bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The Evolutionary Genomics of Grape (*Vitis vinifera* ssp. *vinifera*) Domestication

Yongfeng Zhou[1], Mélanie Massonnet[2], Jaleal Sanjak[1], Dario Cantu[2], Brandon S. Gaut[1,*]

[1]Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, UC Irvine, Irvine, CA 92617

[2]Department of Viticulture and Enology, One Shields Avenue, University of California, Davis, Davis, CA 95616

[*]Corresponding author:

Brandon Gaut

Dept. Ecology and Evolutionary Biology

321 Steinhaus Hall

UC Irvine

Irvine, CA 92697-2525

Phone: 949 825-2564

E-mail: bgaut@uci.edu.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ABSTRACT

The domestication history of grapes (*Vitis vinifera* ssp. *vinifera*) has not yet been investigated with genome sequencing data. We gathered data for a sample of 18 cultivars and nine putatively wild accessions to address three features of domestication history. The first was demography. We estimated that the wild and cultivated samples diverged ~22,000 years ago. Thereafter the cultivated lineage experienced a steady decline in population size (N_e), reaching its nadir near the time of domestication ~8,000 years ago. The long decline may reflect low intensity cultivation by humans prior to domestication. N_e of the wild sample fluctuated over the same timeframe, commensurate with glacial expansion and retraction. Second, we characterized regions of putative selective sweeps, identifying 309 candidate-selected genes in the cultivated sample. The set included genes that function in sugar metabolism, flower development and stress responses. Selected genes in the wild sample were enriched exclusively for functions related to biotic and abiotic stresses. A genomic region of high differentiation between wild and domesticated samples corresponded to the sex determination region, which included a candidate gene for a male sterility factor and additional genes that vary in gene expression among sexes. Finally, we investigated the cost of domestication. Despite the lack of a strong domestication bottleneck, grape accessions contained 5.2% more deleterious variants than wild individuals, and these were more often in a heterozygous state. We confirm that clonal propagation leads to the accumulation of recessive deleterious mutations, which is a likely cause of severe inbreeding depression in grapes.

Keywords: perennial crops, population genomics, pre-domestication, deleterious variants, demographic bottleneck

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

INTRODUCTION

Grapevines (*Vitis vinifera* ssp. *vinifera*) are the most economically important horticultural crop in the world (Myles et al., 2011). The products of grape cultivation include table grapes, raisins, juice, wine and oil, and they contribute an estimated $162B annually to the American economy alone (Initiative, 2007). In addition to its economic value, *V. vinifera* is a model organism for the study of perennial fruit crops, both because it can be transformed and micropropagated via somatic embryogenesis (Kikkert et al., 2005, Wang et al., 2005) and because it has a relatively small genome. Its ~500 Mb genome is similar in size to that of rice (430 Mb) (Goff et al., 2002) and poplar (465Mb) (Tuskan et al., 2006).

Cultivated grapes (hereafter *vinifera*) have been a source of food and wine since their domestication ~8000 years ago (8.0kya) from their wild progenitor, *V. vinifera* ssp. *sylvestris* (hereafter *sylvestris*) (McGovern et al., 2003). The exact location of domestication remains uncertain, but most lines of evidence point to a primary domestication event in the Near-East (McGovern et al., 2003, Myles et al., 2011). Domestication caused morphological shifts that include larger berry and bunch sizes, higher sugar content, altered seed morphology, and a shift from dioecy to a hermaphroditic mating system (This et al., 2007). There is interest in identifying the genes that contribute to these morphological shifts. For example, several papers have attempted to identify the gene(s) that are responsible for the shift to hermaphroditism, which are located in a ~150kb region on chromosome 2 (Fechter et al., 2012, Picq et al., 2014).

Historically, genetic diversity among *V. vinifera* varieties has been studied with simple sequence repeats (SSRs) (Bowers et al., 1999). More recently, a group genotyped 950 *vinifera* and 59 *sylvestris* accessions with a chip containing 9,000 SNPs (Myles et al., 2011). Their data suggest that grape domestication led to a noticeable, but mild, reduction of genetic diversity. Still more recent studies have used whole-genome sequencing (WGS) to assess structural variation among grape varieties (Di Genova et al., 2014, Cardone et al., 2016, Xu et al., 2016). Surprisingly, however, WGS data have not been used to investigate the population genomics of grapes.

Several population genomic studies have focused on annual plants, but few have focused on perennial crops like grapes (Velasco et al., 2016, Wang et al., 2017). The distinction between annual and perennial crops is important, because perennial domestication is expected to differ from annual domestication in at least three aspects (Miller and Gross, 2011, Gaut et al., 2015).

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The first is time. Long-lived perennials have extended juvenile stages and are often propagated clonally. As a result, the number of sexual generations is much reduced for perennials relative to annual crops, even for perennials like grapes that were domesticated relatively early in human agricultural history. The second is the severity of the domestication bottleneck. Many (and perhaps most) perennial crops differ from annual crops in that they have not experienced a severe domestication bottleneck (Miller and Gross, 2011). The third is clonal propagation; many perennials are propagated clonally but most annuals are not. Clonal propagation maintains genetic diversity in desirous combinations but also limits opportunities for sexual recombination (Miller and Gross, 2011, Ramu et al., 2017).

The distinct features of perennial domestication likely affect genome-wide patterns of genetic diversity. For example, perennials may differ from annuals with respect to the 'cost of domestication' (Lu et al., 2006), which refers to an increased genetic load within cultivars. This cost originates partly from the fact that the decreased effective population size (N_e) during a domestication bottleneck reduces the efficacy of genome-wide selection (Charlesworth and Willis, 2009), which may increase the frequency and number of slightly deleterious variants (Lohmueller, 2014, Henn et al., 2015). The characterization of deleterious variants is important, because they are potential targets for crop improvement (Morrell et al., 2011). Thus far, putatively deleterious variants been studied in several annual crops (Renaut and Rieseberg, 2015, Kono et al., 2016, Liu et al., 2017, Ramu et al., 2017) but not perennials. If perennials do not typically experience strong domestication bottlenecks (Miller and Gross, 2011), their domestication may not result in an associated cost (Gaut et al., 2015).

Here we perform WGS on a sample of *vinifera* cultivars and putatively wild *sylvestris* accessions to focus on three sets of questions. First, what do the data reveal about the demographic history of cultivated grapes, specifically the timing and severity of a domestication bottleneck? Second, what genes bear the signature of selection in *vinifera*? Finally, has there been an accumulation of putatively deleterious variants in grapes relative to *sylvestris*, or have the unique features of perennial domestication permitted an escape from this cost?

RESULTS

Plant samples and population structure: We collected WGS data from nine putatively wild *sylvestris* individuals from the Near-East that represent a single genetic group (Myles et al., 2011), 18 *vinifera* individuals representing 14 cultivars, and one outgroup (*Muscadinia*

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

rotundifolia) (Table S1). Our *sylvestris* accessions are a subset of the wild sample from reference (Myles et al., 2011), which was filtered for provenance and authenticity. We nonetheless label the *sylvestris* sample as 'putatively' wild, because it can be difficult to identify truly wild individuals. Reads were mapped to the Pinot Noir reference genome PN40024 (Jaillon et al., 2007), resulting in the identification of 3,963,172 and 3,732,107 SNPs across the *sylvestris* and *vinifera* samples (see Methods).

To investigate population structure, we applied principal component analysis (PCA) to genotype likelihoods (Korneliussen et al., 2014). Only the first two principal components (PCs) were significant (P<0.001); they explained 23.03% and 21.88% of the total genetic variance, respectively (Fig. 1A). PC1 separated samples of wine and table grapes, except for two accessions (Italia and Muscat of Alexandria) positioned between the two groups. PC2 divided wild and cultivated samples. Wine, table and wild grapes clustered separately in a neighbor joining tree, except for Muscat of Alexandria, which has been used historically for both wine and table grapes (Fig. 1B). Finally, STRUCTURE analyses revealed an optimal grouping of K=4 which separated *sylvestris* accessions, table grapes, wine grapes and the Zinfandel/Primitivo subgroup of wine grapes, and also identified admixed individuals (Fig. S1).

Nucleotide diversity and demographic history: We estimated population genetic parameters based on the *sylvestris* accessions (n=9) and on a cultivated sample of n=14 that included only one Thompson clone and one Zinfandel/Primitivo clone (Table S1). Both samples harbored substantial levels of nucleotide diversity across all sites (*sylvestris*: π_w = 0.0147 ± 0.0011; *vinifera*: π_c = 0.0139 ± 0.0014; Fig. S2). Although π was higher in *sylvestris* (π_c/π_w = 0.94 ± 0.14), *vinifera* had higher levels of heterozygosity and Tajima's D values (0.5421 ± 0.0932; *sylvestris*, D = -0.4651 ± 0.1577; Fig. S2). Linkage disequilibrium (LD) decayed to r^2 < 0.2 within 20 kilobases (kb) in both samples, but it declined more slowly for *vinifera* after ~20kb (Fig. S2).

We inferred the demographic history of the *vinifera* sample using MSMC, which requires phased SNPs (Schiffels and Durbin, 2014). Assuming a generation time of 3 years (McGovern et al., 2003) and a mutation rate of 2.5×10^{-9} mutations per nucleotide per year (Koch et al., 2000), we converted scaled population parameters into years and individuals (N_e). Based on these analyses, *vinifera* experienced a continual reduction of N_e starting ~22.0ky until its nadir from ~7.0kya to 11.0kya (Fig. 2A), which corresponds to the time of domestication and implies a

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

mild domestication bottleneck. Notably, there was no evidence for a dramatic expansion of N_e since domestication. MSMC results were similar across two separate analyses (Fig. 2A), based on n=4 samples of either table or wine grapes (Table S1), suggesting that analyses captured shared aspects of the samples' histories. We also used MSMC to compute divergence times. The divergence between *sylvestris* and *vinifera* was estimated to be ~21kya (Fig. 2B), which corresponds to the onset of the decline of *vinifera* N_e. Divergence between wine and table grapes was estimated to be ~2.5kya, which is well within the hypothesized period of *vinifera* domestication (Fig. 2B).

We repeated demographic analyses with SMC++, which estimates population histories and divergences without phasing (Terhorst et al., 2017)(Fig. 2C). This method yielded no evidence for a discrete bottleneck from ~7.0 to 10kya, but SMC++ and MSMC analyses had four similarities: *i*) an estimated divergence time of ~30kya that greatly predates domestication; *ii*) a slow decline in *vinifera* N_e since divergence; *iii*) no evidence for a rapid expansion in N_e after domestication; and *iv*) a ~2.6kya divergence of wine and table grapes (Fig. 2C, Fig. S3). We also used SMC++ to infer the demographic history of our *sylvestris* sample, revealing a complex N_e pattern that corresponds to features of climatic history (see Discussion).

Sweep mapping: We investigated patterns of selection and interspecific differentiation across the grape genome, using two metrics: CLR (Pavlidis et al., 2013) and F_{ST}. The CLR test identifies potential selection regions by detecting skews in the site frequency spectrum (sfs) within a single taxon, while F_{ST} focuses on regions of high divergence between taxa. All sweep analyses focused on sliding 20kb windows, reflecting the genome-wide pattern of LD decline (Fig. S2). Windows that scored in the top 0.5% were considered candidate sweep regions.

CLR analyses within *vinifera* identified 117 20kb windows encompassing 309 candidate-selected genes (Table S2, Fig. S4). Among these genes, nine functional categories were identified as significantly overrepresented ($P \leq 0.01$), including the "Alcohol dehydrogenase superfamily", "Monoterpenoid indole alkaloid biosynthesis" and "Flower development" (Table S3). We identified genes involved in berry development and/or quality, including the *SWEET1* gene, which encodes a bidirectional sugar transporter (Chong et al., 2014). *SWEET1* is expressed primarily in young and adult leaves (Chong et al., 2014), but it was also overexpressed in full-ripe berries compared to immature berries (adj. P=9.4E-3; Fig. S5), suggesting its involvement in sugar accumulation during berry ripening. *SWEET1* was also identified when we contrasted non-

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

admixed table and wine grapes using F_{ST}. Based on haplotype structure, we hypothesize that the skewed *SWEET1* sfs within *vinifera* is due to elevated divergence between wine and table grapes. Additional genes of interest within *vinifera* included: *i*) a leucoanthocyanidin dioxygenase (LDOX) gene (*VIT_08s0105g00380*) that peaks in expression at the end of veraison (adj. *P*=8.9E-10; Fig. S5) and may be involved in proanthocyanidins accumulation (Bogs et al., 2005, Blanco-Ulate et al., 2015, Savoi et al., 2016); *ii*) genes potentially involved in berry softening, such as two pectinesterase-coding genes and a xyloglucan endotransglucosylase/hydrolase gene that exhibited maximal expression in post-veraison berry pericarps (Fig. S5); and *iii*) flowering time genes, including a *Phytochrome C* homolog.

CLR analyses of the *sylvestris* sample were notable for three reasons. First, the top 0.5% of windows yielded far fewer (88 vs. 309) genes (Table S2). Second, candidate-selected regions within *sylvestris* were distinct from those in *vinifera* (Fig. 3A); none of the putatively selected regions overlapped between taxa. Third, candidate-selected genes were enriched primarily for stress resistance (Table S4), including flavonoid production (*P*=6.27E-3), ethylene-mediated signaling pathways (*P*=8.76E-6), and the stilbenoid biosynthesis pathway (*P*=1.93E-50). Stilbenoids accumulate in response to biotic and abiotic stresses (Bézier et al., 2002, Wang et al., 2010, Amrine et al., 2015).

We also applied F_{ST} to regions of divergent selection between taxa, yielding an additional 929 candidate-selected genes (Tables S2 & S5). One peak of divergence was located from 4.90Mb to 5.33Mb on chromosome 2, which coincides with the sex determination region. F_{ST} analyses detected two peaks in this region; the first was identified previously by QTL analyses (Riaz et al., 2006, Fechter et al., 2012), and the wider region was identified from diversity data (Picq et al., 2014). The two peaks contain 13 and 32 genes, respectively. In the first peak, six genes were overexpressed in female (F) compared to both male (M) and hermaphroditic (H) flowers (adj. $P \leq 0.05$; Fig. S6; Table S6), representing a non-random enrichment of F expression under the peak (binomial; $P<10^{-7}$). One of these genes had been identified as a candidate male sterility gene (*VviFSEX*) (Coito et al., 2017). The second peak included four genes with biased sex expression: one with higher F expression, two with higher H expression and one with higher M expression (Table S6). Altogether, sweep mapping identified numerous candidate genes that may prove useful both for understanding the phenotypic consequences of domestication and for targeted improvement.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Deleterious variants: Domesticated species accumulate more deleterious variants than their progenitors (Marsden et al., 2016, Liu et al., 2017, Ramu et al., 2017). This phenomenon appears to be common for annual crops, but it is not clear whether the unique features of perennial domestication produce a similar effect (Gaut et al., 2015). To examine the potential increase in the number and frequency of deleterious variants at nonsynonymous sites between our *vinifera* and *sylvestris* samples, we predicted deleterious SNPs using SIFT (Ng and Henikoff, 2003). A total of 3,3653 nonsynonymous mutations were predicted to be deleterious in both samples. The number of derived deleterious variants was 5.2% higher, on average, for *vinifera* individuals than *sylvestris* individuals (Fig. 4A), and the ratio of deleterious to synonymous variants was also elevated in *vinifera* (Fig. S7). Most (~77%) deleterious variants were found in a heterozygous state in both samples, but the distribution by state differed between taxa, because deleterious variants were more often homozygous in *sylvestris* ($P < 0.001$, Fig. 4A). Cultivated accessions had a higher proportion of heterozygous deleterious variants ($P = 0.002$, Fig. 4A) and an elevated ratio of deleterious to synonymous variants ($P < 0.001$, Fig. S7).

Cassava, another clonally propagated crop, also had high levels of heterozygous deleterious variants (Ramu et al., 2017). To determine whether clonal propagation can drive this phenomenon, we performed forward simulations under two mating systems: outcrossing and clonal propagation. Each mating system was considered under three demographic models: a constant size population, a long ~30ky population decline similar to that inferred from SMC++ analysis, and a discrete bottleneck (see Methods). As expected (Simons et al., 2014), neither the mating system nor the demographic model had an effect on the accumulation of deleterious variants under an additive model (Fig. 5). Under a recessive model with outcrossing, a population bottleneck led to purging of deleterious variants (Fig. 5), as shown previously (Simons et al., 2014). However, clonal propagation drove the accumulation of deleterious variants under all demographic scenarios, with a bottleneck exacerbating the effect (Fig. 5). We conclude that clonal propagation drives the accumulation of deleterious, recessive alleles in heterozygous regions.

Selective sweeps can also drive the accumulation of deleterious variants, because they drag linked sites to high frequency (Fay and Wu, 2000, Hartfield and Otto, 2011). We examined the distribution of putatively deleterious variants in sweep regions compared to the remainder of the genome (i.e., the 'control'). Sweep regions contained a significantly lower number of

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

deleterious mutations when corrected for length ($P < 0.001$, Fig. 4D), but these variants were also found at significantly higher frequencies ($P < 0.001$, Fig. 4E) and in higher numbers relative to synonymous variants ($P < 0.001$; Fig. 4F), indicative of hitchhiking. All of these trends – including the number of deleterious variants per individual, the distribution by state, and the effects in sweep regions - were qualitatively similar using PROVEAN (Choi et al., 2012) to identify deleterious variants.

DISCUSSION

The Eurasian wild grape (*Vitis vinifera* subsp. *sylvestris*) is a dioecious, perennial, forest vine that was widely distributed in the Near East and the northern Mediterranean prior to its domestication (Zohary and Spiegel-Roy, 1975). The earliest archaeological evidence of wine production suggests that domestication took place in the Southern Caucasus between the Caspian and Black Seas ~6.0-8.0kya (McGovern et al., 1996, 2003). After domestication, the cultivars spread south by 5.0kya to the western side of the Fertile Crescent, the Jordan Valley, and Egypt, and finally reached Western Europe by ~2.8kya (Olmo, 1995, McGovern et al., 2003). Here, however, we are not concerned with the spread of modern grapes, but rather demographic history before and during domestication, the identity of genes that may have played a role in domestication, and the potential effects of domestication and breeding on the accumulation of deleterious variants.

A weak domestication bottleneck: We have gathered genome-wide resequencing data from a sample of table grapes, wine grapes and putatively wild grapes to investigate population structure and demographic history. These analyses lead to our first conclusion, which is that our *sylvestris* sample represents *bona fide* wild grapes, as opposed to feral escapees from domestication. This conclusion is evident from the fact that the *sylvestris* accessions cluster together in population structure analyses (Fig. 1), that they are estimated to have diverged from cultivated grapes ~22 to 30kya (Fig. 2), and that the set of putatively selected genes differ markedly between the *vinifera* and *sylvestris* samples (Fig. 3A). The divergence time between wild and cultivated samples suggests, however, that our *sylvestris* accessions likely do not represent the progenitor population of domesticated grapes.

Analyses of the *vinifera* data suggest that its historical population size has experienced a long decline starting from ~22.0 to ~30.0kya. MSMC analyses suggest that this decline

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

culminated in a weak bottleneck around the estimated time of domestication (Fig. 2A). The potential bottleneck corresponds to the estimated time of grape domestication and the shift from hunter–gatherer to agrarian societies (Purugganan and Fuller, 2009). We note, however, that SMC++ analysis found no evidence for a distinct bottleneck, but instead inferred a consistent N_e decline (Fig. 2C). The question becomes, then, whether the domestication of *vinifera* included a discrete bottleneck. The evidence is mixed. The positive *D* for *vinifera* superficially suggests a population bottleneck, but forward simulations show that positive *D* values also result from a long population decline (Fig. S8). If there was a discrete bottleneck for grapes, we join previous studies in concluding that it was weak (Grassi et al., 2003, Barnaud et al., 2010, Myles et al., 2011), based on two lines of evidence. First, the diversity level in our *vinifera* sample is 94% that of *sylvestris*, representing a far higher cultivated-to-wild ratio than that of maize (83%) (Hufford et al., 2012), *indica* rice (64%) (Liu et al., 2017), soybean (83%) (Lam et al., 2010), cassava (71%) (Ramu et al., 2017) and tomato (54%) (Lin et al., 2014). This high ratio is consistent with a meta-analysis documenting that perennial crops retain 95% of neutral variation from their progenitors, on average, while annuals retain an average of 60% (Miller and Gross, 2011). Second, MSMC analyses suggest a ~2 to 3-fold reduction in N_e at the time of domestication (Fig. 2A). This implies that 33%-50% of the progenitor population was retained during domestication, a percentage that contrasts markedly with the <10% estimated for maize (Wright et al., 2005, Beissinger et al., 2016) and ~2% for rice (Zhu et al., 2007).

A protracted pre-domestication history: The protracted decline in N_e for *vinifera* prompts a question about its cause(s). One possibility is that it reflects natural processes that acted on *vinifera* progenitor populations. For example, climatic shifts may have contributed to the long N_e decline, because the the Last Glacial Maximum (LGM) occurred between 33.0 and 26.5kya (Clark et al., 2009). If the LGM caused *vinifera*'s population decline, one might expect to see population recovery during glacial retraction from 19.0 to 20.0kya. We detect evidence of such recovery in *sylvestris* but not *vinifera* (Fig. 2).

A second possibility is that proto-*vinifera* populations experienced a long period of human-mediated management. The crop domestication literature espouses two views about the speed of domestication. One view, supported primarily by the fossil record, is that domestication is a slow process that takes millennia (Purugganan and Fuller, 2009, 2011, Fuller et al., 2014). A second view argues that domestication occurred much more rapidly, based on genetic evidence

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and population modeling (Gaut, 2015). The gap between these two views has been bridged, in part, by a recent study of African rice (Meyer et al., 2016). The study inferred an N_e decline that started ~15kya and reached its nadir ~4kya, when African rice was thought to have been domesticated. The authors hypothesized that the N_e decline reflects a protracted period of low-intensity management and/or cultivation prior to modern domestication. We propose that a similar period of low-intensity usage by humans may have contributed to the long N_e decline in *vinifera*, especially given the contrasting historical pattern of the *sylvestris* sample (Fig. 3C). It is difficult to prove this proposition, but we do note that humans have inhabited the Southern Caucasus mountains, with some sites bearing evidence of human habitation for > 20k years (Adler and Tushabramishvili, 2004).

A surprising feature of demographic inference is that there is no evidence for a post-domestication expansion of *vinifera* (Fig. 2). This observation contrasts sharply with studies of maize (Beissinger et al., 2016) and African rice (Meyer et al., 2016), both of which > 5-fold had N_e increases following domestication. We hypothesize that the lack of expansion in grapes relates to the dynamics of perennial domestication, specifically clonal propagation and the short time frame (in generations). Data from peach are consistent with our hypothesis, but peach also has extremely low historical levels of N_e (Velasco et al., 2016). Almond, which is another clonally-propagated perennial, exhibits ~2-fold N_e expansion after domestication (Velasco et al., 2016), but it also may have been propagated sexually prior to the discovery of grafting (Zohary and Hopf, 2000). Clearly more work needs to be done to compare demographic and domestication histories across crops with varied histories.

Our demographic inference has caveats. First, our study – along with all previous studies - has likely not measured genetic diversity from the precise progenitor population to *vinifera*. Indeed, such a population may be extinct or at least substantially modified since domestication. Second, our sample size is modest, but it is sufficient to infer broad historical patterns (Schiffels and Durbin, 2014). Consistent with this supposition, the two runs of MSMC with two different samples of n=4 yielded qualitatively identical inferences about the demographic history of *vinifera*. Larger samples will be necessary for investigating more recent population history. Finally, demographic calculations assume a mutation rate and a generation time that may be incorrect, and they also treat all sites equivalently. We note that masking selected regions

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

provide similar inferences (Fig. S3). We also gain confidence from the fact that our observations are consistent with independent estimates about domestication times and glacial events.

Selective sweeps and agronomically important genes: Selective sweep analyses have identified genes and regions that have been previously suspected to mediate agronomic changes. One example is that of the *SWEET1* gene, which is detected within a potential sweep regions identified by *vinifera* CLR analyses and by F_{ST} between table vs. wine grapes (Figs. 3C & S4). The latter comparison needs to be considered with care, because we contrasted four Thompson clones, which exhibited no evidence of admixture, against a similarly non-admixed subset of wine grape accessions (Table S1). Even so, these analyses suggest that at least one difference between wine and table grapes is attributable to the *SWEET1* sugar transporter, that 657 genes differentiate wine and table grapes (Table S2) and that the two groups diverged ~2.5kya (Fig. 2B and Tables S2 & S7).

One major change during domestication was the switch from dioecy to hermaphroditism (This et al., 2006). The sex-determining region resides on chromosome 2, based on QTL analyses that fine-mapped the sex locus between ~4.90 and 5.05 Mbp (Fechter et al., 2012, Picq et al., 2014). The region corresponds to a larger chromosomal segment from 4.75 Mb to 5.39 Mb, based on diversity data and segregation in multiple families (Hyma et al., 2015). With population genomic data, we identify a region from 4.90Mb to 5.33Mb that contains two discrete peaks of divergence (from ~4.90 to 5.05Mb and from ~5.2 to 5.3 Mb; Fig. 4B). We posit that the two peaks are meaningful, because a shift to hermaphroditism is thought to require two closely-linked loci: one that causes loss of M function and another that houses a dominant F sterility mutation (Charlesworth et al., 2005, Charlesworth, 2013). The first peak contains six genes overexpressed in F flowers, including *VviFSEX*, which may abort stamen development (Coito et al., 2017). We hypothesize that the second peak houses a dominant F sterility factor. The leading candidates for this function are four genes that are differentially expressed among sexes (Table S2), but none of the four are annotated with an obvious function in sex determination (Ramos et al., 2017) (Table S6).

The cost of domestication in a clonally propagated perennial: It has been unclear whether perennial crops demonstrate an increased burden of slightly deleterious mutations (Gaut et al., 2015), especially given that most have experienced only moderate bottlenecks (Miller and Gross, 2011). We find, however, that there is a cost associated with grape domestication: each

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

vinifera accession contains 5.2% more putatively deleterious SNPs, on average, than the wild individuals in our sample. This difference exceeds that observed in dogs (2.6%) (Marsden et al., 2016) and rice (~3-4%) (Liu et al., 2017) but pales in comparison to cassava (26%), a clonally propagated annual (Ramu et al., 2017).

We have shown that clonal propagation leads to the accumulation of deleterious recessive mutations (Fig. 5 & S9), but it does not explain why cassava has had a more dramatic increase in deleterious variants (i.e., 26% vs. 5.2%) as a consequence domestication. There are at least two reasons. The first is that cassava underwent a strong domestication bottleneck (Ramu et al., 2017). Forward simulations indicate that the deleterious burden is exacerbated with a discrete bottleneck compared to a protracted population decline (Fig. 5). Second, mutation rates vary between annuals and perennials (Gaut et al., 2011), so that cassava may harbor more post-domestication mutations that contribute to cost. Altogether, the high level of deleterious, recessive, heterozygous mutations in cultivated grape provides a genomic explanation to a well-known feature of grape-breeding, which is severe inbreeding depression (Kole, 2011).

MATERIALS AND METHODS

For full materials and methods, please see SI Appendix, Supplementary Text. We collected leaf tissue for 13 individuals from 11 *vinifera* cultivars, nine *sylvestris* accessions and one accession of *V. rotundifolia* (subgenus *Muscadinia*) from the USDA grape germplasm collections in Davis, California (Table S1). DNA was extracted from leaf samples, Illumina paired-end sequencing libraries were constructed (TrueSeq), and libraries were sequenced as 150-bp paired reads. Illumina raw reads for five other cultivars were gathered from the Short Read Archive (SRA) at NCBI (Table S1).

Reads were trimmed, filtered and mapped to the PN40024 reference (12X) (Jaillon et al., 2007). Local realignment was performed around indels; reads were filtered for PCR duplicates; and sites with extremely low or high coverage were removed. For population structure analyses, we used ANGSD (Korneliussen et al., 2014) to generate a BEAGLE file for the variable subset of the genome, and then applied NGSadmix (Skotte et al., 2013). To measure genome-wide genetic diversity and other population parameters, we estimated a genome-wide sfs from genotype likelihoods (Korneliussen et al., 2014). PCA analyses were also based on the sfs.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Functional regions were based on the *V. vinifera* genome annotation in Ensembl (v34). Nonsynonymous SNPs were predicted to be deleterious based on a SIFT score □≤□0.05 (Kumar et al., 2009). The V. *rotundifolia* outgroup allele was submitted to prediction programs to avoid reference bias (Kono et al., 2016, Liu et al., 2017). The number of deleterious or synonymous alleles per individual or region was calculated as 2 × the number of homozygous variants + heterozygous variants (Henn et al., 2016).

We employed MSMC 2.0 to estimate N_e over time (Li and Durbin, 2011, Schiffels and Durbin, 2014), based on SNPs called in GATK v3.5 (DePristo et al., 2011)(see SI Appendix, Supplementary Text). Segregating sites within each sample were phased and imputed using Shapeit (Delaneau et al., 2013) based on a genetic map (Hyma et al., 2015). Demographic history was also inferred with SMC++, which analyzes multiple genotypes without phasing (Terhorst et al., 2017). SweeD (Pavlidis et al., 2013) was used to detect selective sweeps. F_{ST} values were averaged within 20 kbp non-overlapping windows using ANGSD (Korneliussen et al., 2014).

Functional categories were assigned to genes using VitisNet functional annotations (Grimplet et al., 2009). We tested functional category enrichment using Fisher's Exact Test, with $P \leq 0.01$ considered as significant. Gene expression data used SRA data for berry (SRP049306) and flower (SRP041212) samples. Reads were trimmed for qualitynd then mapped onto the PN40024 transcriptome (v.V1 from http://genomes.cribi.unipd.it/grape/) using Bowtie2 (Langmead and Salzberg, 2012). DESeq2 (Love et al., 2014) was used to normalize read counts and to test for differential expression.

Forward in time simulations were carried out using fwdpy11 (Thornton, 2014). Five hundred replicate simulations were run for each demographic and mating scheme model. The population decline model was based on SMC++ results and rescaled for computational performance. Three demographic models were simulated: constant population size, a linear population decline, and a discrete bottleneck. Two mating schemes were simulated: strict outcrossing for the whole simulation, and outcrossing with clonal propagation for the final 100 generations. Additional details are available from SI Appendix, Supplementary Text.

ACKNOWLEDGEMENTS

We thank R. Gaut and R. Figueroa-Balderas for generating the data and sampling and also D. Seymour, Q. Liu, K. Roessler and E. Solares for comments. YZ is supported by the International

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Postdoctoral Exchange Fellowship Program; JS is supported by the NSF-GRFP; BSG is supported by the Borchard Foundation. DC is supported by J. Lohr Vineyards and Wines and by E. & J. Gallo Winery.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FIGURE LEGENDS

Fig. 1 Population structure of cultivated and wild samples. (A) PCA plot based on genetic covariance among all individuals of wild samples (red) and *cultivars* (green for wine grapes and blue for table grapes). (B) Neighbor-Joining (NJ) tree across all samples, rooted by *Vitis rotundifolia.*

Fig. 2 Intra- and intertaxon analyses of demographic history and divergence. (A) MSMC estimates of the effective population size (N_e) *vinifera* based on two separate runs of four individuals. The solid line represents wine grapes and the dashed line is table grapes. (B) MSMC analysis of cross-coalescence based on comparisons between *sylvestris* and either wine or table grapes. The Y–axis is relative cross-Coalescent rate, a parameter related to migration rates. (C) Divergence time and past N_e changes inferred by the SMC ++ analyses, based on unphased genotypes. For all analyses, the time scale on the *x* -axis was calculated assuming a neutral mutation rate per generation (μ) = 2.5×10^{-8} and a generation time of 3 years.

Fig. 3 The CLR statistic computed for 20 Kbp windows along chromosomes separately for wild samples and cultivars. (A) The scatterplot shows the CLR statistic for corresponding windows for both wild (X-axis) and cultivated samples (Y-axis). The dashed line represents the 99.5% cutoff, and red dots represent outlier regions. There are no shared significant regions between the two populations. (B) F_{ST} analyses between *vinifera* and *sylvestris* identify two peaks in the sex determination region that encompass 45 annotated genes.

Fig. 4 The number and frequency of derived deleterious alleles in cultivars and wild samples. A) Comparisons between *vinifera* and *sylvestris* for the number of deleterious variants per individual overall (left), as homozygotes (middle) and as heterozygotes (right). B) Comparisons between sweep regions and the rest of the genome (control) for the number (left), population frequency (middle) and ratio (right) of the number deleterious to synonymous variants per individual.

Fig. 5 Forward simulations over time under three demographic scenarios and two mating systems. The top and bottom graphs depict additive and recessive modes, respectively. Each line represents the average number of deleterious alleles per accession. Abbreviations for

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

demographic models are: Con = constant population size; Dec = declining population size; Bot = bottleneck. Abbreviations for mating schemes are: Out = outcrossing; Clo = clonal propagation.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

REFERENCES

Adler DS, Tushabramishvili N. 2004. Middle Palaeolithic patterns of settlement and subsistence in the southern Caucasus. Middle Palaeolithic settlement dynamics. 91-32.

Amrine KC, Blanco-Ulate B, Riaz S, Pap D, Jones L, Figueroa-Balderas R, Walker MA, Cantu D. 2015. Comparative transcriptomics of Central Asian Vitis vinifera accessions reveals distinct defense strategies against powdery mildew. Hortic Res. 2:15037.

Barnaud A, Laucou V, This P, Lacombe T, Doligez A. 2010. Linkage disequilibrium in wild French grapevine, Vitis vinifera L. subsp. silvestris. Heredity (Edinb). 104:431-437.

Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J. 2016. Recent demography drives changes in linked selection across the maize genome. Nat Plants. 2:16084.

Bézier A, Lambert B, Baillieul F. 2002. Cloning of a grapevine Botrytis-responsive gene that has homology to the tobacco hypersensitivity-related hsr203J. J Exp Bot. 53:2279-2280.

Blanco-Ulate B, Amrine KC, Collins TS et al. 2015. Developmental and Metabolic Plasticity of White-Skinned Grape Berries in Response to Botrytis cinerea during Noble Rot. Plant Physiol. 169:2422-2443.

Bogs J, Downey MO, Harvey JS, Ashton AR, Tanner GJ, Robinson SP. 2005. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. Plant Physiol. 139:652-663.

Bowers J, Boursiquot JM, This P, Chu K, Johansson H, Meredith C. 1999. Historical Genetics: The Parentage of Chardonnay, Gamay, and Other Wine Grapes of Northeastern France. Science. 285:1562-1565.

Cardone MF, D'Addabbo P, Alkan C et al. 2016. Inter-varietal structural variation in grapevine genomes. Plant J. 10.1111/tpj.13274

Charlesworth D. 2013. Plant sex chromosome evolution. J Exp Bot. 64:405-420.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity (Edinb). 95:118-128.

Charlesworth D, Willis JH. 2009. The genetics of inbreeding depression. Nat Rev Genet. 10:783-796.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 7:e46688.

Chong J, Piron MC, Meyer S, Merdinoglu D, Bertsch C, Mestre P. 2014. The SWEET family of sugar transporters in grapevine: VvSWEET4 is involved in the interaction with Botrytis cinerea. J Exp Bot. 65:6589-6601.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The Last Glacial Maximum. Science. 325:710-714.

Coito JL, Ramos MJ, Cunha J, Silva HG, Amâncio S, Costa MM, Rocheta M. 2017. VviAPRT3 and VviFSEX: Two Genes Involved in Sex Specification Able to Distinguish Different Flower Types in Vitis. Front Plant Sci. 8:98.

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 10:5-6.

DePristo MA, Banks E, Poplin R et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43:491-498.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Di Genova A, Almeida AM, Muñoz-Espinoza C, Vizoso P, Travisany D, Moraga C, Pinto M, Hinrichsen P, Orellana A, Maass A. 2014. Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. BMC Plant Biol. 14:7.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. Genetics. 155:1405-1413.

Fechter I, Hausmann L, Daum M, Sörensen TR, Viehöver P, Weisshaar B, Töpfer R. 2012. Candidate genes within a 143 kb region of the flower sex locus in Vitis. Mol Genet Genomics. 287:247-259.

Fuller DQ, Denham T, Arroyo-Kalin M, Lucas L, Stevens CJ, Qin L, Allaby RG, Purugganan MD. 2014. Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. Proc Natl Acad Sci U S A. 111:6147-6152.

Gaut B, Yang L, Takuno S, Eguiarte LE. 2011. The Patterns and Causes of Variation in Plant Nucleotide Substitution Rates. Annu Rev Ecol Evol S. 42:245-266.

Gaut BS. 2015. Evolution Is an Experiment: Assessing Parallelism in Crop Domestication and Experimental Evolution: (Nei Lecture, SMBE 2014, Puerto Rico). Mol Biol Evol. 32:1661-1671.

Gaut BS, Díez CM, Morrell PL. 2015. Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. Trends Genet. 31:709-719.

Goff SA, Ricke D, Lan TH et al. 2002. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. 296:92-100.

Grassi F, Labra M, Imazio S, Spada A, Sgorbati S, Scienza A, Sala F. 2003. Evidence of a secondary grapevine domestication centre detected by SSR analysis. Theor Appl Genet. 107:1315-1320.

Grimplet J, Cramer GR, Dickerson JA, Mathiason K, Van Hemert J, Fennell AY. 2009. VitisNet: “Omics” integration through grapevine molecular networks. PLoS One. 4:e8365.

Hartfield M, Otto SP. 2011. Recombination and hitchhiking of deleterious alleles. Evolution. 65:2421-2434.

Henn BM, Botigue LR, Bustamante CD, Clark AG, Gravel S. 2015. Estimating the mutation load in human genomes. Nat Rev Genet. 16:333-343.

Henn BM, Botigue LR, Peischl S et al. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci U S A. 113:E440-9.

Hufford MB, Xu X, van Heerwaarden J et al. 2012. Comparative population genomics of maize domestication and improvement. Nat Genet. 44:808-811.

Hyma KE, Barba P, Wang M, Londo JP, Acharya CB, Mitchell SE, Sun Q, Reisch B, Cadle-Davidson L. 2015. Heterozygous Mapping Strategy (HetMappS) for High Resolution Genotyping-By-Sequencing Markers: A Case Study in Grapevine. PLoS One. 10:e0134880.

Jaillon O, Aury JM, Noel B et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 449:463-467.

Kikkert JR, Striem MJ, Vidal JR, Wallace PG, Barnard J, Reisch BI. 2005. Long-term study of somatic embryogenesis from anthers and ovaries of 12 grapevine (Vitis sp.) genotypes. In Vitro Cellular & Developmental Biology-Plant. 41:232-239.

Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol Biol Evol. 17:1483-98.

Kole C. 2011. Genetics, Genomics, and Breeding of Grapes. Science Publishers. 160-185.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Kono TJ, Fu F, Mohammadi M, Hoffman PJ, Liu C, Stupar RM, Smith KP, Tiffin P, Fay JC, Morrell PL. 2016. The Role of Deleterious Substitutions in Crop Genomes. Mol Biol Evol.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15:356.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4:1073-1081.

Lam HM, Xu X, Liu X et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet. 42:1053-1059.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357-359.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature. 475:493-496.

Lin T, Zhu G, Zhang J et al. 2014. Genomic analyses provide insights into the history of tomato breeding. Nat Genet. 46:1220-1226.

Liu Q, Zhou Y, Morrell PL, Gaut BS. 2017. Deleterious Variants in Asian Rice and the Potential Cost of Domestication. Mol Biol Evol. 34:908-924.

Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 29:139-146.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Lu J, Tang T, Tang H, Huang J, Shi S, Wu CI. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet. 22:126-131.

Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vila C, Marques-Bonet T, Schnabel RD, Wayne RK, Lohmueller KE. 2016. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci U S A. 113:152-157.

McGovern PE, Fleming SJ, Katz SH. 2003. The origins and ancient history of wine: food and nutrition in history and antropology. Routledge.

McGovern PE, Glusker DL, Exner LJ, Voigt MM. 1996. Neolithic resinated wine. Nature. 381:481-481.

Meyer RS, Choi JY, Sanches M et al. 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. Nat Genet. 48:1083-1088.

Miller AJ, Gross BL. 2011. From forest to field: perennial fruit crop domestication. Am J Bot. 98:1389-1414.

Morrell PL, Buckler ES, Ross-Ibarra J. 2011. Crop genomics: advances and applications. Nat Rev Genet. 13:85-96.

Myles S, Boyko AR, Owens CL et al. 2011. Genetic structure and domestication history of the grape. Proc Natl Acad Sci U S A. 108:3530-3535.

Initiative NGAW. 2007. The Impact of Wine, Grapes and Grape Products on the American Economy: Family Businesses Building Value. http://wwwngwiorg/economic-impact-study_226html.

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 31:3812-3814.

Olmo H. 1995. *Evolution of Crop Plants*: Grapes
. New York: Longman.

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Pavlidis P, Zivkovic D, Stamatakis A, Alachiotis N. 2013. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol Biol Evol. 30:2224-2234.

Picq S, Santoni S, Lacombe T et al. 2014. A small XY chromosomal region explains sex determination in wild dioecious V. vinifera and the reversal to hermaphroditism in domesticated grapevines. BMC Plant Biol. 14:229.

Purugganan MD, Fuller DQ. 2009. The nature of selection during plant domestication. Nature. 457:843-848.

Purugganan MD, Fuller DQ. 2011. Archaeological data reveal slow rates of evolution during plant domestication. Evolution. 65:171-183.

Ramos MJ, Coito JL, Fino J, Cunha J, Silva H, de Almeida PG, Costa MM, Amâncio S, Paulo OS, Rocheta M. 2017. Deep analysis of wild Vitis flower transcriptome reveals unexplored genome regions associated with sex specification. Plant Mol Biol. 93:151-170.

Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Lu F. 2017. Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nat Genet.

Renaut S, Rieseberg LH. 2015. The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops. Mol Biol Evol.

Riaz S, Krivanek AF, Xu K, Walker MA. 2006. Refined mapping of the Pierce's disease resistance locus, PdR1, and Sex on an extended genetic map of Vitis rupestris x V. arizonica. Theor Appl Genet. 113:1317-1329.

Savoi S, Wong DC, Arapitsas P, Miculan M, Bucchetti B, Peterlunger E, Fait A, Mattivi F, Castellarin SD. 2016. Transcriptome and metabolite profiling reveals that prolonged drought modulates the phenylpropanoid and terpenoid pathway in white grapes (Vitis vinifera L.). BMC Plant Biol. 16:67.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 46:919-925.

Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nat Genet. 46:220-224.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. Genetics. 195:693-702.

Terhorst J, Kamm JA, Song YS. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 49:303-309.

This P, Lacombe T, Cadle-Davidson M, Owens CL. 2007. Wine grape (Vitis vinifera L.) color associates with allelic variation in the domestication gene VvmybA1. Theor Appl Genet. 114:723-730.

This P, Lacombe T, Thomas MR. 2006. Historical origins and genetic diversity of wine grapes. Trends Genet. 22:511-519.

Thornton KR. 2014. A C++ template library for efficient forward-time population genetic simulation of large populations. Genetics. 198:157-166.

Tuskan GA, Difazio S, Jansson S et al. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science. 313:1596-1604.

Velasco D, Hough J, Aradhya M, Ross-Ibarra J. 2016. Evolutionary Genomics of Peach and Almond Domestication. G3 (Bethesda). 6:3985-3993.

Wang Q, Li P, Hanania U, Sahar N, Mawassi M, Gafny R, Sela I, Tanne E, Perl A. 2005. Improvement of Agrobacterium-mediated transformation efficiency and transgenic plant

bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

regeneration of Vitis vinifera L. by optimizing selection regimes and utilizing cryopreserved cell suspensions. Plant Science. 168:565-571.

Wang W, Tang K, Yang HR, Wen PF, Zhang P, Wang HL, Huang WD. 2010. Distribution of resveratrol and stilbene synthase in young grape plants (Vitis vinifera L. cv. Cabernet Sauvignon) and the effect of UV-C on its accumulation. Plant Physiol Biochem. 48:142-152.

Wang X, Xu Y, Zhang S et al. 2017. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 49:765-772.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310-1314.

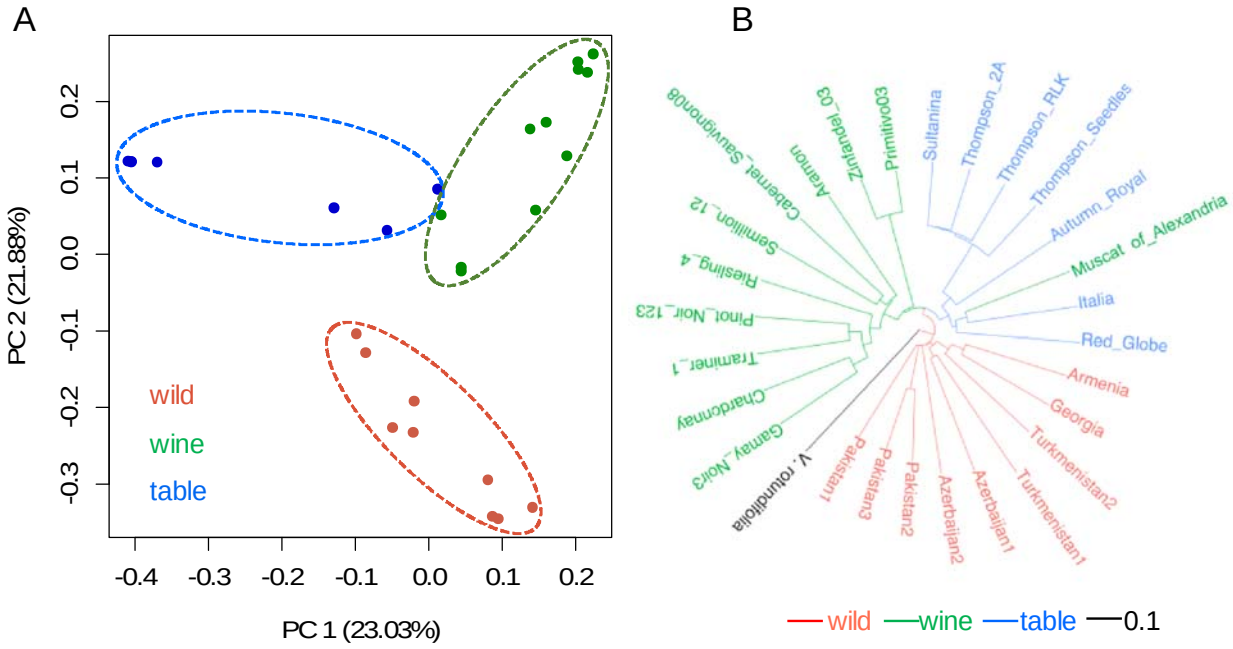
Xu Y, Gao Z, Tao J, Jiang W, Zhang S, Wang Q, Qu S. 2016. Genome-Wide Detection of SNP and SV Variations to Reveal Early Ripening-Related Genes in Grape. PLoS One. 11:e0147749.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of Oryza sativa and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol. 24:875-888.

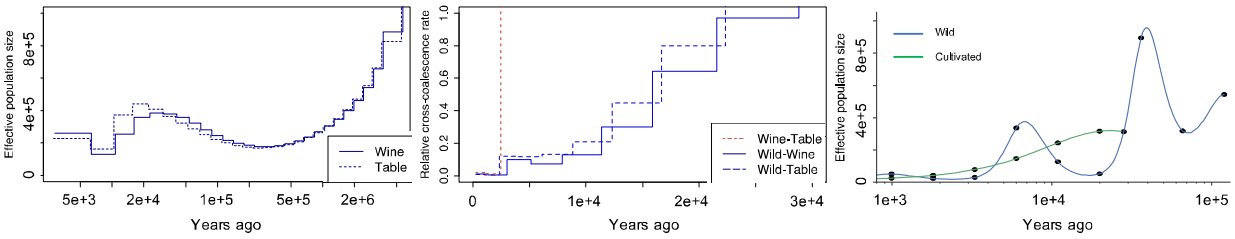
Zohary D, Spiegel-Roy P. 1975. Beginnings of Fruit Growing in the Old World. Science. 187:319-327.

Zohary D, Hopf M. 2000. Domestication of plants in the Old World. Oxford: Oxford University Press.

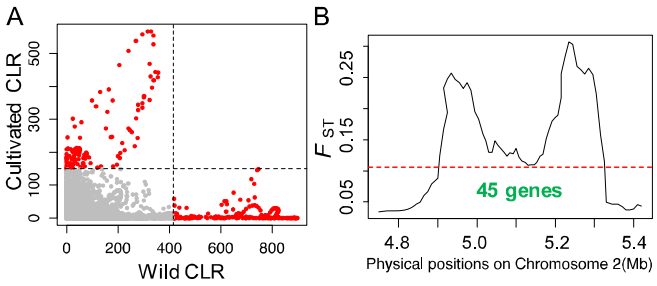
bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



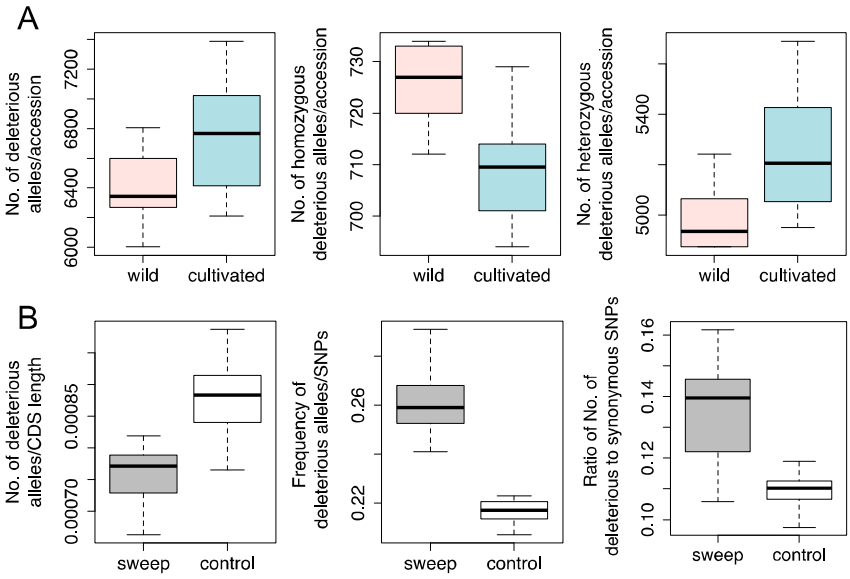
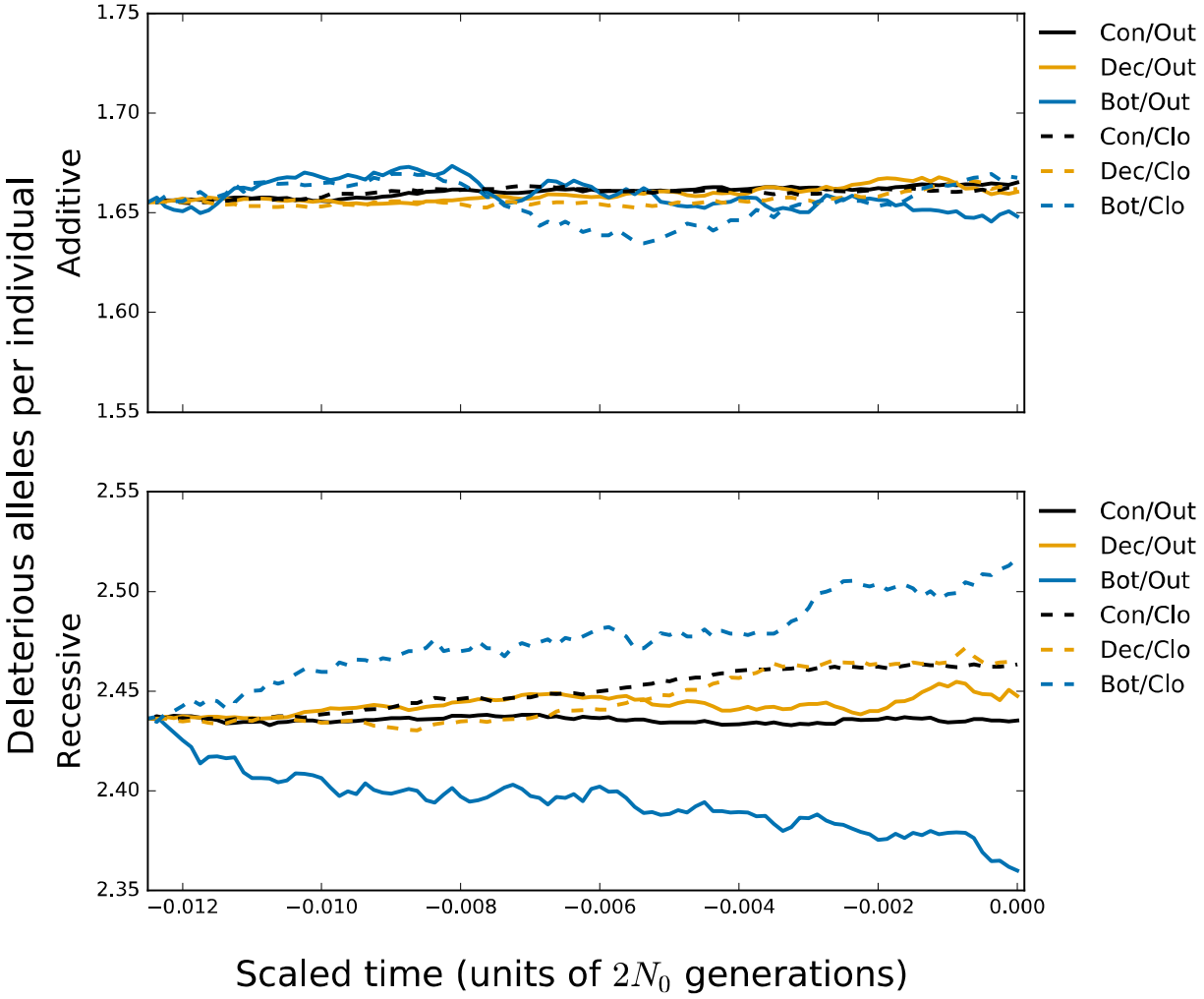
bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/146373; this version posted June 5, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.