from
162 the Global Biodiversity Information Facility (www.gbif.org; 2020;
163 DOI10.15468/dl.4emr87). We removed duplicated locations and those that were
164 potential misclassifications due to being obvious outliers. We also downloaded 19
165 bioclimatic variables from Worldclim 2 (Fick & Hijmans, 2017) for a period representing
166 the present, which is an average of observations from 1970 to 2010, and to 54 forecasts of
167 climate change (FCCs). Our rationale for including 54 FCCs was to incorporate uncertainty
168 in climate projections. The forecasts were downloaded at a 2.5 minute resolution from
169 Worldclim 2 (last accessed May 2022), based on the CMIP6 project (Eyring et al., 2016).

The 54 FCCs corresponded to five circulation models (GFDL-ESM4, IPSL-CM6A-LR, MPI-ESM1-2-HR, MRI-ESM2-0, UKESM1-0-LL), three time periods (2041-2060 [mean 2050]; 2061-2080 [mean 2070]; 2081-2100 [mean 2090]) and four Shared Socio-economic pathways (SSPs) that model different trajectories of greenhouse effects (SSPs 126, 245, 370, 585). [We were unable to acquire the data for the GFDL-ESM4 model with SSPS 245 and 585 for any time period, resulting in 54 instead of total 60 FCCs.]

To build an SDM for each species, we used previously published scripts (Aguirre-Liguori et al., 2021) that performed four steps. First, the script removed correlated bioclimatic variables (r>0.8) with the highest Variance Inflation Factor, to moderate multicollinearity between variables. Second, it identified a calibration or background area, which corresponded to the migration layer in the BAM (Biotic, Abiotic and Migration) model (Peterson et al., 2011; Soberón, 2010) and to terrestrial eco-regions of the world (Olson et al., 2001). Third, the script employed the BIOMOD2 package in R (Thuiller, Georges, Engler, Georges, & Thuiller, 2016) and the Maxent algorithm (SJ Phillips, Anderson, & Schapire, 2006; Steven Phillips & Dudík, 2008) to build and project the models. Finally, we performed 20 bootstrap replicates per model. For each replicate we used 70% of occurrences to train the model and 30% to test the model with True Skill Statistics for tenfold internal cross-validation.

Based on the present-day SDMs, we projected the potential distribution of species to the 54 FCCs. For each species and each FCC we estimated how the geographic distribution (i.e., the number of pixels in the projection) was expected to change from the present. Ultimately, we evaluated the mean change and the dispersion across the FCCs in

192 each of the three projected time-periods. We also estimated whether each of the 105
193 accessions was expected to persist in a given FCC. That is, for each individual at a given
194 time period, we counted how many times FCCs predicted the persistence of the
195 individual's sample location. If the location was predicted to be outside the species'
196 estimated niche for all 20 FCC models at a specific time-period, we coded that population
197 as extinct by that time period and into the future.
198
199 **Investigating patterns of local adaptation**
200 *Identification of outlier SNPs:* For each CWR species sample, we used Baypass
201 (Gautier, 2015) to identify outlier loci that had significant associations with the first four
202 principal components (PC) of bioclimatic variables (Figure S1). [Figure S1 provides a
203 schematic overview of methods that incorporated genomic and environmental data, as
204 applied to the FOLDS model.] The PCs were determined with the *prcomp* function in R,
205 based on the 19 bioclimatic variables from the present-day. Baypass analyzes patterns of
206 covariation among accessions to account for population structure and then analyzes the
207 correlation between a variable (in this case each PC) and the allelic frequencies of
208 individual SNPs across individuals. A SNP was deemed an outlier if it had a significant
209 correlation with a variable that was not explained by the covariance between populations.
210 Following Jeffreys' rule (Gautier, 2015), we characterized outlier SNPs as those that had a
211 Bayes Factor (BF) > 20.
212 *Local Genomic Offsets:* To calculate genetic offsets, we used gradient forest (GF;
213 Figure S1) (Ellis, Smith, & Roland Pitcher, 2012), a machine learning method that models

214 the turnover in genetic composition across the landscape (Fitzpatrick & Keller, 2015). GF
215 identifies which bioclimatic variables contribute importantly to the construction of the
216 model, the SNPs that are associated with bioclimatic variables, and the adaptive genetic
217 composition across the predicted climate (Capblancq et al., 2020; Fitzpatrick & Keller,
218 2015; Waldvogel et al., 2019). When applied to bioclimatic data from both the present
219 and the future, GF also estimates the local genetic offset (or 'local offset'), an estimate of
220 the amount of genetic change necessary to adapt in the future at the same locality
221 (Capblancq et al., 2020; Fitzpatrick & Keller, 2015). The local offset depends on the genetic
222 composition of populations and the magnitude of the environmental change that is
223 expected to occur; populations with higher offsets are expected to be more vulnerable to
224 climate change (Fitzpatrick & Keller, 2015).

225 For each species, we obtained the GF model using the *gradient forest* package in R
226 (Ellis et al., 2012) for all the outlier SNPs identified by Baypass (**Table 1**), based on allelic
227 frequencies within individuals (Figure S1). These frequencies were either 0 or 1 for
228 homozygotes or 0.5 for inferred heterozygotes. The local offset was then calculated as
229 the Euclidian distance of the genetic compositions between the present and different time
230 periods (Fitzpatrick & Keller, 2015) (Figure S1). We used the GF models to predict how
231 genetic compositions were expected to change in each of the 54 FCCs.

232 GF is traditionally applied to allele frequencies across populations, but here we
233 applied it to individual genotypes. To investigate whether results were applicable for
234 individuals, we performed two tests. First, we transformed allelic frequencies within
235 individuals to pseudo-population frequencies, using a probabilistic model, and reanalyzed

236 the data (see Supplemental Methods). Our simulations suggest that the sampling of
237 individuals, instead of populations, had little impact on the construction of the GF model,
238 the estimation of genomic offsets and the identification of important bioclimatic variables
239 (See **Supplementary Methods** and **Figure S2**). Second, we used a population dataset that
240 had been used to calculate genetic offsets previously (Aguirre-Liguori et al. 2021) and
241 evaluated the effect of subsampling from 1 to 11 individuals per population on the
242 estimation of local offsets (see Supplementary Methods). We found that using individuals
243 instead of allelic frequencies reduced the estimated local offsets across populations but
244 also that the rank correlation among offsets was >0.99. This high correlation indicated
245 that the relative magnitude and order of local offsets remained similar across locations
246 and sampling strategies (Figure S3), further suggesting that use of individuals in GF is a
247 suitable for our comparisons.

248 *Calculating the Adaptive Score:* In addition to local offsets, we calculated other
249 parameters from the data to help summarize genetic information that can be used to
250 evaluate climate persistence. For example, to summarize information about adaptive
251 alleles we calculated an adaptive score (S_a). S_a measured the proportion of alleles across
252 all candidate SNPs that were 'pre-adapted' to climate, based on climate projections
253 (Figure S1). As used here, S_a is modelled after the population adaptive index (PAI) of
254 Bonin et al. (2007). The PAI measured the proportion of loci that have a selective signal
255 within populations – i.e., the proportion of loci that were outliers due to high allelic
256 frequency deviations among populations. In our case, we identified the alleles that --

257 according to turnover functions from GF analyses -- were predicted to be adaptive in
258 future climates (Figure S4).

259 To identify pre-adapted alleles, we assessed the projected direction of
260 environmental change for important bioclimatic variables (i.e., change across the *x*-axis in
261 **Figure S4**) by identifying 'pre-adapted' individuals that, according to turnover functions,
262 require small change in their genetic composition in the future (i.e., small changes across
263 the y-axis, **Figure S4**). After identifying pre-adapted individual(s), for each candidate SNP
264 we built a linear model (using *lm* in R) to confirm that the genotypic state correlated
265 significantly with the expected bioclimatic variable. Significantly correlated alleles were
266 deemed adaptive, and the high frequency allele was categorized as the adaptive state.
267 Given adaptive alleles, we calculated S_a as the total number adaptive alleles per individual
268 across all candidate SNP sites, divided by 2x the number of candidate SNPs (The factor of
269 2x reflected the diploid state). Thus S_a reflected the proportion of adaptive alleles that
270 were present in an individual across all candidate SNPs. It ranged from 0 to 1, where 0
271 indicates an individual with no adaptive alleles and 1 indicates homozygosity of adaptive
272 alleles at all candidate SNPs.

273 For each species, we calculated S_a for each of the four bioclimatic variables that
274 had the strongest contribution to building the GF model and then calculated the mean
275 across the four variables (Figure S1). Further details, tests for potential biases and
276 rationale for S_a are provided in the **Supplementary Methods**. We note that S_a overlaps
277 with, but differs from, local offsets in two ways. First, high local offsets can be caused
278 either by strong predicted environmental shifts or by extensive expected genetic shifts. In

279 contrast, S_a focuses solely on genetic composition by counting whether an inferred
280 adaptive allele is present at each position. Second, local offsets cannot be compared
281 across species (Láruson, Fitzpatrick, Keller, Haller, & Lotterhos, 2022), but S_a can because
282 it reports a proportion – i.e., of adaptive alleles across candidate climate-related SNPs.
283
284 **Additional Population Genetic Parameters**

285 *Demographic history:* To incorporate information about genetic drift into the
286 FOLDS model, we estimated the effective population (*Ne*) size of each species using
287 MSMC2 v2.1.1 (Mallick et al., 2016) with unphased SNPs. To include only the most
288 informative genomic regions, we created a mappability mask and a coverage mask. We
289 created the mappability mask for the *V. arizonica* b40-14 v1.0 genome using the software
290 SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml). We generated 150 bp mers
291 moving in 1 bp increments across the genome, mapped the sequences back to the
292 genome with BWA v0.7.8-r455 (Li & Durbin, 2009) and identified "mappable" genomic
293 regions where the majority of sequences mapped uniquely without mismatches. To
294 include only the regions with sufficient sequencing coverage, we calculated the coverage
295 from the alignment file of each sample with the bedcov program from the samtools
296 (v1.10) package, supplying a bed file of 10 kb non-overlapping windows across the
297 genome. We then created a mask per sample that only included regions with sequencing
298 coverage higher than 5x. We then used unphased SNPs from the six samples per species
299 with the highest average coverage as input for the run in MSMC2. For analyses, we
300 assumed a mutation rate of 5.4e-9 (Liang et al., 2019) and a generation time of 3 years
301 (Zhou et al., 2017). In addition to MSMC2 with unphased data, we applied MSMC2 to

302 phased data and also used SMC++ as a another source for *Ne* estimation (see
303 Supplementary Methods). The results of the most recent time index from SMC++ and
304 MSC2-unphased were highly similar (**Figure S5**). The results were also very consistent
305 among individual MSMC2 runs (**Figure S6),** and so we showed the average per species
306 (**Figure S5**).

307 *Genomic load*: Genomic load is a measure of the number of predicted deleterious
308 alleles in an individual. To identify deleterious alleles, we first determined the functional
309 context (exonic, intronic, and intergenic) of SNPs based on *V. arizonica* B40-14 genome
310 annotations ((Morales-Cruz et al., 2021), https://doi.org/10.5281/zenodo.4977234 and
311 www.grapegenomics.com/pages/Vari/). Exonic SNPs were annotated to be synonymous,
312 nonsynonymous or frameshift mutations using SnpEff version 5.0e (Cingolani et al., 2012).
313 Nonsynonymous SNPs were predicted as deleterious or tolerated using the SIFT score (Ng
314 & Henikoff, 2003), as computed in the program SIFT 4G (Vaser, Adusumalli, Leng, Sikic, &
315 Ng, 2016). SIFT scores ≤ 0.05 were interpreted as deleterious and SIFT scores > 0.05 were
316 considered to be tolerant. Because the reference genome biases the distribution of
317 missing data (Lohmueller et al., 2008), we only used polymorphic sites without missing
318 data that were predicted across all five wild grape species, resulting in a set of 397,723
319 SNPs.

320

321 **Applying the FOLDS model**

322 To assess the potential effects of climate change, we employed the FOLDS model
323 (Aguirre-Liguori et al., 2021), based on six layers of genomic, phenotypic and climate

324 information described above (Figure S1). We used UpsetR (Conway, Lex, & Gehlenborg,
325 2017) to plot intersections among the six layers of information for each of the 54 FCCs.
326 For each UpsetR model, we evaluated which accessions passed specific filters for the six
327 layers. As described below (see Results), we selected different thresholds and filters for
328 inclusion and synthesis. For example, accessions were included based on SDMs if the
329 sampling location was predicted to be in the species' niche in the future and also when
330 they were resistant to *X. fastidiosa* (i.e., had concentrations < 13 Least Square Means of
331 CFUs/mL). We also selected accessions as persistent if they had local offsets and genomic
332 load values below 90%, 50% and 25% of their distribution across species. Since local
333 offsets are not comparable between species (Láruson et al., 2022), the thresholds for local
334 offsets were considered only within species. Finally, accessions were included when they
335 had S_a or *Ne* above the 10%, 50% and 75% of the distributions across species.
336
337 **Migration Load**
338 The FOLDS model defined a set of CWR accessions that were best situated to
339 persist in the face of climate change. We focused on these accessions to evaluate their
340 genetic offset at locations where *V. vinifera* is currently cultivated in the United States.
341 This measure of offset, which has been called either the "forward offset" (Gougherty,
342 Keller, & Fitzpatrick, 2021) or the "migration load" (Rhoné et al., 2021), estimates
343 whether an accession has the allelic complement to persist when moved to a different
344 location. Thus, the migration load is the genetic offset for an individual calculated
345 between its present-day sampling site and for predicted climates at other locations. In

346 this case, the other locations represent current locations of viticulture. The migration load
347 is a tool to evaluate whether CWR accessions might be useful in future climates where
348 grapevines are currently cultivated.

349 We estimated migration load for accessions that passed the FOLDs threshold for
350 each of the 54 FCC models at the 1278 United States locations of *V. vinifera* cultivation
351 available from gbif.org. Following previous work (Aguirre-Liguori et al., 2021; Rhoné et al.,
352 2020), the migration load was calculated as the Euclidian distance of the genetic
353 composition between the present location of an accession and the future estimate in the
354 projected climate of *V. vinifera* cultivation. We identified locations where the migration
355 load was within the current range of the species' predicted local offsets over the same FCC
356 period. Finally, for each *V. vinifera* location, we counted how many times an accession had
357 a migration load within the range of its local offsets. We interpreted these accessions as
358 candidates for having climate-related alleles that could be useful to viticulture in the face
359 of climate change.

360

361 **RESULTS**

362 **Species Sampling and Data:**

363 In this study, we used previously published resequencing data from 105 individuals
364 representing five North American *Vitis* species: *V. arizonica*, *V. mustangensis* (synonym *V.*
365 *candicans*), *V. berlandieri*, *V. girdiana* and *V. riparia* (Morales-Cruz et al., 2021). These
366 species are either currently used as rootstocks, represent the parents of hybrid rootstocks
367 and scions, or have phenotypic properties that make them potentially valuable to

368 viticulture (Heinitz et al., 2019). The two most closely related species in this study (*V.*
369 *girdiana* and *V. arizonica*) split at least 10 million years ago (Wan et al., 2013) and perhaps
370 even earlier (Morales-Cruz et al., 2021), so they have not diverged especially recently.
371 For each of the five species, 18 to 24 accessions were resequenced (**Table 1**). We
372 employed the previously determined SNPs (Morales-Cruz et al., 2021), which were used
373 previously to explore the divergence dynamics among species and their history of
374 hybridization. Here, however, we focused on a subset of SNPs that had no missing data
375 within each species. The resulting number of SNPs ranged from 3.6 million in *V. girdiana*
376 to 5.6 million in *V. riparia* (**Table S2**). The accessions, their sampling locations, and their
377 SNPs provided the basis for evaluating species' persistence under projected climate
378 change.
379
380 **Evaluating the Persistence of Wild *Vitis* Under Predicted Climate Change**
381 To assess the potential effects of climate change, we combined six layers of
382 genomic and climate information within the FOLDS framework (Aguirre-Liguori et al.,
383 2021). The six layers were chosen to include different types of information, including: i)
384 SDMs to incorporate niche information; ii) patterns of local adaptation and predicted
385 allelic responses to climate change; iii) insights into genetic drift and genomic load, both of
386 which may affect evolvability and persistence; and iv) phenotypic characterizations of
387 resistance to PD. Below we describe each of the six layers before synthesizing information
388 within FOLDS framework.

389 *Layer 1 - Species distribution models:* We estimated SDMs to examine expected
390 changes in the geographic distribution of the five wild *Vitis* species over 54 climate models
391 (FCCs), focusing on four time periods (present-day, 2050, 2070 and 2090), five circulation
392 models and four shared socio-economic pathways (see Methods). For each of the five
393 species, we projected the SDM onto the 54 FCCs. Then, for each SDM we measured the
394 projected habitable area of each species (**Figure 1a**), as has been done previously for wild
395 Vitis (Callen et al. 2016; Heibnitz et al, 2019) but also the predicted trends over time
396 (**Figure 1b**). Strikingly, the average amount of habitable area was predicted to increase
397 over time for all species, but with a marginal reduction for *V. girdiana* after 2050 (**Figure**
398 **1b, Table S3**).
399 Although the overall expansion of niche area is promising, there was substantial
400 variation in the details across the 54 FCCs. For example, the dispersion across models
401 increased as a function of time, and the SSPs that predicted higher greenhouse gas
402 emissions predicted larger CWR distributions (**Figure 1b**). Overall, the SDMs predicted
403 many *V. arizonica*, *V. mustangenis* and *V. berlandieri* locations were likely to persist until
404 2090 (**Figure 1c, Table S3**). However, persistence varied substantially among species,
405 because from 20% to 43% of sampling localities (depending on the species and climate
406 models), were predicted to disappear from species' niches by 2090. Overall, SDMs predict
407 that distribution of the wild *Vitis* species is expected to increase in the future, but the
408 results also suggest that many populations represented by the samples in our study will
409 need to either adapt or migrate to survive.

410 *Layer 2 - Local Offsets and Adaptative Variation:* We used GF to estimate the local
411 offsets of *Vitis* individuals within each species. Following common practice (Capblancq et
412 al., 2020; Fitzpatrick, Keller, & Lotterhos, 2018), we ran GF with a subset of outlier SNPs
413 that were associated with environmental variables. To identify these outlier SNPs, we
414 performed genome environment associations (GEA) with Baypass (Gautier, 2015),
415 identifying SNPs correlated to the first four principal components (PCs) of 19 present-day
416 bioclimatic variables (Figure S1). [The first 4 principal components explained 99.9% of
417 bioclimatic data for *arizonica;* 97.3% for *berlandieri* 97.3%; 92.9% for *mustangensis*; 98.8%
418 for *girdiana* 98.8% and 97.4% for *riparia 97.4%*]. Our GEA detected from 5,650 to 13,531
419 outlier SNPs within each species (**Table 1**). We applied GF to the set of outlier SNPs from
420 the GEA, yielding a subset of 2,730 to 8,720 candidate SNPs within each species that had
421 significant non-linear associations across samples (**Table 1**), based on present-day
422 bioclimatic data.
423 GF analyses provided three additional outcomes (Figure S1). The first was a set of
424 bioclimatic variables that were important for building the model, as measured by the
425 contribution of the variable to the construction of the model ($R^2 > 0$). For simplicity, we
426 focused on the four most important variables for each species, some of which (**Table 1**)
427 were shared across species. For example, BIO10 (the mean temperature of the warmest
428 quarter) had a strong contribution to the building of GF models in *V. berlandieri, V.*
429 *mustangensis* and *V. riparia,* whereas BIO4 (temperature seasonality) was a top
430 contributor for all species but *V. arizonica* (**Table 1**).

431 The second outcome was estimates of local genetic offsets, which we computed
432 for each individual in each species based on contrasts between present-day climate and
433 each of 54 FCC models (see Methods), resulting in 5,670 (= 54 models x 105 individuals)
434 estimates. We compared rank correlations among offsets between all FCC models to
435 determine if there was a consistent offset signal. Pairwise comparison revealed that >95%
436 of pairs had rank correlations of *rho*>0.5, suggesting that different climatic models
437 generally showed similar offset trends (Figure S7). For simplicity, in Figure 2a we show the
438 genetic offsets in each species for a single model (IPSL-CM6A-LR_ssp585_2061-2080),
439 because this model had the highest average rank correlation (mean rho = 0.83) compared
440 to the remaining 53 models. The corresponding results for the remaining models are
441 presented in Table S1.
442 Local offsets cannot be compared quantitatively across species (Láruson et al.
443 2022), because they are estimated with different SNPs and bioclimatic information. They
444 do, however, help predict whether specific alleles are expected to be adaptive in future
445 climates. To build a metric for cross-species comparison, we utilized local offsets to
446 inform a measure, S_a, that measures the proportion of adaptive alleles within each
447 individual and that can be compared across individuals and species (see Methods). For
448 each species, we calculated the mean adaptive score (S_a) (see Methods & Figure S1). Like
449 offsets, S_a varied within species (**Figure 2b**, **Table S1**); comparing S_a across species
450 indicated that our *V. girdiana* and *V. riparia* samples have pre-adaptive alleles in lower
451 proportions than the other species (**Figure 2b**).

452 Because S_a had not been applied previously, we performed tests to evaluate
453 whether S_a was biased by geographic, demographic or climatic variables. As expected, S_a
454 was strongly and significantly correlated with at least one of the four most important
455 bioclimatic variables identified by GF (**Figure 2c**); in fact, it was correlated with all four for
456 three of 5 species (**Figure S8**). As expected, S_a was more highly correlated to bioclimatic
457 variables than a similar score (S_{ref}) calculated from a set of reference SNPs (**Figure S8;** see
458 Methods), suggesting it contains additional information. The notable exception was for *V.*
459 *girdiana* samples, where S_{ref} was strongly correlated with all four of the top 4 associated
460 bioclimatic variables (**Figure S8**), likely reflecting genetic structure (Morales-Cruz et al.,
461 2021) or other features that complicate interpretation of S_a in this species.
462 *Layer 4- Effective population size:* The efficacy of selection depends on adaptive
463 diversity but also on population factors, like genetic drift and genomic load, that are
464 typically ignored in climate studies (Aguirre-Liguori et al., 2021; Brady et al., 2019; Nadeau
465 & Urban, 2019). To assess the potential influence of genetic drift, we estimated historical
466 effective population sizes (N_e) over time, using MSMC2 (Schiffels & Wang, 2020). For all
467 species we found that N_e decreased from the distant past until ~20,000 years ago,
468 followed by recent moderate recoveries (**Figure 2d**). From these analyses, we retrieved
469 the current estimate of N_e, which ranged from ~35,000 individuals in *arizonica* to ~75,000
470 in *V. mustangensis* (**Table 1**); *V. mustangensis* also yielded the highest N_e estimate based
471 on another method, SMC++ (Terhorst, Kamm, & Song, 2017) (**Figure S5**) and the results
472 were consistent among individual MSMC2 runs (**Figure S6**). The relatively high N_e for *V.*
473 *mustangensis* was surprising, because it currently has a narrow geographic distribution

474 (**Figure 1a**), but it nonetheless suggests that it may be less prone to drift effects than other
475 species.

476 *Layer 5 - Genomic Load:* Genomic load, the accumulation of deleterious mutations
477 in a population, can predict fitness (Frankham, 2005). We estimated genomic load within
478 and between species by measuring the number of potentially deleterious alleles per
479 accession (see Methods). To account for differences in the number of SNPs between
480 species and potential reference biases, we focused only on SNPs shared between species,
481 with no missing data. Overall, genomic load indicated that *V. mustangensis*, *V. arizonica*
482 and *V. riparia* individuals had relatively low genomic load among species, but there was
483 also notable variation among individuals within species (**Figure 2e**).

484 *Layer 6 - Resistance to Pierce's Disease:* Biotic interactions between species can
485 play a key role in shaping population's responses to climate (Bascompte, García, Ortega,
486 Rezende, & Pironon, 2019; Zamora-Gutiérrez, Rivera-Villanueva, Martínez Balvanera,
487 Castro-Castro, & Aguirre-Gutiérrez, 2021). As a case study for including such information,
488 we included data about resistance to the bacterium *X. fastidiosa* (Morales-Cruz et al.,
489 2021; Summaira Riaz et al., 2020). We plotted CFU counts for each individual, revealing
490 variation in PD resistance among individuals and species (**Figure 2f**) – e.g., 100% of *V.*
491 *arizonica* and 87% of *V. mustangensis* accessions were PD resistant while the *V. girdiana*
492 distribution was notably bimodal, with ~50% of accessions resistant (CFU < 13). In
493 contrast, only one *V. riparia* and two *V. berlandieri* accessions were resistant.

494

495 **The FOLDS framework:**

496 Ideally, assessing the response of populations to climate change should be based
497 on multiple layers of information (Razgour et al., 2018; Waldvogel et al., 2019); we relied
498 on the preceding six layers to surmise whether an accession is relatively well situated to
499 adapt to climate change. Such accessions should have low offsets, low genomic loads,
500 high S_a values, high N_e (low drift), PD resistance and predicted persistence based on SDMs.
501 We applied the FOLDS (Aguirre-Liguori et al., 2021) using different thresholds
502 (Figure S1). For example, we first filtered accessions based on all six layers, with
503 thresholds defined by: PD resistance (CFU < 13), SDM persistence, among the lowest 25%
504 of local offset and genomic load within a species, and among the highest 25% of Ne and Sa
505 within a species. We illustrated the intersection of layers for the IPSL-CM6A-
506 LR_ssp585_2061-2080 climatic model (Figure 3a), based on the set of summary statistics
507 (Table S4). The figure shows some of the combinations of layers, and the number of
508 accessions that pass that combination of filters. We can see, for example, that five
509 accessions passed all six layers, suggesting that they are best poised to contribute alleles
510 to adapt climate change, based on location, climate and genomic data. In addition, results
511 for the model in Figure 3a, we provided the complete set of summary statistics used for
512 the UpsetR analyses, with results for the additional 53 climate models (Table S1). These
513 additional tables can be used to determine which accessions are suited to respond
514 adequately to climate change for a particular FCC model.
515 Since we examined 54 FCCs, we calculated the number of times an accession
516 passed the threshold for all six layers for each FCC model (Figure 3b-c). Overall, we found
517 that for the conservative thresholds (0.25 and 0.5), only accessions of *V. mustangensis*

518 passed all six filters for at least one climate model (**Figure 3b-c**). For the 25% threshold,
519 four *V. mustangensis* accessions (Vmus02, Vmus03, Vmus14 and Vmus16) passed all filters
520 for at least 48% of the 54 climate models. Based on this information, we predict that *V.*
521 *mustangensis* accessions are relatively well situated to persist to predicted shifts in
522 climate or to contribute adaptive variation in the future.

523 We also relaxed the thresholds, so that (for example) accessions remained when
524 they retained among the lowest 90% of offset and genomic load within a species, and
525 among the highest 10% of N_e and S_a within a species (**Figure 3c-d**). At these threshold
526 levels, a few accessions of *V. girdiana* and *V. arizonica* passed the filters. Importantly, the
527 exercise also provided potential insight into relative risks among taxa. For example, only
528 one *V. riparia* accession remained, even after applying 90% thresholds, for most models
529 (Figure 3d). These results suggest that this species is particularly susceptible to population
530 extinction within our sampled areas.

531

532 **Assessing the potential of wild *Vitis* to mitigate climate change in viticulture**

533 Given insights into the potential persistence of wild species, our next task was to
534 evaluate their potential to aid viticulture in the face of climate change. To evaluate which
535 CWRs might be best suited for *in situ* adaptation of the crop, we relied on the concept of
536 migration load (Rhoné et al. 2021), which is the genetic offset for an individual calculated
537 between its present-day sampling site and other locations. In this case, the 'other
538 locations' were all the 1278 *V. vinifera* locations in the United States identified in GBIF
539 (**Figure 4a, Table S5**).

540 The approach and results are perhaps most easily represented by examples. *V.*
541 *mustangenesis* accession vmus16 was one of six accessions that passed the FOLDS criteria
542 at the strictest threshold level (Figure 3). As an illustration, we calculated the migration
543 load for this accession in one FCC model (IPSL-CM6A-LR_ssp585_2061-2080), (Figure 4a).
544 We then asked: which, if any, of the migration loads fell within the range of local offsets
545 for the wild species? If the migration offset fell within the range of local offsets, we
546 surmised that vmus16 is a potential bioclimatic fit for crop adaptation at that location.
547 Altogether, vmus16 had low migration offsets in many locations of the Northeastern USA
548 and in one location in the Northwest for this single FCC model (**Figure 4a**). These are, then,
549 candidate regions for the use of this CWR, where it may have climate alleles that could
550 contribute to agricultural applications.
551 We extended these analyses to the six *V. mustangenesis* accessions that passed
552 the filters of the FOLDs model (Figure 4b), calculating the migration load for each
553 accession and each FCC model at all 1278 *V. vinifera* locations. After assessing whether
554 the migration offset fell within the range of local offsets for each FCC model, we counted
555 the proportion of times the *V. mustangenesis* accessions passed the assessment at each
556 location. The resulting map illustrated the potential suitability of *V. mustangensis* at each
557 location (**Figure 4b**). Overall, *V. mustangenensis* was projected to be more suitable for
558 predicted climates in the Northeast US than in the Western US.
559 We expanded these analyses to all 48 accessions that passed filters for all six layers
560 of the FOLDS model across the three threshold levels. We found 2630 instances in which
561 an accession passed the FOLDS model for at least one FCC model and one threshold level,

562 including 2096 cases for 90% FOLDS threshold, 391 cases for the 50% threshold, and 143
563 cases for strictest (25%) threshold (**Figure 3b-d**). From these, we estimated 2,630 x
564 1,278=3,361,140 migration loads. Tables S5 & S6 provide the calculated migrations loads
565 and indicates whether they fell within the range of local offsets. These Tables thus provide
566 a preliminary evaluation of the potential climate adaptability of wild *Vitis* accessions in
567 regions of *V. vinifera* cultivation.
568
569 **DISCUSSION**
570 Climate change is a threat to all biodiversity (Parmesan, 2006; Parmesan & Yohe,
571 2003), but the fate of CWRs is particularly important. A growing literature has predicted
572 the fate of CWRs under climate change, based on SDMs, but without explicit consideration
573 of evolutionary change, evolutionary history, or phenotypes. Here we have focused on
574 five *Vitis* CWRs as exemplars to address three goals. The first has been to incorporate
575 genomic data to assess the relative potential of CWRs to persist under climate change. The
576 second has been to assess which CWRs, if any, have bioclimatic and genomic profiles that
577 are of potential agronomic interest, pending further experimental evaluation. Finally, we
578 offer this work as a framework to apply to other CWRs and crop systems.
579
580 **Evaluating CWR Persistence: insights and caveats**
581 Predicting how populations and species will respond to climate change is not a
582 simple task because multiple factors --- including evolutionary, human, and ecological
583 processes –impact the response of populations to climate change (Waldvogel et al., 2019).

584 We have combined individual measures, including SDMs and measures of genomic
585 variability, to assess the relative ability of CWRs to respond to climate change. To our
586 knowledge, no formal statistical framework exists to perform this task; instead, we relied
587 on the existing FOLDS framework (Aguirre-Liguori et al., 2021).

588 We recognize that this approach is not complete for many reasons. For example,
589 myriad additional biotic and abiotic phenomena likely affect the ability of populations to
590 evolve in the face of climate change, but not all have been (or can be) measured. In
591 theory, an advantage of the FOLDS approach is that it can incorporate unlimited layers of
592 information, including additional biotic interactions, genetic and environmental data, and
593 potential human impact data (e.g., urbanization or human-mediated fire threats) (Aguirre-
594 Liguori et al. 2021). However, a corresponding shortcoming is that individual layers are
595 weighted equally and treated independently, which may be inaccurate. For example, we
596 have found that N_e and genomic load are correlated across NA *Vitis* accessions (r = -0.41,
597 p<0.001; **Figure S9**). While this correlation is not unexpected -- because selection is less
598 efficient in small populations -- it implies that information from the two layers is partially
599 redundant. We have also unexpectedly found that genomic load and PD resistance are
600 correlated (r=-0.26; p< 0.01; Figure S9). Perhaps the simplest explanation for the
601 correlation is that less-fit plants (as measured by genomic load) are more susceptible to
602 disease, but it again demonstrates that two seemingly unrelated layers may not provide
603 fully independent information. One important extension of the FOLDS (or similar) models
604 will be the ability to weight the contribution of individual layers to maximize predictive
605 power. This may not be possible, however, without experimental data from climate

606 experiments that can empirically test the accuracies of predictions. Such experiments are
607 an obvious and important next step.

608 Additional shortcomings of our analyses relate to our treatment of the *Vitis*
609 dataset. For example, local offsets are typically calculated on estimated allele frequencies
610 within populations (Fitzpatrick & Keller, 2015), but we have applied it to diploid
611 frequencies within individuals. While this may decrease predictive power, and is likely
612 conservative for that reason, both simulations and resampling analyses suggest that our
613 conclusions based on single genotypes are robust (See **Supplementary Methods**). We
614 have also ignored the potential for natural migration, which may also impact
615 opportunities to adapt to climate change.

616 Nonetheless, our synthesis provides a framework to begin to evaluate the relative
617 risk among CWRs to predicted climate change that goes beyond SDMs and niche
618 modeling. By considering evolutionary information, we predict that *V. mustangensis*
619 accessions are the most apt to persist in the face of predicted climate change. The *V.*
620 *mustangensis* accessions have good persistence in SDMs, high *Ne*, low genomic load, low
621 genomic offsets and PD resistance. The framework also provides additional information
622 for conserving other wild *Vitis* species. For example, some individuals scored well for five
623 of six layers but have either low *Ne* or low adaptive scores **(Figure 3a**). Based on this
624 information, it may be necessary to aid management of source populations by adding
625 genetic diversity from different populations or species. Finally, we accentuate again that
626 any predictions should, and can, be validated experimentally (Kardos & Shafer, 2018) –

627 e.g., by growing suitable wild *Vitis* accessions in common gardens that simulate climate
628 change and testing for fitness effects.
629
630 **Crop adaptation and climate change**
631 Given genomic information from CWRs, we have used the GF framework to assess
632 the potential of specific accessions to grow in the projected climates where domesticated
633 grapevines are currently grown in the United States. The motivation for this approach is
634 that viticulture is under climate risk. For example, Morales-Castilla et al. (2020) have used
635 phenological data to estimate that 56% of current, worldwide growing locations will be
636 lost under 2Cº of warming, although this value decreased to 24% when they considered
637 intraspecific variety in phenology among grapevine cultivars. Similarly, Hannah et al.
638 (2013) projected a 25% to 73% decrease in suitable land for viticulture by 2050 among
639 major wine producing regions. Hence, we view our analyses as a step toward assessing
640 whether CWRs have alleles that can help mediate the effects of climate change on
641 viticulture.
642 An important question is how best to counteract climate threats to crops.
643 Generally, there are three non-exclusive strategies. One is human-assisted migration that
644 shift the locations of cultivation for a specific crop (Sloat et al., 2020). A second is to
645 develop new agricultural regions – i.e., to expand arable land. However, land availability is
646 limited, and the ecological and economic costs of this strategy is especially large (Fita,
647 Rodríguez-Burruezo, Boscaiu, Prohens, & Vicente, 2015). The third strategy is *in situ*
648 adaptation – i.e., breeding cultivars that tolerate stresses associated with climate change

649 in their current locations of cultivation (Sloat et al., 2020). This is a strategy that has been
650 advocated previously for *Vitis* rootstocks (Callen et al, 2016) and one for which genomic
651 data may be helpful. In fact, Genomic data have been used to inform this strategy, as
652 illustrated in a landmark study of pearl millet (Rhoné et al., 2020). In the study, Rhoné et
653 al. (2020) generated genomic data from landraces across the African continent, identified
654 geographic regions where cultivation was at risk due to climate change, and then
655 determined the migration load to determine which landraces have genomic profiles that
656 may thrive if moved to at-risk regions.

657 Using a similar approach, we have asked which of our sampled CWRs have genetic
658 and bioclimatic profiles that may make them suitable for growing in viticulture locations
659 within the United States. Our CWR sample from the Southwestern U.S. may be
660 particularly suitable for this study, because the accessions are adapted to warm and dry
661 environments and hence may be needed for future breeding programs (Hietnitz et al.
662 2019). Since the FOLDS model predicted that six *V. mustangensis* accessions are
663 particularly likely to persist under climate change, we focused on estimating their
664 migration loads across a range of climate models. Migration load was often within the
665 range of local offsets for that species (**Figure 4**). Based on these results, we hypothesize
666 that some of these wild accessions may prove useful for *in situ* adaptation of cultivated
667 grapevine and specifically that *V. mustangensis* may prove vital agronomically.

668 It is important, however, to discuss this hypothesis critically. First, we recognize
669 that *V. mustangensis* is not currently used as a rootstock or for scion breeding. In fact, it
670 may not even be the best candidate CWR for *in situ* adaptation at vulnerable sites, given

671 that we have studied only five of 30 species and 105 accessions. In this context, it is worth
672 noting that *V. riparia* is commonly used as a rootstock but none of the *V. riparia*
673 accessions in our sample fared particularly well in the FOLDS model. We suspect this
674 reflects the fact that our sample is taken from the geographic limits of the species, where
675 genomic load is often higher and fitness may be lower (Willi et al. 2018).

676 Second, our analyses are clearly only a first step, because many additional features
677 of wild germplasm must be characterized before assessing agronomic utility. For example,
678 although our study is rare for including one potential biotic interactor, there are many
679 more other potential biotic interactors (Griggs, Steenwerth, Mills, Cantu, & Bokulich,
680 2021) -- especially resistance to phylloxera -- that are crucial for agricultural use. We also
681 are not considering important complexities about viticulture, like irrigation and soil type.
682 It is also an open question whether *in situ* adaption to climate change in grapevine will be
683 better achieved by identifying new rootstocks or by using CWRs for scion breeding. The
684 latter has already been shown to be effective for discrete traits that are governed by a few
685 major loci, such as disease resistance. Indeed, PD resistance at the *Pdr1* locus (Krivanek,
686 Riaz, & Walker, 2006) has been backcrossed into various varieties to introduce PD
687 resistance. Bioclimatic adaptation is polygenic, however, as evidenced by the thousands
688 of SNPs associated with bioclimatic variables in each of these species (**Table 1**) and by the
689 fact that associated SNPs are distributed throughout the chromosomes (Morales-Cruz et
690 al., 2021). Given, the polygenic nature of bioclimatic adaptation, we suspect that *Vitis*
691 CWRs will typically be more effective as rootstocks. That said, rootstocks can have a
692 multitude of effects on the scion, including yield, phenology and drought tolerance that

693 may also display environmental interactions (Warschefsky et al., 2016). Moreover,
694 potential rootstock and scion combinations can be incompatible, for reasons that are not
695 yet well understood genetically (Gaut, Miller, & Seymour, 2019).
696 Finally, we have assumed that the climate models are accurate. Although we see
697 no way around this assumption, we have tried to represent uncertainty in climate
698 predictions by including 54 FCC models representing different time scales, circulation
699 models and SSPs. Some accessions have low migration loads in some locations across
700 most climate models, providing some consistency to the hope that these can be sources
701 for climate-related alleles that may prove useful. Altogether, we have applied a unique
702 combination of both climate and genomic information to identify potential and
703 unexpected candidates for priority evaluation. We believe that this methodological
704 approximation could be valuable for identifying CWRs of other crop for conservation or
705 improvement in the future.
706
707
708 **ACKNOWLEDGEMENTS**
709 The authors thanks Yongfeng Zhou, Summaira Riaz, Andrew Walker and Dario Cantu for
710 sharing data from previous manuscripts. This work was improved by the comments of
711 three anonymous reviewers and was supported by National Science Foundation grant no.
712 1741627 to B.S.G. J. A-L was supported in part by a UC-Mexus postdoctoral fellowship.
713
714 **REFERENCES**
715 Aguirre-Gutiérrez, J., van Treuren, R., Hoekstra, R., & van Hintum, T. J. L. (2017). Crop wild

relatives range shifts and conservation in Europe under climate change. *Diversity and Distributions*, *23*(7), 739–750. doi: 10.1111/ddi.12573

Aguirre-Liguori, Jonas A., Ramírez-Barahona, S., & Gaut, B. S. (2021). The evolutionary genomics of species' responses to climate change Jonás. *Nature Ecology & Evolution*. doi: 10.1038/s41559-021-01526-9

Aguirre-Liguori, Jonás A., Ramírez-Barahona, S., Tiffin, P., & Eguiarte, L. E. (2019). Climate change is predicted to disrupt patterns of local adaptation in wild and cultivated maize. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1906). doi: 10.1098/rspb.2019.0486

Bascompte, J., García, M. B., Ortega, R., Rezende, E. L., & Pironon, S. (2019). Mutualistic interactions reshuffle the effects of climate change on plants across the tree of life. *Science Advances*, *5*(5), 1–9. doi: 10.1126/sciadv.aav2539

Brady, S. P., Bolnick, D. I., Angert, A. L., Gonzalez, A., Barrett, R. D. H., Crispo, E., … Hendry, A. P. (2019). Causes of maladaptation. *Evolutionary Applications*, *12*(7), 1229–1242. doi: 10.1111/eva.12844

Callen, S. T., Klein, L. L., & Miller, A. J. (2016). Climatic niche characterization of 13 North American vitis species. *American Journal of Enology and Viticulture*, *67*(3), 339–349. doi: 10.5344/ajev.2016.15110

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic Prediction of (Mal)Adaptation across Current and Future Climatic Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*, 245–269. doi: 10.1146/annurev-ecolsys-020720-042553

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., … Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, *2*(4), 1–6. doi: 10.1038/NPLANTS.2016.22

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, *6*(2), 80–92. doi: 10.4161/fly.19695

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940. doi: 10.1093/bioinformatics/btx364

Des Roches, S., Pendleton, L. H., Shapiro, B., & Palkovacs, E. P. (2021). Conserving intraspecific variation for nature's contributions to people. *Nature Ecology and Evolution*, *5*(5), 574–582. doi: 10.1038/s41559-021-01403-5

Ellis, N., Smith, S. J., & Roland Pitcher, C. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168. doi: 10.1890/11-0252.1

Exposito-Alonso, M., Burbano, H. A., Bossdorf, O., Nielsen, R., & Weigel, D. (2019). Natural selection on the Arabidopsis thaliana genome in present and future climates. *Nature*, *573*(7772), 126–129. doi: 10.1038/s41586-019-1520-9

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in Arabidopsis thaliana. *Nature Ecology and Evolution*, *2*(2), 352–358. doi: 10.1038/s41559-017-0423-0

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Feeley, K. J., Rehm, E. M., & Machovina, B. (2012). perspective: The responses of tropical forest species to global climate change: acclimate, adapt, migrate, or go extinct? *Frontiers of Biogeography*, *4*(2). doi: 10.21425/f5fbg12621

Fick, S. E., & 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. doi: 10.1002/joc.5086

Fita, A., Rodríguez-Burruezo, A., Boscaiu, M., Prohens, J., & Vicente, O. (2015). Breeding and domesticating crops adapted to drought and salinity: A new paradigm for increasing food production. *Frontiers in Plant Science*, *6*(NOVEMBER), 1–14. doi: 10.3389/fpls.2015.00978

Fitzpatrick, M. C., Chhatre, V. E., Soolanayakanahally, R. Y., & Keller, S. R. (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources*, (August 2020), 1–17. doi: 10.1111/1755-0998.13374

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16. doi: 10.1111/ele.12376

Fitzpatrick, M. C., Keller, S. R., & Lotterhos, K. E. (2018). Comment on “Genomic signals of selection predict climate-driven population declines in a migratory bird.” *Science*, *361*(6401), 1–2. doi: 10.1126/science.aat7956

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*(2), 131–140. doi: 10.1016/j.biocon.2005.05.002

Gaut, B. S., Miller, A. J., & Seymour, D. K. (2019). Living with Two Genomes: Grafting and Its Implications for Plant Genome-to-Genome Interactions, Phenotypic Variation, and Evolution. *Annual Review of Genetics*, *53*, 195–215. doi: 10.1146/annurev-genet-112618-043545

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*(4), 1555–1579. doi: 10.1534/genetics.115.181453

Gougherty, A. V., Keller, S. R., & Fitzpatrick, M. C. (2021). Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nature Climate Change*, *11*(2), 166–171. doi: 10.1038/s41558-020-00968-6

Griggs, R. G., Steenwerth, K. L., Mills, D. A., Cantu, D., & Bokulich, N. A. (2021). Sources and Assembly of Microbial Communities in Vineyards as a Functional Component of Winegrowing. *Frontiers in Microbiology*, *12*(April). doi: 10.3389/fmicb.2021.673810

Hällfors, M. H., Liao, J., Dzurisin, J., Grundel, R., Hyvärinen, M., Towle, K., … Hellmann, J. J. (2016). Addressing potential local adaptation in species distribution models: Implications for conservation under climate change. *Ecological Applications*, *26*(4), 1154–1169. doi: 10.1890/15-0926

Heinitz, C. C., Uretsky, J., Peterson, J. C. D., Acosta, K. G. H.-, & Walker, M. A. (2019). Crop

Wild Relatives of Grape (Vitis vinifera L .) Throughout North America. In *North American Crop Wild Relatives* (pp. 329–351).

Ikeda, D. H., Max, T. L., Allan, G. J., Lau, M. K., Shuster, S. M., & Whitham, T. G. (2017). Genetically informed ecological niche models improve climate change predictions. *Global Change Biology*, *23*(1), 164–176. doi: 10.1111/gcb.13470

Janzen, G. M., Wang, L., & Hufford, M. B. (2019). The extent of adaptive wild introgression in crops. *New Phytologist*, *221*(3), 1279–1288. doi: 10.1111/nph.15457

Jarvis, A., Lane, A., & Hijmans, R. J. (2008). The effect of climate change on crop wild relatives. *Agriculture, Ecosystems and Environment*, *126*(1–2), 13–23. doi: 10.1016/j.agee.2008.01.013

Kardos, M., & Shafer, A. B. A. (2018). The Peril of Gene-Targeted Conservation. *Trends in Ecology and Evolution*, *33*(11), 827–839. doi: 10.1016/j.tree.2018.08.011

Khoury, C. K., Carver, D., Greene, S. L., Williams, K. A., Achicanoy, H. A., Schori, M., … Frances, A. (2020). Crop wild relatives of the United States require urgent conservation action. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(52), 33351–33357. doi: 10.1073/PNAS.2007029117

Krivanek, A. F., Riaz, S., & Walker, M. A. (2006). Identification and molecular mapping of PdR1 a primary resistance gene to Pierce's disease in Vitis. *Theoretical and Applied Genetics*, *112*(6), 1125–1131. doi: 10.1007/s00122-006-0214-5

Láruson, Á. J., Fitzpatrick, M. C., Keller, S. R., Haller, B. C., & Lotterhos, K. E. (2022). Seeing the forest for the trees: Assessing genetic offset predictions from gradient forest. *Evolutionary Applications*, *15*(3), 403–416. doi: 10.1111/eva.13354

Lawler, J. J., White, D., Neilson, R. P., & Blaustein, A. R. (2006). Predicting climate-induced range shifts: Model differences and model reliability. *Global Change Biology*, *12*(8), 1568–1584. doi: 10.1111/j.1365-2486.2006.01191.x

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., … Dong, Y. (2019). Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses. *Nature Communications*, *10*(1), 1–12. doi: 10.1038/s41467-019-09135-8

Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., Hubisz, M. J., … Bustamante, C. D. (2008). Proportionally more deleterious genetic variation in European than in African populations. *Nature*, *451*(7181), 994–997. doi: 10.1038/nature06611

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. doi: 10.1038/nature18964

Marín, D., Armengol, J., Carbonell-Bejerano, P., Escalona, J. M., Gramaje, D., Hernández-Montes, E., … de Herralde, F. (2021). Challenges of viticulture adaptation to global change: tackling the issue from the roots. *Australian Journal of Grape and Wine Research*, *27*(1), 8–25. doi: 10.1111/ajgw.12463

McCouch, S. (2013). Feeding the future. *Nature*, *499*, 13–17.

Morales-castilla, I., Cook, B. I., Lacombe, T., Parker, A., Leeuwen, C. Van, Nicholas, K. A., & Wolkovich, E. M. (2020). Diversity buffers winegrowing regions from climate change

losses. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1906731117

Morales-Cruz, A., Aguirre-Liguori, J. A., Zhou, Y., Minio, A., Riaz, S., Walker, A. M., … Gaut, B. S. (2021). Introgression among North American wild grapes (Vitis) fuels biotic and abiotic adaptation. *Genome Biology*, *22*(1), 1–27. doi: 10.1186/s13059-021-02467-z

Nadeau, C. P., & Urban, M. C. (2019). Eco-evolution on the edge during climate change. *Ecography*, *42*(7), 1280–1297. doi: 10.1111/ecog.04404

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814. doi: 10.1093/nar/gkg509

Ollat, N., Peccoux, A., Papura, D., Esmenjaud, D., Marguerit, E., Tandonnet, J. P., … Delrot, S. (2016). Rootstocks as a component of adaptation to environment. *Grapevine in a Changing Environment: A Molecular and Ecophysiological Perspective*, 68–108. doi: 10.1002/9781118735985.ch4

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., … Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, *51*(11), 933–938. doi: 10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Parmesan, C. (2006). Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 637–669. doi: 10.1146/annurev.ecolsys.37.091305.110100

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change. *Nature*, *421*, 37–42. doi: 10.1038/nature01286

Peterson, A. T., Soberón, J., Richard G, P., Robert P, A., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions*.

Phillips, SJ, Anderson, R., & Schapire, R. (2006). Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. *Ecological Modelling*, *190*, 231–259. doi: 10.1016/j.ecolmodel.2005.03.026

Phillips, Steven, & Dudík, M. (2008). Modeling of species distributions with axent: new extensions and comprehensive evaluation. *Ecography*, *31*, 161–175. doi: 10.1111/j.2007.0906-7590.05203.x

Rapicavoli, J., Ingel, B., Blanco-Ulate, B., Cantu, D., & Roper, C. (2018). Xylella fastidiosa: an examination of a re-emerging plant pathogen. *Molecular Plant Pathology*, *19*(4), 786–800. doi: 10.1111/mpp.12585

Razgour, O., Taggart, J. B., Manel, S., Juste, J., Ibáñez, C., Rebelo, H., … Park, K. (2018). An integrated framework to identify wildlife populations under threat from climate change. *Molecular Ecology Resources*, *18*(1), 18–31. doi: 10.1111/1755-0998.12694

Rhoné, B., Defrance, D., Berthouly-Salazar, C., Mariac, C., Cubry, P., Couderc, M., … Vigouroux, Y. (2020). Pearl millet genomic vulnerability to climate change in West Africa highlights the need for regional collaboration. *Nature Communications*, *11*(1). doi: 10.1038/s41467-020-19066-4

Riaz, S., Huerta-Acosta, K., Tenscher, A. C., & Walker, M. A. (2018). Genetic characterization of Vitis germplasm collected from the southwestern US and Mexico to expedite Pierce's disease-resistance breeding. *Theoretical and Applied Genetics*, *131*(7), 1589–1602. doi: 10.1007/s00122-018-3100-z

Riaz, Summaira, Pap, D., Uretsky, J., Laucou, V., Boursiquot, J. M., Kocsis, L., & Andrew Walker, M. (2019). Genetic diversity and parentage analysis of grape rootstocks. *Theoretical and Applied Genetics*, *132*(6), 1847–1860. doi: 10.1007/s00122-019-03320-5

Riaz, Summaira, Tenscher, A. C., Graziani, R., Krivanek, A. F., Ramming, D. W., & Walker, M. A. (2009). Using marker-assisted selection to breed pierce's disease-resistant grapes. *American Journal of Enology and Viticulture*, *60*(2), 199–207.

Riaz, Summaira, Tenscher, A. C., Heinitz, C. C., Huerta-Acosta, K. G., & Andrew Walker, M. (2020). Genetic analysis reveals an east-west divide within North American Vitis species that mirrors their resistance to Pierce's disease. *PLoS ONE*, *15*(12 December), 7–10. doi: 10.1371/journal.pone.0243445

Schiffels, S., & Wang, K. (2020). MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In *Methods in molecular biology*. doi: 10.1007/978-1-0716-0199-0

Sloat, L. L., Davis, S. J., Gerber, J. S., Moore, F. C., Ray, D. K., West, P. C., & Mueller, N. D. (2020). Climate adaptation by crop migration. *Nature Communications*, *11*(1), 1–9. doi: 10.1038/s41467-020-15076-4

Soberón, J. M. (2010). Niche and area of distribution modeling: A population ecology perspective. *Ecography*, *33*(1), 159–167. doi: 10.1111/j.1600-0587.2009.06074.x

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303–309. doi: 10.1038/ng.3748

Thuiller, A. W., Georges, D., Engler, R., Georges, M. D., & Thuiller, C. W. (2016). The biomod2 package: the updated object-oriented version of BIOMOD package. *CRAN*, (2), 1–104. Retrieved from https://cran.r-project.org/web/packages/biomod2/index.html

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, *11*(1), 1–9. doi: 10.1038/nprot.2015.123

Waldvogel, A., Rellstab, C., Bataillon, T., Feldmeyer, B., Rolshausen, G., Exposito-alonso, M., … Pfenninger, M. (2019). Evolutionary genomics can improve prediction of species ' responses to climate change. *Evoution Letters*, 1–15. doi: 10.1002/evl3.154

Wan, Y., Schwaninger, H. R., Baldo, A. M., Labate, J. A., Zhong, G. Y., & Simon, C. J. (2013). A phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change. *BMC Evolutionary Biology*, *13*(1). doi: 10.1186/1471-2148-13-141

Warschefsky, E. J., Klein, L. L., Frank, M. H., Chitwood, D. H., Londo, J. P., von Wettberg, E. J. B., & Miller, A. J. (2016). Rootstocks: Diversity, Domestication, and Impacts on Shoot Phenotypes. *Trends in Plant Science*, *21*(5), 418–437. doi: 10.1016/j.tplants.2015.11.008

Zamora-Gutiérrez, V., Rivera-Villanueva, A. N., Martínez Balvanera, S., Castro-Castro, A., & Aguirre-Gutiérrez, J. (2021). Vulnerability of bat-plant pollination interactions due to environmental change. *Global Change Biology*, (March), 1–16. doi: 10.1111/gcb.15611

Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D., & Gaut, B. S. (2017). Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication. *Proceedings of the*

936 *National Academy of Sciences of the United States of America*, *114*(44), 11715–
937 11720. doi: 10.1073/pnas.1709257114
938
939

940 **DATA ACCESSIBILITY STATEMENT**

941 Sequence data that supports the findings of this study were downloaded from the Short

942 Read Archive at NCBI under BioProject ID: PRJNA731597.

943

944

945

946 **CONFLICT OF INTEREST**

947 The authors declare no conflicts of interest.

948

949 **AUTHOR CONTRIBUTIONS**

950 All authors contributed to the ideas of the study; J.A-L. & A.M-C. performed analyses; all

951 authors wrote the manuscript.

952

953 **TABLES**

954

TABLE 1 Five *Vitis* CWR samples, with summary statistics

Species	N_{ind}[1]	*Ne* (95% CI)[2]	N_{Bay}[3]	N_{GF}[4]	BIO[5]
V. arizonica	22	35,448 (33,658-37,237)	13,531	5,792	3,15,2,9
V. berlandieri	22	40,112 (37,008-43,215)	8,425	3,371	4,7,3,10
V. mustangensis	24	74,839 (59,722-89,957)	8,787	2,730	18,5,10,4
V. girdiana	18	53,666 (47,065-60,267)	5.650	5,254	4,7,3,11
V. riparia	19	49,279 (45,454-53,105)	10,626	8,720	10,4,8,1

[1] N_{ind} is the number of samples per species
[2] the mean effective population size estimated across 6 samples with 95% confidence intervals
[3] N_{Bay} refers to the number of outlier SNPs identified with Baypass
[4] N_{GF} refers to the number of outlier SNPs identified with Gradient Forest
[5] BIO includes the list of the four top bioclimatic variables identified with GF, in decreasing order of importance

955

956

957

958 **FIGURE LEGENDS:**

959 **FIGURE 1 Results based on SDMs of the five *Vitis* species in this study.** (a) Present-day
960 SDMs showing the estimated bioclimatic niche of each of the five species. Species are
961 colored according to the key. The map also provides the sampling locations of the 105
962 accessions, with shapes and colors of different species provided by the key. (b) The
963 geographic area of the niche for each species graphed into the future for 54 future climate
964 change (FCC) models and their mean. The symbols and colors represent the estimated
965 geographic size for different circulation models and shared socio-economic pathways,
966 respectively. The central line and symbols across the years 2050, 2070 and 2090,
967 represents the mean of SDMs based on the different FCC models. The colors represent
968 the species, as indicated in the key of panel A. (c) Each column represents one of the 105
969 accessions in this study. The color in the column indicates the proportion of times that an
970 accession was predicted to exist for a given time period across the 18 FCC models. If the
971 location of sampled individual did not fall within the future niche for all FCC models for a
972 given time period, as defined by the SDM, then it is no longer represented in the figure
973 (white column). Hence, by 2090, SDMs predict that most *V. arizonica* and *V. mustangensis*
974 samples will be able to persist in their current locations, as will a few *V. berlandieri*. In
975 contrast, most *V. riparia* in our sample are not predicted to persist.
976
977 **FIGURE 2 Layers of information used in the FOLDS model.** (a) Distribution of local
978 genomic offset for each species for future climate change (FCC) model
979 IPSL−CM6A−LR_ssp585_2061−2080. We darkened circles based on their status above and
980 below the median within each species. (b) Variation in the adaptive scores (S_a) across
981 species. Each circle represents a sampled accession. The dashed line shows the median of
982 S_a across all accessions; circles are darkened if they have S_a above the median. (c) An
983 example of correlations between S_a and bioclimatic variables, in this case the maximum
984 temperature of the warmest month (BIO 5) for *V. mustangensis* samples. The dashed line
985 shows the correlation for reference SNPs. For all results and test for potential biases see
986 **Figure S8.** (d) Change in effective population size across time for each species. The color of
987 each species is indicated by the key. (e) The measure of genomic load – i.e., the
988 proportion of predicted deleterious SNPs per genome across a shared set of SNPs –
989 plotted for each accession and species. [Note that the proportion of deleterious SNPs is
990 equivalent to the number of deleterious SNPs per genome, since all inferences were based
991 on a shared set of SNPs.] The dashed line represents the median of the genomic load.
992 Darker circles are below the median. (f) Distribution of the concentration of *X. fastidiosa*
993 in different wild *Vitis* species after exposure to the bacterium. The dashed line represents
994 the value 13 CFUs, which has been the boundary between resistance and susceptibility.
995 Darker circles represent accessions with CFUs < 13.
996
997

998 **FIGURE 3 The FOLDS model applied to wild *Vitis*.** (a) The upsetR plot shows the
999 intersection across all 105 individuals that pass thresholds (at the strictest 25% threshold)
1000 for the six layers of information and for future climate change model
1001 IPSL−CM6A−LR_ssp585_2061−2080. Black dots joined by a line indicate that individuals
1002 pass the threshold for those layers of information. The number of individuals presenting
1003 different intersections are at the right side of the plot. The orange line at the top,
1004 indicated that 5 individuals pass the thresholds for 6 layers of information. **(**b-d) The maps
1005 plot the individuals that remained after applying each of the three thresholds (25%, 50%,
1006 90%). The color of the symbols indicate the number of times an individual passes the 6
1007 layers of information based on the 54 FCC models. Table S7 reports the number of times
1008 an individual passed the FOLDS test for different thresholds FCC models.
1009
1010 **FIGURE 4 Climate change and in situ adaptation of *V. vinifera***. (a) An example of the
1011 locations where the migration load of *Vmus16* (see Figure 3) would be higher (white
1012 circles) or lower (blue circles) than the local genetic offset of *V. mustangesis* if it was
1013 moved to locations where *V. vinifera* currently grow. This was performed using the future
1014 climate change (FCC) model IPSL−CM6A−LR_ssp585_2061−2080. *Vmus16* would be more
1015 likely to be a suitable candidate for *in situ* adaptation of grapevine in the NE of the
1016 distribution. (b) percentage of times that the migration load for 6 *V. mustangensis*
1017 individuals identified by the FOLDS model at a 0.25 threshold (Figure 3b) would be lower
1018 than the local genetic offset of *V. mustangensis* across the 54 (FCC) models. The colors of
1019 the symbols indicate, from all the possible events, the percentage of times than assisted
1020 migration could potentially be used for *in situ* adaptation (out of 143 possible events).
1021

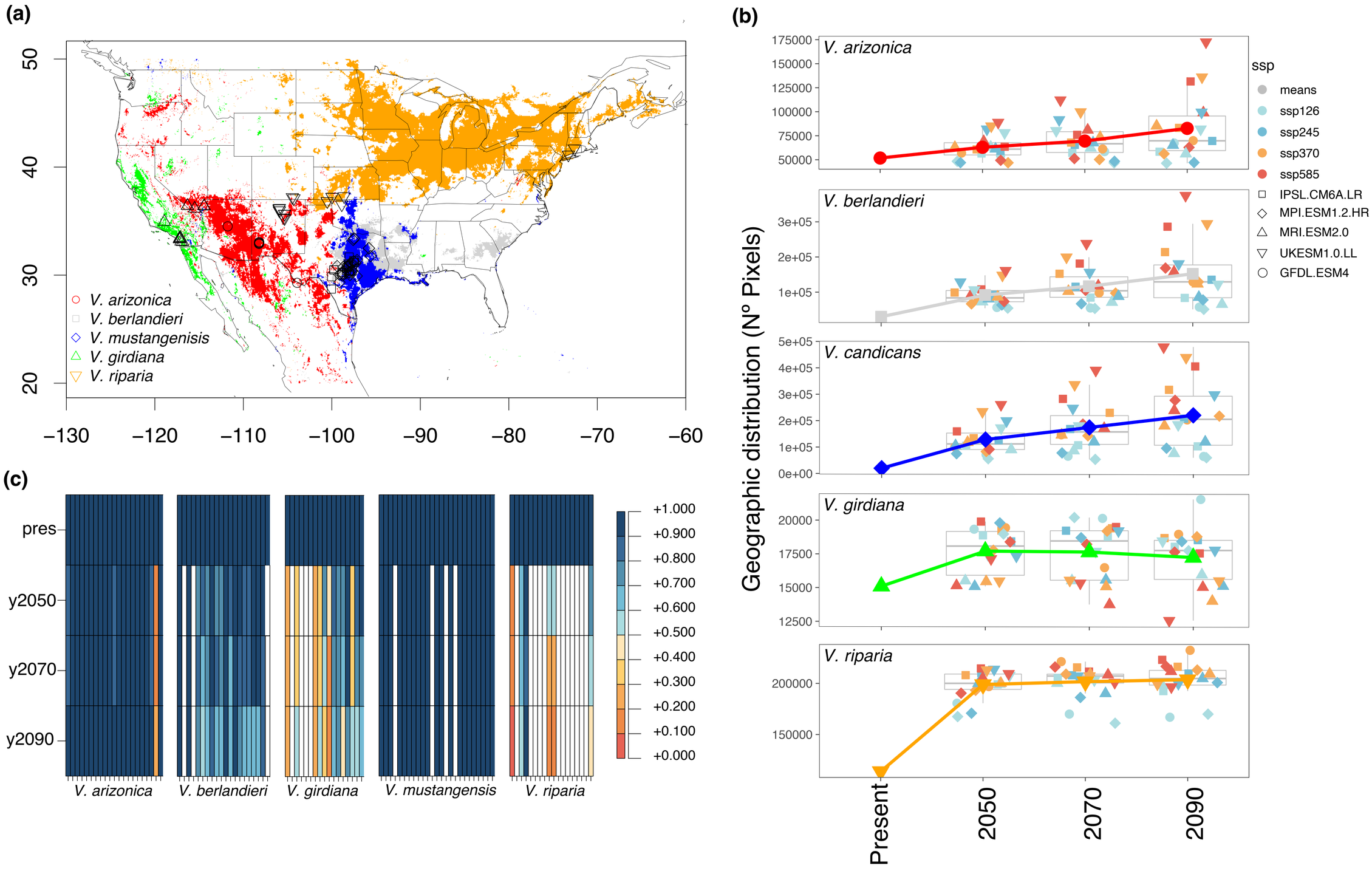

(a)
V. arizonica
V. berlandieri
V. mustangenisis
V. girdiana
V. riparia
50
40
30
20
−130
−120
−110
−100
−90
−80
−70
−60
(b)
V. arizonica
V. berlandieri
V. candicans
V. girdiana
V. riparia
Geographic distribution (Nº Pixels)
ssp
means
ssp126
ssp245
ssp370
ssp585
IPSL.CM6A.LR
MPI.ESM1.2.HR
MRI.ESM2.0
UKESM1.0.LL
GFDL.ESM4
Present
2050
2070
2090
(c)
pres
y2050
y2070
y2090
V. arizonica
V. berlandieri
V. girdiana
V. mustangensis
V. riparia
+1.000
+0.900
+0.800
+0.700
+0.600
+0.500
+0.400
+0.300
+0.200
+0.100
+0.000

Figure 1

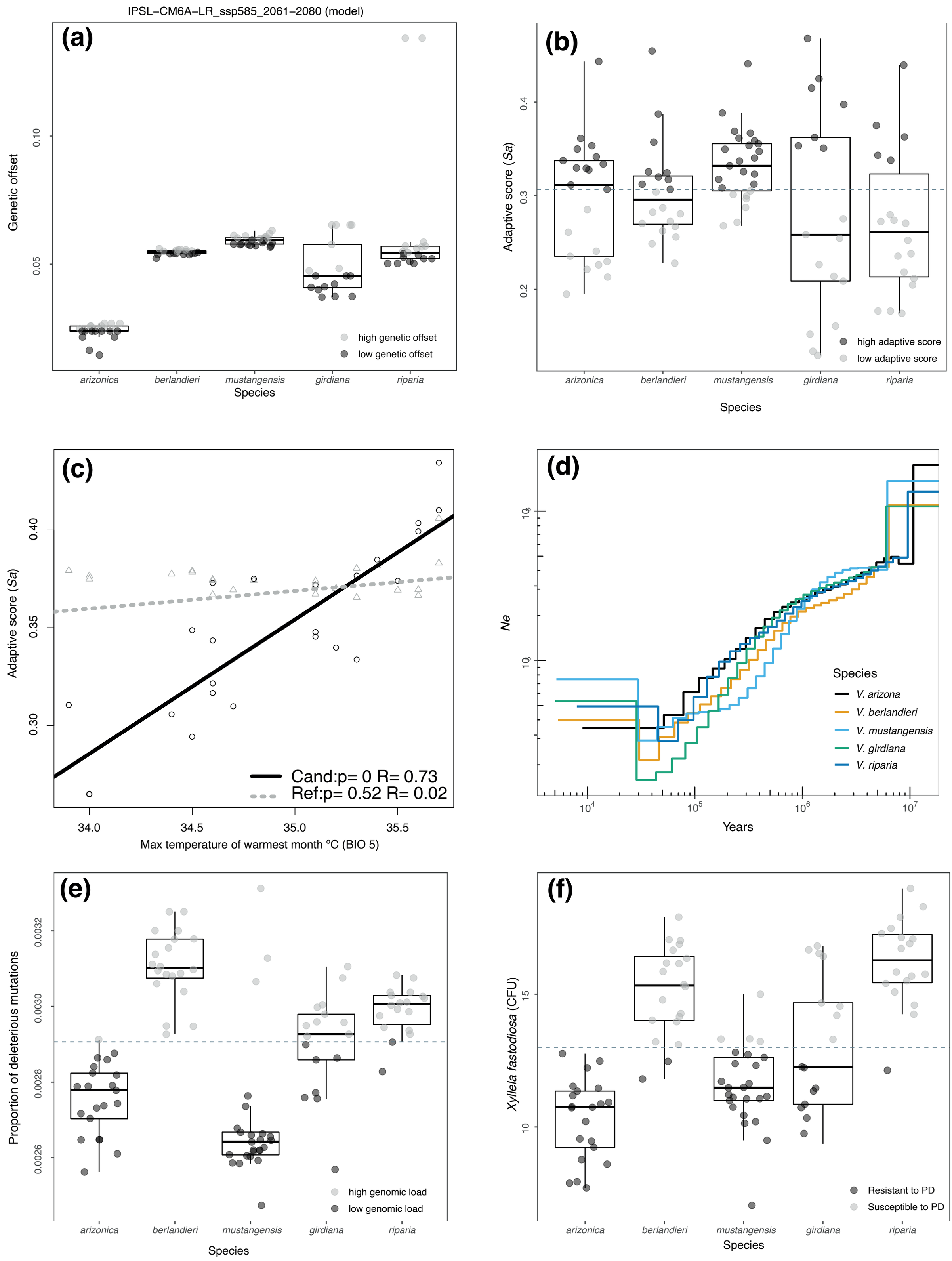

IPSL−CM6A−LR_ssp585_2061−2080 (model)
(a)
Genetic offset
0.05
0.10
high genetic offset
low genetic offset
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
(b)
Adaptive score (Sa)
0.2
0.3
0.4
high adaptive score
low adaptive score
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
(c)
Adaptive score (Sa)
0.30
0.35
0.40
Cand:p= 0 R= 0.73
Ref:p= 0.52 R= 0.02
34.0
34.5
35.0
35.5
Max temperature of warmest month °C (BIO 5)
(d)
Ne
Species
V. arizona
V. berlandieri
V. mustangensis
V. girdiana
V. riparia
10^4
10^5
10^6
10^7
Years
(e)
Proportion of deleterious mutations
0.0026
0.0028
0.0030
0.0032
high genomic load
low genomic load
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
(f)
Xyllela fastodiosa (CFU)
10
15
Resistant to PD
Susceptible to PD
arizonica
berlandieri
mustangensis
girdiana
riparia
Species

Figure 2

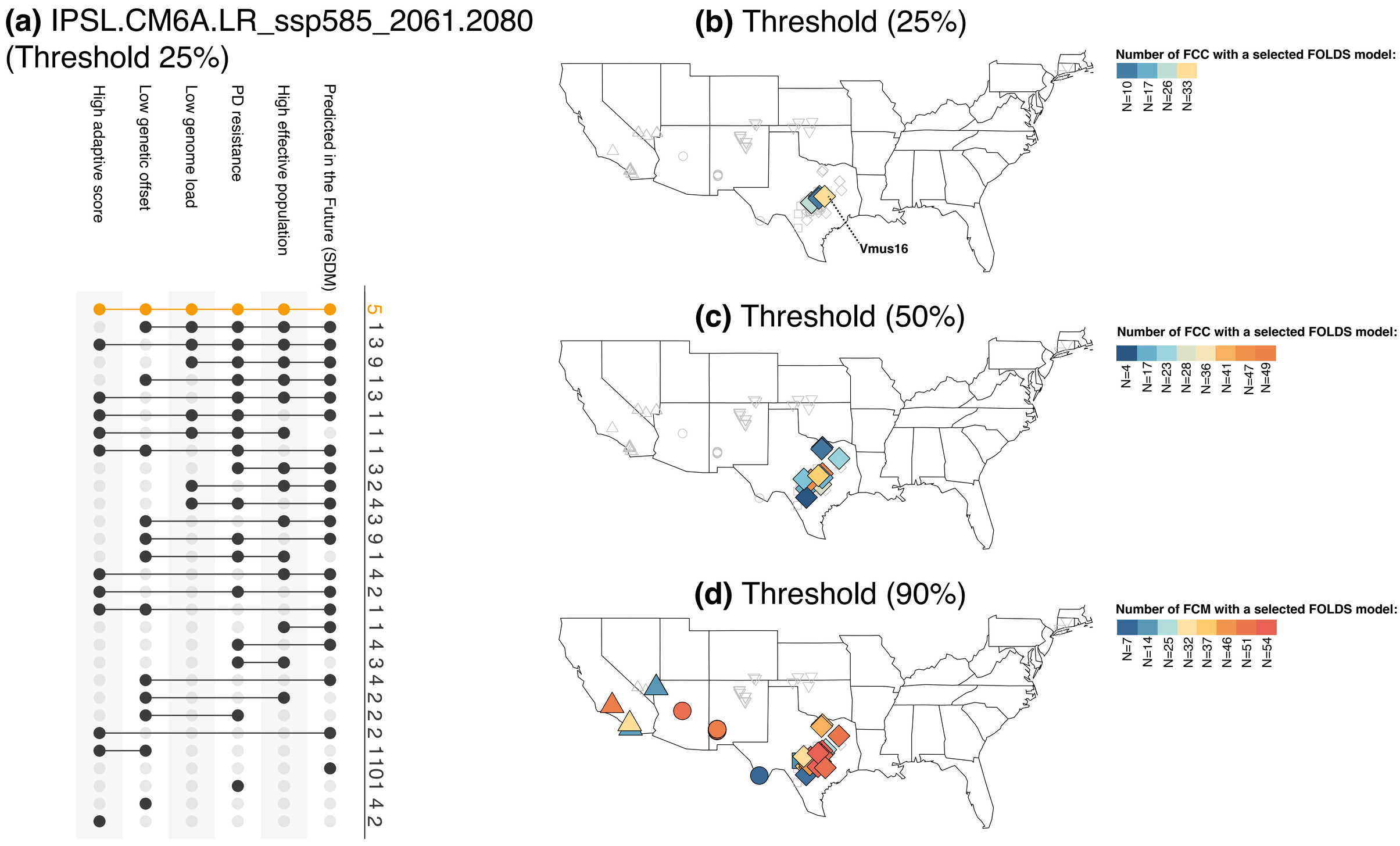

(a) IPSL.CM6A.LR_ssp585_2061.2080 (Threshold 25%)
Predicted in the Future (SDM)
High effective population
PD resistance
Low genome load
Low genetic offset
High adaptive score
5 1 3 9 1 3 1 1 1 3 2 4 3 9 1 4 2 1 1 4 3 4 2 2 2 1 1 0 1 4 2
(b) Threshold (25%)
Vmus16
Number of FCC with a selected FOLDS model:
N=10
N=17
N=26
N=33
(c) Threshold (50%)
Number of FCC with a selected FOLDS model:
N=4
N=17
N=23
N=28
N=36
N=41
N=47
N=49
(d) Threshold (90%)
Number of FCM with a selected FOLDS model:
N=7
N=14
N=25
N=32
N=37
N=46
N=51
N=54

Figure 3

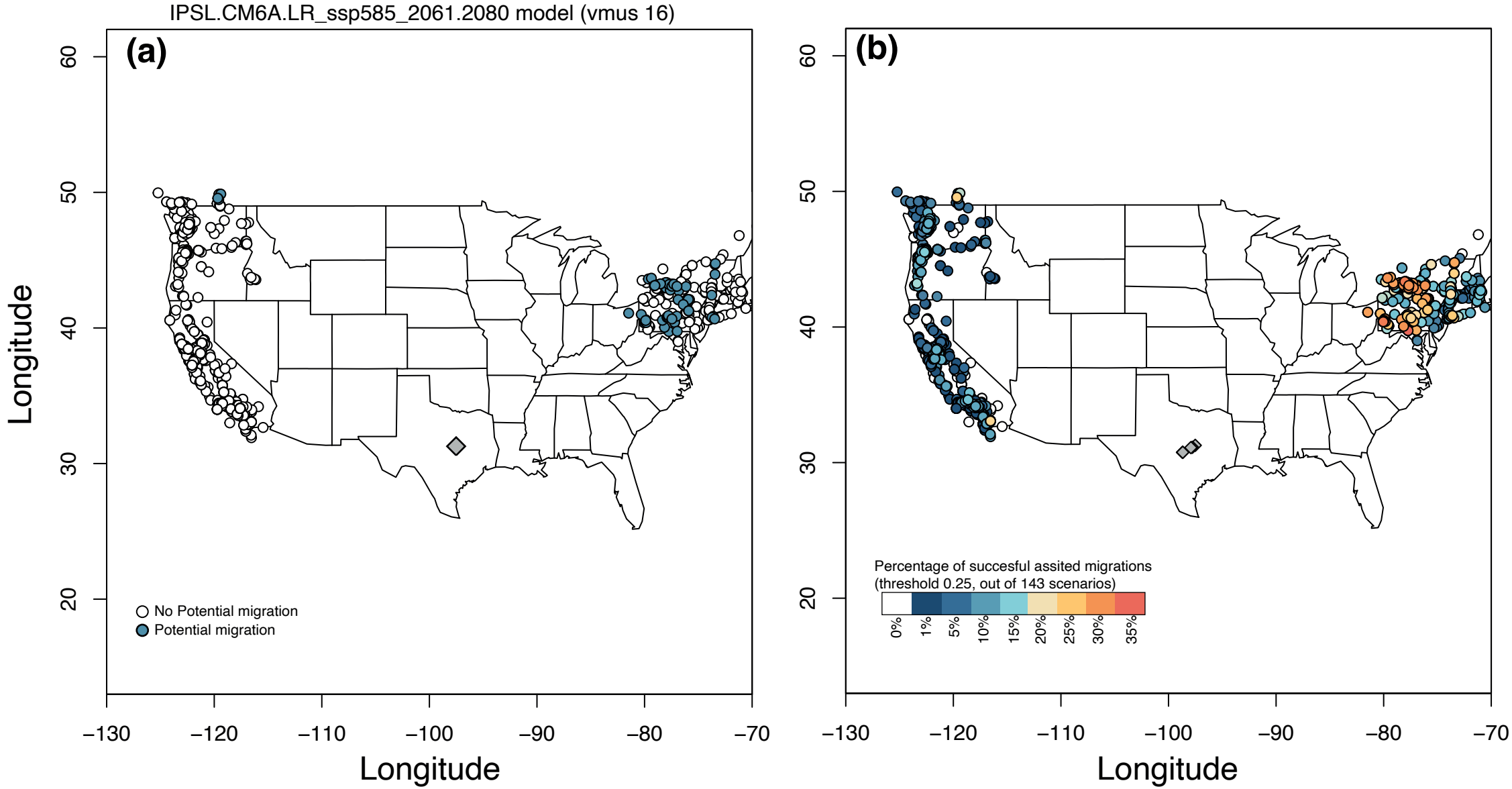

IPSL.CM6A.LR_ssp585_2061.2080 model (vmus 16)
(a)
(b)
Longitude
Longitude
Longitude
Longitude
No Potential migration
Potential migration
Percentage of succesful assited migrations
(threshold 0.25, out of 143 scenarios)
0%
1%
5%
10%
15%
20%
25%
30%
35%

Figure 4

SUPPORTING INFORMATION

Evaluating the persistence and utility of five wild *Vitis* species in the context of climate change

Aguirre-Liguori JA*, A Morales-Cruz* and BS Gaut
*joint first authors

Supplemental Methods

Adding random variation to allele frequencies to evaluate the application of GF to individuals: GF has been applied using allelic frequencies from populations instead of individual genotypes. We evaluated how sampling individuals may impact GF results. To do so we simulated "pseudo-frequencies" by changing the frequency (Δ) of each individual genotype ($0+\Delta$, $0.5\pm\Delta$ or $1-\Delta$) in each accession. The deviation Δ in each SNP and each individual was obtained by randomly sampling a value that ranged from 0 to 1, with a beta distribution and means 0.1 and 0.3 (**Figure S2a**). The idea of using two means was to increase the magnitude of Δ to account for a stronger bias in the sampling of individuals. For each pseudo-frequency dataset, we ran the GF model and estimated the 6 strongest BIO variable predicted by the model ($R^2>0$), plotted the turnover function for the BIO with the strongest R^2 and estimated the genetic offset of the pseudo-populations in 2070. These simulations were performed for all species 100 times.

Figure S2 shows the results for all simulations in *mustangensis,* but we found quantitatively similar results for all species. First, we found that generating datasets with pseudo-frequencies did not affect the general patterns of the turnover functions (**Figure S2b**). However, as expected, we did find variance between replicates and also that the power was lower for the simulations with Δ distributions with mean 0.3 than 0.1. Second, we found that in general the four most important variables identified for the real dataset were also found with the simulated datasets (**Figure S2c**). However, as Δ increased we found lower consistency. Importantly, however, the BIO variables that were predicted for the simulated data corresponded to the top 6 variables identified in the real dataset. Finally, when we compared the distributions of the simulated genetic offsets with the real genetic offsets, we found similar patterns (**Figure S2d**), indicating that populations with low offsets in the real dataset also had low offsets in the simulated datasets. While we cannot know the real frequencies in the populations, our results suggest that sampling individuals instead of populations may have a small effect on the estimation of the GF models, and that the general patterns are retained.

Testing the effect subsampling different number of individuals to estimate the genetic offsets: We evaluated if subsampling different number of individuals could modify the genetic offset rank between populations. For this, we downloaded the dataset of *Zea mays* ssp. *mexicana* that was used by Aguirre-Liguori et al. (2021) to propose the FOLDS model. We used this dataset because it consists of 23 populations, with between 11 and 15 individuals per population. To test the effect of sampling fewer individuals we wrote a script that 1)

subsamples 1 to 11 random individuals from each of the population, 2) performs an outlier analysis (LFMM2); 3) performs GF, 4) estimates the genetic offsets of subsampled populations, 5) calculates the rank correlation between the genetic offsets of the subsampled populations and the genetic offsets of the populations using all the sampled individuals. This was done for each subsampling N1 to N11, 1,000 times. Figure S3 shows the boxplots showing the range of the *r* values of the rank correlations for N1 to N11. We found that for all subsamples the rank correlation between the genetic offsets of the subsampled populations and entire populations was higher than 0.995, strongly suggesting that sampling one individual per population does not dramatically affect the rank order of the population's genetic offsets. In our study we define the persistence of populations based on thresholds that rely on relative rankings; based on these results, sampling individuals to estimate the genetic offsets may not heavily bias our results.

Calculating and Evaluating Sa: We calculated the adaptive score (S_a), which was motivated by an analogous measure, the population adaptive index (Bonin, Nicole, Pompanon, Miaud, & Taberlet, 2007). However, instead of analyzing the proportion of adaptive SNPs in a populations, like the adaptive index, S_a estimates the proportion of alleles that will be adaptive in the future. To find adaptive alleles, we first identified pre-adapted individuals and then ran a linear model for each candidate SNP across all individuals, to evaluate if the genotypic state correlated significantly with bioclimatic variables. If the linear model was not significant, we discarded the SNP. If it was significant, we considered the allelic state that was frequent in pre-adapted individuals to be the adaptive allele. For each individual, we summed the number of pre-adaptive alleles present in the accession (2 for each adaptive allele in the homozygous state, 1 for the heterozygous state) across all candidate loci, and then divided that number by the maximum number of potential adaptive alleles. The maximum number of potential adaptive alleles was calculated as 2 x N_{Sp}, where N_{Sp} is the number of candidate loci for a given species. We called this value the adaptive score (S_a), which ranges from 0 to 1, where 0 indicates that the population does not have adaptive alleles and 1 indicates that all their alleles are adaptive. For each species we estimated the linear association between the adaptive score across individuals and the 4 bioclimatic variables that have a strong impact on the GF models. Since the correlation between S_a were in general high (r>0.7), we estimated the mean S_a across individuals for the 4 bioclimatic variables.

Demographic inference: In the main text, we reported MSMC2 results based on unphased SNPs, but we also calculated demography based on phased SNPs. To phase, we extracted phase-informative SNPs (PIRs) for each sample as described in Delaneau et al. (2013) (Delaneau, Howie, Cox, Zagury, & Marchini, 2013), which identifies reads that span at least 2 heterozygous SNPs, and used the PIRs and the VCF files for each species by chromosome as input to shapeit2 v2.r904 (Delaneau, Zagury, & Marchini, 2013) to phase the SNPs. We used the "generate_multihetsep.py" script from the MSMC tools (https://github.com/stschiff/msmc-tools) to create the input files for MSMC2 taking into account the mappability mask and the coverage masks. For the phased-SNPs runs, we extracted the phased SNPs of the two individuals with the highest coverage per species to run MSMC2 (**Figure S5**).

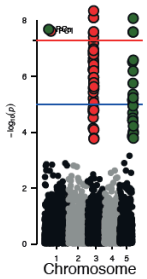
We also used SMC++ v1.15.4.dev18+gca077da (Terhorst, Kamm, & Song, 2017). To create a low coverage mask, we calculated sequencing coverage from the alignment files using the bedcov program from the samtools (v.1.10) package, supplying a bed file of 10 Kb non-overlapping windows across the genome. We calculated the sequencing coverage per species by summing the coverage of all individuals in the same genomic window. We classified regions as low coverage per species when they had a coverage lower than 5x. The low coverage masks were supplied together with the VCF files to the *vcf2smc* program of SMC++ for species and each chromosome. The demographic history was calculated per species using the *estimate* program from SMC++.

Comparing among methods, MSMC2 with phased and unphased SNPs yielded similar demographic trajectories over time, with the exception of current *Ne*s, which were up to two orders of magnitude higher with phased data. Since these estimates were much higher than the other two results (i.e., MSMC2 with unphased data and SMC++), we believe errors in the phasing process led to overestimation of Ne. In contrast to the two MSMC2-based analyses, the SMC++ results suggested that each taxon reached their lowest Ne at different time periods (**Figure S5**).

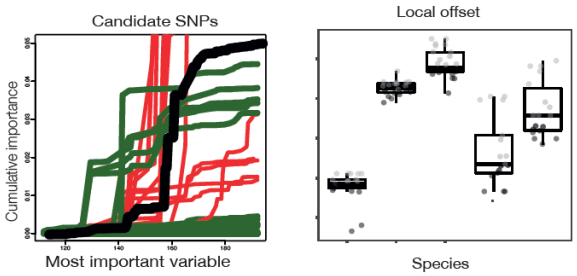
Supplemental Figures

APPLYING THE FOLDS MODEL

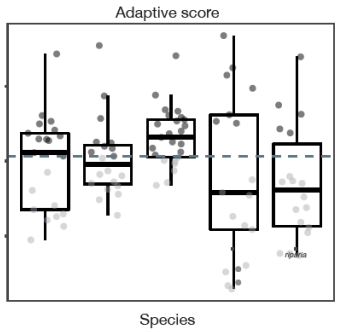
Step 1: Local adaptation
Identify outlier SNPs based on principal components of bioclimatic variables.



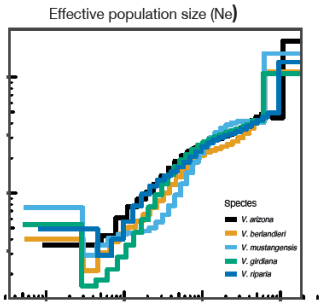
Step 2: Gradient Forest
Estimate local offsets and turnover functions, based on outlier SNPs and bioclimatic data.



Step 3: Adaptive Score
Calculate the polygenic adaptive score from candidate SNPs and turnover functions from step 2.



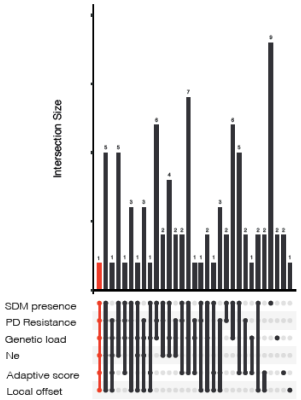
Step 4: Additional Summaries
Estimate summary statistics, such as Ne and genome load.



Step 5: Incidence Table
Combine the six layers into a single table with entries of 0 and 1 to indicate whether a particular accession and layer passes the relevant threshold.

Layer	Genetic offset	Mean adaptive score	Ne	Genome load	Resistance to PD	SDM
Acc 1	1	0	0	1	1	0
Acc 2	0	0	0	0	0	0
Acc 3	1	1	1	0	0	0
Acc 4	1	1	1	1	1	1
Acc 5	0	0	0	0	0	0
Acc 6	0	0	0	0	0	0

Step 6: UpsetR
UpsetR is an alternative to a Venn diagram, for plotting complex patterns of overlap among layers.



Step 7: Repeat
Repeat Steps #1 to #6 on all 60 climate models and three sets of thresholds.

Figure S1. A graphical summary of the steps used in the FOLDS model. Details for each step are provided in the Materials and Methods of the main text.

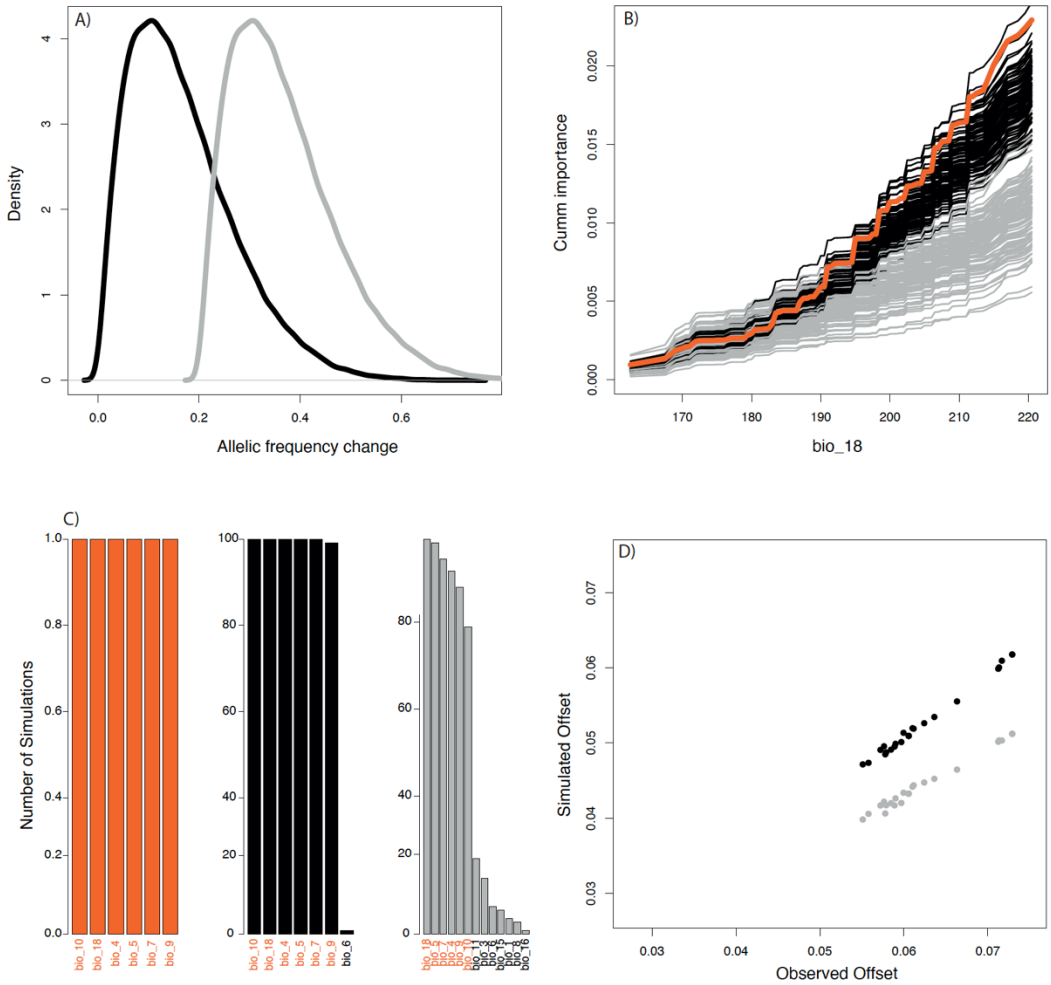


Figure S2. **The results of a probabilistic simulation to investigate the effects of using individuals in GF analyses.** A) Beta distribution of the magnitude of change of the simulated pseudo-frequencies (Δ) with means 0.1 (black) and 0.3 (grey) for the beta distribution. B) Turnover functions for Bio-18. Black and grey curves show 100 simulations with Δ using a beta distribution with mean 0.1 and 0.3, respectively. The orange curve shows the turnover function for the individual based dataset. C) Histograms showing the number of times the bioclimatic variables were found among the top 6 variables with strong importance (R^2>0) in the building of the GF models. The orange histogram shows the results obtained with the individual based dataset (1 model). Black and grey histograms show the 100 simulated datasets with different means. The y-axis denoting the proportion of resampled datasets that include the bioclimatic parameter as an important variable. The same bioclimatic variables were routinely identified as important in all three datasets. D) Scatterplot showing genetic offset for one 2070 FCC model of the individual dataset in the x-axis and the mean of the simulated datasets for the different Δ distributions. There is a strong correlation.

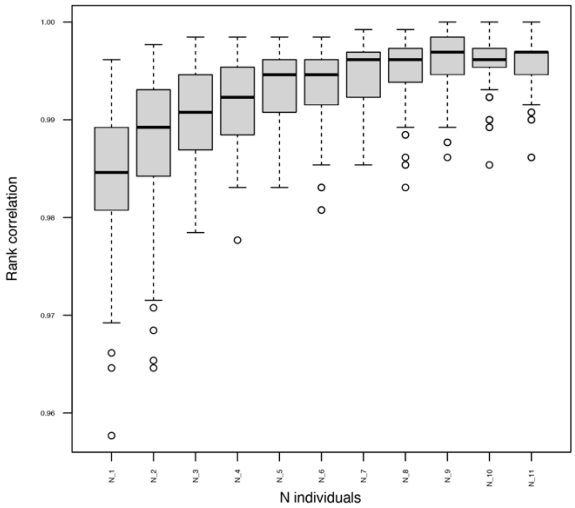


Figure S3. **A test to evaluate the effect of subsampling different number of individuals for estimating the genetic offsets across populations, using puplished data of *Zea mays* ssp. *mexicana* (Aguirre-Liguori et al. 2021).** Each boxplot shows the rank correlation between the genetic offset across populations subsampling 1 to 11 individuals per population and the genetic offsets using all individuals across all populations. Note that the range of y-axis represents correlations from 0.96 to 1.00. The subsampling were performed 1,000 times per N1 to N11. The results show that for all subsamplings, the rank correlations were >0.96, reflecting that sampling fewer individuals per population – including just one individual - does not strongly affect the order of the genetic offsets across populations.

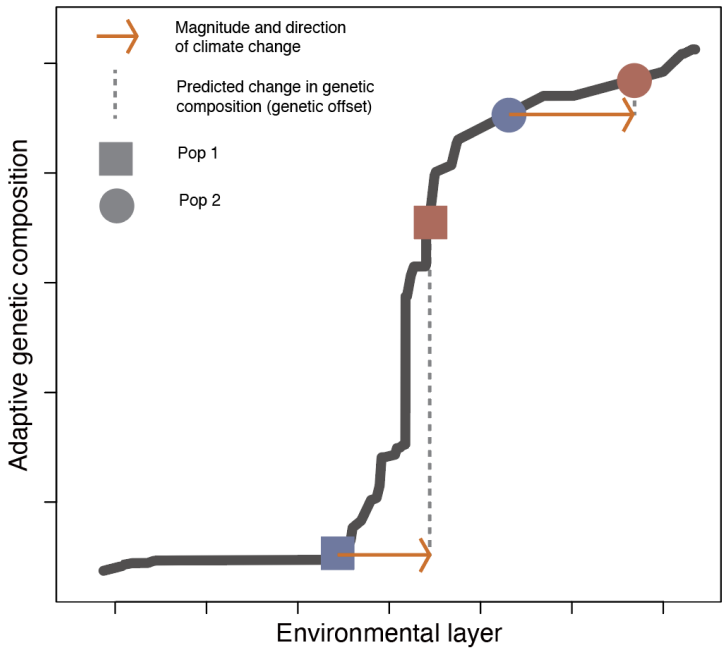


Figure S4. Description of the turnover function and how it can be used to identify the population that is pre-adapted to climate change. The circle and square show two populations along their turnover functions for present (blue) and future (red) climates. The arrows show how climate change will move in the future and the dotted line shows how the genetic composition is expected to change because of climate change (genetic offset). In this scenario, the circled population will have a low offset in the future, and it also grows currently toward the direction where climate will move in the future. Based on this, we consider that from the two populations, the circled one is more pre-adapted to climate change. The importance of this approach is that we infer that pre-adapted populations have the favored allele, which then is included in calculations of *Sa*.

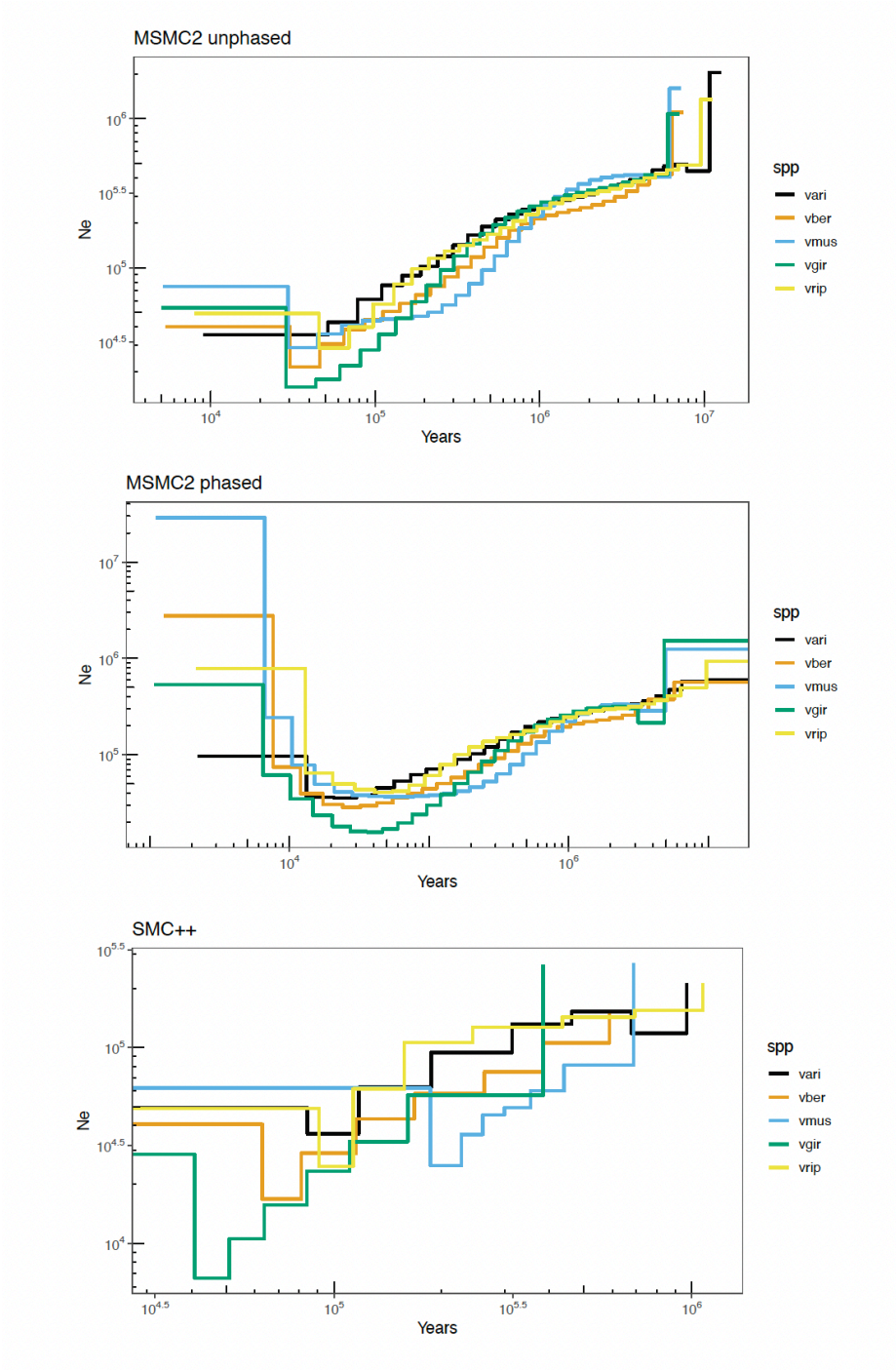


Figure S5. Outcomes from demographic inference. We inferred demographic history based on three different methods: MSMC2 with unphased data (top), MSMC2 with unphased data (middle) and SMC++ (bottom).

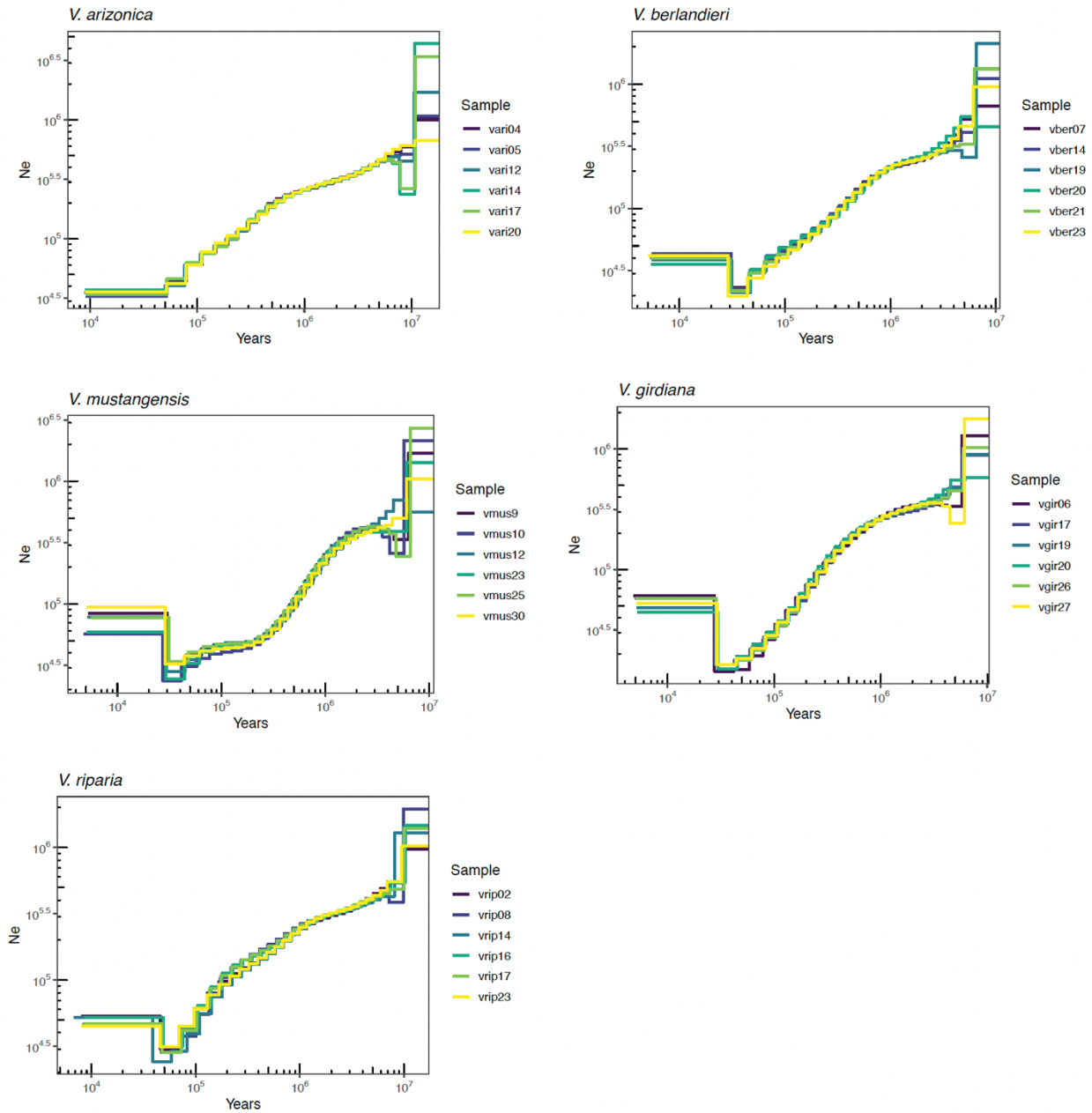


Figure S6: Demographic inference based on individuals in each of the five CWR species. Each graph represents a single species, as labelled, with separate individuals represented by different colors. All inferences were based on MSMC2 with unphased data.

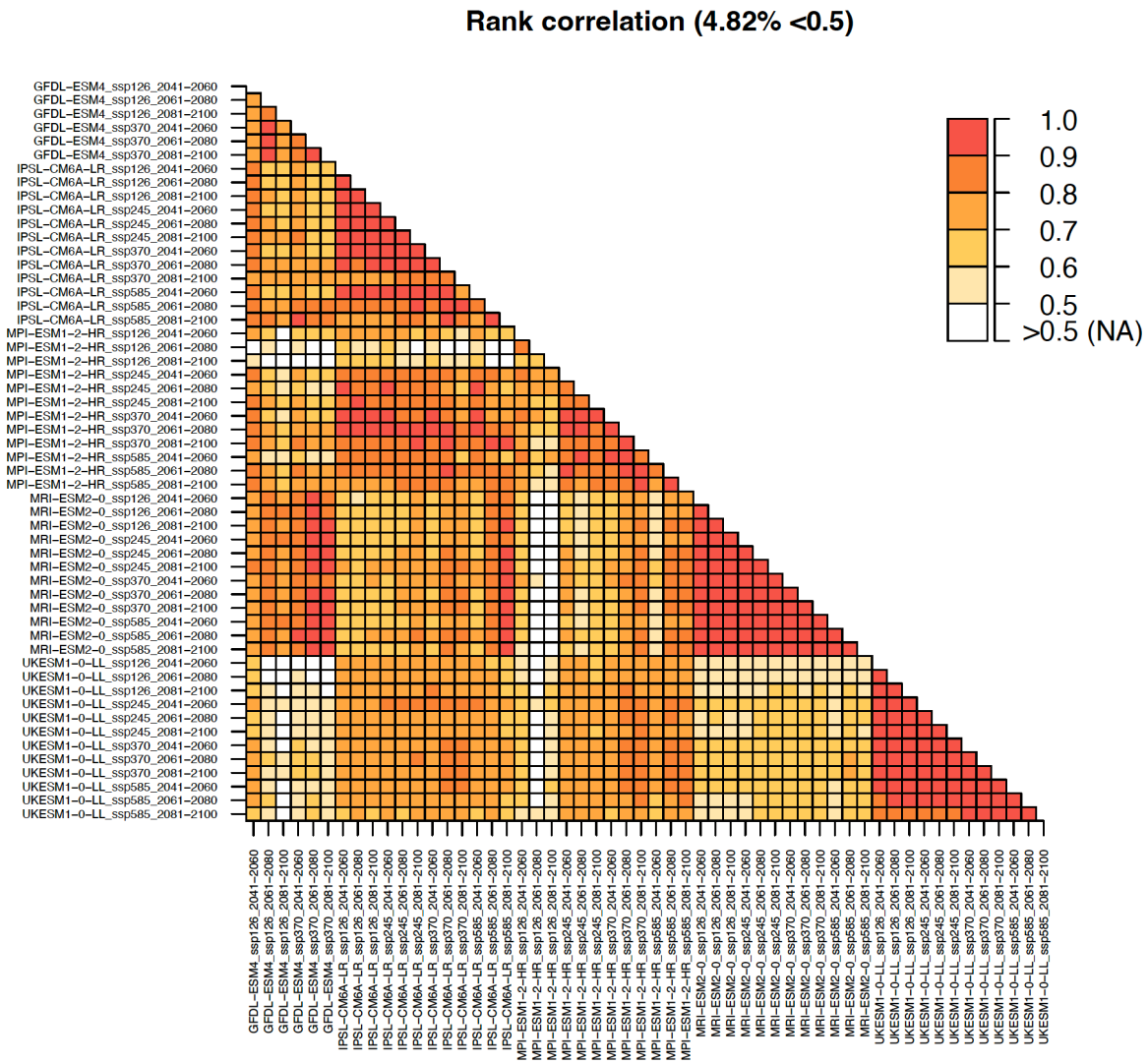


Figure S7. Rank correlations between the genetic offsets of populations using different FCC models. Higher correlations (warmer colors) indicate that the populations have similar offsets between climatic models. White colors indicate that the correlation is below 0.5. We found that only <5% of models were not correlated above 0.5.

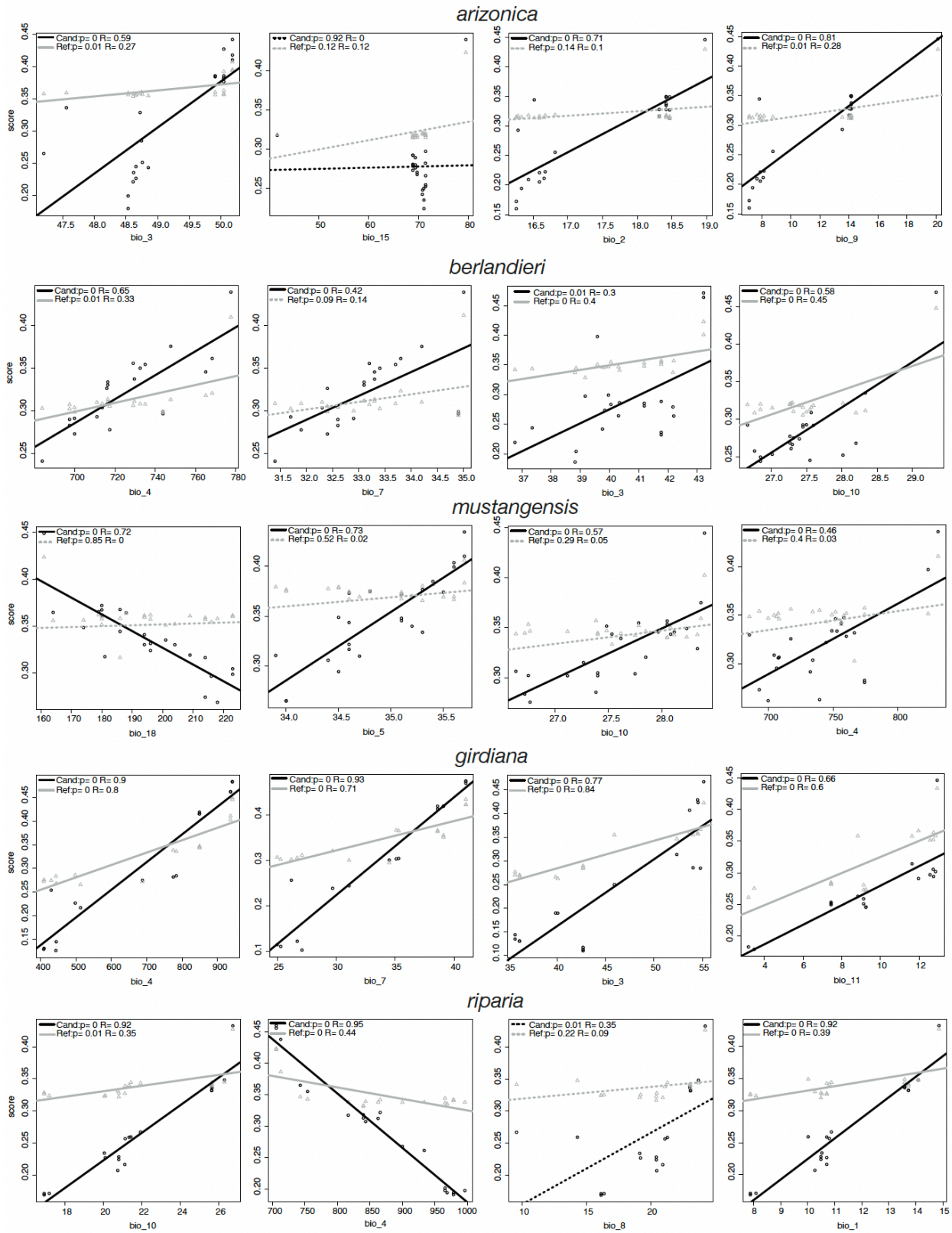


Figure S8. Correlations between adaptive scores and environmental variables. Scatter plots showing the correlation between the top 4 Bioclimatic variables that have a strong contribution

to the construction of the GF models (x-axis) and the adaptive scores (y-axis) using the candidate dataset (black, Sa) and the reference dataset (grey, Sref). Continuous lines indicate that the correlation is significant ($p<0.05$).

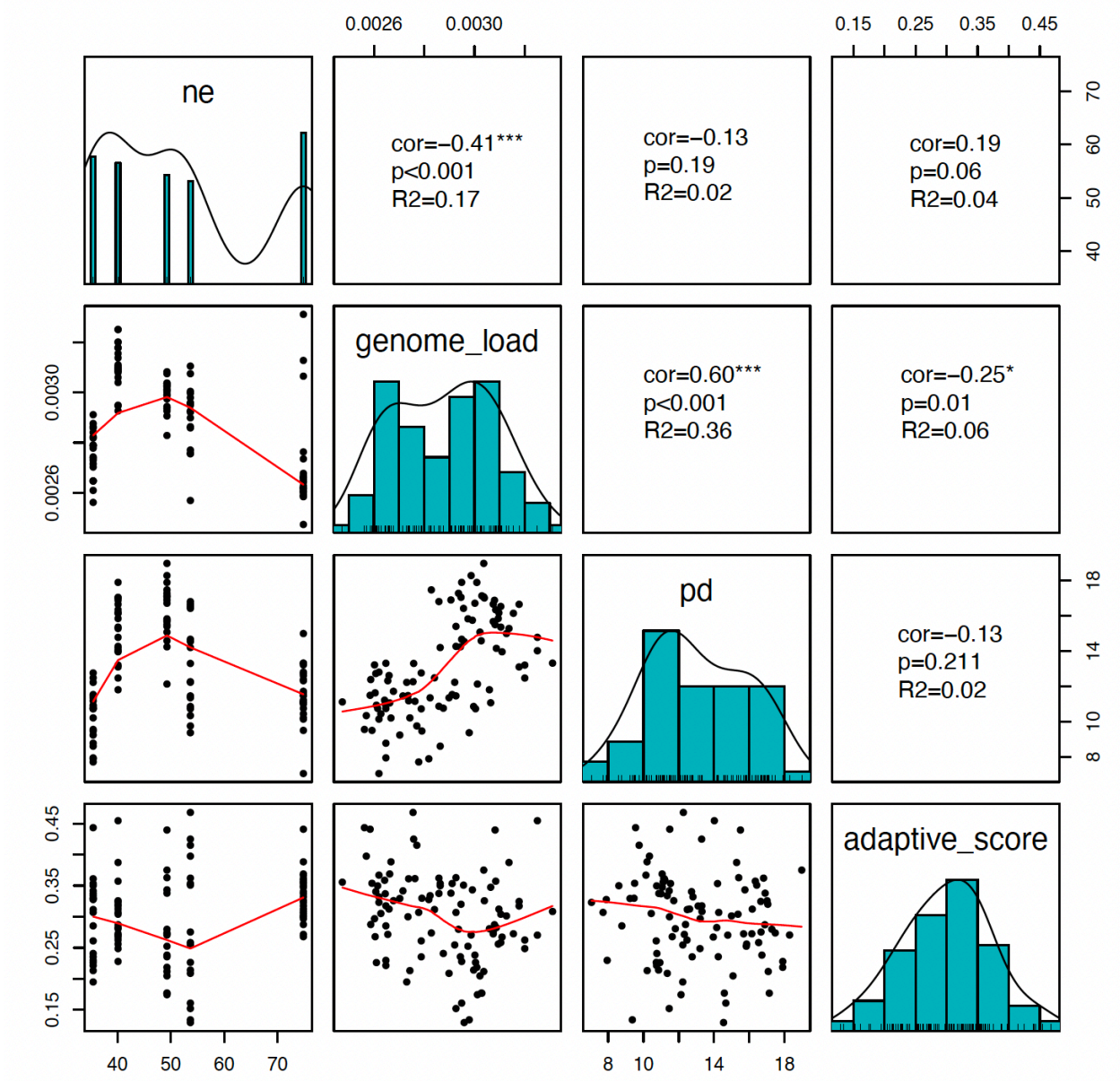


Figure S9. Correlations between variables. Scatterplot matrix showing the correlations between variables across all 105 samples in the study. The diagonal shows the distribution across the samples for each variable. The upper triangle shows the correlations (cor), significance (p) and fit (R^2) between variables. The lower triangle shows scatterplots for the two sets of variables in question.

Supplemental tables

Table S1. Table containing the geographic information, summary statistics and upsetR categories for all the analyses.

Table S2. Number of SNPs without missing data per species

Species	Nº of SNPs
Vitis arizonica	5,456,474
Vitis berlandieri	3,859,330
Vitis girdiana	3,598,951
Vitis mustangensis	4,971,805
Vitis riparia	5,630,334

Table S3. Species distribution models projected to 60 different climatic models. The number indicates the number of 2.5 minute pixels estimated from the raster model.

sp	ari	ari	ari	ari	ber	ber	ber	ber	gir	gir	gir	gir	mus	mus	mus	mus	rip	rip	rip	rip
Year	2050	2070	2090	pres	2050	2070	2090	pres	2050	2070	2090	pres	2050	2070	2090	pres	2050	2070	2090	pres
GFDL.ESM4 ssp126	54485	58518	67411	51805	56468	54963	57453	30029	19323	20126	21529	15095	70937	65132	64209	19755	180616	169985	166925	113839
GFDL.ESM4 ssp370	56969	61218	69913	51805	79395	97830	129101	30029	19436	16474	18944	15095	100011	146263	202641	19755	196864	221273	232327	113839
IPSL.CM6A.LR ssp126	56548	57252	58406	51805	80437	84458	83543	30029	18866	18772	18020	15095	105059	106821	102604	19755	199021	205215	192474	113839
IPSL.CM6A.LR ssp245	59221	65550	69563	51805	92228	115360	135250	30029	18419	19217	18514	15095	125830	168168	207271	19755	210284	207974	209618	113839
IPSL.CM6A.LR ssp370	60823	67893	83174	51805	98932	152333	213863	30029	19223	19396	18669	15095	133663	229631	316748	19755	208067	215631	216282	113839
IPSL.CM6A.LR ssp585	63613	75980	131522	51805	107811	181300	286122	30029	19891	19488	17528	15095	159352	282401	405209	19755	214600	211546	223038	113839
MPI.ESM1.2.HR ssp126	48393	46764	46390	51805	54360	53540	51109	30029	18970	20190	17727	15095	55562	53150	59322	19755	167537	161073	169915	113839
MPI.ESM1.2.HR ssp245	47003	48431	47108	51805	63428	66987	77823	30029	19792	18694	18406	15095	74735	77700	95324	19755	170836	186244	200740	113839
MPI.ESM1.2.HR ssp370	46872	50240	56193	51805	66609	97229	129313	30029	17764	19183	18767	15095	82908	142492	217062	19755	193220	209029	213107	113839
MPI.ESM1.2.HR ssp585	49419	51224	62836	51805	73381	105149	168470	30029	18384	18209	17629	15095	91240	186752	276891	19755	190428	216060	216682	113839
MRI.ESM2.0 ssp126	63244	62216	56194	51805	72505	71338	65541	30029	15504	15509	15944	15095	90187	85932	75711	19755	199367	200138	198048	113839
MRI.ESM2.0 ssp245	61273	67748	65164	51805	80997	88103	85485	30029	15084	15566	15097	15095	105616	119560	119361	19755	200003	190123	204767	113839
MRI.ESM2.0 ssp370	62961	73352	85255	51805	86352	102518	123293	30029	15445	15069	13988	15095	117808	145807	180052	19755	200162	209022	209300	113839
MRI.ESM2.0 ssp585	69086	80980	98989	51805	95279	117911	157513	30029	15147	13738	15019	15095	133227	170705	238317	19755	205155	208343	211722	113839
UKESM1.0.LL ssp126	78063	80285	81850	51805	106447	117678	122283	30029	17442	17674	18445	15095	172072	180072	184207	19755	207764	203306	204821	113839
UKESM1.0.LL ssp245	81863	91580	99449	51805	128707	155980	180361	30029	17322	19197	17761	15095	199045	246452	297574	19755	213758	207142	199803	113839
UKESM1.0.LL ssp370	85063	99814	136035	51805	147855	199399	294411	30029	15475	15532	15486	15095	233745	335846	438705	19755	213096	207323	199326	113839
UKESM1.0.LL ssp585	88831	112125	172534	51805	161286	238264	374755	30029	17127	15319	12543	15095	260358	390313	478728	19755	209513	200856	196519	113839
means	62985	69509	82665	51805	91804	116685	151982	30029	17700	17630	17223	15095	128408	174066	219996	19755	198905	201682	203634	113839

Table S4. Example showing the table used to create the upsetR analysis. The table shows the summary statitics, and then the future categories for each layer and for each population, based on on the 0.5 threhold. Here we are showing an example using the IPSL.CM6A.LR_ssp585_2061.2080 model The names of the columns indicate the threshold and the layer. In the excel table the genetic offsets and categories are found for all the climatic models and for all the thresholds (names in the columns start with thresh_XX to indicate the threshold).

	spp	lon	Lat	Adaptive score	Genoma load	pd	ne	Mxt IPSL.CM6A.LR ssp585 2061.2080	Thresh 0.25 adaptive_score	thresh_0.25 genome_load	thresh_pd	thresh_0.25 Ne	thresh_0.25 offset PSL.CM6A.LR ssp585_2061.2080
vrip15	ari	-111.77	34.52	0.31	0	12.49	35.45	1	0	0	1	0	1
vari05	ari	-108.2	32.93	0.21	0	10.21	35.45	1	0	0	1	0	1
vber02	ber	-99.77	30.49	0.28	0	14.16	40.11	0	0	0	0	0	0
vber14	ber	-98.35	30.44	0.3	0	15.36	40.11	1	0	0	0	0	0
vmus15	mus	-99.14	29.35	0.35	0	10.75	74.84	0	1	1	1	1	0
vmus23	mus	-97.94	30.01	0.29	0	12.39	74.84	1	0	1	1	1	1
vgir25	gir	-116.32	36.42	0.35	0	14.2	53.67	1	1	0	0	1	0
vgir10	gir	-114.43	36.39	0.47	0	12.26	53.67	1	1	0	1	1	0
vrip06	rip	-105.33	35.27	0.24	0	14.24	49.28	0	0	0	0	0	0
vrip02	rip	-104.28	37.11	0.27	0	15.68	49.28	0	0	0	0	0	0

Table S5. Migration load projected from all accessions defined by the FOLDS test across 54 FCC models to all the vinifera locations in the US (lon and lat). The names of the columns indicate the threshold, the FCC model and the accession used. The column with @cat@ always shows the category 1 and 0, which indicates that the accession could be used or not in the future since they have a migration load below or above the species offset, respectively. This table is in the excel file.

Table S6. Example showing the migration load projected from a accesion to all the vinifera locations in the US (lon and lat). . The table shows the migration load of the accesion using the IPSL.CM6A.LR_ssp585_2061.2080 model . The names of the columns indicate the threshold, the FCC model and the accession used. The second column always shows the category 1 and 0, which indicates that the accession could be used or not in the future since they have a migration laod below or above the species offset, respectively.

lon	lat	tresh_0.25.vmus16.IPSL.CM6A.LR_ssp585_2061.2080	tresh_0.25.vmus16.cat.IPSL.CM6A.LR_ssp585_2061.2080	tresh_0.5.vmus16.IPSL.CM6A.LR_ssp585_2061.2080	tresh_0.5.vmus16.cat.IPSL.CM6A.LR_ssp585_2061.2080	tresh_0.9.vmus16.IPSL.CM6A.LR_ssp585_2061.2080	tresh_0.9.vmus16.cat.IPSL.CM6A.LR_ssp585_2061.2080
-79.64	43.61	0.06	1	0.06	1	0.06	1
-73.73	40.78	0.06	1	0.06	1	0.06	1
-77.85	40.81	0.06	1	0.06	1	0.06	1
-77.64	43.07	0.05	1	0.05	1	0.05	1
-75.9	41.26	0.06	1	0.06	1	0.06	1
-119.55	34.41	0.12	0	0.12	0	0.12	0
-76.15	43.19	0.08	0	0.08	0	0.08	0
-117.17	33.63	0.1	0	0.1	0	0.1	0
-121.99	37.32	0.09	0	0.09	0	0.09	0
-119	37	0.1	0	0.1	0	0.1	0

Table S7. Number of times that an accession is identified by the FOLDS test for 54 FCC models and 3 different thresholds. This table is in the excel file.