Genetic structure and domestication history of the grape

Sean Myles, Adam R. Boyko, Christopher L. Owens, Patrick J. Brown, Fabrizio Grassi, Mallikarjuna K. Aradhya, Bernard Prins, Andy Reynolds, Jer-Ming Chia, Doreen Ware, Carlos D. Bustamante, and Edward S. Buckler

The grape is one of the earliest domesticated fruit crops and, since antiquity, it has been widely cultivated and prized for its fruit and wine. Here, we characterize genome-wide patterns of genetic variation in over 1,000 samples of the domesticated grape, Vitis vinifera subsp. vinifera, and its wild relative, V. vinifera subsp. sylvestris from the US Department of Agriculture grape germplasm collection. We find support for a Near East origin of vinifera and present evidence of introgression from local sylvestris as the grape moved into Europe. High levels of genetic diversity and rapid linkage disequilibrium (LD) decay have been maintained in vinifera, which is consistent with a weak domestication bottleneck followed by thousands of years of widespread vegetative propagation. The considerable genetic diversity within vinifera, however, is contained within a complex network of close pedigree relationships that has been generated by crosses among elite cultivars. We show that first-degree relationships are rare between wine and table grapes and among grapes from geographically distant regions. Our results suggest that although substantial genetic diversity has been maintained in the grape subsequent to domestication, there has been a limited exploration of this diversity. We propose that the adoption of vegetative propagation was a double-edged sword: Although it provided a benefit by ensuring true breeding cultivars, it also discouraged the generation of unique cultivars through crosses. The grape currently faces severe pathogen pressures, and the long-term sustainability of the grape and wine industries will rely on the exploitation of the grape’s tremendous natural genetic diversity.

The grape is the most valuable horticultural crop in the world. The fruit from the world’s ~8 million ha of vineyard is mostly processed into wine, but some is destined for fresh consumption as table grapes, dried into raisins, processed into nontoxicolohic juice, and distilled into spirits (http://faostat.fao.org). The archeological record suggests that cultivation of the domesticated grape, Vitis vinifera subsp. vinifera, began 6,000–8,000 y ago in the Near East from its wild progenitor, Vitis vinifera subsp. sylvestris (1). The thousands of grape cultivars in use today have been generated since then by vegetative propagation and by crosses. Wine and table grapes currently receive intense chemical applications to combat severe pathogen pressures. This susceptibility to disease, however, is not attributable to a lack of genetic diversity. Vinifera harbors levels of genetic variation an order of magnitude greater than humans and is comparable in diversity to maize (2, 3), with polymorphism that dates back tens of millions of years (4). Thus, an environmentally sustainable grape-growing industry will rely on accessing and using the grape’s tremendous genetic diversity to develop improved disease-resistant grape cultivars through marker-assisted breeding (5). Traditionally, grape breeding programs have sought genotype-phenotype associations using linkage mapping. Because of the grape’s long generation time (generally 3 y), however, establishing and maintaining linkage-mapping populations is time-consuming and expensive. Thus, genome-wide association (GWA) (6) and genomic selection (GS) (7) are attractive alternatives to traditional linkage mapping in the grape and other long-lived perennial fruit crops.

Well-powered GWA and GS require a genome-wide assessment of genetic diversity, patterns of population structure, and the decay of linkage disequilibrium (LD). To this end, we recently discovered over 70,000 high-quality SNPs in the grape using next-generation DNA sequencing (4). From this SNP set, we developed and validated a 9,000-SNP genotyping array (the Vitis9kSNP array). Here, we present an analysis of genotype data from 950 vinifera and 59 sylvestris accessions using the Vitis9kSNP array as part of an effort to characterize an entire US Department of Agriculture (USDA) germplasm collection on a genome-wide scale. We provide a refined model of the domestication and breeding history of vinifera by evaluating levels of haplotype diversity, the decay of LD, and patterns of population structure in vinifera and its progenitor, sylvestris. In addition, our analyses reveal extensive clonal relationships among cultivars and a complex pedigree structure within vinifera that are the result of widespread vegetative propagation. We suggest that the last several thousand years of grape breeding explored only a small fraction of possible genetic combinations and that future marker-assisted breeding efforts therefore have tremendous diversity at their disposal to produce desirable wine and table grapes with resistance to existing and future pathogens.

Results

Pedigree Analysis Within vinifera. We used the Vitis9kSNP array (4) to generate 5,387 SNP genotypes from 950 vinifera accessions (451 table grape accessions, 469 wine grape accessions, and 30 accessions of unknown type) from the grape germplasm collection of the USDA, one of the most comprehensive repositories of grape diversity in the world. Currently, there are over 10,000 grape cultivar names in use worldwide (8), and their classification is often confusing because of homonyms, synonyms, scarce

Author contributions: S.M., D.W., C.D.B., and E.S.B. designed research; S.M. and J.-M.C. performed research; F.G., M.K.A., and B.P. contributed new reagents/analytic tools; S.M., A.R.B., C.L.O., P.J.B., and A.R. analyzed data; and S.M., C.D.B., and E.S.B. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

†To whom correspondence should be addressed. E-mail: smm367@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1009363108/-/DCSupplemental.
or incorrect historical information, and curation error. Some cultivars have been differentiated into several through the vegetative propagation of somatic mutants (9), and we expect clones derived from the same cultivar to be genetically identical at the tested marker loci. We find that 551 (58%) of the 950 *vinifera* accessions are clones of at least 1 other accession in the USDA grape germplasm collection. Many clonal relationships are restricted to pairs of accessions, but groups of up to 17 accessions were found to be clonally related (Fig. 1). We determined that there are 583 unique *vinifera* cultivars in the USDA grape germplasm collection: 399 accessions with no clonal relationships and an additional 184 accessions composed of one accession from each set of clones.

The grape is a vegetatively propagated outcrossing perennial species, which means that old cultivars propagated for hundreds or even thousands of years may coexist with cultivars generated from recent crosses. The potential for selfing and for crosses across interleaved generations, including crosses between related cultivars, makes accurate genealogical reconstruction from genomic data intractable. Nevertheless, we used patterns of identity-by-descent (IBD) and predictions of population genetic theory to infer simple pedigree relationships among the 583 unique *vinifera* cultivars that remained after excluding clonal relationships (Materials and Methods). We found that 74.8% of the cultivars are related to at least one other cultivar by a first-degree relationship (Fig. 24). The resulting complex pedigree structure of our sample can be visualized as a set of networks (Fig. 2B).

The pedigree structure of *vinifera* is characterized primarily by first-degree relationships between grapes of the same type: 89.3% of edges in the network connect table grapes to table grapes (Fig. 2B). A similar trend was found for geography; only 6.1% of connections are between eastern and western cultivars (SI Appendix, Fig. S1 and Table S1). These two categories of connections in the network occur far more often than expected by chance (binomial test, \( P < 1 \times 10^{-15} \)).

We infer that about half (47.6%) of the first-degree relationships in our sample are likely parent-offspring. The other half (52.4%) we refer to as “sibling or equivalent,” because complex crossing schemes can generate IBD values that are indistinguishable with our data from sibling relationships (SI Appendix, Fig. S1 and Table S2). By evaluating Mendelian inconsistencies, we assigned two parents to a cultivar wherever possible and thereby resolved 83 trios (SI Appendix, Fig. S1 and Table S3). A network of some well-known cultivars that includes several trios is shown in Fig. 3. Table S5 contains a list of inferred clonal and first-degree relationships for each cultivar. The assignment of clonal and pedigree relationships from the present study will be verified in the field by germplasm curators and used to improve the accuracy of the USDA grape germplasm collection.

### Haplotype Diversity and LD Decay

To evaluate the effects of domestication on levels of genetic diversity, haplotype diversity was measured in nonoverlapping sliding windows of varying sizes across the grape genome. Although we observe a statistically significant reduction in haplotype diversity in *vinifera* compared with *sylvestris* \( (P < 1 \times 10^{-12}) \) for eight or more SNP haplotypes, the observed reduction is relatively minor (SI Appendix, Fig. S2).

Moreover, the decay of LD is very rapid in *vinifera* and appears unchanged between the wild ancestor and the domesticated grape (SI Appendix, Fig. S2). The identification of *sylvestris* is notoriously difficult because of the morphological similarity between *sylvestris* and *vinifera* and the ease with which they cross (10, 11). All studies that make use of *sylvestris* samples should therefore be interpreted with caution. Details of the *sylvestris* samples used in the present study and the impact of potential misidentification are provided in the SI Appendix.

### Grape Domestication History

Although the reduction of diversity attributable to domestication and breeding appears to be weak on a genome-wide scale \( (i.e., \text{much weaker than in the tomato} \ (12) \ \text{and likely even weaker than in highly diverse maize} \ (13)) \), a few notable changes in morphology have emerged since grape domestication, including perfect flowers, larger berry sizes, higher sugar content, and a wide range of berry colors (14). To identify genomic regions potentially responsible for these domestication traits, we scanned the genome for signatures of selection and identified a 5-Mb candidate domestication locus on chromosome 17 (SI Appendix, Fig. S3). Although the long-range LD generated by strong selection over a few generations prevents gene-level dissection of such loci, extended LD may also be exploited to identify genotype-phenotype associations. To test this, we performed a GWA study for berry color and identified a 5-Mb region on chromosome 2 that encompasses a group of MYB transcription factor genes known to be the major determinants of grape color (15, 16) \( (P = 4.8 \times 10^{-12}) \), SI Appendix, Fig. S4). We also observe a strong signal of positive selection for white grapes around this locus, consistent with intense breeding for lighter berry color and

---

**Fig. 1.** Clonal relationships within the USDA grape germplasm collection. (A) Number of clonal relationships was evaluated for each of the 950 *vinifera* accessions. Most of the accessions \( (551 \ (58\%) \ of \ 950 \) accessions) have a clonal relationship with at least 1 other accession. (B) Degree of clonal relatedness among all 950 *vinifera* accessions is represented as a set of clusters. The 399 accessions that do not have a clonal relationship with another accession are shown as lone black dots. Accessions with six or fewer clonal relationships are grouped together with their clones and shown in gray. Clusters of clones with \( \geq 7 \) accessions are colored, and their names are indicated in the legend. Names listed in the legend are the prime names from the Vitis International Variety Catalogue (http://www.vivc.de).

**Fig. 2.** First-degree relationships within the USDA grape germplasm collection. (A) Number of first-degree relationships was evaluated for each of the 583 unique *vinifera* cultivars. A total of 74.8% of the unique cultivars are related to at least 1 other cultivar by a first-degree relationship, and some cultivars have many first-degree relationships \( (i.e., \geq 10; \ SI \ Appendix, \ Table \ S4) \). (B) Pedigree structure of *vinifera* is represented as a set of networks. Edges in the network represent inferred first-degree relationships. The vertices, or dots, represent grape cultivars and are colored by grape type (legend). The sample size of each grape type is shown in parentheses. Lone dots represent cultivars with no first-degree relatives in the dataset. Note that one single interconnected network is clearly visible and includes 384 (58.3%) of the 583 unique cultivars that are interconnected by a series of first-degree relationships.
the rapid spread of the MYB mutations responsible for reduced pigmentation (17) (SI Appendix, Fig. S4).

Relatedness among our geographically diverse sample of vinifera and sylvestris provides strong support for an origin of vinifera in the Near East: All vinifera populations are genetically closer to eastern sylvestris than to western sylvestris (Table 1; SI Appendix, Table S1). After domestication, grape growing and winemaking expanded westward, reaching Western Europe by 2.800 y ago (1). We find that haplotype diversity in western vinifera is slightly reduced compared with eastern vinifera (SI Appendix, Fig. S5), suggesting that the grape experienced a modest reduction in genetic diversity as it was brought to Western Europe.

Based on morphological and genetic evidence, it has been suggested that Western European vinifera cultivars experienced introgression from local Western European sylvestris. Our finding that western vinifera are more closely related to western sylvestris than are other vinifera populations is consistent with gene flow between wild and cultivated grapes in Western Europe (Table 1). To examine this in more detail, we used principal components analysis (PCA) to visualize relationships among individual accessions. Fig. 4 shows the first two principal components (PCs) calculated from sylvestris accessions only, with vinifera cultivars subsequently projected onto the axes. Whereas PC2 differentiates a subset of geographically isolated sylvestris accessions (a subpopulation from southern Spain and two samples from Georgia), PC1 reflects a clear west-east gradient in sylvestris that is recapitulated in the vinifera that have been projected onto PC space. The observation that relationships among vinifera mirror patterns of relatedness in its wild progenitor supports a scenario in which Western European cultivars experienced introgression from local wild sylvestris grapes. Alternatively, the western sylvestris may have experienced gene flow from western vinifera. To

![Fig. 3. Network of first-degree relationships among common grape cultivars. Solid vertices represent likely parent-offspring relationships. Dotted vertices represent sibling relationships or equivalent. Arrows point from parents to offspring for inferred trios (details are provided in Materials and Methods).]

<table>
<thead>
<tr>
<th>Table 1. Population pairwise Fst estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sylvestris west</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Sylvestris west</td>
</tr>
<tr>
<td>Sylvestris east</td>
</tr>
<tr>
<td>Vinifera west</td>
</tr>
<tr>
<td>Vinifera central</td>
</tr>
<tr>
<td>Vinifera east</td>
</tr>
</tbody>
</table>

Geographic regions are defined as follows: “east” includes locations east of Istanbul, Turkey; “west” includes locations west of Slovenia, including Austria; and “central” refers to locations between them (details are presented in SI Appendix, Table 1).
are a mixture of using 36 microsatellites, however (26). This results are in agreement with previous studies showing no reduction in genetic diversity in vinifera compared with its wild ancestor (22, 23, 25, 26) and with the relatively minor changes in morphology observed between sylvestris and vinifera (11). A recent study found a significant increase in LD in vinifera compared with sylvestris using 36 microsatellites, however (26). This result is reconcilable with our findings because the use of microsatellites, which evolve more rapidly than SNPs, will amplify the signal of the weak domestication bottleneck that we detect (SI Appendix, Fig. S2). Many of the SNPs assayed by the Vitis9kSNP array have likely been segregating for millions of years (4), and our measure of LD decay thus captures recombination events deeper into the grape’s evolutionary history than LD measured from microsatellites. Although we are unable to provide precise estimates of the severity of the grape domestication bottleneck with the current dataset, it is evident that LD, as measured by $r^2$ between SNPs, decays rapidly in vinifera, far more quickly than in humans (27), Arabidopsis (28), rice (29), and maize (30, 31). For this reason, we conclude that well-powered GWA studies in the grape will require whole-genome sequencing, as in other high-diversity plants with rapid LD decay (9). Nevertheless, for traits that were selected during domestication and/or breeding, the resulting increase in LD means that a relatively low marker density will often suffice to identify diffuse association peaks using GWA. We demonstrate this by mapping the locus responsible for lighter berry pigmentation, a trait that experienced strong artificial selection, using GWA (SI Appendix, Fig. S4).

Our weak bottleneck model for grape domestication also fits well with the recorded widespread use of vegetative propagation during the history of vinifera: Many cultivars in use today may be only a small number of generations removed from the wild progenitor (21, 24). Vegetative propagation immortalizes a cultivar by allowing numerous genetically identical copies to be produced, but it also enables clones with unique traits to be generated by propagating tissue from the mother plant, which carries somatic mutations leading to unique phenotypes (9). For example, Pinot has been extensively propagated into clones with diverse phenotypes, such as lighter berry color (e.g., Pinot Blanc, Pinot Gris) and pigmented pulp (Pinot Teinturier). We find that Pinot has the most clonal diversity in the USDA grape germplasm collection, with 17 accessions that are clonally related (Fig. S4, B). Although we observe misclassification in our dataset (SI Appendix, SI Text), we mostly verify well-known clonal relationships among cultivars like Haneepoot and Muscat of Alexandria; Sultanina, Kishmish, and Thompson Seedless; and Sauvignon Blanc and Sauvignon Gris. Thus, our data suggest that a considerable proportion of the morphological diversity maintained by the USDA grape germplasm collection is the result of spontaneous somatic mutations captured through vegetative propagation rather than that of segregating polymorphism. Identifying the causal genetic variants underlying phenotypic variation among clones will be a challenging task for which deep resequencing will be required. As is the case for grape berry color (16, 32), the locus underlying a particular phenotypic difference between clones may be the same locus involved in segregating variation for that phenotype across cultivars. Thus, deep resequencing of the clones we have identified here can be used to complement future genetic mapping efforts in the grape.

Although the adoption of vegetative propagation contributed to the maintenance of high levels of genetic diversity in vinifera and enabled the generation of clones with unique traits, we suggest that it also reduced grape growers’ motivation to undertake extensive crossing and to breed new cultivars. In support of this scenario, we find that 75% of the vinifera cultivars in the USDA grape germplasm collection are related to at least one observed in sylvestris and the decay of LD appears unchanged between vinifera and sylvestris (SI Appendix, Fig. S2).
other cultivar by a first-degree relationship (Fig. 2). In fact, 384 (58.3%) of the 583 cultivars form a single-pedigree network: Over half of the cultivars in the USDA grape germplasm collection are interconnected by a series of first-degree relationships (Fig. 2B). With more extensive sampling of *vinifera*, lone vertices and disjoint networks from Fig. 2B are likely to become increasingly interconnected. Moreover, the pairwise IBD distribution (SI Appendix, Fig. S1) suggests that there are a vast number of higher order pedigrees relationships (e.g., second and third degree) within the present set of *vinifera* cultivars. Thus, the genetic structure of *vinifera* can be largely understood as one large complex pedigree. We propose that this pedigree structure is the result of a limited number of crosses made among elite cultivars that were immortalized and sometimes vegetatively propagated for centuries.

Most cultivars have only one or two first-degree relatives in our sample, but a small number of cultivars are highly connected in the network, and thus likely represent ancient cultivars widely used during grape breeding (SI Appendix, Table S4). This is consistent with the previous finding that Pinot and Gouais Blanc are the parents of 16 common French cultivars (33). Most of the highly connected cultivars are table grapes, including Muscat of Alexandria and the world’s preeminent raisin grape, Sultanina (Thompson’s Seedless). Table grapes show more first-degree relationships than wine grapes (Mann–Whitney U test, \(P = 3.72 \times 10^{-8}\)), and the high degree of connectivity among table grapes suggests that there was more intense breeding in table grapes than in wine grapes. This result follows from the ease with which table grapes are evaluated compared with wine grapes. The most highly connected wine grape in the present sample is Traminer, which has 20 first-degree relatives and is believed to be an ancient cultivar widely used during the history of grape breeding (34). Fig. 3 provides a more detailed view of relationships among some well-known cultivars and includes several inferred trios. Particularly noteworthy is our discovery that Chenin Blanc and Sauvignon Blanc are likely siblings and both share a parent-offspring relationship with Traminer. Also, we find that two of the most common cultivars of the Rhône Valley in France, Viognier and Syrah, are likely siblings. Not only do we find that elite cultivars were often reused in different crosses but that most first-degree relationships are between grapes of the same type (i.e., wine grapes from the same geographic region, i.e., east, west). Together, these observations suggest that grape breeding has been restricted to a relatively small number of cultivars that have a small number of the possible genetic combinations within *vinifera* have been explored.

We propose that the adoption and widespread use of vegetative propagation has been a double-edged sword during grape breeding. Although the production of fine wine would be impossible without the control over genetic variability that vegetative propagation offers, vegetative propagation has also discouraged the breeding of new cultivars and is at least partially responsible for a worldwide grape industry dominated by cultivars sharing extensive coancestry. Other factors that have contributed to the small number of cultivars in use today include the devastation of European vineyards in the second half of the 19th century by mildews and phylloxera and the development of the global wine industry. Currently, grapes face intense pathogen pressures, and are thus intensely chemically treated. There are numerous examples of sources of resistance to these pathogens, both from wild *Vitis* species and from *vinifera* cultivars that are often found in marginal areas of cultivation and remain largely unexploited (5, 35). The grape is clearly exceptional in terms of its domestication and breeding history compared with most crops studied to date. The *vinifera* grape has retained high levels of genetic diversity since its domestication ~7,000 y ago, yet its genetic variation remains relatively unshuffled within an extended pedigree. Developing an environmentally sustainable wine and grape industry will rely on tapping into this tremendous diversity by genetically characterizing the world’s germplasm collections and using marker-assisted breeding approaches to generate improved cultivars.

### Materials and Methods

#### Sample Collection and Genotype Calling

Leaf tissue was collected from the USDA grape germplasm collections in Davis, California and Geneva, New York, and DNA was extracted using standard protocols. Some sylvestris DNA samples were provided by F.G. Genotype data were generated from the custom Illumina Vitis9KSNP array, which assays 8,898 SNPs (4). After quality filters, 5,387 SNPs genotyped in 950 *vinifera* and 59 sylvestris samples remained for analysis (SI Appendix, SI Materials and Methods).

#### Pedigree Construction

We calculated IBD for all pairwise comparisons among the 950 *vinifera* accessions using PLINK (36). We considered pairs of accessions to be genetically identical (i.e., clones, sports) if they had an IBD >0.95. For inferences of first-degree relationships based on IBD, we reduced the samples to a set of “unique cultivars,” which included 399 accessions without clonal relationships and an additional 184 accessions composed of one randomly chosen accession from each of the clones. We refer to this set of samples as 583 unique *vinifera* cultivars. Our power to estimate IBD reliably is reduced by our relatively small number of SNPs and the ascertainment bias introduced during our SNP discovery procedure. We therefore used known pedigree relationships to calibrate our IBD values. From 43 confirmed parent-offspring relationships (SI Appendix, Table S2), the lowest pairwise IBD value was 0.466; therefore, we considered all pairwise relationships in our data to be likely first-degree relatives if they had IBD values \(>0.466\). (SI Appendix, Fig. S1). To differentiate between parent-offspring and other pedigree relationships, we used the 20 values and 21 values observed from our confirmed parent-offspring pairs as thresholds for defining other parent-offspring pairs in the data (SI Appendix, Fig. S1 and Table S2).

For each cultivar that was related to at least 2 other cultivars with IBD, 20 and 21 scores consistent with a parent-offspring relationship, we calculated the number of SNPs consistent with Mendelian inheritance (3) and used the number of SNPs consistent with Mendelian inheritance (Table S3). Adjacency matrices and network images were generated using the “network” package in R (37).

#### Population Structure Analyses

Samples with >20% missing data were removed, and we ensured that no two samples had an Identity-By-State >0.95, which resulted in 570 *vinifera* cultivars and 59 sylvestris accessions included in the analysis. Average population-pairwise Fs\(t\) estimates were calculated from all 5,387 SNPs weighted by allele frequency (equation 10 in (38)). For PCA, SNPs with >20% missing data and minor allele frequency (MAF) <0.05 were excluded. We then pruned the SNPs for LD using PLINK (36) by considering a window of 10 SNPs, removing 1 of a pair of SNPs if LD >0.5, and then shifting the window by 3 SNPs repeating the procedure. After these filters, 2,958 SNPs remained. PCA was performed using SMARTPCA (39).

#### 3-Population Test for Mixture

We used the 3-population test for mixture (18) to compare a focal population X with two putatively parental populations Y and W to determine whether X, Y, and W are related in a simple tree or not. For PCA, SNPs with >20% missing data and minor allele frequency (MAF) <0.05 were excluded. We then pruned the SNPs for LD using PLINK (36) by considering a window of 10 SNPs, removing 1 of a pair of SNPs if LD >0.5, and then shifting the window by 3 SNPs repeating the procedure. After these filters, 2,958 SNPs remained. PCA was performed using SMARTPCA (39).

#### Population Structure Analyses

Samples with >20% missing data were removed, and we ensured that no two samples had an Identity-By-State >0.95, which resulted in 570 *vinifera* cultivars and 59 sylvestris accessions included in the analysis. Average population-pairwise Fs\(t\) estimates were calculated from all 5,387 SNPs weighted by allele frequency (equation 10 in (38)). For PCA, SNPs with >20% missing data and minor allele frequency (MAF) <0.05 were excluded. We then pruned the SNPs for LD using PLINK (36) by considering a window of 10 SNPs, removing 1 of a pair of SNPs if LD >0.5, and then shifting the window by 3 SNPs repeating the procedure. After these filters, 2,958 SNPs remained. PCA was performed using SMARTPCA (39).
VINIFERA cultivars and 59 sylvestris accessions with the following fastPHASE options: -T10 -C0 -K10 -KU20 -H -Ki1 -Ks5 -Kp.05. We chose K = 10 as values above this threshold provided little improvement in performance. This gives an imputation error rate of 0.0638, which is comparable to values reported when phasing SNP data in humans (40). We phased each of the 19 chromosomes using fastPHASE with the K = 10 option and the default settings for the remaining options. The analysis included 3,397 SNPs from the 19 assembled chromosomes. Haplotype diversity was calculated as (n(n−1)) (1−Σx2), where x is the haplotype frequency of each haplotype and n is the sample size (41). Chromosomes with larger numbers of SNPs had lower imputation error rates, but results remain unchanged when haplotype diversity was calculated using all chromosomes vs. only the 6 chromosomes with imputation error rates <0.08. Haplotype diversity was evaluated across a range of window sizes (SI Appendix, Fig. S2 and Fig. S5).

LD. Samples with >20% missing data were removed, and we ensured that no two samples had an IBS <0.95, which resulted in 570 vinifera cultivars and 59 sylvestris accessions included in the analysis. We only considered SNPs from the 19 assembled chromosomes with MAF >0.05 and <20% missing data, which resulted in 3,558 SNPs in vinifera and 3,349 SNPs in sylvestris. LD, as measured by r2, was calculated using PLINK (36).

Grape Color Association. Grape color was recorded on September 8–15, 2008, using the USDA grape germplasm in Davis, California. The following scale was used: 1 = gray/white, 2 = yellow, 3 = green, 4 = rose, 5 = red, 6 = red black, and 7 = blue/black. We performed a GWA study for grape color in 289 vinifera accessions for which genotype and phenotype data were available. After excluding SNPs with MAF <0.05 and >10% missing data, 5,110 SNPs remained for analysis. GWA was performed using the mixed model (42) implemented in EMMA (43), with the IBS matrix from PLINK (36) as a random effect. Signatures of selection were evaluated by dividing cultivars into white (scores = 1, 2, and 3; n = 139) and red (scores = 5, 6, and 7; n = 112).

ACKNOWLEDGMENTS. We thank Heidi Schwaninger, Dianne Velasco, Amy Szewc-McFadden, and Chuck Simon for DNA preparation and distribution; Bill Srmack, Dave Beckhorn, Greg Noden, Robert Martens, and Dawn Delleske for germplasm maintenance and distribution; Gan-Yuan Zhong, Lance Cadaval-Jean, Jean-Luc Jannink, Peter Bradbury, and Martha Hamblin for discussion; Mehmet Somel, Elliot Heffner, and Aaron Lorenz for comments on drafts of the manuscript; and Kristy Goodwin for assistance with figures. C.D.B. and A.R.B. were partially supported by National Science Foundation Grant 0701382 and A.R.B. by National Science Foundation Grant 0948510.
