Title
Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication

Permalink
https://escholarship.org/uc/item/7wb6x2zc

Journal
Proceedings of the National Academy of Sciences of the United States of America, 114(44)

ISSN
0027-8424

Authors
Zhou, Y
Massonnet, M
Sanjak, JS
et al.

Publication Date
2017-10-31

DOI
10.1073/pnas.1709257114

License
CC BY 4.0

Peer reviewed
Evolutionary genomics of grape (Vitis vinifera ssp. vinifera) domestication

Yongfeng Zhou, Mélanie Massonnet, Jaleal S. Sanjak, Dario Cantù, and Brandon S. Gaut

Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, CA 92697; and Department of Viticulture and Enology, University of California, Davis, CA 95616

Edited by Andrew G. Clark, Cornell University, Ithaca, NY, and approved September 25, 2017 (received for review June 5, 2017)

We generated genomic data from grapes (Vitis vinifera ssp. vinifera), a clonally propagated perennial crop, to address three ongoing mysteries about plant domestication. The first is the duration of domestication; archaeological evidence suggests that domestication occurs over millennia, but genetic evidence indicates that it can occur rapidly. We estimated that our wild and cultivated grape samples diverged ~22,000 years ago and that the cultivated lineage experienced a steady decline in population size (N_e) thereafter. The long decline may reflect low-intensity management by humans before domestication. The second mystery is the identification of genes that contribute to domestication phenotypes. In cultivated grapes, we identified candidate-selected genes that function in sugar metabolism, flower development, and stress responses. In contrast, candidate-selected genes in the wild sample were limited to abiotic and biotic stress responses. A genomic region of high divergence corresponded to the sex determination region and included a candidate male sterility factor and additional genes with sex-specific expression. The third mystery concerns the cost of domestication. Annual crops accumulate putatively deleterious variants, in part due to strong domestication bottlenecks. The domestication of perennial crops differs from that of annuals in several ways, including the intensity of bottlenecks, and it is not yet clear if they accumulate deleterious variants. We found that grape accessions contained 5.2% more deleterious variants than wild individuals, and these were more often in a heterozygous state. Using forward simulations, we confirm that clonal propagation leads to the accumulation of recessive deleterious mutations but without decreasing fitness.

The study of crop domestication has long been used as a proxy for studying evolutionary processes, such as the genetic effects of bottlenecks (1) and the detection of selection to identify agronomically important loci (2–4). Several crops have been studied in this evolutionary context (5), but there are at least two emerging issues. The first is the speed at which domestication occurs. One view, supported primarily by archaeological evidence, is that domestication is a slow process that takes millennia (6–8). Another view, based on genetic evidence and population modeling (9, 10), argues that domestication occurs much more rapidly. The gap between these two views has been bridged, in part, by a recent study of African rice. The study used population genomic data to infer that a bottleneck occurred during domestication ~3.5 kya and also that the bottleneck was preceded by a long, ~14,000-y decline in the effective population size (N_e) of the progenitor population (11). The authors hypothesized that the protracted N_e decline reflects a period of low-intensity management and/or cultivation before modern domestication. While an intriguing hypothesis, it is not yet clear whether other crops also have demographic histories marked by protracted N_e declines.

The second emerging issue is the “cost of domestication” (12), which refers to an increased genetic load within cultivars. This cost originates partly from the fact that the decreased N_e during a domestication bottleneck reduces the efficacy of genome-wide selection (13), which may in turn increase the frequency and number of slightly deleterious variants (14, 15). The characterization of deleterious variants is important because they may be fitting targets for crop improvement (16). Consistent with a cost of domestication, annual crops are known to contain an increase in derived, putatively deleterious variants relative to their wild progenitors (17–20). However, it is not yet clear whether these deleterious variants increase genetic load and whether this phenomenon applies to perennial crops.

The distinction between annual and perennial crops is crucial because perennial domestication is expected to differ from annual domestication in at least three aspects (21, 22). The first is clonal propagation; many perennials are propagated clonally but most annuals are not. Clonal propagation maintains genetic diversity in desirable combinations but also limits opportunities for sexual recombination (20, 22). The second aspect is time. Long-lived perennials have extended juvenile stages. As a result, the number of sexual generations is much reduced for perennials relative to annual crops, even for perennials that were domesticate early in human agricultural history. The third aspect is the severity of the domestication bottleneck. A meta-analysis has documented that perennial crops retain 95% of neutral variation from their progenitors, on average, while annuals retain an average of 60% (22). This observation suggests that many (and perhaps most) perennial crops have not experienced severe domestication bottlenecks; as a consequence, their domestication may not come with a cost.

Here we study the domestication history of the grapevine (Vitis vinifera ssp. vinifera), which is the most economically important of grapes, the most economically important horticultural crop in the world. Domesticated grapes experienced a protracted, 22,000-y population decline prior to domestication; we hypothesize that this decline reflects low-intensity cultivation by humans prior to domestication. Domestication altered the mating system of grapes. The sex determination region is detectable as a region of heightened genetic divergence between wild and cultivated accessions. Based on gene expression analyses, we propose candidate genes that alter sex determination. Finally, grapes contain more deleterious mutations in heterozygous states than do their wild ancestors. The accumulation of deleterious mutations is due in part to clonal propagation, which shelters deleterious recessive mutations.

Significance

We generated genomic data to estimate the population history of grapes, the most economically important horticultural crop in the world. Domesticated grapes experienced a protracted, 22,000-y population decline prior to domestication; we hypothesize that this decline reflects low-intensity cultivation by humans prior to domestication. Domestication altered the mating system of grapes. The sex determination region is detectable as a region of heightened genetic divergence between wild and cultivated accessions. Based on gene expression analyses, we propose candidate genes that alter sex determination. Finally, grapes contain more deleterious mutations in heterozygous states than do their wild ancestors. The accumulation of deleterious mutations is due in part to clonal propagation, which shelters deleterious recessive mutations.

Author contributions: Y.Z., D.C., and B.S.G. designed research; Y.Z. and J.S.S. performed research; Y.Z. and B.S.G. wrote the paper.

The authors declare no conflict of interest.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1709257114/-/DCSupplemental.

This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The sequence reported in this paper has been deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) database (accession no. PRJNA388292).

1To whom correspondence should be addressed. Email: bgaut@uci.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1709257114/-/DCSupplemental.
horticultural crop in the world (23). Grapes (hereafter vinifera) have been a source of food and wine since their hypothesized domestication ∼8.0 kya from their wild progenitor, V. vinifera ssp. sylvestris (hereafter sylvestris) (24). The exact location of domestication remains uncertain, but most lines of evidence point to a primary domestication event in the Near East (23, 24). 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 (25). 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 were mapped to an ∼150-kb region on chromosome 2 (26, 27).

Historically, genetic diversity among V. vinifera varieties has been studied with simple sequence repeats (28). More recently, a group genotyped 950 vinifera and 59 sylvestris accessions with a chip containing 9,000 SNPs (23). Their data suggest that grape domestication led to a mild reduction of genetic diversity, indicating that grape is a reasonable perennial model for studying the accumulation of deleterious variation in the absence of a pronounced bottleneck. Still more recent studies have used whole-genome sequencing (WGS) to assess structural variation among grape varieties (29–31). Surprisingly, however, WGS data have not been used to investigate the population genomics of grapes. Here we perform WGS on a sample of vinifera cultivars and on 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, and do they provide insights into the agro- nomic shifts associated with domestication? Finally, do domesticated grapes have more derived, putatively deleterious variants relative to sylvestris, or have the unique features of perennial domestication permitted an escape from this potential 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 (23), 18 vinifera individuals representing 14 cultivars, and one outgroup (Vitis rotundifolia) (SI Appendix, Table S1). Our sylvestris accessions are a subset of the wild sample from ref. 23, 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 used to investigate the population genomics of grapes. Here PNA0024 (32), resulting in the identification of 3,963,172 and 3,732,107 SNPs across the sylvestris and vinifera samples (Materials and Methods).

To investigate population structure, we applied principal component analysis (PCA) to genotype likelihoods (33). 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 while also identifying admixed individuals (SI Appendix, 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 (SI Appendix, Table S1). Both samples harbored substantial levels of nucleotide diversity across all sites (sylvestris: πs = 0.0147 ± 0.0011; vinifera: πv = 0.0139 ± 0.0014; SI Appendix, Fig. S2). Although π was higher in sylvestris (πs/πv = 0.94 ± 0.14), vinifera had higher levels of heterozygosity and Tajima’s D values (vinifera, D = 0.5421 ± 0.0932; sylvestris, D = −0.4651 ± 0.1577; SI Appendix, Fig. S2). Linkage disequilibrium (LD) decayed to r2 < 0.2 within 20 kb in both samples, but it declined more slowly for vinifera after ∼20 kb (SI Appendix, Fig. S2).

We inferred the demographic history of the vinifera sample using MSMC, a method that infers both population size and gene flow using phased SNPs (34). Assuming a generation time of 3 y (24) and a mutation rate of 2.5 × 10−9 mutations per nucleotide per year (35), we converted scaled population parameters into years and individuals (N). Based on these analyses, vinifera experienced a continual reduction of N after ∼2.0 kya until its nadir from ∼7.0 kya to 11.0 kya (Fig. 2A), which corresponds to the time of domestication and implies a mild domestication bottleneck. Notably, there was no evidence for a dramatic expansion of N after domestication. MSMC results were similar across two separate analyses (Fig. 2A), based on n = 4 samples of either table or wine grapes (SI Appendix, 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 ∼22 kya (Fig. 2B), which corresponds to the onset of the decline of vinifera N after domestication; and (iv) an ∼2.6-kya divergence of wine and table grapes (Fig. 2C and SI Appendix, Fig. S3). We also used SMC++ to infer the demographic history of our sylvestris sample, revealing a complex N pattern that corresponds to features of climatic history (Discussion).

Sweep Mapping. We investigated patterns of selection and interspecific differentiation across the grape genome. All sweep analyses focused on sliding 20-kb windows, reflecting the genome-wide pattern of LD decline (SI Appendix, Fig. S2). Windows that scored in the top 0.5% were considered candidate sweep regions.

We began with CLR (37), which identifies potentially selected regions by detecting skews in the site frequency spectrum (sfs) within a single taxon, and XP-CLR (38), which detects sfs skews relative to a reference taxon (sylvestris). Within vinifera, CLR identified 117 20-kb windows encompassing 309 candidate-selected genes (SI Appendix, Table S2). Among those detected by CLR, nine functional categories were identified as significantly
overrepresented ($P \leq 0.01$), including the “alcohol dehydrogenase superfamily,” “monoterpenoid indole alkaloid biosynthesis,” and “flower development” (SI Appendix, Table S3). XP-CLR identified a similar number of genes (367); both tests identified genes involved in berry development and/or quality, including the SWEET1 gene (SI Appendix, Fig. S4), which encodes a bidirectional sugar transporter (39). SWEET1 was overexpressed in full-ripe berries compared with immature berries [adjusted (adj.) $P = 9.4E-3$; SI Appendix, Fig. S6], suggesting an involvement in sugar accumulation during berry ripening. Additional genes of interest detected by both tests included: (i) a leucoanthocyanidin dioxygenase ($LDox$) gene (VIT_08s0105g00380) that peaks in expression at the end of veraison (adj. $P = 8.9E-10$; SI Appendix, Fig. S6) and may be involved in proanthocyanidin accumulation (40–42); (ii) genes potentially involved in berry softening, such as two pectinesterase-coding genes and a xyloglucan endotransglycosylase/hydrolase gene that exhibited maximal expression in postveraison berry pericarps (SI Appendix, Fig. S6); and (iii) flowering-time genes, including a Phloctype B homolog.

As a comparison, we applied CLR analyses to the sylvestris sample, which were notable for three reasons. First, the top $0.5\%$ of windows yielded far fewer (88 vs. 309) genes (SI Appendix, 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 (SI Appendix, 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 and may be involved in proanthocyanidin accumulation (40–42); (ii) genes potentially involved in berry softening, such as two pectinesterase-coding genes and a xyloglucan endotransglycosylase/hydrolase gene that exhibited maximal expression in postveraison berry pericarps (SI Appendix, Fig. S6); and (iii) flowering-time genes, including a Phloctype B homolog.

As a comparison, we applied CLR analyses to the sylvestris sample, which were notable for three reasons. First, the top $0.5\%$ of windows yielded far fewer (88 vs. 309) genes (SI Appendix, 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 (SI Appendix, 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 (43).

We also detected regions of high divergence between wild and cultivated samples using $F_{ST}$ and $D_{xy}$, which identified 929 and 546 candidate-selected genes (SI Appendix, Tables S2 and S6). A prominent region of divergence was identified by both methods from $\sim 4.90$ Mb to $\sim 5.33$ Mb on chromosome 2 (Fig. 3B and SI Appendix, Fig. S5), which coincides with the sex determination region (44). With both methods, the region contained two peaks of divergence. In $F_{ST}$ analyses, the two peaks contain 13 and 32 genes, respectively. In the first peak, six genes were overexpressed in female (F) compared with both male (M) and hermaphroditic (H) flowers (adj. $P \leq 0.05$; SI Appendix, Fig. S7 and Table S6), representing a nonrandom enrichment of F expression under the peak (binomial; $P < 10^{-7}$). One of these genes had been identified as a candidate male sterility gene (VitiFSEX) (45). 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 (SI Appendix, Table S6).

**Deleterious Variants.** Domesticated annual crops accumulate more deleterious variants than their progenitors (17, 20, 46). To examine the potential increase in the number and frequency of deleterious variants at nonsynonymous sites between vinifera and sylvestris samples, we predicted deleterious SNPs using SIFT (47). A total of 33,653 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 for sylvestris individuals (Fig. 4), and the ratio of deleterious to synonymous variants was also elevated in vinifera (SI Appendix, Fig. S8). Most ($\sim 77\%$) 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. 4). Cultivated accessions had a higher proportion of heterozygous deleterious variants ($P = 0.002$, Fig. 4) and an elevated ratio of deleterious to synonymous variants ($P < 0.001$, SI Appendix, Fig. S8).

We also examined the distribution of putatively deleterious variants for vinifera in sweep regions compared with the remainder of the genome (i.e., the “control”). Sweep regions contained a significantly lower number of deleterious mutations when corrected for length ($P < 0.001$, Fig. 4), but these variants were also found at significantly higher frequencies ($P < 0.001$, Fig. 4) and in higher numbers relative to synonymous variants ($P < 0.001$; Fig. 4). 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 (48) to identify deleterious variants (SI Appendix, Fig. S9).

Like grapes, cassava is clonally propagated, and it also has high levels of heterozygous deleterious variants (20). To determine whether clonal propagation can contribute to the accumulation of deleterious variants, we performed forward simulations under two mating systems: outcrossing and clonal propagation that began at the time of domestication ($\sim 8$ kya). Each mating system was considered under three demographic models: a constant size population, a long ~30,000-y population decline similar to that inferred from SMC++ analyses, and a discrete bottleneck (Materials and Methods and SI Appendix, Fig. S10). Under an additive model without back mutation, the discrete bottleneck increased the number of deleterious

---

**Fig. 2.** Intra- and intertaxon analyses of demographic history and divergence. (A) MSMC estimates of the effective population size ($N_e$) of 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 ($y$ axis) based on comparisons between sylvestris and either wine or table grapes. (C) Divergence time and past $N_e$ changes inferred by the SMC++ analyses, based on unphased genotypes.
alleles under both mating systems but with little effect on load (SI Appendix, Fig. S12). Under a recessive model, an outcrossing, bottlenecked population purged deleterious variants (49, 50) (Fig. 5), and clonal propagation increased the number of deleterious variants under all demographic scenarios (Fig. 5). Despite the increase in deleterious variants, clonal propagation decreased load under the recessive model (Fig. 5) because clonality hides deleterious, recessive 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 before its domestication (51). The earliest archaeological evidence of wine production suggests that domestication took place in the Southern Caucasus between the Caspian and Black Seas ∼6.0–8.0 kya (24, 52). After domestication, the cultivars spread south by 5.0 kya to the western side of the Fertile Crescent, the Jordan Valley, and Egypt and finally reached Western Europe by ∼2.8 kya (24, 53). Here, however, we are not concerned with the spread of modern grapes, but rather with 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 Protracted Predomestication History? 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 escapes. 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 kya to 30 kya (Fig. 2), and that the set of putatively selected genes differs 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 vinifera data suggest that its historical Ne has experienced a long decline starting from ∼22.0 kya to ∼30.0 kya. MSMC analyses indicate that this decline 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 (6). We note, however, that SMC++ analysis found no evidence for a distinct bottleneck, but instead inferred a consistent Ne decline (Fig. 2C). The question becomes, then, whether the domestication of vinifera included a discrete bottleneck. The evidence is mixed. The positive Tajima’s D for vinifera superficially suggests a population bottleneck, but forward simulations show that positive D values also result from a long population decline (SI Appendix, Fig. S11). If there was a discrete bottleneck for grapes, we join previous studies in concluding that it was weak (23, 54, 55), 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%) (4), indica rice (64%) (17), soybean (83%) (56), cassava (71%) (20), and tomato (54%) (57). Second, MSMC analyses suggest an approximately two- to threefold reduction in Ne 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 (3, 58) and ∼2% for rice (59).

The protracted decline in Ne 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 Ne decline because the Last Glacial Maximum (LGM) occurred between 33.0 and 26.5 kya (60). If the LGM caused vinifera’s population decline, one might expect to see population recovery during glacial retraction from 19.0 kya to 20.0 kya. We detect evidence of recovery in sylvestris but not in vinifera (Fig. 2C). A second possibility is that the domesticated germplasm is derived from a single deme of a larger metapopulation because population structure can produce a signal of apparent Ne decline (61). Finally, it is possible that proto-vinifera populations experienced a long period of human-mediated management, as suggested in the study of African rice (11). It is difficult to prove this proposition, but three factors are consistent with this possibility: (i) the contrasting historical pattern of the wild sample, (ii) the fact that some sites in the Southern Caucasus mountains have evidence of human habitation for >20,000 y (62), and (iii) a growing consensus that humans altered ecosystems long before the onset of agriculture (63).

A surprising feature of demographic inference is the lack of evidence for a postdomestication expansion of vinifera (Fig. 2). This
observation contrasts sharply with studies of maize (58) and African rice (11), both of which had greater than fivefold $N_e$ 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 $N_e$ (64). Almond, which is another clonally propagated perennial, exhibits an approximately twofold $N_e$ expansion after domestication (64), but it also may have been propagated sexually before the discovery of grafting (65). Clearly more work needs to be done to compare demographic histories across crops with varied demographic and life histories.

Our demographic inferences have 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 (34). 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 and may provide further insights into the potential for population expansion after domestication. Finally, demographic calculations assume a mutation rate and a generation time that may be incorrect, and they also treat all sites equivalently. Note, however, that masking selected regions provides similar inferences (SI Appendix, Fig. S3) and also that our observations are consistent with independent estimates about domestication times and glacial events.

Selective Sweeps and Agronomically Important Genes. Selective sweep analyses identified genes and regions that have been previously suspected to mediate agronomic change. One example is that of the SWEET1 gene, which is within a potential vinifera sweep region. The same gene is also within a region of differentiation between nonadmixed table and wine grapes (SI Appendix, Fig. S4). Based on haplotype structures, we hypothesize that at least one difference between wine and table grapes is attributable to the SWEET1 sugar transporter.

A major change during grape domestication was the switch from dioecy to hermaphrodism (66). The sex-determining region resides on chromosome 2, based on quantitative trait locus analyses that fine-mapped the sex locus between $4.90$ and $5.05$ Mbps (26, 27). The region corresponds to a larger chromosomal segment from $4.75$ Mb to $5.39$ Mb based on Genotype-by-Sequencing data and on segregation patterns from multiple families (44). With WGS data, we have identified a similar region that contains two discrete divergence peaks, from $4.90$ to $5.05$ Mb and from $5.2$ Mb to $5.3$ Mb (Fig. 2B and SI Appendix, Fig. S5). We posit that the two peaks are meaningful because the evolution of dioecy requires two closely linked loci: one that causes loss of $M$ function and another that houses a dominant $F$ sterility mutation (67, 68). The first peak contains six genes overexpressed in F flowers, including VviFSX, which may abort stamen development (45). We predict that the second peak houses a dominant $F$ sterility factor. The leading candidates are four genes that are differentially expressed among sexes (SI Appendix, Table S2), but none of the four are annotated with an obvious function in sex determination (69) (SI Appendix, Table S6).

Putatively Deleterious Mutations in a Clonally Propagated Perennial. Like grapes, most perennials have experienced moderate bottleneck (22), raising the question as to whether they typically have an increased burden of slightly deleterious mutations (21). We find that each vinifera accession contains 5.2% more putatively deleterious SNPs, on average, than the wild individuals in our sample. This difference exceeds that observed for dogs (2.6%) (46) and rice (~3.5%, 17) but pales in comparison with cassava (28%), a clonally propagated annual (20). Our simulations show that clonal propagation can lead to the accumulation of deleterious recessive mutations and a reduction of load under a recessive model (Fig. 5). We do not know the dominance of variants in grapes, but we predict that most heterozygous, putatively deleterious mutations are recessive and hence do not contribute to increased load or to a cost associated with domestication. These same mutations, however, do provide a genomic explanation of a well-known feature of grape breeding: severe inbreeding depression (70).

Materials and Methods

For full materials and methods, see SI Appendix, Supplementary Text. We collected leaf tissue for 13 individuals from 11 vinifera cultivars, 9 sylvestris accessions, and 1 accession of V. rotundifolia (SI Appendix, 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 the National Center for Biotechnology Information (SI Appendix, Table S1).

We employed MSMC 2.0 to estimate $N_e$ over time (34, 74), based on SNPs (64). Almond, which is a model species, was simulated (78). DESeq2 (79) was used to identify regions with differential expression that fine-mapped the sex locus between $4.75$ Mb to $5.39$ Mb (Fig. 3, SI Appendix, Figs. S5 and Table S5) using Bowtie2 (78). Functional regions were based on the V. vinifera genome annotation in Ensmb (v34). Nonsynonymous SNPs were predicted to be deleterious based on a SIFT score of $0.05$ (72). The V. rotundifolia outgroup allele was submitted to prediction programs to avoid reference bias (17, 18). The number of deleterious or synonymous alleles per individual or region was calculated as $2 \times$ the number of homozygous variants + heterozygous variants (73).

We employed MSMC 2.0 to estimate $N_e$ over time (34, 74), based on SNPs called in GATK v3.5 (75) (SI Appendix, Supplementary Text). Segregating sites within each sample were phased and imputed using Shapeit (76) based on a genetic map (44). Demographic history was also inferred with SMC++, which analyzes multiple genotypes without phasing (36). SweeD (37) and XP-CLR (38) were used to detect selective sweeps. $D_S$ and $D_R$ values were averaged within 20 kbp nonoverlapping windows using ANGSD (33).

We employed MSMC 2.0 to estimate $N_e$ over time (34, 74), based on SNPs called in GATK v3.5 (75) (SI Appendix, Supplementary Text). Segregating sites within each sample were phased and imputed using Shapeit (76) based on a genetic map (44). Demographic history was also inferred with SMC++, which analyzes multiple genotypes without phasing (36). SweeD (37) and XP-CLR (38) were used to detect selective sweeps. $D_S$ and $D_R$ values were averaged within 20 kbp nonoverlapping windows using ANGSD (33).

We employed MSMC 2.0 to estimate $N_e$ over time (34, 74), based on SNPs called in GATK v3.5 (75) (SI Appendix, Supplementary Text). Segregating sites within each sample were phased and imputed using Shapeit (76) based on a genetic map (44). Demographic history was also inferred with SMC++, which analyzes multiple genotypes without phasing (36). SweeD (37) and XP-CLR (38) were used to detect selective sweeps. $D_S$ and $D_R$ values were averaged within 20 kbp nonoverlapping windows using ANGSD (33).

Functional regions were assigned to genes using VitisNet functional annotations (77). We tested functional category enrichment using Fisher’s exact test with $P \leq 0.01$ as significant. Gene expression data used SRA data for berry (SRP049306) and flower (SRP041212) samples. Reads were trimmed for quality and mapped onto the PN40024 transcriptome (v1.1 from genomes.criibi.unipd.it/grape) using Bowtie2 (78). DESeq2 (79) was used to normalize read counts and to perform differential expression analysis.

Forward-in-time simulations were carried out using fdpYd11 (80). 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 at the onset of domestication. Additional details are available in SI Appendix, Supplementary Text.

ACKNOWLEDGMENTS. We thank R. Gaut and R. Figueroa-Balderas for generating the data and sampling; two anonymous reviewers, D. Seymour, Q. Liu, K. Rosessler, and E. Solares provided comments. Y.Z. is supported by the International Postdoctoral Exchange Fellowship Program, J.S. is supported by J. Lohr Vineyards and Wines, E. & J. Gallo Winery, and the Louis P. Martini Endowment.