Climate Change, Crop Wild Relatives and Genomic Data: The persistence and agronomic utility of five *Vitis* species

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

* These authors contributed equally to this work

Department of Ecology and Evolutionary Biology

University of California, Irvine

[1] 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 perennial crops, because perennials are often clonally propagated and consequently do not evolve rapidly. By studying the landscape genomics of 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 assessing the relative potential of different CWRs to persist in the context 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 of *V. mustangensis* are particularly well-suited to persist. The second goal was identifying candidate CWRs to contribute to bioclimatic adaptation for grapevine (*V. vinifera*) cultivation. To do so, we first estimated that ~40% of current viticulture sites in the United States will be vulnerable to climate change, based on species distribution models projected to 2070. We then predicted which CWRs have the genomic profile to contribute to bioclimatic adaptation at these vulnerable sites. We identified rootstock candidates from *V. mustangensis*, *V. riparia* and *V. girdiana* and hypothesized that they may prove useful for mitigating climate impacts on viticulture. By identifying candidate germplasm, this work takes a conceptual step toward mitigating the climate impacts on crops by utilizing the genomic and bioclimatic characteristics of CWRs.

INTRODUCTION

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

Given their importance, numerous studies have focused on the geographic distributions, *ex situ* germplasm collections and *in situ* conservation of CWRs (e.g., (Castañeda-Álvarez et al., 2016)). Surprisingly fewer studies have assessed the potential effects of climate change on CWRs (Jarvis, Lane, & Hijmans, 2008), but recent notable exceptions include studies of CWRs in Europe (Aguirre-Gutiérrez, van Treuren, Hoekstra, & van Hintum, 2017) and the United States (US) (Khoury et al., 2020). These and similar studies typically employ species distribution models (SDMs) to infer the climatic niche of CWRs based on the locations where the species is (or is not) found. This niche is then projected into the future, based on the predicted climate, usually by ignoring biotic interactions (Lawler, White, Neilson, & Blaustein, 2006), by assuming no additional phenotypic plasticity, and by ignoring evolution. The outcome of SDMs is an estimate of the predicted geographic niche under climate change, which helps inform the fate of species. As a concrete example, a recent US-based study modeled the distribution of 600 CWRs and found that ~85% are either vulnerable or endangered, due largely to projected losses of geographic area associated with their bioclimatic niche (Khoury et al., 2020).

Although the number of climate-based studies of CWRs is growing, even fewer studies have assessed the fate of CWRs by integrating landscape genomic data with climate predictions (Jonas A. Aguirre-Liguori, Ramírez-Barahona, & Gaut, 2021; Jonás A.

Aguirre-Liguori, Ramírez-Barahona, Tiffin, & Eguiarte, 2019). The argument for this integration is that SDMs ignore the potential for evolutionary adaptation and may therefore yield biased predictions (Exposito-Alonso, Vasseur, et al., 2018; Hällfors et al., 2016; Ikeda et al., 2017; Razgour et al., 2018). To address this bias, genetic information has been used to study local adaptation to bioclimatic variables and then to model how adaptive genetic variation changes in the context of climate predictions. In this way, one can identify species or populations that appear genetically poised to adapt to climatic change and, conversely, populations that may require especially dramatic genetic changes and are thus genetically endangered.

The integration of genomic data with climate predictions is relatively new (Capblancq, Fitzpatrick, Bay, Exposito-Alonso, & Keller, 2020; Fitzpatrick & Keller, 2015), and hence it has been applied to only a handful taxa (see (Capblancq et al., 2020)). Work on poplars, pearl millet and *Arabidopsis* has been especially notable, because it has shown that predicted shifts in allelic variants predict components of fitness (Exposito-Alonso, Burbano, Bossdorf, Nielsen, & Weigel, 2019; Fitzpatrick, Chhatre, Soolanayakanahally, & Keller, 2021; Rhoné et al., 2020), thus arguing for the relevance of the approach. It is worth noting, however, that genomic data contains information about additional processes that may help inform about the evolutionary fate of species (Jonas A. Aguirre-Liguori et al., 2021; Waldvogel et al., 2019). Such information includes the history of genetic migration among populations, insights into the potential effects of genetic drift and the impact of genetic load, which correlates inversely with fitness (Frankham, 2005).

Here we employ landscape genomic data to assess potential responses to climate change for five *Vitis* CWRs. We focus on *Vitis* because its CWRs are agronomically crucial; ~80% of viticulture worldwide utilizes rootstocks from North American *Vitis* species (Migicovsky et al., 2021). Among ~25 North American *Vitis* species (Wan et al., 2013), those from the American Southwest are of particular interest, because some are resistant to *Xylella fastidiosa*, the causative agent of Pierce's disease (PD), some harbor resistance to diseases like powdery and downy mildew, and others grow under abiotic stresses like extreme temperature that may 'preadapt' them to some aspects of climate change

(Heinitz, Uretsky, Peterson, Acosta, & Walker, 2019). We also focus on *Vitis* because climate change is predicted to disrupt the phenology and production of grapevine (*V. vinifera*) (Morales-castilla et al., 2020), representing substantial climate-related impacts to a ~$5 billion per annum industry (USDA, 2013). Finally, it is worth noting that *Vitis* rootstocks are currently dominated by only a few clones, underscoring a pressing need to identify additional germplasm for rootstocks and breeding (Heinitz et al., 2019; Summaira Riaz et al., 2019).

We first use landscape genomic data to help evaluate the potential for persistence of *Vitis* CWRs under climate change. We then turn to a second question: Which, if any, of the wild *Vitis* species have ecological tolerances that may aid viticulture in the future, particularly in regions where viticulture is endangered by predicted shifts in climate? Genomic data has also been used to address this question, too, as perhaps best illustrated by a landmark study of pearl millet (Rhoné et al., 2020). In the pearl millet study, Rhoné et al generated genomic data from landraces across the African continent. They then focused on adaptive variation to assess the genetic potential of a landrace to adapt to climate shifts where it is grown locally – i.e., its capacity to adapt *in situ*. They identified several existing landraces that may fail to adapt, leading to potential local collapses of agriculture. In a next step, they asked whether existing landraces from other regions in Africa had the genetic capacity to adapt to climate change if they were moved to regions of expected collapse. They identified several potential landrace candidates for human-assisted migration. In other words, Rhoné et al. (2020) predicted matches among predicted climates, geographic locations and the adaptive genetic capacity of specific landraces, in the hope of avoiding agricultural collapse (Rhoné et al., 2020).

Building on this example, we focus on grapevines and the potential of specific CWRs to help mediate the effects of climate on viticulture. To do so, we first use SDMs to identify locations in the United States where viticulture may be at risk under climate change. We then assess the genetic potential of the five *Vitis* CWRs to contribute to bioclimatic adaptation in the regions of risk, based on the CWRs' bioclimatic profiles and their projected evolutionary persistence. In short, we use landscape genomic data to

identify CWR germplasm that that may prove helpful for mitigating climate risks to grapevine cultivation. The candidate germplasm includes individuals with low predicted vulnerability to climate change based on projected patterns of adaptation to bioclimatic variables and also with genomic qualities like low genomic load and high historical effective population sizes. By identifying candidate germplasm, this work takes a conceptual step toward mitigating the climate impacts on viticulture. More importantly, this work represents an effort to integrate genomic information from CWRs with climate models to identify species and accessions that may prove especially useful for *in situ* adaptation of crops.

RESULTS

Species Sampling and Data:

In this study we used previously published resequencing data consisting of 105 individuals from five NA *Vitis* species, including *V. arizonica*, *V. mustangensis* (synonym *V. candicans*), *V. berlandieri*, *V. girdiana* and *V. riparia* (Morales-Cruz et al., 2021). These species are either currently used as rootstocks, represent the parents of commonly used hybrid rootstocks, or have phenotypic properties that make them potentially valuable to viticulture. For example, rootstocks with *V. riparia* parentage are associated with early berry ripening and are sometimes chosen to strategically restrict scion growth in highly fertile soils (Heinitz et al., 2019). Similarly, *V. berlandieri*, which is also known as *V. cinerea* var. *helleri* (Moore, 1991), is notable for growth in limestone soils and hot, dry environments (Heinitz et al., 2019). It is worth noting that there is a history of hybridization among species, but they have not diverged especially recently. The two most closely related species in this study (*girdiana* and *arizonica*) split at least 10 million years ago (ya) (Wan et al., 2013) and perhaps even earlier (Morales-Cruz et al., 2021).

The 105 *Vitis* individuals were sampled from throughout the Southwestern United States over a period of ~30 years and are maintained in a living collection at UC Davis. For each species, 18 to 24 accessions were resequenced (**Table 1**). For our analyses, we employed SNPs determined previously (Morales-Cruz et al., 2021), but we focused on a subset of SNPs that had no missing data within each species. The resulting number of

SNPs ranged from 3.6 million in *V. girdiana* to 5.6 million in *V. riparia.* The accessions, their sampling locations, and their SNPs provided the basis for evaluating species' persistence under projected climate change.

Evaluating the Persistence of Wild *Vitis* Under Predicted Climate Change

To assess the potential effects of climate change, we employed the FOLDS model, a conceptual framework that considers distinct layers of information to evaluate how populations will respond to climate change (Jonas A. Aguirre-Liguori et al., 2021). The FOLDS model was applied based on six layers of genomic and climate information. Below we briefly describe each layer before combining the data to assess relative potentials for persistence.

Species distribution models: We estimated SDMs to examine expected changes in the geographic distribution of the five wild *Vitis* species over time. To build the SDMs, we downloaded all known locations of the five species from the Global Biodiversity Information Facility (GBIF; www.gbif.org) and also 19 bioclimatic variables from Worldclim 2 (Fick & Hijmans, 2017) for the present-day (representing data from 1970 to 2010) and for 12 climate forecasts. These 12 forecasts include four Shared Socio-economic Pathways (SSPs) for three years (2050, 2070 and 2090). For each species, we estimated SDMs by combining results across the four SSPs in each year (see Methods). This procedure yielded an SDM for each species in four time periods: present-day, 2050, 2070 and 2090.

For each SDM, we measured the projected habitable area of each species (**Figure 1A**) and its trend over time. The amount of habitable area was predicted to increase over time for four species (*arizonica, berlandieri, mustangensis* and *riparia*) but reduce marginally for *girdiana* (**Figure 1B**). Although the expansion of niche area sounds promising, we also found that 39% of the 105 individuals were sampled from locations that were predicted to fall outside the species' niche by 2050 (**Figure 1C**). This proportion increased to 62% by 2070 and to 76% by 2090. There were exceptions to this general trend – e.g., SDMs predicted that *berlandieri* is likely to persist in its current locations until

2090 (**Figure 1C**). Overall, however, these models suggest that most of the *Vitis* sampled in our study will need to either adapt or migrate to survive.

Local Offsets and Adaptation: Recently, machine learning approximations have been used to model adaptive genetic diversity across a geographic landscape (Capblancq et al., 2020; Exposito-Alonso, Brennan, Alonso-Blanco, & Picó, 2018; Exposito-Alonso et al., 2019; Fitzpatrick & Keller, 2015). We used a machine learning-based approach called Gradient Forest (GF) (Ellis, Smith, & Roland Pitcher, 2012) to compare SNP allele frequencies to bioclimatic data (Fitzpatrick & Keller, 2015). GF identifies environmental gradients that associate with genetic variation and also estimates the turnover in allele frequencies along the gradient (Capblancq et al., 2020; Fitzpatrick & Keller, 2015; Waldvogel et al., 2019). When applied to bioclimatic data from both the present and the future, GF also estimates the local genetic offset (or 'local offset'), an estimate of the amount of genetic change necessary to adapt in the future at the same locality (Capblancq et al., 2020; Fitzpatrick & Keller, 2015). The local offset depends on the genetic composition of populations and the magnitude of the environmental change that is expected to occur; populations with higher offsets are expected to be more vulnerable to climate change (Fitzpatrick & Keller, 2015).

We used GF to estimate the offsets of *Vitis* individuals. Following common practice (Capblancq et al., 2020; Fitzpatrick, Keller, & Lotterhos, 2018), we ran GF with a subset of outlier SNPs that were associated with environmental variables. To identify these outlier SNPs, we used Baypass (Gautier, 2015), identifying SNPs correlated to the first four principal components (PCs) of 19 present-day bioclimatic variables. [The first four PCs captured from 92.9% to 99.9% of bioclimatic variation across the five species.] Baypass detected from 5,650 to 13,531 outlier SNPs within each species (**Table 1**). We then applied GF to the set of outlier SNPs, yielding a subset of 2,730 to 8,720 SNPs within each species that had significant non-linear associations across samples (**Table 1**), based on present-day bioclimatic data.

GF analyses provided three additional outcomes. The first was a set of bioclimatic variables that were important for building the model, as measured by the contribution of

the variable to the construction of the model ($R^2 > 0$). For simplicity, we focused on the four most important variables for each species, some of which (**Table 1**) were shared across species. For example, BIO10 (the mean temperature of the warmest quarter) had a strong contribution to the building of GF models in *berlandieri, mustangensis* and *riparia,* whereas BIO4 (temperature seasonality) was a top contributor for all species but *arizonica* (**Table 1**). The second outcome was local genetic offsets, which we computed for each individual in each species based on contrasts between present-day climate and 2070 projections (see Methods). Note that offsets cannot be compared quantitatively across species, because they were estimated with different SNPs and bioclimatic information. Local offsets do, however, provide insights into the individuals within a species that may require more adaptive change to survive; for example, in *arizonica*, two individuals had especially low offsets (**Figure 2A**), suggesting they will require relatively little genetic change to adapt to future climates. In contrast, *riparia* and *girdiana* exhibited especially notable variation in local offsets among individuals, suggesting that there is variation in the capacity to adapt (**Figure 2A**).

Finally, GF provided insights into the allelic state of adaptive SNPs. We used GF to identify allelic states that will be favorable in the future based on turnover functions (**Figure S1**) and then calculated an adaptive score (S_a), which is the proportion of future adaptive alleles across all candidate SNPs. This measure overlaps with, but differs from, local offsets, because a high offset can be caused either by strong predicted environmental shifts or by genetic shifts. In contrast, S_a focuses solely on genetic composition by counting whether an inferred adaptive allele is present at each position. Like offsets, S_a varied within species (**Figure 2B**, **Table 1**), but unlike offsets it can be compared across species (as a proportion). Comparing S_a across species suggests that our *girdiana* and *riparia* samples have adaptive allelic states less often than the other species (**Figure 2B**).

Because S_a has not been applied previously, we performed tests to evaluate whether S_a was biased by geographic, demographic or climatic variables. For example, as expected, S_a was strongly and significantly correlated with at least one of the four most

important bioclimatic variables identified by GF (**Figure 2C**); in fact, it was correlated with all four for three of 5 species (**Figure S2**). S_a was more highly correlated to bioclimatic variables than a similar score (S_{ref}) calculated from a set of reference SNPs (**Figure S2;** see Methods). The notable exception was for *girdiana* samples, where S_{ref} was strongly correlated with all four of the top 4 associated bioclimatic variables (**Figure S2**), likely reflecting genetic structure (Morales-Cruz et al., 2021) or other features that complicate interpretation of S_a in this species.

Effective population size and genomic load: The efficacy of selection depends on adaptive diversity but also on population factors, like genetic drift and genomic load, that are typically ignored in climate studies (Jonas A. Aguirre-Liguori et al., 2021; Brady et al., 2019; Nadeau & Urban, 2019). To assess drift, we estimated historical effective population sizes (N_e) over time, using MSMC2 (Schiffels & Wang, 2020). For all species we found that N_e decreased from the distant past until ~20,000 years ago, followed by recent moderate recoveries (**Figure 2D**). From these analyses, we retrieved the current estimate of N_e, which ranged from ~35,000 individuals in *arizonica* to ~75,000 in *mustangensis* (**Table 1**); *mustangensis* also yielded the highest N_e estimate based on another method, SMC++ (Terhorst, Kamm, & Song, 2017) (**Figure S3**). The relatively high N_e for *mustangensis* was surprising, because it currently has a narrow geographic distribution (**Figure 1A**), but it suggests that *mustangensis* may be less prone to drift effects than other species.

Genetic load may also predict fitness (Frankham, 2005). We estimated genetic load within and between species by measuring the number of potentially deleterious alleles per accession (see Methods). To account for differences in the number of SNPs between species and potential reference biases, we focused only on SNPs shared between species, with no missing data. Overall, our measure of load indicates that *mustangensis*, *arizonica* and *riparia* individuals have relatively low load among species, but there was also notable variation among individuals within species (**Figure 2E**).

Resistance to Pierce's Disease: Biotic interactions between species can play a key role in shaping population's responses to climate (Bascompte, García, Ortega, Rezende, & Pironon, 2019; Zamora-Gutiérrez, Rivera-Villanueva, Martínez Balvanera, Castro-Castro, &

Aguirre-Gutiérrez, 2021). Here we included one biotic interaction, based on information about resistance to the bacterium *X. fastidiosa* (Morales-Cruz et al., 2021; Summaira Riaz, Tenscher, Heinitz, Huerta-Acosta, & Andrew Walker, 2020). Two previous studies categorized susceptibility to PD across all 105 accessions by evaluating the concentration of colonizing forming units (CFU) of *X. fastidiosa* after a period of experimental exposure to the bacterium. Although CFU is a qualitative measure, concentrations < 13 CFU are consistent with an individual being PD resistant (S. Riaz, Huerta-Acosta, Tenscher, & Walker, 2018). We plotted CFU counts for each individual, which revealed variation in PD resistance among individuals and species (**Figure 2F**) – e.g., 100% of *arizonica* and 87% of *mustangensis* accessions were PD resistant while the *girdiana* distribution was notably bimodal, with ~50% of accessions resistant. In contrast, only one *riparia* and two *berlandieri* accessions were resistant.

The FOLDS framework: Ideally, assessing the response of populations to climate change should be based on multiple layers of information (Razgour et al., 2018; Waldvogel et al., 2019). We relied on six layers - i.e., local offsets, N_e, S_a, SDMs, load and PD resistance – and surmised that an accession is relatively well situated to adapt to climate change if the accession has a low offset, low genomic load, high S_a, high N_e (low drift), PD resistance and SDM-based persistence by 2070.

We applied the FOLDS (Jonas A. Aguirre-Liguori et al., 2021) approach considering two factors: the number of layers and the threshold on each layer. For example, we first filtered accessions based on all six layers, with thresholds defined by: PD resistance (CFU < 13), SDM persistence until 2070, among the lowest 25% of local offset and load within a species, and among the highest 25% of Ne and Sa within a species. After applying those thresholds, only two accessions remained, both from *V. mustangensis* (**Figure 3A&B**). Based on this information, we predict that these two accessions are relatively well situated to evolve and persist to predicted shifts in climate. We also relaxed the analysis so that only five layers were required to pass thresholds, resulting in eight accessions (all from *V. mustangensis*) predicted to remain by 2070 (**Figure 3A&B**). Finally, we also relaxed the thresholds, so that (for example) accessions remained when they retained

among the lowest 50% (or 90%) of offset and load within a species, and among the highest 50% (or 90%) of N_e and S_a within a species (**Figure 3C&D**). As expected, weaker thresholds led to more accessions and species passing the filters, but the exercise provided potential insight into relative risks among taxa. For example, none of the *girdiana* or *riparia* accessions remained after applying 50% thresholds, suggesting these species may be particularly susceptible to population extinction within our sampled areas.

This approach treats each layer as if it is independent and equally weighted. While there is no obvious way to correct for a lack of independence, it is worthwhile to better understand how the layers relate to one another. We therefore calculated pairwise correlations between layers across all 105 individuals (**Figure S4**). [The pairwise correlations did not include local offsets, because it is not appropriate to compare them across species.] We calculated, for example, negative correlations between genetic load and *Ne* (-0.41, $p < 0.001$); this relationship was reassuring because load is expected to increase with genetic drift (Frankham, 2005). Load was also negatively correlated with S_a (r=-0.26; $p<0.01$), suggesting that accessions with high load tend to have less adaptative potential, and with PD susceptibility (r=0.60, $p < 0.001$) (**Figure S4**). We also assessed concordance between SDMs and genomic measures. To do so, we coded each individual as persistent (or not) within projected SDMs for the year 2070, based on their sampling location. We then used point biserial correlations to test whether persistence correlated with quantitative measures. Persistence was not significantly correlated with *Ne* (p=0.08, r=-0.17), PD levels (p=0.19, r=0.13) or S_a (p=0.92, r=-0.1), but it did correlate with load (r=0.24; p=0.013). These results generally suggest that genetic data provide non-redundant information to SDMs.

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

<u>*Identifying regions of risk to viticulture:*</u> Given insights into the potential persistence of wild species, our next task was to evaluate their potential to aid viticulture in the face of climate change. We began by using SDMs to predict geographic regions where viticulture will be endangered. To do so, we downloaded all 1278 registered

locations for viticulture within the United States, according to the GBIF database. We also obtained the 19 bioclimatic variables for the present and for 2070 (SSP 370). We then used the data to estimate the ecological niche for *V. vinifera* in the present (**Figure 4A**). This SDM performed well, but not perfectly, in that most (84.7%) but not all GBIF locations fell within the predicted niche for *V. vinifera*. Imperfect prediction was not surprising, however, given that our SDM does not consider irrigation, cultivar diversity and other factors that influence the location and productivity of viticulture. The model nonetheless had interesting properties. One was that a large region of the Southern U.S. was predicted to be suitable bioclimatically for grapevine cultivation, but grapevines are generally not grown in that region due in part to the prevalence of *X. fastidiosa* (Heinitz et al., 2019). This observation conveys the importance of considering biotic interactors, where possible, and especially *X. fastidiosa*.

We then projected the SDM to 2070, yielding contrasts between present-day and future niches (**Figure 4B**). In this model, 44.7% of 1278 current locations no longer fell within the predicted bioclimatic niche; in other words, an additional ~25% (or 321) of the 1278 locations in GBIF were no longer predicted to support viticulture by 2070. The 2070 SDM was particularly striking in the Northeastern U.S., because most of the states of Pennsylvania (PA) and New York (NY) were no longer climatically suitable. Additionally, the southernmost region of the Central Valley of California (CA), where table grapes and raisins are traditional products, fell out of the model. In contrast to PA, NY and parts of CA, suitable area for viticulture expanded in the state of Washington and regions of Canada.

Investigating the potential of wild relatives to mitigate crop risk: If grapevine cultivation is to persist in at-risk regions like NY, one potential solution is to employ CWRs that expand bioclimatic capabilities for breeding and rootstocks. But what CWRs might be best suited for *in situ* adaptation of the crop? To evaluate this question, we relied on what we called the 'migration offset' (but which was first called the 'forward offset'; (Gougherty, Keller, & Fitzpatrick, 2021)), which we defined as the genetic offset for an individual calculated between its present-day sampling site and other locations in 2070. In

this case, the 'other locations' were the 517 locations where SDMs predicted that viticulture is at risk (**Figure 4B**). We calculated the migration offset for each wild accessions at each of the 517 locations. Given 517 estimates of the migration offset for each CWR, we then asked: which, if any, fell within the range of local genomic offsets for the wild species? If the migration offset fell within the range of local offsets, we concluded that the accession is a potential bioclimatic fit for *in situ* crop adaptation under 2070 climate projections.

The approach and results are perhaps most easily represented by examples. We focused first on *mustangensis;* recall that *mustangensis* accessions were expected to be relatively resilient under climate change based on both SDM and landscape genomic data (**Figure 3B**). We calculated the migration offset for each *mustangensis* accession across all 517 risk sites (**Figure 4C**). For each accession, migration offsets were typically higher than the current species' range of local offsets, suggesting that human-assisted CWR migration to most of the locations would exceed the CWR's genetic capacity for adaptation. However, some accessions were predicted to have low offsets at some of the 517 sites. Consider, for example, accession #25 from *mustangensis* (vmus25) (**Figure 4D**). For 2070, it was predicted to have low migration offsets (i.e., within the range of local offsets) in 306 of the 517 (or 60%) of at-risk sites. Remarkably, another *mustangensis* accession (vmus20) had low migration offsets at 80 of the 517 sites, none of which overlapped with the vmus25 sites. Together, these results suggest that these two accessions may be able to contribute to *in situ* bioclimatic adaptation of grapevine for ~75% of the predicted at-risk sites. Importantly, these two accessions also have the desirable feature of PD resistance.

While vmus25 and vmus20 were exceptional, we extended the analysis to all 105 accessions in our dataset. Altogether, we found accessions that were predicted to be bioclimatically suitable for 495 of the 517 (95.7%) of the at-risk sites for viticulture. These results were dominated by *mustangensis*, which fitted 403 at-risk sites across all *n*=15 accessions. V. *mustangensis* was followed by *girdiana* (309 sites), *riparia* (182 sites) and *berlandieri* (5 sites) (**Figure S5**). Surprisingly, even though *V. arizonica* is an important source of PD resistance (**Figure 2**; (Krivanek, Riaz, & Walker, 2006)), we found no sites

where it was predicted to fit bioclimatically to the *in situ* adaptation of *vinifera* for the 2070 climate.

DISCUSSION

Climate change is a threat to all biodiversity (Parmesan, 2006; Parmesan & Yohe, 2003), but the fate of CWRs is particularly important. We have focused on five *Vitis* CWRs, with two goals in mind. The first has been to incorporate genomic data to help assess the relative potential of CWRs to persist under climate change. The second has been to identify locations of grapevine cultivation in the U.S. that are vulnerable to climate change and then to assess which CWRs, if any, have bioclimatic and genomic profiles that may be agronomically useful for *in situ* crop adaptation. We accentuate that the approaches used in this study are not limited to *Vitis*. Indeed, we offer this work as a framework for further methodological extension and for extension to additional CWRs and crop systems.

Evaluating CWR Persistence: insights and caveats

Predicting how populations and species will respond to climate change is not a simple task because multiple factors --- including evolutionary, human, and ecological processes –impact the response of populations to climate change (Waldvogel et al., 2019). We have combined individual measures, including SDMs and measures of genomic variability, to assess the relative ability of CWRs to respond to climate change. To our knowledge, no formal modelling framework exists to perform this task; instead, we relied on an existing conceptual framework (Jonas A. Aguirre-Liguori et al., 2021). We recognize that this approach is not complete for many reasons. First, myriad additional biotic and abiotic phenomena likely affect the ability of populations to evolve in the face of climate change, but not all can be included. Second, we used estimated local offsets (and S_a) using individuals instead of populations. Local offsets are typically calculated on estimated allele frequencies within populations (Fitzpatrick & Keller, 2015), but we have applied it to diploid frequencies within individuals. While this may decrease predictive power, and is likely conservative for that reason, simulations suggest that our conclusions are robust

when adding additional allele frequency variation to the data (See **Supplementary Methods**). Third, we have ignored the potential for natural migration, which may also impact opportunities to adapt to climate change. Fourth, individual layers are not always statistically independent. While some correlations were anticipated (N_e and genomic load), we did not expect to find a significant correlation between load and PD. We believe the simplest explanation is that less-fit plants (as measured by load) are more susceptible to disease. This suggests that resistance is influenced by genomic background in addition to the presence or absence of major PD resistance alleles (Krivanek et al., 2006). It is interesting to surmise that this could be a general principle governing plant resistance; to prove generality will require study of more taxa and diseases.

Nonetheless, this synthesis provides a framework to begin to evaluate the relative risk among CWRs to predicted climate change by using genomic information. By considering evolutionary information, we conclude that *V. mustangensis* accessions are the most apt to persist in the face of predicted climate change. Generally, *mustangensis* accessions have good persistence in SDMs, high *Ne*, low load, and PD resistance. Second, the framework provides additional information for management and conservation. For example, at the 50% thresholds, two *arizonica* individuals score well for five of six layers, but *arizonica* has low *Ne* (**Figure 3A&C**). Based on this information, it may be necessary to aid management of *arizonica* populations by adding genetic diversity from different populations or species. Finally, we must accentuate that any analytical method predicting how populations will respond to climate change should, and can, be validated experimentally (Kardos & Shafer, 2018). For example, it will be interesting to grow *Vitis* individuals with different levels of load in common gardens that simulate climate change, both to corroborate that load is associated with fitness and to assess the accuracy of predictions.

Crop adaptation in the face of climate change

Given genomic information from CWRs, we have also assessed their potential utility to mediate the effects of climate change on viticulture. To do so, we first used

SDMs to identify U.S. locations where grapevine cultivation is likely to be at risk. Our SDM approach is simplistic, because it categorizes sites as suitable (or not) for viticulture, without considering qualitative measures like yield, disease resistance, phenology and intraspecific diversity (Foden et al., 2019; Lawler et al., 2006; Razgour et al., 2018). It also does not consider agronomic practices like irrigation. Nonetheless, we estimate that 25% of sites that match the present-day SDM will fall out of the bioclimatic niche by 2070 (**Figure 4**). Our estimate parallels previous studies based on alternative methodologies and data. For example, Morales-Castilla et al. (2020) (Morales-castilla et al., 2020) used phenological data to estimate that 56% of current, worldwide growing locations will be lost under 2Cº of warming; this value decreased to 24% when they considered intraspecific variety in phenology among grapevine cultivars. Similarly, Hannah et al. (2013) projected a 25% to 73% decrease in suitable land for viticulture by 2050 among major wine producing regions (Hannah et al., 2013). Interestingly, Morales-Castilla et al. (2020) also predicted the expansion of growing regions in Pacific Northwest U.S.(Morales-castilla et al., 2020), as do our SDMs (**Figure 4B**).

These studies consistently point to the loss of suitable regions for grapevine cultivation. At question is the response to this threat. Generally, there are three non-exclusive strategies to counter climate threat to crops. One is human-assisted migration that shift the locations of cultivation for a specific crop (Sloat et al., 2020). A second is to develop new agricultural regions – i.e., to expand arable land. However, land availability is limited, and the ecological and economic costs of this strategy is especially large (Fita, Rodríguez-Burruezo, Boscaiu, Prohens, & Vicente, 2015). The third strategy is *in situ* adaptation, which refers to breeding cultivars that tolerate stresses associated with climate change in their current locations of cultivation (Sloat et al., 2020).

An important component of *in situ* adaptation is introducing genetic novelty into crops from CWRs. The use of CWRs may be particularly crucial for breeding perennial crops like grapevines, as opposed to annual crops (Migicovsky, Cousins, & Myles, 2017), for at least two reasons. First, perennial cultivation is often based on only a few elite lines, so that many crops have a dangerously narrow genetic base (Myles, 2013). Second, most

(75%) perennial crops are clonally propagated (Miller & Gross, 2011); clonal propagation reproduces phenotypes but may lead to the accumulation of deleterious variants that affect fitness (Zhou et al., 2019) and susceptibility to disease (Myles, 2013). Accordingly, the wild relatives of perennial crops are especially valuable as sources for disease resistance (Migicovsky et al., 2017).

In this study, we asked: which CWRs have genetic and bioclimatic profiles that may make them useful to viticulture in regions of high vulnerability? To do so, we calculated 'migration offsets' and found that the migration offsets of *V. mustangensis* were consistently within the range of local offsets for that species (**Figure 4**). In fact, *mustangensis* accessions fit our criteria for contributing to *in situ* grapevine adaptation in most (490) of the 517 high-risk locations. *V. mustangensis* was not alone, however, because individuals in our *girdiana* and *riparia* samples also fit these criteria, although at fewer high-risk sites (**Figure S5**). Based on these results, we hypothesize that some of these wild accessions may prove useful for *in situ* adaptation of cultivated grapevine and specifically that *mustangensis* will prove vital agronomically, even though it has not been used widely as a rootstock.

It is important, however, to discuss this hypothesis critically. First, it of course assumes that the climate models are accurate. Second, we rely on ranges for migration offsets based on local genetic offsets; this range may or may not be reasonable. Validation is necessary here, too, by observing growth and physiology in environments similar to the predicted climates (Rellstab, Dauphin, & Exposito-Alonso, 2021). Third, it certainly is not yet proven that *mustangensis* will be agronomically useful. In fact, *mustangensis* may not even be the *best* candidate CWR for in situ adaptation at vulnerable sites, given that we have studied only five of 30 species and 105 accessions. That said, a practical implication of this study is that further *ex situ* sampling of *mustangensis* may be particularly beneficial, especially if focused near the sampling sites where vmus25 and vmus20 were first isolated. Finally, although our study is rare for including one potential biotic interactor, there are many more potential biotic and abiotic interactors, including microbiomes (Griggs, Steenwerth, Mills, Cantu, & Bokulich, 2021). If

CWRs are moved and used for breeding, they may be unexpectedly susceptible to pathogens or other abiotic stressors in their new location.

It is an open question whether *in situ* adaption to climate change in grapevine will be better achieved by identifying new rootstocks or by using CWRs for scion breeding. The latter has already been shown to be effective for discrete traits that are governed by a few major loci, such as disease resistance. Indeed, PD resistance at the *Pdr1* locus (Krivanek et al., 2006) has been backcrossed into various varieties to introgress the trait of PD resistance. Bioclimatic adaptation is polygenic, however, as evidenced by the thousands of SNPs associated with bioclimatic variables in each of these species (**Table 1**) and by the fact that associated SNPs are distributed throughout the chromosomes (Morales-Cruz et al., 2021). Given, the polygenic nature of bioclimatic adaptation, we suspect that *Vitis* CWRs will typically be more effective as rootstocks. That said, rootstocks can have a multitude of effects on the scion, including yield, phenology and drought tolerance (Warschefsky et al., 2016); moreover rootstock and scions can be incompatible, for reasons that are not yet well understood genetically (Gaut, Miller, & Seymour, 2019). Thus, we have at best identified candidates for further study, but our study nonetheless offers steps forward by suggesting candidates for priority evaluation.

METHODS

Plant material, resequencing and variant identification

The plant material used in this study consisted of 105 individuals from six American *Vitis* species (*V. arizonica, n* = 22; *V. mustangensis, n* = 24; *V. berlandieri, n* = 22; *V. girdiana, n* = 18 and *V. riparia; n* = 19), previously described in (Morales-Cruz et al., 2021). Samples were collected from the wild grape germplasm collection at Davis, California. The germplasm reported in this study was collected from 1997 to 2016 as cuttings across the southwestern states and maintained at the Department of Viticulture and Enology, University of California, Davis, CA. A table of sampling locations is available from (Morales-Cruz et al., 2021). The 105 accessions were also assayed previously for *X. fastidiosa* resistance (or tolerance) (Morales-Cruz et al., 2021; S. Riaz et al., 2018). Briefly, PD

resistance was evaluated using greenhouse experiments, where *X. fastidiosa* was inoculated in different individuals and the concentration of *X. fastidiosa* was evaluated using ELISA tests 10 to 14 weeks after the inoculation. Individuals were considered to be resistant to PD if they had concentrations of *X. fastidiosa* below 13 colonizing forming units (CTUs) (Summaira Riaz et al., 2020).

We used the sequencing data generated by Morales-Cruz *et al.* (2021) (Morales-Cruz et al., 2021), which is in the Short Read Archive at NCBI under BioProject ID: PRJNA731597, and their SNP calls. Briefly, they filtered Illumina paired-end reads of 150 base pairs (bp) and mapped them to the *V. arizonica* b40-14 v1.0 genome (Morales-Cruz et al., 2021) (https://doi.org/10.5281/zenodo.4977234 and www.grapegenomics.com/pages/Vari/). Joint SNP calling was conducted using the HaplotypeCaller in the GATK v.4.0 pipeline following ref. (Zhou, Massonnet, Sanjak, Cantu, & Gaut, 2017). The VCF files were split by species, and the raw SNPs were filtered with bcftools v1.9 (https://samtools.github.io/bcftools/) and vcftools v0.1.15 (https://vcftools.github.io/). SNPs sites were kept for downstream analyses if they were biallelic, had no missing data, had quality higher than 30, had a depth of coverage higher than five reads, and also had no more than three times the median coverage depth. Additionally, the following expression was applied under the exclusion argument of the filter function in bcftools: "QD < 2.0 | FS > 60.0 | MQ < 40.0 | MQRankSum < -12.5 | ReadPosRankSum < -8.0 | SOR > 3.0". These steps resulted in a range of high-quality filtered SNPs from 3.6 million in *V. girdiana* to 5.6 million in *V. riparia.*

Species Distribution Models

For each species we downloaded their occurrences from the Global Biodiversity Information Facility (www.gbif.org; last accessed November 2020). We removed duplicated locations and locations that were potential misclassifications due to being obvious outliers. We also downloaded 19 bioclimatic variables from Worldclim 2 (Fick & Hijmans, 2017) for the present, which is an average of observation from 1970 to 2010, and to 12 forecasts of climate change. The forecasts were downloaded from Worldclim 2 (last

accessed November 2020), based on the MIROC circulation model. The 12 forecasts correspond to three time periods (2050, 2070, 2090) and four Shared Socio-economic pathways (SSPs) that model different trajectories of greenhouse effects (SSPs 126, 245, 370, 585).

To build the model for each species, we used scripts that performed four steps to build an SDM (Jonas A. Aguirre-Liguori et al., 2021). First, the script removed correlated bioclimatic variables (r>0.8) using a Variance Inflation Factor. Second, it identified a calibration or background area, which corresponded to the migration layer in the BAM model (Peterson et al., 2011; Soberón, 2010) and corresponds 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, to build the SDMs, 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 with tenfold internal cross-validation.

Based on the present-day SDMs, we projected the potential distribution of species to the 12 predicted climate change scenarios. For each species we estimated how the geographic distribution (i.e., the number of pixels in the projection) is expected to change for years 2050, 2070 and 2090 and for each different SSP (i.e., SSPs 126, 245, 370, 585). For simplicity, we evaluated the change in the distribution of species between years as the mean area predicted across the 4 SSPs (**Figure 1B**). Based on the union of the 4 pathway models at each time period, which was taken based on the mosaic function in the R package *raster* (Hijmans et al., 2013), we also estimated whether each of the 105 accessions of the 5 *Vitis* species would persist (**Figure 1C**). If a sample location was predicted to be outside the species' estimated niche, we coded that population as extinct at that time period and into the future.

The SDMs for cultivated *V. vinifera* was built using the same script and methods. The 1278 locations of viticulture in the U.S. were downloaded from GBIF (last accessed

November 2020). The same 19 bioclimatic data were used, focusing only on present-day models and for the 2070 MIROC model (370 SSP).

Investigating patterns of local adaptation

Identification of outlier SNPs: For each CWR species sample, we used Baypass (Gautier, 2015) to identify outlier loci that had significant associations with the first four principal components (PC) that described the environmental distribution of each species. The principal components were determined with the *prcomp* function in R, based on the 19 bioclimatic variables from the present-day. [The first 4 principal components explained 99.9% of bioclimatic data for *arizonica;* 97.3% for *berlandieri* 97.3%; 92.9% for *mustangensis*; 98.8% for *girdiana* 98.8% and 97.4% for *riparia 97.4%*]. Baypass analyzes patterns of covariation among accessions to account for population structure and then analyzes the correlation between a variable (in this case each PC) and the allelic frequencies of individual SNPs across individuals. A SNP is deemed an outlier if it had a significant correlation with a variable that was not explained by the covariance between populations. Following Jeffreys' rule (Gautier, 2015), we characterized outlier SNPs as those that had a Bayes Factor (BF) > 20.

Local Genomic Offsets: To calculate local genomic offsets, we used gradient forest (GF), a machine learning method that models the turnover in genetic composition across the landscape (Fitzpatrick & Keller, 2015). GF identifies which bioclimatic variables contribute importantly to the construction of the model, the SNPs that are associated with bioclimatic variables, and the adaptive genetic composition across the predicted climate. For each species, we obtained the GF model using the *gradient forest* package in R (Ellis et al., 2012) for all the outlier SNPs identified by Baypass (**Table 1**), based on allelic frequencies within individuals. These frequencies were either 0 or 1 for homozygotes or 0.5 for inferred heterozygotes. The local offset was then calculated as the Euclidian distance of the genetic compositions between the present and all the same location at different time periods (Fitzpatrick & Keller, 2015). For each species, we used the GF models to predict how genetic compositions were expected to change in 2070 (SSP 370).

GF is traditionally applied to allele frequencies across populations. To investigate whether results were applicable for individuals, we transformed allelic frequencies within individuals to pseudo-population frequencies, using a probabilistic model, and reanalyzed the data (See Supplemental methods). Our simulations suggest that the sampling individuals, instead of populations, had little impact on the construction of the GF model, the estimation of genomic offsets and the identification of important bioclimatic variables (See **Figure S6**).

Calculating Sa: Given projections to 2070, we identified adaptive alleles as those that are at significantly higher frequency in populations that could be pre-adapted to climate change according to future projections of the GF models (**Figure S1**). To identify pre-adapted populations, we predicted the turnover composition of GF models using bioclimatic layers forecasted in the present and 2070 (SSP 370). This allowed us to identify in what direction the environment will change in the future for important variables (i.e., change across the x-axis in **Figure S1**) and which populations would be pre-adapted to future conditions. These are populations that will have a small, predicted change in the genetic composition (small change across the y axis for future values in **Figure S1**).

After identifying the pre-adapted population(s), for each candidate SNP we ran a linear model to evaluate if the genotypic state (0,1,2) of populations correlated significantly with the bioclimatic variables in which populations grow. If the linear model was not significant, we discarded the SNP. If it was significant, we considered as a pre-adaptive, the allele that was frequent in pre-adapted individuals. For each individual, we added the number of pre-adaptive alleles present in the accession (2 for each adaptive allele in the homozygous state, 1 for the heterozygous state) and relativized by the maximum number of potential adaptive alleles that populations could have (2 x N_{Sp} adaptive alleles, 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. S_a is analogous to the population adaptive index (Bonin, Nicole, Pompanon,

Miaud, & Taberlet, 2007), but instead of analyzing the proportion of adaptive SNPs in a populations it estimates the proportion of alleles that will be adaptive in the future. 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.

Migration Offsets: Once we identified the viticulture regions at-risk, we used the GF models of each wild subspecies to calculate the migration offset between each wild accession and each at-risk site. The migration offset can be considered the genomic offsets of a wild accession if, in the predicted climate of 2070, it was moved to an at-risk region. To calculate migration offsets, we calculated the Euclidian distance of the genetic compositions between the location where the accession exists in the present and all the locations where the accession could be moved in 2070. Finally, we identified locations where the migration offset would be within the current range of the species' predicted local offsets.

Population Genetic Parameters

Demographic history: We estimated the effective population (Ne) size of each species using MSMC2 v2.1.1 (Mallick et al., 2016) with unphased SNPs. To include only the most informative genomic regions, we created a mappability mask and a coverage mask. We created the mappability mask for the *V. arizonica* b40-14 v1.0 genome using the software SNPable (http://lh3lh3.users.sourceforge.net/snpable.shtml). We generated 150 bp mers moving in 1 bp increments across the genome, mapped the sequences back to the genome with BWA v0.7.8-r455 (Li & Durbin, 2009) and identified "mappable" genomic regions where the majority of sequences mapped uniquely without mistmaches. To include only the regions with sufficient sequencing coverage, we calculated the coverage from the alignment file of each sample with the bedcov program from the samtools (v1.10) package, supplying a bed file of 10 kb non-overlapping windows across the

genome. We then created a mask per sample that only included regions with sequencing coverage higher than 5x. A

An advantage of MSMC2, relative to previous versions, is that by using a simpler pairwise HMM it allows the analysis of unphased genomes, although it may still not be ideal for population separation analysis (Schiffels & Wang, 2020). Since we were mainly interested in the analysis of populations of each individual species, we took advantage of this advancement by using unphased SNPs from the six samples per species with the higheset average coverage (considering the mappability and coverage masks) as input for the run in MSMC2. For both analyses we used a mutation rate of 5.4e-9 estimated by (Liang et al., 2019) and a generation time of 3 years (Zhou et al., 2017). The results of the most recent time index from SMC++ and MSC2-unphased were highly similar (**Figure S3**). The results were very consistent among individual runs (**Figure S7),** and we show the average per species (**Figure 2D & S3**). In addition to MSMC2 with phased data, we also applied MSMC2 to phased data and SMC++ (see Supplementary Methods). All three methods estimated the highest *Ne* for *mustangensis* (**Figure S3**).

Genomic load: We determined the functional region (exonic, intronic, and intergenic) in which each mutation occurred using information from the *V. arizonica* B4014 genome annotations ((Morales-Cruz et al., 2021), https://doi.org/10.5281/zenodo.4977234 and www.grapegenomics.com/pages/Vari/). Exonic SNPs were annotated to be synonymous, nonsynonymous and frameshift using SnpEff version 5.0e (Cingolani et al., 2012). Nonsynonymous SNPs were predicted as deleterious or tolerated using the SIFT score (Ng & Henikoff, 2003), as computed in the program SIFT 4G (Vaser, Adusumalli, Leng, Sikic, & Ng, 2016). SIFT scores ≤ 0.05 were interpreted as deleterious and SIFT scores > 0.05 were considered to be tolerant. The use of the reference genome is expected to decrease the probability of the accessions genetically closed to the reference genome to return low scores (Lohmueller et al., 2008). Thus, we only used polymorphic sites without missing data that were predicted across all five wild grape species, resulting in a set of 397,723 SNPs.

Applying the FOLDS model

We used UpsetR (Conway, Lex, & Gehlenborg, 2017) to plot the intersections among the six layers of information, as described in the text. We selected different thresholds for inclusion. Accessions were included based on SDMs if the location of sampling was predicted to be in the species' niche in the future (i.e., continuously from 2050, 2070 and 2090 projections) and for PD if they had concentrations of *X. fastidiosa* below 13 CFUs. We also selected accessions as resistant if they had local offsets and deleterious load values below 90%, 50% and 25% of their distribution across species. Since local offsets are not comparable between species, the thresholds for local offsets were considered only within species. Finally, accessions were included when they had S_a or *Ne* above the 10%, 50% and 75% of the distributions across species. **Figure 3** reports only the UpsetR when at least five layers of information met the filtering criteria; the complete UpsetR result is in **Figure S8**.

REFERENCES

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

Bonin, A., Nicole, F., Pompanon, F., Miaud, C., & Taberlet, P. (2007). Population adaptive index: A new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, *21*(3), 697–708. doi: 10.1111/j.1523-1739.2007.00685.x

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

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., Brennan, A. C., Alonso-Blanco, C., & Picó, F. X. (2018). Spatio-temporal variation in fitness responses to contrasting environments in arabidopsis Thaliana. *Evolution*, *72*(8), 1570–1586. doi: 10.1111/evo.13508

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

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

Foden, W. B., Young, B. E., Akçakaya, H. R., Garcia, R. A., Hoffmann, A. A., Stein, B. A., … Huntley, B. (2019). Climate change vulnerability assessment of species. *Wiley Interdisciplinary Reviews: Climate Change*, *10*(1), 1–36. doi: 10.1002/wcc.551

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

Hannah, L., Roehrdanz, P. R., Ikegami, M., Shepard, A. V, Shaw, M. R., Tabor, G., … Hijmans, R. J. (2013). *Climate change , wine , and conservation*. *110*(17). doi: 10.1073/pnas.1210127110

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).

Hijmans, R. J., Van Etten, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., Lamigueiro, O. P., … Shortridge, A. (2013). *Package 'raster', Geographic data analysis and modeling, R cran package version 2.2-12*.

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

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

Massonnet, M., Cochetel, N., Minio, A., Vondras, A. M., Lin, J., Muyle, A., … Cantu, D. (2020). The genetic basis of sex determination in grapes. *Nature Communications*, *11*(1), 1–12. doi: 10.1038/s41467-020-16700-z

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

Migicovsky, Z., Cousins, P., Jordan, L. M., Myles, S., Striegler, R. K., Verdegaal, P., & Chitwood, D. H. (2021). Grapevine rootstocks affect growth-related scion phenotypes. *Plant Direct*, *5*(5), 1–11. doi: 10.1002/pld3.324

Migicovsky, Z., Cousins, P., & Myles, S. (2017). Exploiting wild relatives for genomics-assisted breeding of perennial crops. *Frontiers in Plant Science*, *8*(April), 1–16. doi: 10.3389/fpls.2017.00460

Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication.

American Journal of Botany, *98*(9), 1389–1414. doi: 10.3732/ajb.1000522

Moore, M. O. (1991). *CLASSIFICATION AND SYSTEMATICS OF EASTERN NORTH AMERICAN VITIS L . (VITACEAE) NORTH OF MEXICO Author (s): MICHAEL O . MOORE Source : SIDA , Contributions to Botany , Vol . 14 , No . 3 (AUGUST 1991), pp . 339-367 Published by : The Botanical Research*. *14*(3), 339–367.

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

Myles, S. (2013). Improving fruit and wine: What does genomics have to offer? *Trends in Genetics*, *29*(4), 190–196. doi: 10.1016/j.tig.2013.01.006

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

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

Prohens, J., Gramazio, P., Plazas, M., Dempewolf, H., Kilian, B., Díez, M. J., … Vilanova, S. (2017). Introgressiomics: a new approach for using crop wild relatives in breeding for adaptation to climate change. *Euphytica*, *213*(7). doi: 10.1007/s10681-017-1938-9

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

Rellstab, C., Dauphin, B., & Exposito-Alonso, M. (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications*, (February), 1–11. doi: 10.1111/eva.13205

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

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., … Gaut, B. S. (2019). The population genetics of structural variants in grapevine domestication. *Nature Plants*, *5*(9), 965–979. doi: 10.1038/s41477-019-0507-8

Acknowledgements: The authors thanks Yonfeng Zhou, Summaira Riaz, Andrew Walker and Dario Cantu for sharing data from previous manuscripts. This work was supported by National Science Foundation grant no. 1741627 to B.S.G; J. A-L was supported in part by a UC-Mexus postdoctoral fellowship.

Author Contributions: All authors contributed to the ideas of the study; J.A-L. & A.M-C. performed analyses; all authors wrote the manuscript.

FIGURE LEGENDS:

Figure 1. Results based on SDMs of the five *Vitis* species in this study. A) Present-day SDMs showing the estimated bioclimatic niche of each of the five species. Species are colored according to the key. The map also provides the sampling locations of the 105 accessions, with shapes and colors of different species provided by the key. B) The geographic area of the niche for each species graphed into the future. For 2050, 2070 and 2090, each point represents the mean of SDMs based on four distinct (SSPs). The colors represent the species, as indicated in the key of panel A. C) Each column represents one of the 105 accessions in this study, with the color representing the species as defined in panel A. If the location of sampled individual did not fall within the future niche, as defined by the SDM, then it is no longer represented in the figure. Hence, by 2090, SDMs predict that most *berlandieri* samples will be able to persist in their current locations, as will a few *mustangensis* and *riparia*. However, they also predict that most species will need to migrate or adapt to persist.

Figure 2. The layers of information used in the FOLDS model. A) Distribution of local genomic offset for each species. We darkened circles based on their status above and below the median within each species. B) Variation in the adaptive scores (S_a) across species. Each circle represents a sampled accession. The dashed line shows the median of S_a across all accessions; circles are darkened if they have S_a above the median. C) An example of correlations between S_a and bioclimatic variables, in this case the maximum temperature of the warmest month (BIO 5) for *mustangensis* samples. The dashed line shows the correlation for reference SNPs. For all results and test for potential biases see **Figure S2.** D) Change in effective population size across time for each species. The color of each species is indicated by the key. E) The measure of genetic load – i.e., the proportion of predicted deleterious SNPs per genome across a shared set of SNPs – plotted for each accession and species. [Note that the proportion of deleterious SNPs is equivalent to the number of deleterious SNPs per genome, since all inferences were based on a shared set

of SNPs.] The dashed line represents the median of the genomic load. Darker circles are below the median. F) Distribution of the concentration of *X. fastidiosa* in different wild *Vitis* species after exposure to the bacterium. The dashed line represents the value 13 CFUs, which has been the boundary between resistance and susceptibility. Darker circles represent accessions with CFUs < 13.

Figure 3. The FOLDS model applied to wild *Vitis*. A) The upsetR plot shows the intersection across all 105 individuals that pass thresholds for at least five of the six layers of information for 3 different thresholds defined in the text. **B-D)** The maps plot the individuals that remain after applying each of the three thresholds; comparing across the maps provides an indication of relative persistence and evolvability under climate change for the different accessions.

Figure 4. Climate change and in situ adaptation of *V. vinifera*. A) The circles represent current locations of grapevine cultivation, based on GBIF. The blue shading shows the SDM projection of suitable niches for viticulture, based on the current locations (circles) and present-day bioclimatic models. Clear circles represent locations that do not match the SDM projection. **B)** An SDM of grapevine cultivation niche based on 2070 climate projections. The red shading shows the projected niche in 2070, with clear circles indicating locations that are expected to be outside of the niche and are therefore at risk. **C)** The migration offset for each *mustangensis* accessions (e.g., vcan02, vcan03, etc.) estimated at each of the 517 at-risk viticulture locations (i.e., represented by clear circles in panel B). The dashed lines show the range of the local genetic offset across all accessions, based on 2070 climate projections. **D)** An example of the migration offsets for vmus25 layered onto the SDM projection. Gray and orange diamonds show sampling locations of *mustangensis*; the remaining diamonds correspond to the 517 clear circles in panel B. Black diamonds are locations where the migration offset exceeds the local offsets for *mustangensis*, suggesting that vmus25 may not be a suitable candidate for *in situ* adaptation of grapevine at those sites. Clear and colored diamonds show locations

where migrations offsets are within the range of local offsets; darker colors show higher offsets.

TABLES

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

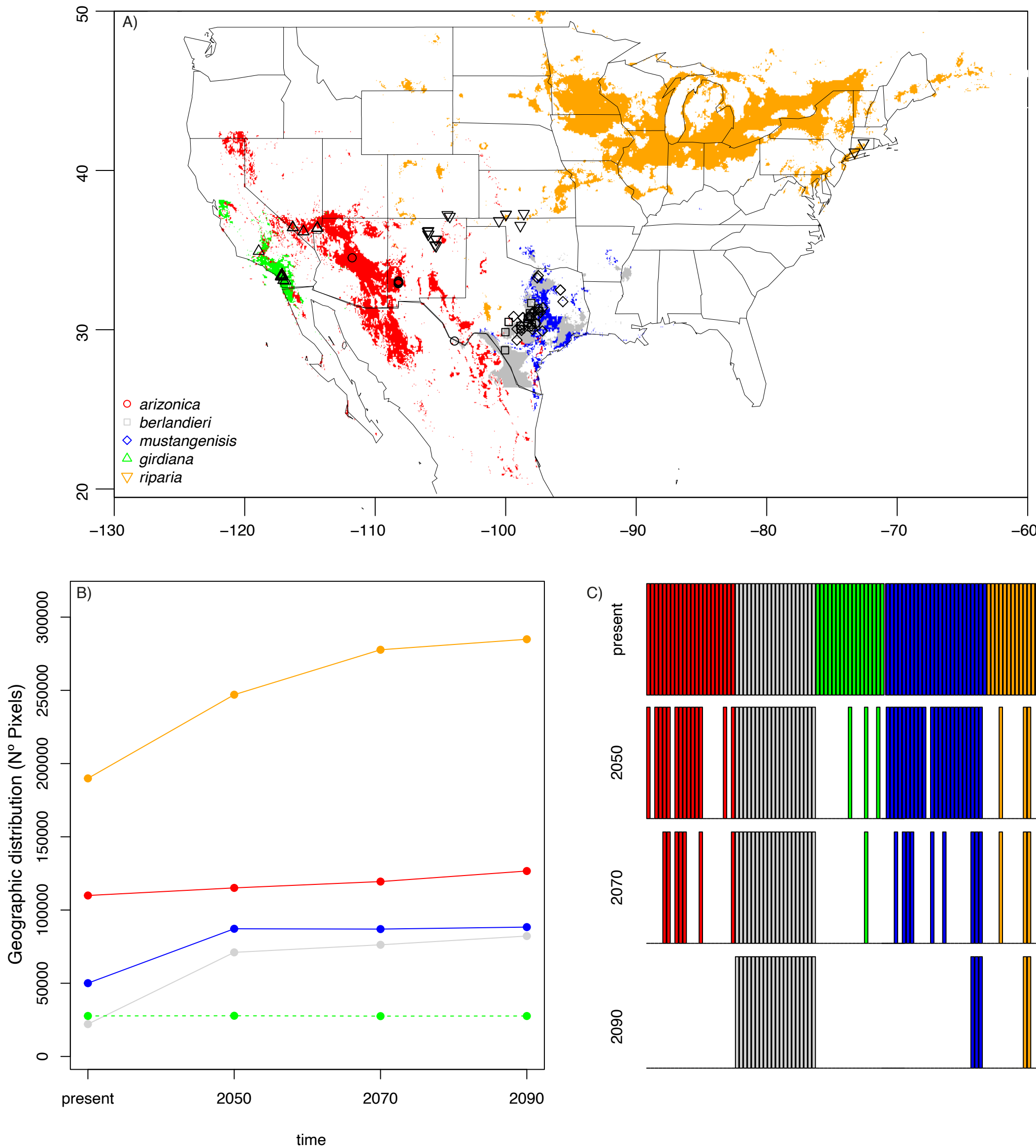

A)
50
40
30
20
−130
−120
−110
−100
−90
−80
−70
−60
arizonica
berlandieri
mustangenisis
girdiana
riparia
B)
Geographic distribution (N° Pixels)
0
50000
100000
150000
200000
250000
300000
present
2050
2070
2090
time
C)
present
2050
2070
2090

Figure 1

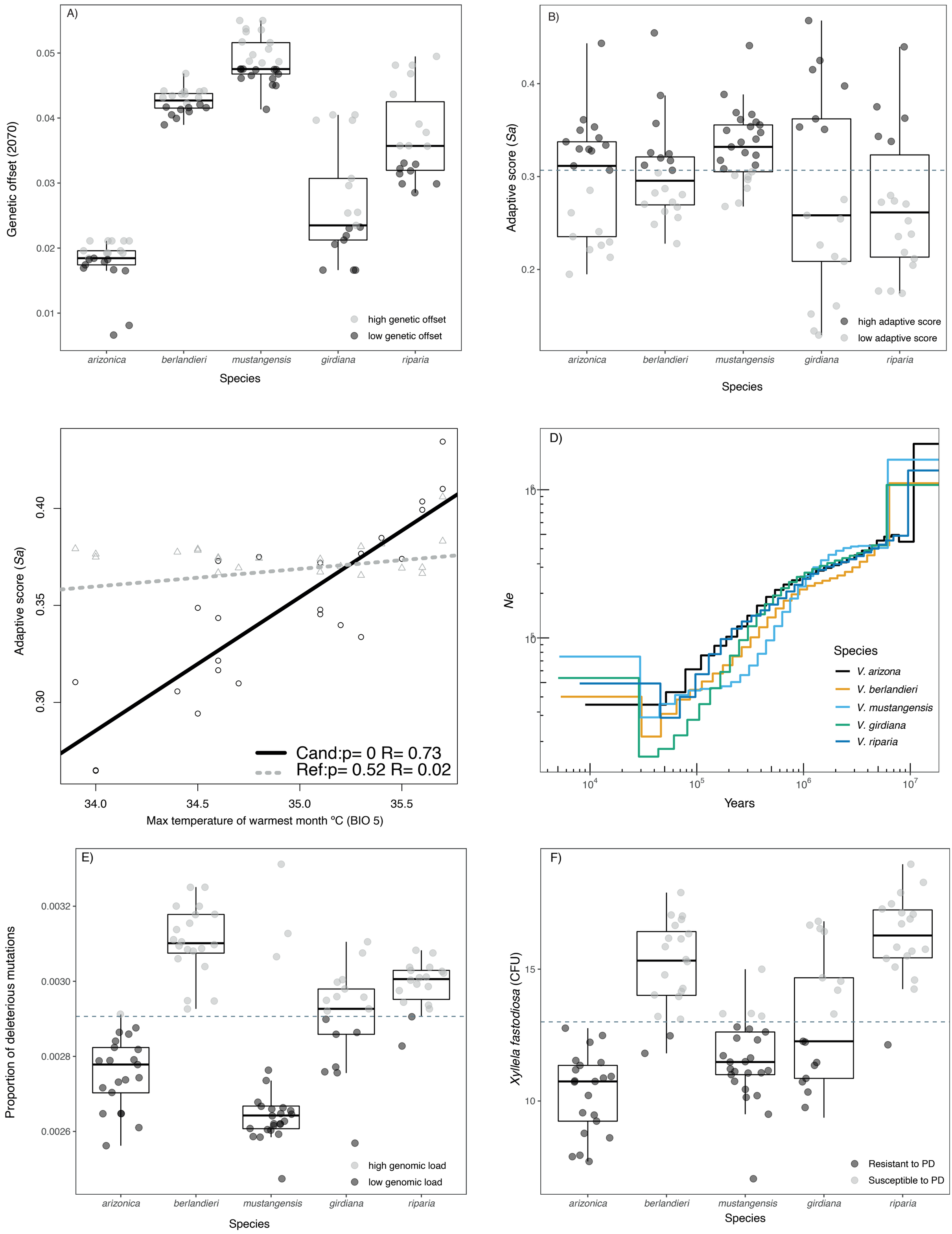
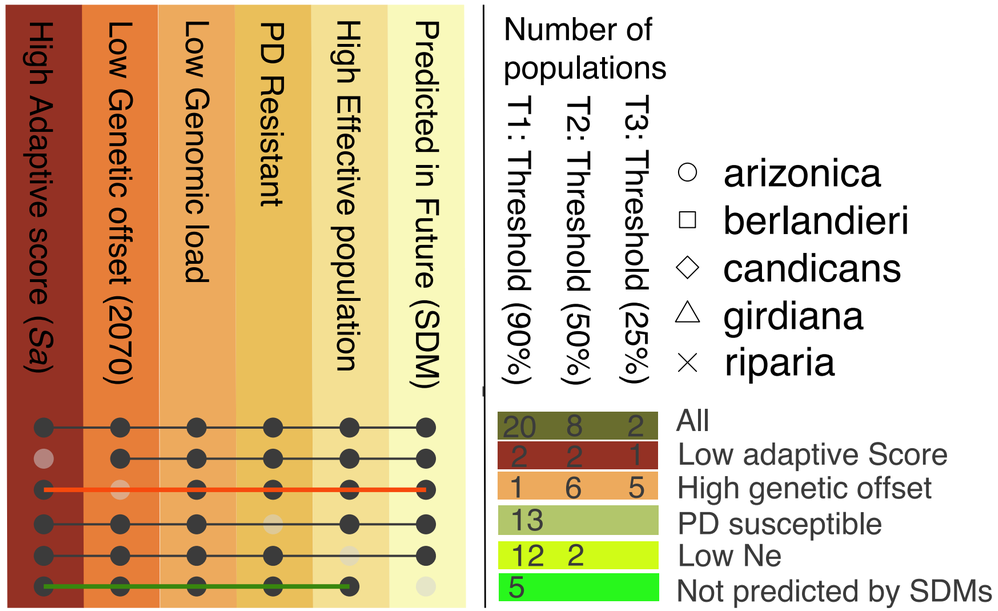
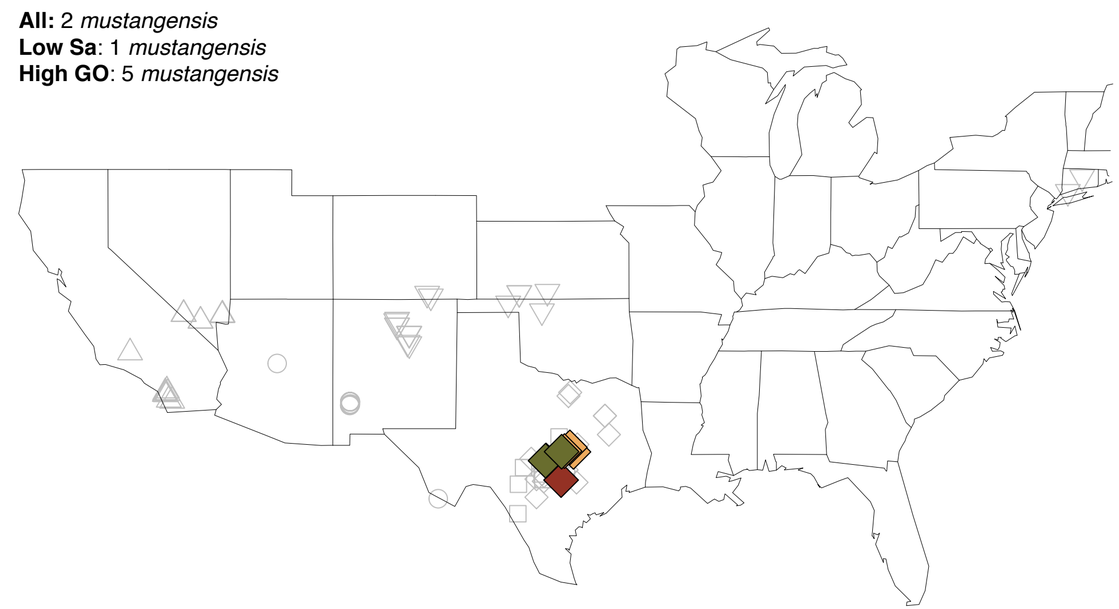

A)
Genetic offset (2070)
0.05
0.04
0.03
0.02
0.01
high genetic offset
low genetic offset
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
B)
Adaptive score (Sa)
0.4
0.3
0.2
high adaptive score
low adaptive score
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
Adaptive score (Sa)
0.40
0.35
0.30
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
Years
E)
Proportion of deleterious mutations
0.0032
0.0030
0.0028
0.0026
high genomic load
low genomic load
arizonica
berlandieri
mustangensis
girdiana
riparia
Species
F)
Xyllela fastodiosa (CFU)
15
10
Resistant to PD
Susceptible to PD
arizonica
berlandieri
mustangensis
girdiana
riparia
Species

Figure 2

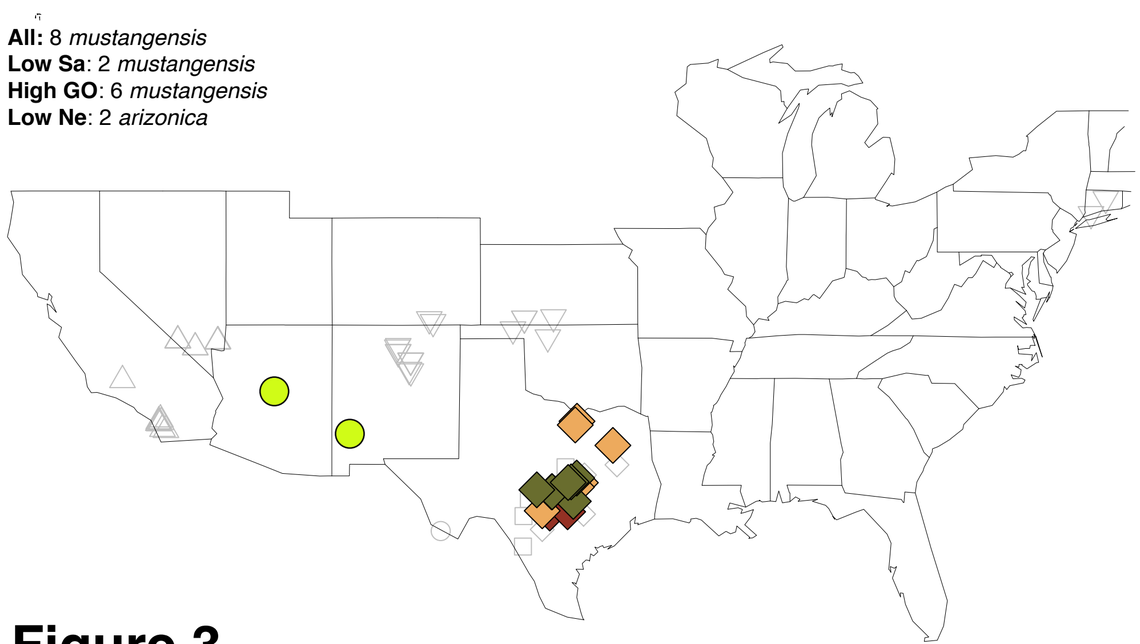
A)



B) Threshold (25%)



C) Threshold (50%)



D) Threshold (90%)

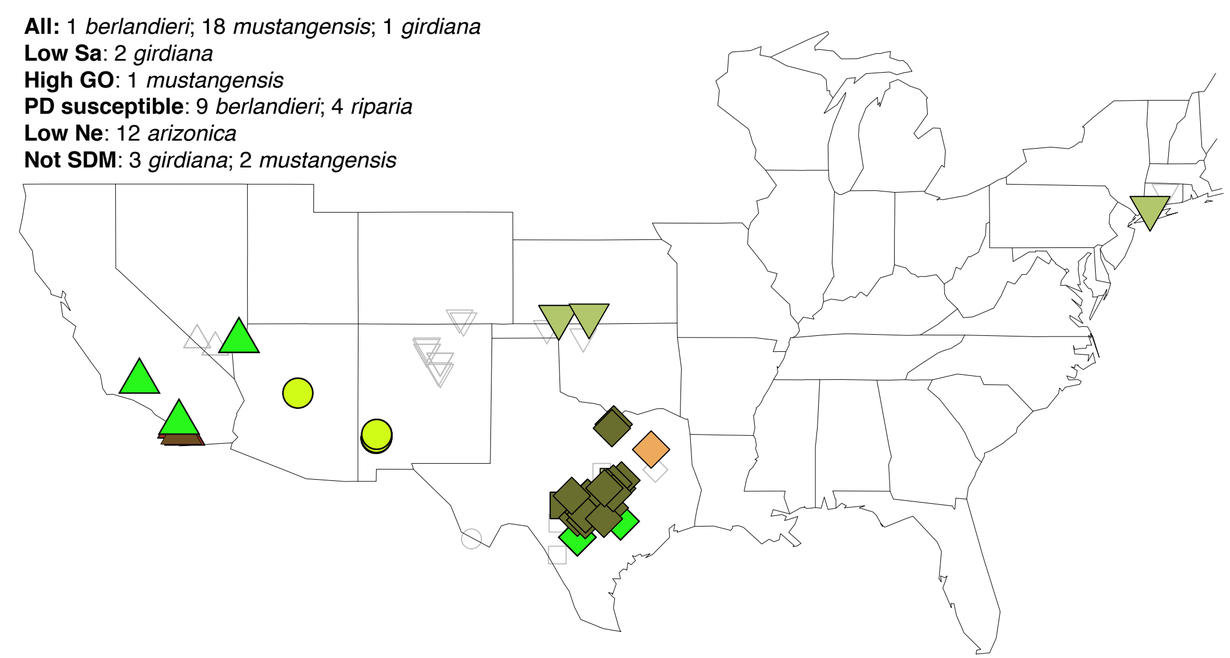


Figure 3

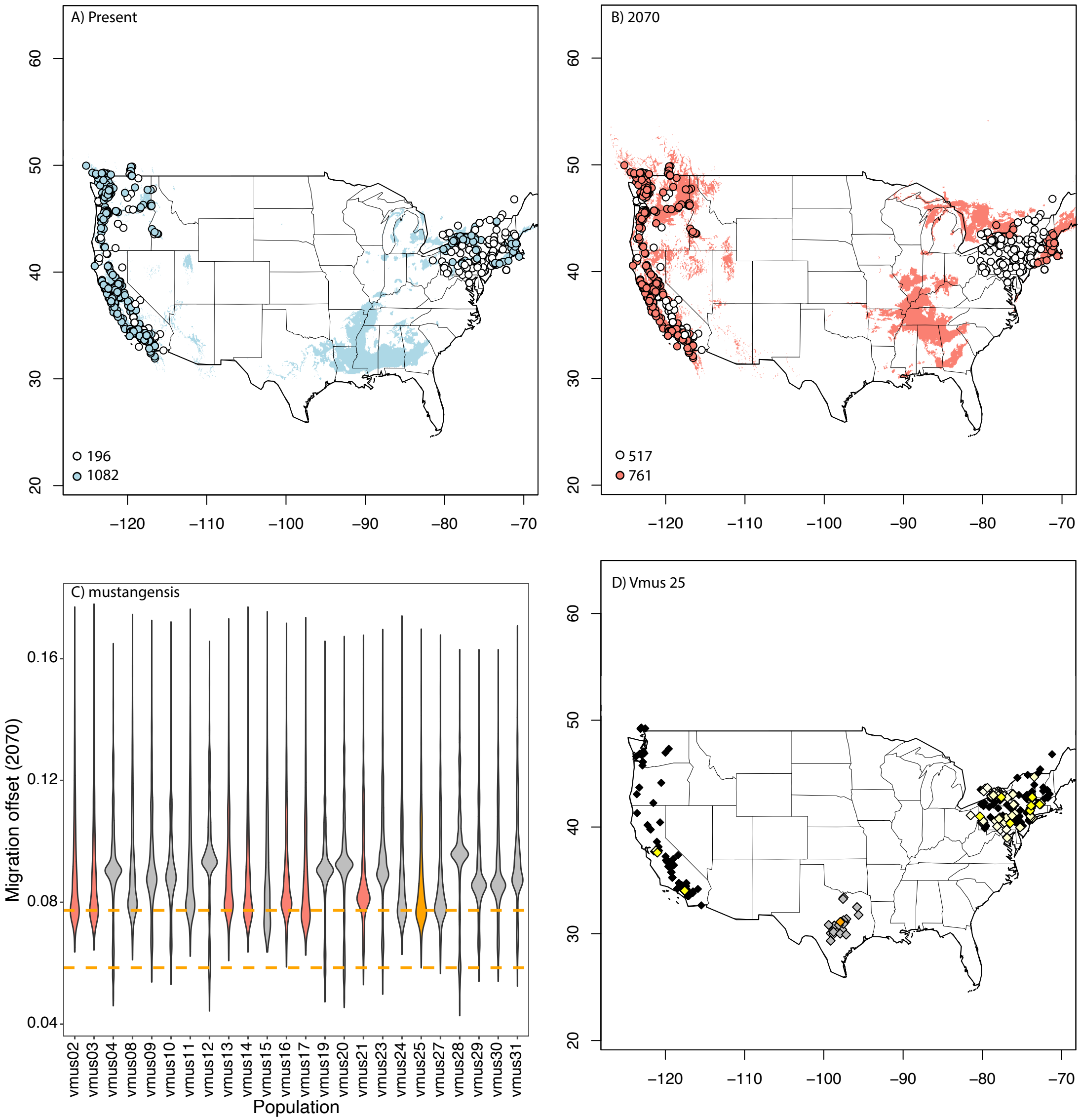

A) Present
196
1082
B) 2070
517
761
C) mustangensis
Migration offset (2070)
0.04
0.08
0.12
0.16
vmus02
vmus03
vmus04
vmus08
vmus09
vmus10
vmus11
vmus12
vmus13
vmus14
vmus15
vmus16
vmus17
vmus19
vmus20
vmus21
vmus23
vmus24
vmus25
vmus27
vmus28
vmus29
vmus30
vmus31
Population
D) Vmus 25
20
30
40
50
60
−120
−110
−100
−90
−80
−70

Figure 4

SUPPORTING INFORMATION

Climate Change, Crop Wild Relatives and Genomic Data: The persistence and agronomic utility of five wild *Vitis* species

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

Supplemental Methods

Examining consistency of offset Gradient Forest inference after adding random variation to allele frequencies. 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+Δ, 0.5±Δ or 1-Δ) 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 S6a**). 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 S6 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 S6b**). 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 S6c**). 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 S6d**), 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 might have a small effect on the estimation of the GF models, and that the general patterns are retained.

Evaluating Sa: Methods that detect local adaptation are sensitive to geographic and demographic biases, especially when the environment correlates with geographic distance. To examine potential biases, we repeated the previous analysis with a set of N_{Sp} reference SNPs with a BF<10 (S_{ref}), where N_{Sp} is the number of candidate SNPs in a given species (**Figure S2**). Strong associations for both the reference and adaptive scores indicate that there are biases associated with genetic structure, such as Isolation by Distance (See **Figure S2**).

Demographic inference: In the main text, we reported MSMC 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 S3**).

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 S3**).

Supplemental Figures

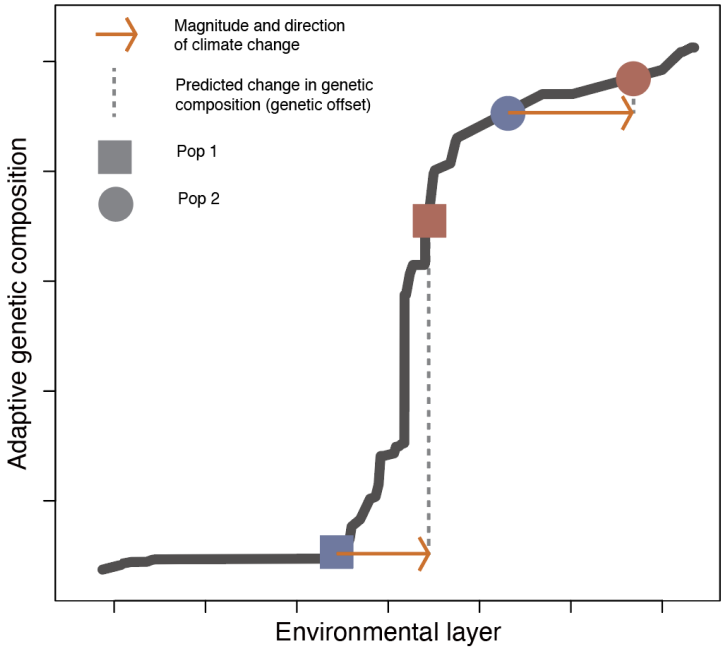


Figure S1. Description of the turnover function and how it can be used to identify the population that is pre-adapted to climate change. The circle and square show two populations along their turnover functions for present (blue) and future (red) climates. The arrows show how climate change will move 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.

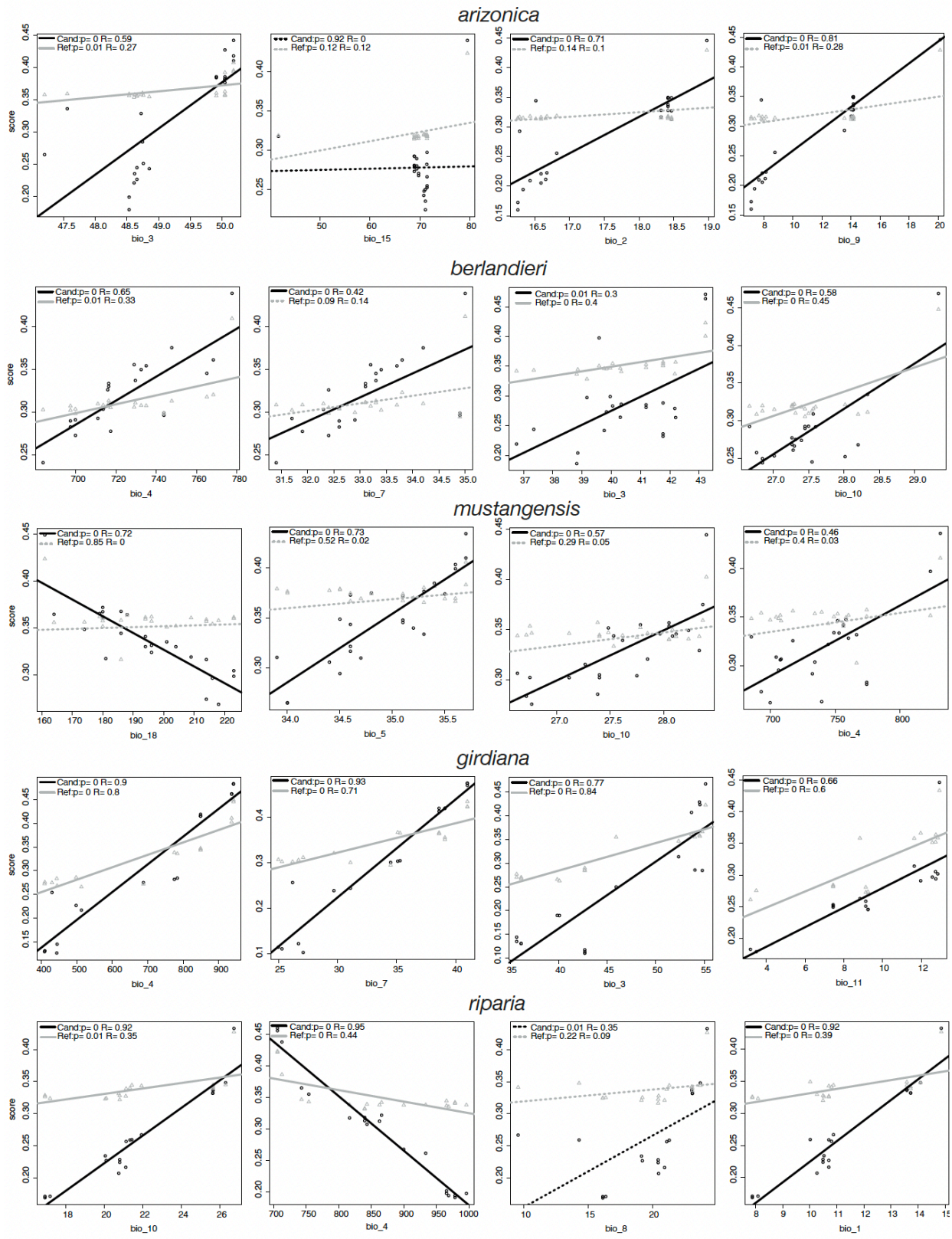

arizonica
Cand:p= 0 R= 0.59
Ref:p= 0.01 R= 0.27
score
bio_3
Cand:p= 0.92 R= 0
Ref:p= 0.12 R= 0.12
bio_15
Cand:p= 0 R= 0.71
Ref:p= 0.14 R= 0.1
bio_2
Cand:p= 0 R= 0.81
Ref:p= 0.01 R= 0.28
bio_9
berlandieri
Cand:p= 0 R= 0.65
Ref:p= 0.01 R= 0.33
score
bio_4
Cand:p= 0 R= 0.42
Ref:p= 0.09 R= 0.14
bio_7
Cand:p= 0.01 R= 0.3
Ref:p= 0 R= 0.4
bio_3
Cand:p= 0 R= 0.58
Ref:p= 0 R= 0.45
bio_10
mustangensis
Cand:p= 0 R= 0.72
Ref:p= 0.85 R= 0
score
bio_18
Cand:p= 0 R= 0.73
Ref:p= 0.52 R= 0.02
bio_5
Cand:p= 0 R= 0.57
Ref:p= 0.29 R= 0.05
bio_10
Cand:p= 0 R= 0.46
Ref:p= 0.4 R= 0.03
bio_4
girdiana
Cand:p= 0 R= 0.9
Ref:p= 0 R= 0.8
score
bio_4
Cand:p= 0 R= 0.93
Ref:p= 0 R= 0.71
bio_7
Cand:p= 0 R= 0.77
Ref:p= 0 R= 0.84
bio_3
Cand:p= 0 R= 0.66
Ref:p= 0 R= 0.6
bio_11
riparia
Cand:p= 0 R= 0.92
Ref:p= 0.01 R= 0.35
score
bio_10
Cand:p= 0 R= 0.95
Ref:p= 0 R= 0.44
bio_4
Cand:p= 0.01 R= 0.35
Ref:p= 0.22 R= 0.09
bio_8
Cand:p= 0 R= 0.92
Ref:p= 0 R= 0.39
bio_1

Figure S2. 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 ($p<0.05$).

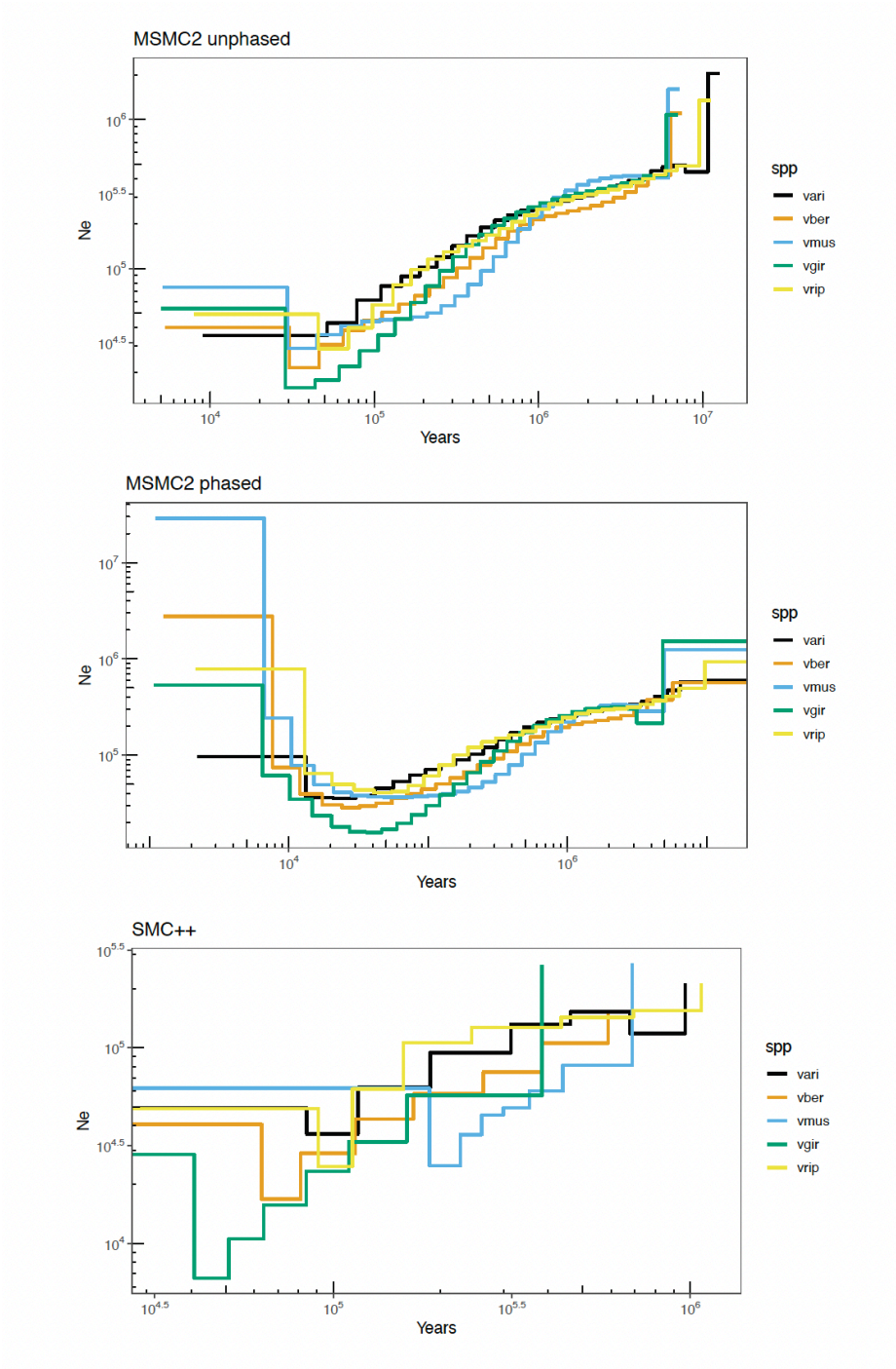


Figure S3. Outcomes from demographic inference. We inferred demographic history based on three different methods: MSMC2 with unphased data (top), MSMC2 with unphased data (middle) and SMC++ (bottom).

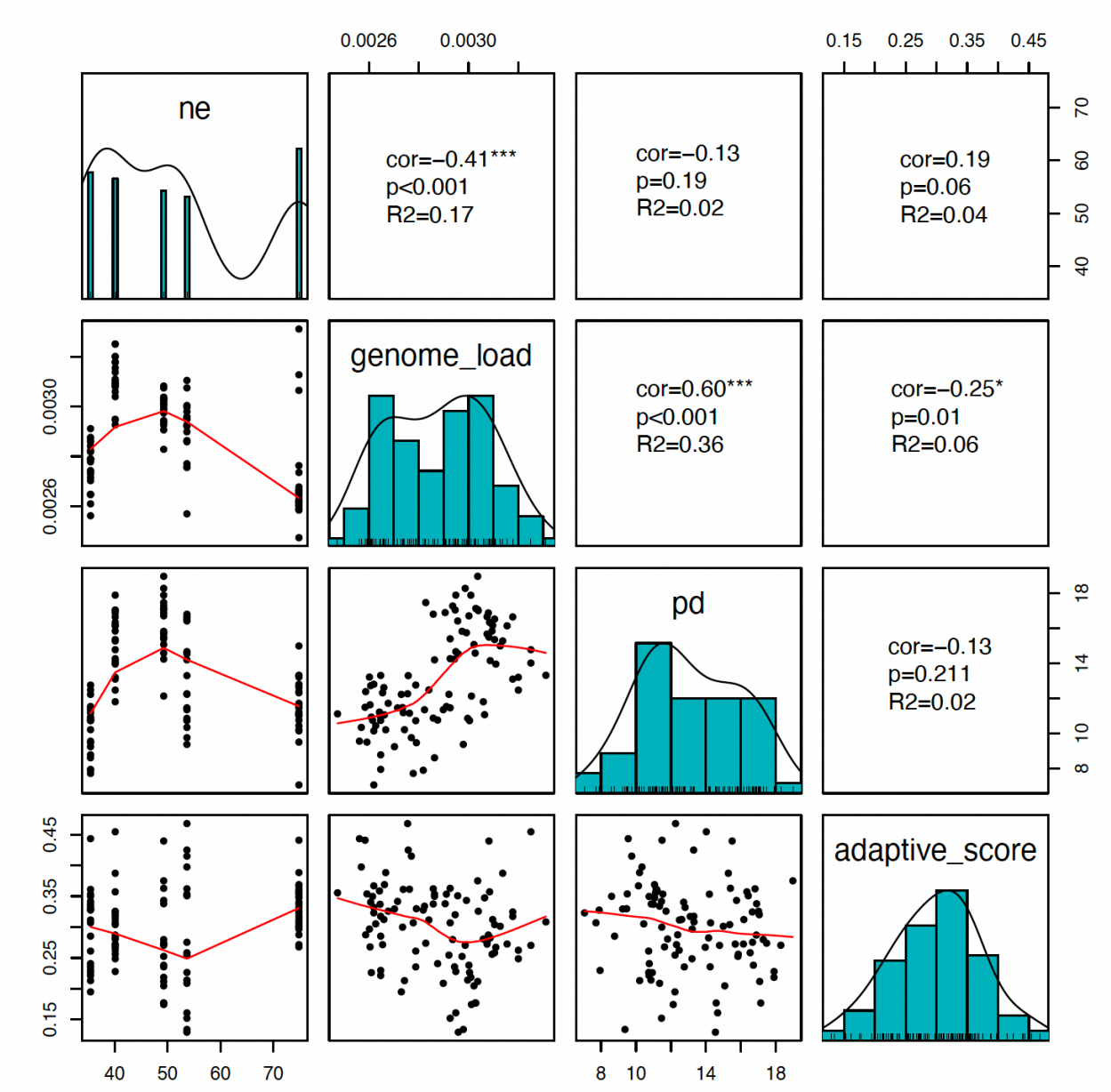


Figure S4. Correlations between variables. Scatterplot matrix showing the correlations between variables across all 105 samples in the study. The diagonal shows the distribution across the samples for each variable. The upper triangle shows the correlations (cor), significance (p) and fit (R^2) between variables. The lower triangle shows scatterplots for the two sets of variables in question.

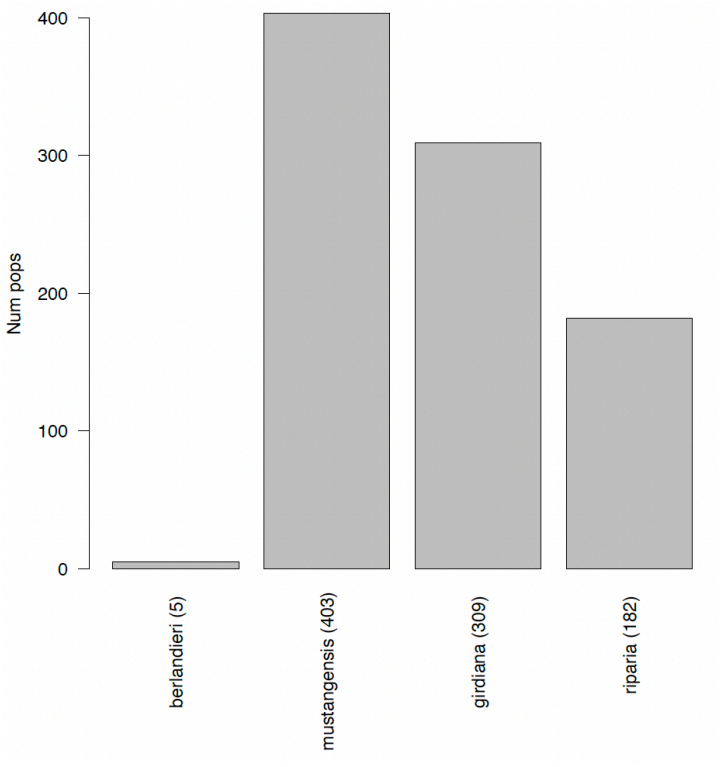


Figure S5. Potential CWR to at-risk viticulture sites. The barplots report the number of 517 at-risk *V. vinifera* populations that have CWRs that may contribute to crop in situ adaptation, based on the criterion that the migration offsets is less than the maximum local offsets in each species. This plot shows that the *mustangenesis* samples match the most at-risk locations based on our criteria, with 403 of 517. However, *girdiana* (309 sites) and *riparia* (182) also perform well.

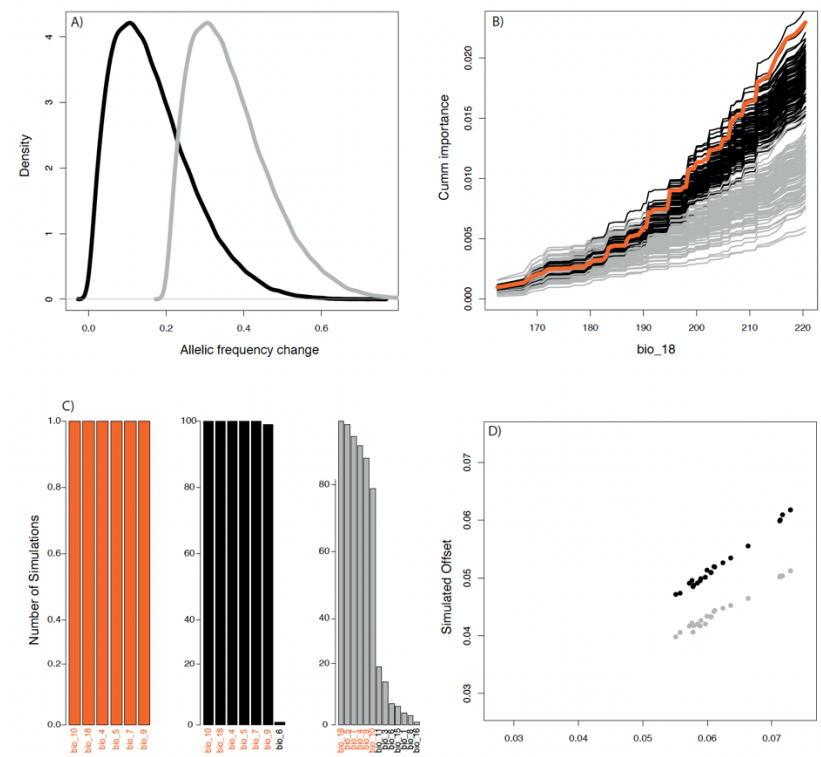


Figure S6. Effect of using individual and frequency data for estimating the gradient forest models in *Vitis mustangensis*. A) Beta distribution of the magnitude of change of the simulated pseudo-frequencies (Δ) with means 0.1 (black) and 0.3 (grey). B) Turnover functions for Bio-18. Black and grey curves 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 gradient forest 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 y-axis denoting the proportion of resampled datasets that include the bioclimatic parameter as an important variable. D) Scatterplot showing the 2070 genetic offset of the individual dataset in the x-axis and the mean of the simulated datasets for the different Δ distributions.

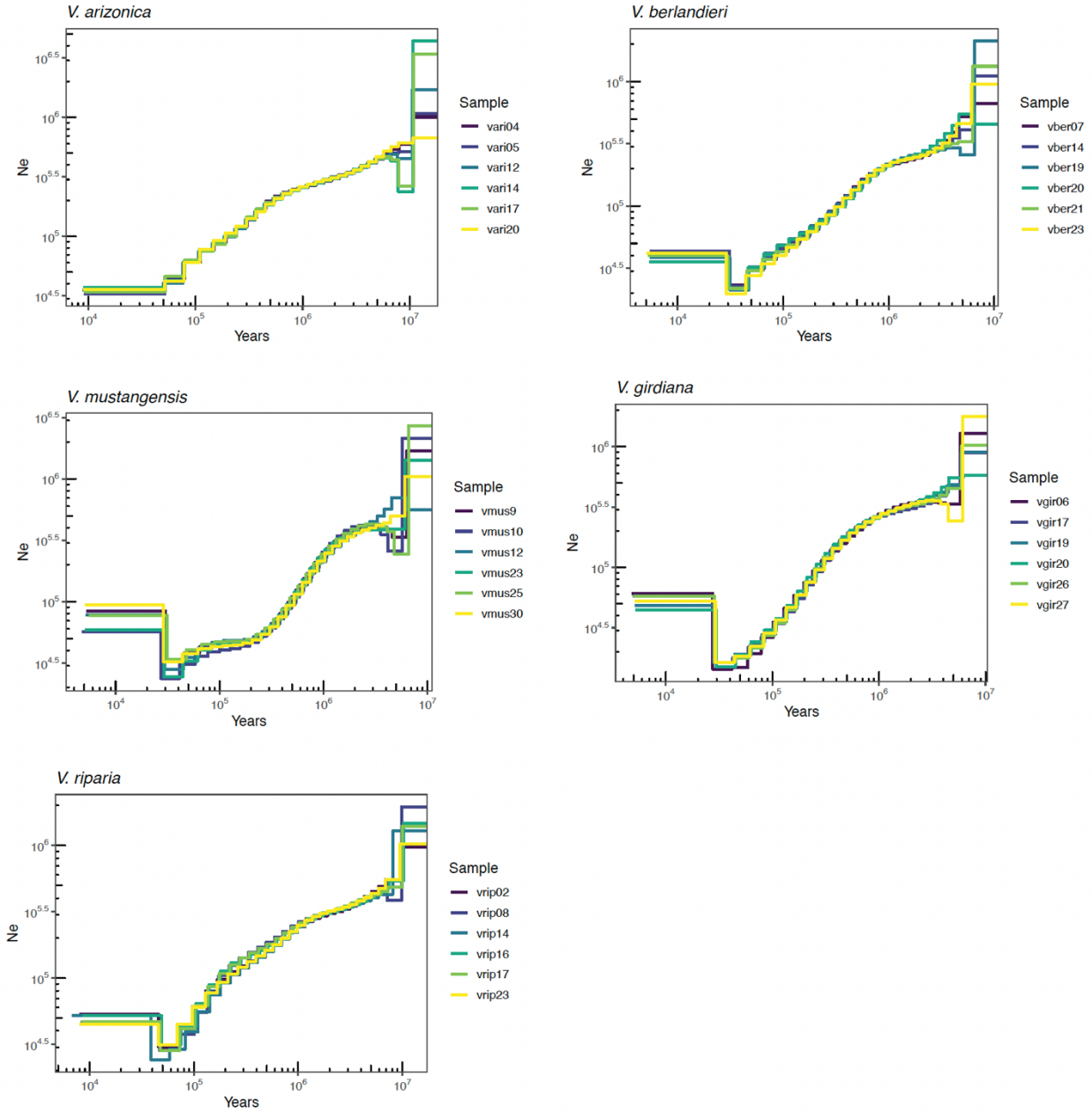


Figure S7: Demographic inference based on individuals in each of the five CWR species. Each graph represents a single species, as labelled, with separate individuals represented by different colors. All inferences were based on MSMC2 with unphased data.

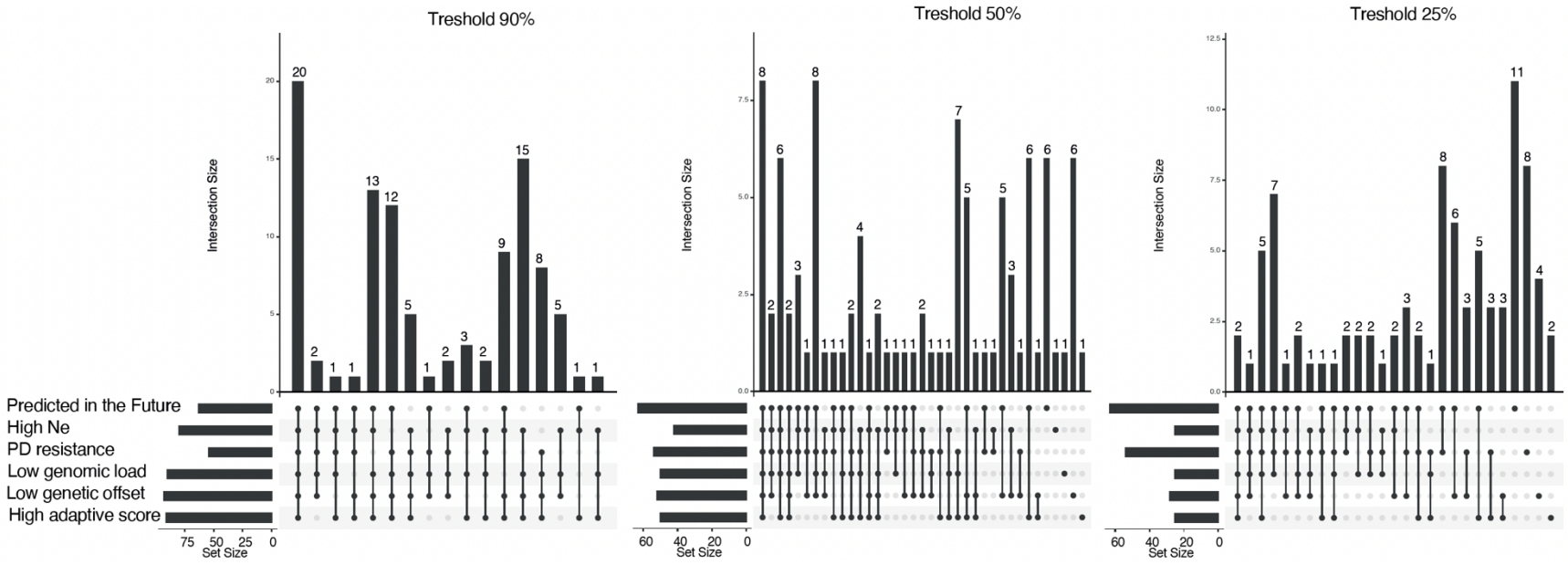


Figure S8. UpsetR plots showing the intersection between layers of information using three different thresholds. The thresholds indicate the individuals that have low offsets and load (below 90%, 50% and 25%) and high adaptive scores and Ne (above 10%, 50% and 75%). With more stringent thresholds, fewer populations are identified by many layers.