Supplementary Materials for

Pervasive hybridization with local wild relatives in Western European grapevine varieties

Sara Freitas, Małgorzata A. Gazda, Miguel Â. Rebelo, Antonio J. Muñoz-Pajares, Carlos Vila-Viçosa, Antonio Muñoz-Mérida, Luís M. Gonçalves, David Azevedo-Silva, Sandra Afonso, Isaura Castro, Pedro H. Castro, Mariana Sottomayor, Albano Beja-Pereira, João Tereso, Nuno Ferrand, Elsa Gonçalves, Antero Martins, Miguel Carneiro, Herlander Azevedo*

*Corresponding author. Email: hazevedo@cibio.up.pt

Published 19 November 2021, Sci. Adv. 7, eabi8584 (2021)
DOI: 10.1126/sciadv.abi8584

The PDF file includes:

Supplementary Materials and Methods
Figs. S1 to S6
Tables S1 to S6
Legends for data files S1 to S3
References

Other Supplementary Material for this manuscript includes the following:

Data files S1 to S3
Supplementary Materials and Methods

Sampling and sequencing

Individual *vinifera* varieties were sampled from two separate Portuguese germplasm collections (PORVID - Associação Portuguesa para a Diversidade da Videira, and UTAD) during spring 2015 and 2016. The *sylvestris* samples were collected in the field in the southwestern region of the Iberian Peninsula, during spring 2016. Samples represent multiple populations across the complete species distribution range in Portugal, they are contiguous to the populations from Southwest Spain, which collectively form the largest distribution range in the Iberian Peninsula. Genomic DNA (gDNA) was extracted from leaf material using the CTAB-based variation of the NucleoSpin Plant II kit (Macherey-Nagel), which includes digestion with *RNA*se A. DNA purity and concentration were inferred from spectrophotometry (Nanodrop 2000, ThermoFisher Scientific) and fluorometric quantitation (Quant-iT PicoGreen dsDNA Assay Kit, ThermoFisher Scientific). Paired-end sequencing libraries for Illumina were generated using the TruSeq DNA PCR-free Library Preparation Kit (Illumina, San Diego, CA), according to the manufacturer’s instructions. Sequencing was carried out using an Illumina HiSeq1500 with 2x125 bp reads. WGS data was deposited in the European Nucleotide Archive (PRJEB44017). Sequencing data from 37 additional genotypes were downloaded from NCBI’s SRA depository (Table S1) and processed similarly to our newly generated sequencing data. Information on genotypes and sequencing effort is summarized in Table S1.

Mapping

Illumina sequences/fastqs were checked for read quality using *FastQC* (v1.7.119) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor sequences, low quality reads and short reads were removed using *Trimmomatic* v0.32 (70), with the following settings: TRAILING = 15, SLIDINGWINDOW = 4:20, and MINLEN = 30. Sequencing reads were mapped to the *Vitis vinifera* PN40024 reference genome release 102 [assembly GCA_000003745.2, concatenated with mitochondrial (GenBank ID FM179380.1) and chloroplast (GenBank ID DQ424856.1) reference genomes], using *BWA-MEM* (71) with default settings. Alignment and coverage summary statistics were calculated from FASTQ files using a shell script, and from BAM files using *BAMStats v1.25* (http://bamstats.sourceforge.net/).

Population structure analysis

To characterize population structure, we performed three independent analyses (PCA, phylogenetic tree, and ancestry analysis), taking into account the uncertainty in genotype calling due to the medium-to-low sequencing depth. Rather than using individual genotype calls, analyses used genotype likelihoods or posterior probabilities (73). For Principal Component Analysis (PCA), we estimated genotype posterior probabilities using ANGSD v0.916 (16), accepting a maximum sequencing depth of 3-fold the sum of the mean coverage of all genomes. Reads with a mapping quality lower than 20 and bases with a sequencing quality lower than 20 were discarded. The number of polymorphic sites was then calculated using a shell script and totaled 9,016,782 positions (for the 86 *Vitis vinifera* samples) or 3,424,962 positions (all 100 samples). We next employed the ngsCovar feature of the ngsPopGen package (https://github.com/mfumagalli/ngsPopGen) to compute the expected correlation matrix between individuals from genotype posterior probabilities (74). The .covar file was then plotted using the plotPCA.R R script from the ngsPopGen package. For the phylogenetic tree, the same genotype posterior probabilities were used to calculate pairwise genetic distances in ngsDist from ngsTools (75). We computed a distance-based minimal evolution tree by inputting the genetic distance matrix into FastME 2.1.5 (http://www.atge-
montpellier.fr/fastme/) with 100 bootstraps for branch support (76). FastME is based on balanced minimum evolution, which is the very principle of Neighbor Joining (NJ). Finally, the Newick tree file was visualized in FigTree v1.4.3 (http://evomis.org/resources/software/molecular-evolution-software/figtree/). For ancestry analysis we used NgsAdmix (http://www.popgen.dk/software/index.php), which is designed to infer admixture proportions from low-coverage NGS data based on genotype likelihoods (77). First, we generated a beagle format file containing genotype likelihoods, using ANGSD (16) with a set p-value of 1x10^-6 for calling SNPs. NgsAdmix was run assuming 2 to 8 ancestral populations with the default minor allele frequency of 0.05.

Nucleotide diversity and genetic differentiation

Although underpowered to make precise genotype calls or compute certain statistics reliably (e.g. linkage disequilibrium - LD), low-to-medium sequencing depth is sufficient to infer various metrics under a probabilistic framework, taking the uncertainty of genotype's assignment into account, by avoiding genotype calling and using genotype likelihoods or posterior probabilities (16, 78, 79). In the present report, V. rotundifolia (9) was used as the outgroup to polarize the ancestral state of alleles at each polymorphic site.

For multiple population genetic statistics, we first estimated the site allele frequency (SAF) per study group, using the doSAF option in ANGSD (-minMapQ 30 -minQ 20) (16). The realSFS feature in ANGSD was then used to estimate the unfolded site frequency spectrum (SFS) of each study group, using GL and the expectation maximization (EM) algorithm (78). Subsequently, the -doThetas option and the thetaStat feature in ANGSD were used to estimate and summarize multiple population genetic statistics, using SFS as prior information. They included Watterson’s Theta (θw; number of segregating sites) (80), pairwise nucleotide differences within study groups (π) (81), Tajima’s D (82), and Fay and Wu’s H (83). Statistics were averaged for all sites in 100 Kbp non-overlapping windows (nucleotide diversity statistics and genetic differentiation analysis), or 100 Kbp windows with 50 Kbp steps (DCMS analysis). Statistical significance between the statistics of different study groups was estimated using a paired Mann-Whitney test with the wilcox.test() R function.

Genetic differentiation was summarized using Wright’s fixation index (FST) (24). For this analysis, we first estimated the allele frequency in ANGSD (-minMapQ 30 -minQ 20 -minInd 60 -SNP_pval 1e-6), while incorporating all the individuals of all study groups. This estimation allowed us to extract information about chromosomes and positions of all sites to be used in the following analysis. The realSFS feature in ANGSD was then used to estimate bi-dimensional SFS between two study groups (2DSFS). Statistics were averaged for all sites in 100 Kbp non-overlapping windows (nucleotide diversity statistics and genetic differentiation analysis), or 100 Kbp windows with 50 Kbp steps (DCMS analysis). MDS of the matrix of mean FST values was carried out using the cmdscale() function from RStudio v1.0.143 (http://www.rstudio.com/), setting the eigen values as true (eig=TRUE) and maximum dimension as two (K=2).

IBD estimation and SNP calling

To look at the relationship between different grape cultivars, we performed Identity-By-Descent (IBD) analysis. Since our depth of coverage was low to moderate, we applied probabilistic methods implemented in ANGSD (16) to perform a SNP call using the doPlink option. We applied stringent criteria for the SNP call, with post-cutoff of 0.95 and a SNP P value of 1x10^-9, to include only highly supported SNPs. Unmapped scaffolds from the reference genome were excluded from this analysis. Subsequently, the SNPs were used to calculate IBD for all pairwise comparisons among the 100 samples using PLINK v1.90b3.26 (72) and applying the following filters: maf 0.05 and geno 0.05. The results were validated by looking
at: 1) known clones that were incorporated into the sampling; 2) samples expected to represent
the same variety based on name or literature. These confirmed a very high IBD score, falling
into the 0.8-1 range, followed by a gap in the distribution of pairwise IBD score, which reflects
the move from clonal to parent-offspring pairs, siblings, and other complex pedigree
relationships.

For study group differentiation analysis, IBD scores that represented comparisons
between genotypes of any two groups of interest were summarized in a violin plot performed
with RStudio v1.0.143 using Hmisc, ggplot2, and RColorBrewer packages.

For the heterozygosity estimation, we used the same SNP calling strategy just reported,
while implementing less stringent filters in PLINK (maf 0.05 and geno 0.2). The frequency of
heterozygosity in SNPs was estimated per individual, and summarized at the study group and
wild/cultivated levels.

**Admixture test using Patterson’s D statistic**

Patterson’s D statistics (or ABBA-BABA test) was originally developed to estimate the
genome-wide fraction of admixture present across human genomes (23). It assumes three
populations (P1,P2,P3) and one outgroup (O), which are phyletically related as
(((P1,P2),P3),O). This test looks for two specific SNP patterns (ABBA and BABA), where A is
the ancestral allele and B is the derived allele. Under a neutral coalescent model with no gene
flow, ABBA and BABA should have similar frequencies (D = 0). An excess of either ABBA
(D>0) or BABA (D>0) patterns is expected when gene flow from P3 has occurred after the split
of the two target populations (P1 and P2). This statistic detects introgression by comparing the
observed difference in the number of ABBA and BABA patterns to that expected from a
complete replacement of native alleles by introgressed ones, taking into account the fact that
gene flow may occur in both directions.

In the present analysis, we estimated all permutations of the six groups of interest
(WEAST, WIBERIA, CTABLE, CWWCE, CWIB1, CWIB2) as P1, P2, and P3. To polarize the ancestral state
of mutations, we used the *Vitis rotundifolia* sample as an outgroup (O). Comparisons that were
biologically meaningful are summarized in Table S3. We computed Patterson’s D statistic
using allele frequencies instead of binary counts of fixed ABBA-BABA sites, as implemented
in the ABBABABA2 (Multipopulation) function in ANGSD v0.917 (16), using non-overlapping
20 Kbp windows. Maximum depth was set as 3-fold the estimated coverage depth of the study
groups, while quality parameters were set as -minQ 20 -minMapQ 30. D values across genomic
windows were summarized using the DSTAT R script available in ANGSD, which estimates
average D values, jackknife, variance, Z, and P-value scores, plus the number of ABBA and
BABA patterns. No Error Correction or Ancient Transition removal was implemented. The
bulk data output file (.abbababa2) was used to retrieve per-window information on numerator
(ABBA-BABA counts), denominator (ABBA+BABA counts), and number of informative
sites. This information was used to plot chromosome-level estimates.

Unlike Patterson’s D statistic, which is inappropriate for quantifying introgression over
small genomic windows, we can quantify introgression using f statistics, which estimates the
proportion of haplotypes in the recipient population (P2) that trace their ancestry through the
donor population (P3). More precisely we calculated $f^d$, which has the best performance to
assess the fraction of the genome shared through introgression (25). In this study, $f^d$
was determined for comparison P1=CWIB1, P2=CWIB2, P3=WIBERIA, O=V. rotundifolia. Values for $f^d$
were calculated as presented by Martin and co-workers (25):

$$f^d = \frac{S(P_1,P_2,P_3,O)}{S(P_1,P_2,P_3,O)}$$  

(1)

The numerator of the D statistics, i.e. the difference between sums of ABBA and BABAs, is
defined as S. This equation identifies a donor population P_D, which can be either P2 or P3. We
estimated $D(P_1, P_2, P_2, O)$ and $D(P_1, P_3, P_3, O)$ in ANGSD using the $ABBABABA2$ ($Multipopulation$) function and the same parameters just reported for Patterson’s $D$ statistic. This allowed us to obtain $S(P_1, P_2, P_2, O)$ and $S(P_1, P_3, P_3, O)$ for each genomic window based on the bulk data output file. The donor population $P_D$ was then defined as the population with the higher frequency of the derived allele, meaning the one that maximized the denominator in equation (1). Values for $f^d$ were calculated for all windows with positive $D$ (25). Because $D$ values occur disproportionately in genomic regions with lower diversity (25), we tested the minimum threshold of informative sites at various stringencies, which resulted in a cut-off of 1000 informative sites per 20 Kb window, to account for a minimum presence of polymorphic sites. This led to the exclusion of 414 windows (1.8%), homogenously distributed throughout the genome. The $f^d$ statistic should vary between 0 (no introgression) and 1 (complete replacement). Values of $f^d$ above 1, which are minimal in $f^d$ when compared to other $f$ statistics (25), were also spurious in our dataset, suggesting the conservative nature of our analysis. Introgression tracts were defined using criteria that implemented stringency by selecting windows in the 95th percentile of the empirical distribution (including windows with negative $D$), but took into consideration their tract nature by collapsing regions distancing <100 Kb and accepting adjoining windows in the 90th percentile. Isolate windows were considered outliers as previously implemented (84).

**DCMS analysis of positive selection signatures**

Demographic events such as admixture of populations, population bottleneck, migration, introgression, inbreeding, and genetic drift, make the detection of selection signatures fairly complex. This can be mitigated by combining different statistics into a composite of signals (29, 85). However, combining various tests to detect selection signatures can be statistically challenging. Here, we used de-correlated composite of multiple signals (DCMS) as a summarizing statistic that combines the outcome of several complementary methods taking the covariance of the statistics into account. We expect this strategy to outperform single-statistic inference, by increasing positional resolution and minimizing the proportion of false positives (29, 85). More specifically, we used DCMS to combine four separate statistics that differ in their approach to detect selection events based on the type of selection signals that are targeted: genetic differentiation from local losses of heterozygosity ($F_{ST}$), shifts in the allele frequency spectrum of mutations (Delta Tajima’s $D$ – $\Delta TD$ - and Fay and Wu’s $H$) and reduction of genetic diversity from pairwise nucleotide diversity measures (reduction of diversity - ROD).

First, a sliding window approach was used to calculate $F_{ST}$, Tajima’s $D$, Fay and Wu’s $H$, and nucleotide diversity ($\pi$) statistics (24, 81-83). These were computed across the genome in ANGSD to account for uncertainty in medium-low depth sequencing data (78, 79). Statistics were calculated as 100 Kbp windows with 50 kb overlapping steps.

In this study, we investigated selection from the wild grapevine center of origin towards either the range expansion of wild populations, or domestication. Therefore, $\Delta TD$ and ROD statistics were designed to take advantage of information given by the existence of two contrasting study groups (86). For the between-wild grapevine comparison, the $\Delta TD$ statistic was estimated by subtracting $TD_{WEAST} – TD_{WIBERIA}$; ROD was calculated as $1 – \pi_{WIBERIA}/\pi_{WEAST}$; we used the $F_{ST}$ of the pairwise comparison between $WEAST$ and $WIBERIA$; for $H$, which is an absolute measure of selection, we used values for the $WIBERIA$ study group. For the wild vs cultivated grapevine comparisons, the $\Delta TD$ statistic was estimated by subtracting $TD_{WEAST} – TD_{Cultivated}$; ROD was calculated as $1 – \pi_{Cultivated}/\pi_{WEAST}$; we used the $F_{ST}$ of the pairwise comparison between $WEAST$ and $Cultivated$; for $H$ which is an absolute measure of selection, we used values for the $Cultivated$ group.
Sliding windows of the four statistics ($F_{ST}$, ROD, $\Delta T_D$, Fay and Wu's $H$) were combined using the DCMS method, which is expected to increase our capacity to delineate regions under positive selection, while reducing the proportion of false positives. First, we normalized the statistics with a two-step approach to transform continuous variables into normality (87). In the first step, the original variable was transformed towards statistical uniformity by calculating the percentile rank of each score at each window, resulting in uniformly distributed probabilities. The second step involved the inverse-normal transformation from uniformity to normality of the percentile ranks, generating a variable consisting of normally distributed Z-scores. Then a $P$ value was computed from this transformation, and finally, individual $P$ values for each statistic and correlation factors were used to calculate the final DCMS score (Table S5). DCMS uses the ratio of $(1-p_{lt})/p_{lt}$ for hypothesis testing directly. The $p_{lt}$ is the $P$ value in each window ($l$) for each statistic $t$. Correspondingly, the ‘de-correlated composite of multiple signals’ at each window is calculated as:

$$DCMS(l) = \sum_{i=1}^{n} \frac{i \log(1-p_{lt})}{\sum_{i=1}^{n} \frac{i}{n}}$$

where $r_{it}$ is the genome-wide correlation coefficient between the test statistic of the $i$th and the $t$th used method, and $n$ represents the total number of used methods (29). Comparisons-of-interest were defined as those that confronted the $W_{EAST}$ group with the remaining five groups. Genes-of-interest were considered for genomic windows above the 95th percentile of the distribution.

**Analysis of genes of interest**

Genes present in introgression tracts and DCMS comparison windows were retrieved using a standard Perl script. Gene lists were cross-referenced in VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html). Gene annotation was retrieved from PANTHER (http://pantherdb.org) and UniProtKB (https://www.uniprot.org/uniprot/). Also, for the 76 cross-referenced genes-of-interest, protein Fasta sequences were retrieved from UniProtKB, and used to perform a BlastP search in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi), against the *Non-redundant protein sequences* (nr) database, and the *Arabidopsis Thaliana RefSeq* database, outputting the first 50 and 10 hits, respectively. GO terms (*GO biological process complete*) were subjected to statistical overrepresentation testing in PANTHER (http://pantherdb.org), with Fisher's Exact test with False Discovery Rate correction ($P < 0.05$), by comparison against the full *Vitis vinifera* gene database (88).

**Genome alignment**

The collinearity between the *Vitis vinifera* ssp. *sylvestris* (48) and *Vitis vinifera* ssp. *vinifera* (*Vitis vinifera* PN40024 release 102; assembly GCA_000003745.2) genome assemblies, was determined using D-GENIES (http://dgenies.toulouse.inra.fr/) (89) a WEB application that performs large genome alignments using the minimap2 software package. Analysis used default visualization parameters and the hide noise option.
Fig. S1. Percentage of the reference genome with at least one mapped read, plotted as a function of the sample’s sequencing depth.
Fig. S2. Detailed PCA plots for population structure analysis.

PCA of the 100 sampled genotypes \([Vitis]\) sp. as well as wild (\textit{sylvestris}) and cultivated (\textit{vinifera}) grapevine genotypes, for eigenvectors 1 and 2 (A) and 1 and 3 (B). PCA of \textit{Vitis vinifera} wild and cultivated genotypes, for eigenvectors 1 and 2 (C) and 1 and 3 (D). In C.D, open symbols depict genotypes excluded from the six study groups due to admixture or clonal redundancy.
Fig. S3. Plot and histogram of Identical-by-Descent (IBD) pairwise analysis of cultivated grapevine varieties.

In total, 23 genotypes evidenced clonal relationships (Table S2). Robustness of the analysis was validated by the identification, within clonal pairs, of: 1) three clone pairs (Fernão Pires, Touriga Franca and Alvarinho) purposefully incorporated into the sampling effort; 2) well established sports (e.g. Pinot Noir, Pinot Gris, Pinot Blanc). Since presence of clones will skew allele frequencies specific to an individual, ultimately altering study group and nucleotide diversity statistics, just one of the clones was incorporated into the six study groups defined in the present study. Dashed line represented the threshold for assignment of clonal nature between pairwise compared genotypes. All 18 comparisons with IBD>0.8 represented expected clones.
Fig. S4. Ancestry and phylogenetic tree analysis of population structure. (A) Ancestry proportions of all *Vitis vinifera* genotypes following admixture analysis for K equaling to 2-8. (B) Phylogenetic tree of *Vitis vinifera* genotypes, with bootstrap values indicated for all nodes (100 replicates).
Fig. S5. Nucleotide diversity, genetic differentiation and heterozygosity of the six study groups.

(A,B) Violin plot distribution of Watterson’s Theta (A) and Fay and Hu’s $H$ (B). (C) Boxplot analysis of heterozygosity/SNP in individuals belonging to wild and cultivated grapes, and the six study groups.
Fig. S6. Dot plot alignment of the genomes from *Vitis vinifera* ssp. *sylvestris* and *Vitis vinifera* ssp. *vinifera*, highlighting the high collinearity between both assemblies. The different ranges of the proportion of identity are shown in different colors.
Supplementary tables

Table S1. Plant material and sequencing summary statistics.

| Sample origin (SRR code) | Genotype label | Estimated provenance | Assigned core study group | Sequencing depth - X (Before subsampling) | % of Genome with mapped positions (Before subsampling) |
|--------------------------|----------------|----------------------|---------------------------|------------------------------------------|-------------------------------------------------------|
| Present work             | Moscatel_Graudo | vinifera table       | C_TABLE*                  | 3,1                                      | 78,4%                                                  |
| Present work             | Airén           | vinifera table       | C_TABLE                   | 6,0                                      | 89,7%                                                  |
| Present work             | Moscatel_Galego_Tinto | vinifera table     | C_TABLE                   | 4,2                                      | 84,7%                                                  |
| Present work             | Dona_Maria      | vinifera table       | C_TABLE                   | 5,8                                      | 89,1%                                                  |
| Present work             | Bastardo        | vinifera wine WCE    | Ca_WCE                    | 3,1                                      | 83,1%                                                  |
| Present work             | Riesling        | vinifera wine WCE    | Ca_WCE                    | 3,8                                      | 84,1%                                                  |
| Present work             | Gewürztraminer  | vinifera wine WCE    | Ca_WCE*                   | 4,4                                      | 88,4%                                                  |
| Present work             | Albarceiro      | vinifera wine IB     | Ca_WCE                    | 3,8                                      | 86,4%                                                  |
| Present work             | Sauvignon_Blan  | vinifera wine WCE    | Ca_WCE*                   | 4,8                                      | 87,3%                                                  |
| Present work             | Pinot_Blan      | vinifera wine WCE    | Ca_WCE*                   | 4,4                                      | 89,8%                                                  |
| Present work             | Sylvaner        | vinifera wine WCE    | Ca_WCE*                   | 4,9                                      | 89,5%                                                  |
| Present work             | Aligote         | vinifera wine WCE    | Ca_WCE*                   | 3,9                                      | 87,2%                                                  |
| Present work             | Chenin          | vinifera wine WCE    | Ca_WCE                    | 4,3                                      | 88,3%                                                  |
| Present work             | Colombard       | vinifera wine WCE    | Ca_WCE                    | 4,1                                      | 86,6%                                                  |
| Present work             | Roussanne_Blan  | vinifera wine WCE    | Ca_WCE                    | 6,0                                      | 90,7%                                                  |
| Present work             | Pinot_Gris      | vinifera wine WCE    | Ca_WCE*                   | 6,0                                      | 91,8%                                                  |
| Present work             | Gamay           | vinifera wine WCE    | Ca_WCE*                   | 3,6                                      | 85,0%                                                  |
| Present work             | Ibérica_1       | sylvestris           | W_IBERIA                  | 4,3                                      | 83,3%                                                  |
| Present work             | Ibérica_2       | sylvestris           | W_IBERIA                  | 3,3                                      | 82,0%                                                  |
| Present work             | Ibérica_3       | sylvestris           | W_IBERIA                  | 3,5                                      | 81,4%                                                  |
| Present work             | Ibérica_4       | sylvestris           | W_IBERIA                  | 5,8                                      | 86,4%                                                  |
| Present work             | Ibérica_5       | sylvestris           | W_IBERIA                  | 6,6                                      | 89,6%                                                  |
| Present work             | Ibérica_6       | sylvestris           | W_IBERIA                  | 4,7                                      | 87,1%                                                  |
| Present work             | Ibérica_7       | sylvestris           | W_IBERIA                  | 4,1                                      | 84,8%                                                  |
| Present work             | Ibérica_8       | sylvestris           | W_IBERIA                  | 5,0                                      | 85,4%                                                  |
| Present work             | Ibérica_9       | sylvestris           | W_IBERIA                  | 3,7                                      | 83,9%                                                  |
| Present work             | Escaped_sativa  | vinifera wine IB     | Ca IB2                    | 5,2                                      | 89,5%                                                  |
| Present work             | Alvarinho_1     | vinifera wine IB     | Ca IB2                    | 4,9                                      | 88,3%                                                  |
| Present work             | Melhorio        | vinifera wine IB     | Ca IB2                    | 3,8                                      | 85,8%                                                  |
| Present work             | Espadeiro       | vinifera wine IB     | Ca IB2                    | 4,4                                      | 83,8%                                                  |
| Present work             | Alvarelhão      | vinifera wine IB     | Ca IB2*                   | 5,0                                      | 86,5%                                                  |
| Present work             | Alvarinho_2     | vinifera wine IB     | Ca IB2*                   | 3,8                                      | 86,2%                                                  |
| Present work             | Amaral          | vinifera wine IB     | Ca IB2                    | 4,1                                      | 86,2%                                                  |
| Present work             | Borraça         | vinifera wine IB     | Ca IB2                    | 4,4                                      | 84,6%                                                  |
| Present work             | Loureiro        | vinifera wine IB     | Ca IB2                    | 4,8                                      | 85,6%                                                  |
| Present work             | Touriga_Nacional| vinifera wine IB     | Ca IB2*                   | 3,6                                      | 83,7%                                                  |
| Present work             | Tinto_Cao       | vinifera wine IB     | Ca IB2                    | 4,8                                      | 88,4%                                                  |
| Work | Variety | Type | IB | Absorbance | Percentage |
|------|---------|------|----|------------|------------|
| Present work | Touriga Franca_1 | vinifera wine | IB | CwIB2 | 5.2 | 89.0% |
| Present work | Touriga Franca_2 | vinifera wine | IB | CwIB2* | 4.4 | 82.5% |
| Present work | Vinhao | vinifera wine | IB | CwIB2 | 4.9 | 86.8% |
| Present work | Arinto | vinifera wine | IB | CwIB1 | 4.6 | 85.9% |
| Present work | Fernão Pires_1 | vinifera wine | IB | CwIB1* | 3.6 | 84.3% |
| Present work | Tinta Miuda | vinifera wine | IB | CwIB1 | 3.0 | 79.7% |
| Present work | Assaraky | vinifera wine | IB |  | 4.1 | 87.5% |
| Present work | Vitis sp_1 | Vitis sp | IB |  | 7.1 | 88.2% |
| Present work | Vitis sp_2 | Vitis sp | IB |  | 4.4 | 76.7% |
| Present work | Vitis sp_3 | Vitis sp | IB |  | 3.3 | 82.1% |
| Present work | Bical | vinifera wine | IB | CwIB1 | 5.0 | 87.6% |
| Present work | Camarate | vinifera wine | IB | CwIB1 | 4.0 | 85.2% |
| Present work | Cercial | vinifera wine | IB | CwIB1 | 4.6 | 86.4% |
| Present work | Cerseal_Branco | vinifera wine | IB | CwIB1 | 4.8 | 85.1% |
| Present work | Siria | vinifera wine | IB | CwIB1 | 3.8 | 85.2% |
| Present work | Fernão Pires_2 | vinifera wine | IB | CwIB1 | 4.7 | 87.3% |
| Present work | Gouveio | vinifera wine | IB |  | 4.5 | 88.7% |
| Present work | Malvasia Fina | vinifera wine | IB | CwIB1 | 5.0 | 87.5% |
| Present work | Moreto | vinifera wine | IB | CwIB1 | 4.1 | 82.5% |
| Present work | Negra Mole | vinifera wine | IB | CwIB1 | 5.0 | 88.1% |
| Present work | Castelão | vinifera wine | IB | CwIB1 | 4.6 | 84.1% |
| Present work | Aragonez | vinifera wine | IB |  | 5.7 | 88.9% |
| Present work | Trajadura | vinifera wine | IB | CwIB1 | 6.0 | 89.8% |
| Present work | Trincadeira | vinifera wine | IB | CwIB1 | 5.2 | 86.4% |
| Present work | Verdelho | vinifera wine | IB |  | 4.1 | 86.9% |
| Present work | Grenache | vinifera wine | IB |  | 4.6 | 87.4% |
| Zhou et al. (9) | Primitivo_03 | vinifera table | C_TABLE* | 7.9 (29,8) | 92.9% (93,5%) |
| Zhou et al. (9) | Thompson_RLK | vinifera table | C_TABLE* | 7.8 (13,5) | 91,6% (92,4%) |
| Zhou et al. (9) | Zinfandel_03 | vinifera table | C_TABLE | 7.6 (43,2) | 92,1% (93,7%) |
| Zhou et al. (9) | Muscat_of_Alexandria | vinifera table | C_TABLE | 6.1 (14,8) | 90,7% (93,1%) |
| Zhou et al. (9) | Thompson_2A | vinifera table | C_TABLE* | 8.0 (15,9) | 91,1% (92,7%) |
| Zhou et al. (9) | Semillion_12 | vinifera wine WCE | CwWCE | 7.8 (13,8) | 91,5% (92,9%) |
| Zhou et al. (9) | Riesling_4 | vinifera wine WCE | 7.8 (14,4) | 91,9% (92,9%) |
| Zhou et al. (9) | Cabernet_Sauvignon_08 | vinifera wine WCE | 4.1 (13,5) | 87,2% (93,1%) |
| Zhou et al. (9) | Pinot Noir_123 | vinifera wine WCE | CwWCE | 7.8 (17,2) | 91,8% (94,1%) |
| Zhou et al. (9) | Gamay Noir_3 | vinifera wine WCE | CwWCE | 7.8 (12,7) | 91,2% (92,8%) |
| Zhou et al. (9) | Chardonnay_04 | vinifera wine WCE | CwWCE | 7.9 (56,3) | 91,7% (93,3%) |
| Zhou et al. (9) | Traminer_1 | vinifera wine WCE | CwWCE | 7.9 (11,4) | 91,8% (93,0%) |
| Zhou et al. (9) | Aramon | vinifera wine WCE | CwIB2* | 7.6 (17,9) | 90,9% (93,3%) |
| Zhou et al. (9) | Georgia | sylvestris | W_EAST | 6.7 (15,5) | 89,8% (92,0%) |
| Zhou et al. (9) | Azerbaijan_2 | sylvestris | W_EAST | 6.1 (13,5) | 88,8% (91,3%) |
| Author(s)                        | Sample Location          | Species          | Geographic Region | Uninformative Sites | Consensus Coverage |
|--------------------------------|--------------------------|------------------|-------------------|---------------------|--------------------|
| Zhou et al. (9)                | Pakistan_3               | *sylvestris*     | W EAST            | 6,7 (19,6)          | 91% (93,3%)        |
| Zhou et al. (9)                | Pakistan_2               | *sylvestris*     | W EAST            | 6,8 (21,8)          | 90,9% (93,4%)      |
| Zhou et al. (9)                | Armenia                  | *sylvestris*     | W EAST            | 5,0 (6,3)           | 87,2% (87,2%)      |
| Zhou et al. (9)                | Pakistan_1               | *sylvestris*     | W EAST            | 6,7 (19,0)          | 91% (93,1%)        |
| Zhou et al. (9)                | Turkmenistan_2           | *sylvestris*     | W EAST            | 6,7 (15,4)          | 90,1% (92,2%)      |
| Zhou et al. (9)                | Azerbaijan_1             | *sylvestris*     | W EAST            | 6,6 (14,1)          | 90,1% (92,2%)      |
| Zhou et al. (9)                | Vitis_rotundifolia       | *Vitis sp.*      | Outgroup          | 5,9 (20,9)          | 71,3% (79,3%)      |
| Zhou et al. (9)                | Turkmenistan_1           | *sylvestris*     |                   | 6,5 (14,4)          | 90,2% (92,5%)      |
| Cardone et al. (90)            | Italia                   | *vinifera table* | C TABLE           | 8,2 (12,4)          | 79,7% (81,9%)      |
| Cardone et al. (90)            | Autumn_Royal             | *vinifera table* | C_TABLE           | 6,5 (7,9)           | 75,7% (75,7%)      |
| Cardone et al. (90)            | Thompson_Seedless        | *vinifera table* | C_TABLE*          | 7,8 (12,3)          | 79,9% (82,5%)      |
| Cardone et al. (90)            | Red_Globe                | *vinifera table* | C_TABLE           | 7,6 (11,5)          | 70,3% (76,1%)      |
| Di Genova et al. (91)          | Sultanina                | *vinifera table* | C TABLE           | 7,4 (105,7)         | 87,8% (94%)        |
| NCBI (SRR769844)               | Vitis_amurensis_amurensis| *Vitis sp.*      |                   | 4,1                 | 79,5%              |
| NCBI (SRR2603972)              | Vitis_flexuosa           | *Vitis sp.*      |                   | 5,7                 | 79,5%              |
| NCBI (SRR769837)               | Vitis_davidii            | *Vitis sp.*      |                   | 3,9                 | 77,8%              |
| NCBI (SRR769838)               | Vitis_thunbergii         | *Vitis sp.*      |                   | 4,8                 | 82,6%              |
| NCBI (SRR769839)               | Vitis_aestivalis         | *Vitis sp.*      |                   | 4,9                 | 81,8%              |
| NCBI (SRR769840)               | Vitis_cinerea            | *Vitis sp.*      |                   | 4,9                 | 82,3%              |
| NCBI (SRR769841)               | Vitis_coignetiae         | *Vitis sp.*      |                   | 4,1                 | 79,7%              |
| NCBI (SRR769842)               | Vitis_amurensis_dissecta | *Vitis sp.*      |                   | 4,2                 | 76,6%              |
| NCBI (SRR769843)               | Vitis_riparia            | *Vitis sp.*      |                   | 4,3                 | 76,6%              |

WCE, Western and Central Europe; IB, Iberian Peninsula; * Excluded from the study group due to presence of clonal relationship
| Clones                          | Purposefully incorporated clonal pairs |
|--------------------------------|----------------------------------------|
| Sultanina, Thompson_Seedless; Thompson_2A; Thompson_RLK |                                        |
| Pinot_Noir_123; Pinot_Gris; Pinot_Blanc |                                        |
| Zinfandel_03; Primitivo_03       |                                        |
| Touriga_Nacional, Aramon        |                                        |
| Fernão_Pires_2; Fernão_Pires_1 | ×                                      |
| Muscat_of_Alexandria; Moscatel_Graudo |                                        |
| Traminer_1; Gewürztraminer      |                                        |
| Alvarinho_1; Alvarinho_2        | ×                                      |
| Touriga_Franca_1; Touriga_Franca_2 | ×                                    |
| Riesling_4; Riesling            |                                        |
| Gamay_Noir3; Gamay              |                                        |
### Table S3. Genome-wide results and significance of selected comparisons for Patterson’s $D$ statistic.

| $P_1$  | $P_2$  | $P_3$   | $D$ Statistic | $Z$ score | $P$ value |
|--------|--------|---------|---------------|-----------|-----------|
| C_TABLE | CwIB2  | WEST    | -0.033087     | -25.539736| 0.000000  |
| C_TABLE | CwIB2  | IBERIA  | 0.177209      | 114.249097| 0.000000  |
| C_TABLE | CwWCE  | WEST    | -0.026247     | -17.966964| 0.000000  |
| C_TABLE | CwWCE  | IBERIA  | 0.144810      | 76.765866 | 0.000000  |
| C_TABLE | CwIB1  | WEST    | -0.011705     | -9.555438 | 0.000000  |
| C_TABLE | CwIB1  | IBERIA  | 0.092798      | 59.475533 | 0.000000  |
| CwIB1  | CwIB2  | WEST    | -0.023259     | -20.448190| 0.000000  |
| CwIB1  | CwIB2  | IBERIA  | 0.094978      | 69.026545 | 0.000000  |
| CwIB1  | CwWCE  | WEST    | -0.015487     | -11.816794| 0.000000  |
| CwIB1  | CwWCE  | IBERIA  | 0.056776      | 33.590551 | 0.000000  |
| CwIB2  | CwWCE  | WEST    | 0.008206      | 6.243488  | 0.000000  |
| CwIB2  | CwWCE  | IBERIA  | -0.040772     | -25.183210| 0.000000  |
Table S4. GO term statistical enrichment for Biological Process, of genes present within introgression tracts of the CwIB2 study group.

| GO Biological process category; GO term hierarchy is represented | Observed no. of genes | Expected no. of genes | Fold enrichment | FDR corrected P value |
|---------------------------------------------------------------|------------------------|-----------------------|----------------|----------------------|
| abscisic acid-activated signaling pathway (GO:0009738)        | 18                     | 3.31                  | 5.43           | 2.19E-04            |
| cellular response to abscisic acid stimulus (GO:0071215)      | 19                     | 3.75                  | 5.06           | 1.83E-04            |
| cellular response to alcohol (GO:0097306)                     | 19                     | 3.75                  | 5.06           | 1.46E-04            |
| response to alcohol (GO:0097305)                              | 28                     | 9.72                  | 2.88           | 1.18E-03            |
| response to oxygen-containing compound (GO:1901700)           | 49                     | 27.61                 | 1.78           | 4.80E-02            |
| cellular response to organic substance (GO:0071310)           | 42                     | 21.50                 | 1.95           | 2.89E-02            |
| cellular response to oxygen-containing compound (GO:1901701)  | 33                     | 11.63                 | 2.84           | 3.03E-04            |
| cellular response to hormone stimulus (GO:0032870)            | 37                     | 15.46                 | 2.39           | 2.27E-03            |
| cellular response to endogenous stimulus (GO:0071495)         | 37                     | 16.56                 | 2.23           | 6.93E-03            |
| cellular response to lipid (GO:0071396)                       | 25                     | 6.11                  | 4.09           | 1.38E-04            |
| response to lipid (GO:0033993)                                | 33                     | 12.96                 | 2.55           | 1.58E-03            |
| response to abscisic acid (GO:0009737)                        | 28                     | 9.72                  | 2.88           | 1.12E-03            |
| hormone-mediated signaling pathway (GO:0009755)               | 36                     | 14.87                 | 2.42           | 1.69E-03            |
| regulation of protein serine/threonine phosphatase activity (GO:0080163) | 17                     | 2.28                  | 7.45           | 4.03E-05            |
| regulation of phosphoprotein phosphatase activity (GO:0043666) | 21                     | 5.67                  | 3.70           | 7.89E-04            |
| regulation of protein dephosphorylation (GO:0035304)          | 21                     | 5.67                  | 3.70           | 8.46E-04            |
| regulation of dephosphorylation (GO:0035303)                  | 21                     | 5.96                  | 3.52           | 1.22E-03            |
| regulation of phosphate metabolic process (GO:00199220)        | 31                     | 14.21                 | 2.18           | 2.92E-02            |
| regulation of phosphorus metabolic process (GO:0051174)        | 31                     | 14.28                 | 2.17           | 4.23E-02            |
| regulation of protein modification process (GO:0031399)        | 33                     | 15.75                 | 2.09           | 4.89E-02            |
| regulation of phosphatase activity (GO:0010921)               | 21                     | 5.82                  | 3.61           | 1.04E-03            |
| negative regulation of phosphoprotein phosphatase activity (GO:0032515) | 17                     | 2.94                  | 5.77           | 1.05E-04            |
| negative regulation of protein dephosphorylation (GO:0035308)  | 17                     | 2.94                  | 5.77           | 1.22E-04            |
| negative regulation of protein modification process (GO:0031400) | 20                     | 4.34                  | 4.60           | 1.21E-04            |
| negative regulation of cellular protein metabolic process (GO:0032269) | 22                     | 8.91                  | 2.47           | 4.92E-02            |
| negative regulation of protein metabolic process (GO:0051248)  | 22                     | 8.91                  | 2.47           | 4.77E-02            |
| negative regulation of dephosphorylation (GO:0035305)         | 17                     | 3.02                  | 5.63           | 1.08E-04            |
| negative regulation of phosphate metabolic process (GO:0045936) | 19                     | 4.12                  | 4.61           | 1.98E-04            |
| negative regulation of phosphorus metabolic process (GO:0010563) | 19                     | 4.12                  | 4.61           | 2.16E-04            |
| GO:0010923          | 17  | 3.02 | 5.63 | 1.21E-04 |
|---------------------|-----|------|------|----------|
| GO:0051346          | 18  | 5.23 | 3.44 | 6.05E-03 |
| GO:0043086          | 27  | 11.04| 2.45 | 2.29E-02 |
| GO:0044092          | 27  | 11.26| 2.40 | 2.51E-02 |
| GO:0008610          | 59  | 32.54| 1.81 | 9.94E-03 |
Table S5. Correlation values estimated for the four statistics employed in DCMS analysis of two comparisons of interest.

| Pairwise comparison | Statistics     | $F_{ST}$       | Fay and Wu's H | $\Delta T_D$ | ROD    |
|---------------------|----------------|----------------|----------------|---------------|--------|
| W EAST VS WIBERIA   | $F_{ST}$       | 1              |                |               |        |
|                     | Fay and Wu's H | -0.3458341     |                | -0.6106718    | 1      |
|                     | $\Delta T_D$   | 0.18727539     | -0.6106718    | 1             |        |
|                     | ROD             | 0.17600912     | -0.5833357    | 0.569436      | 1      |
| W EAST VS CW1       | $F_{ST}$       | 1              |                |               |        |
|                     | Fay and Wu's H | -0.0666451     |                | 1             |        |
|                     | $\Delta T_D$   | 0.03267892     | -0.6020114    | 1             |        |
|                     | ROD             | -0.0148182     | -0.5893929    | 0.53223704    | 1      |
| W EAST VS CW2       | $F_{ST}$       | 1              |                |               |        |
|                     | Fay and Wu's H | -0.2437957     |                | 1             |        |
|                     | $\Delta T_D$   | 0.11921819     | -0.5306787    | 1             |        |
|                     | ROD             | 0.11026983     | -0.5209194    | 0.47058521    | 1      |
| W EAST VS CW1       | $F_{ST}$       | 1              |                |               |        |
|                     | Fay and Wu's H | -0.3319414     |                | 1             |        |
|                     | $\Delta T_D$   | 0.13796094     | -0.6363812    | 1             |        |
|                     | ROD             | 0.21849512     | -0.6348185    | 0.48077335    | 1      |
| W EAST VS CW2       | $F_{ST}$       | 1              |                |               |        |
|                     | Fay and Wu's H | -0.233644      |                | 1             |        |
|                     | $\Delta T_D$   | 0.07533273     | -0.4911239    | 1             |        |
|                     | ROD             | 0.21943832     | -0.539944     | 0.38281576    | 1      |
| Gene ID          | Annotation                                                                 | Chr | Start  | End    |
|-----------------|------------------------------------------------------------------------------|-----|--------|--------|
| VIT_00s0194g00070 | ZINC FINGER PROTEIN CONSTANS-LIKE 9 FAMILY PROTEIN                           | 10  | 2260252| 2272206|
| VIT_00s0194g00080 | PHOTOTROPIC-RESPONSIVE NPH3                                                  | 10  | 2272444| 2275038|
| VIT_00s0194g00090 | KETOACYL-ACP SYNTHASE 1                                                     | 10  | 2280897| 2285357|
| VIT_00s0194g00100 | C2 AND GRAM DOMAIN-CONTAINING PROTEIN                                        | 10  | 2288893| 2297485|
| VIT_00s0194g00110 | 2-PHOSPHOGLYCOLATE PHOSPHATASE 2                                             | 10  | 2300759| 2305096|
| VIT_00s0194g00120 | SERINE ACETYLTRANSFERASE 1, CHLOROPLASTIC                                   | 10  | 2305715| 2307059|
| VIT_00s0194g00130 | TRANSCRIPTIONAL ADAPTER ADA2B                                                | 10  | 2313677| 2318858|
| VIT_00s0194g00140 | TRAPPC3                                                                      | 10  | 2319562| 2321948|
| VIT_00s0194g00150 | TRAPPC3                                                                      | 10  | 2323170| 2323607|
| VIT_00s0194g00160 | CYTOCHROME P450, FAMILY 707, SUBFAMILY A, POLYPEPTIDE 1                     | 10  | 2341847| 2346605|
| VIT_00s0194g00170 | KILLING ME SLOWLY 2                                                         | 10  | 2345752| 2346605|
| VIT_00s0194g00180 | HEPATIC PROTEIN                                                             | 10  | 2351931| 2352567|
| VIT_00s0194g00190 | PUTATIVE ENDONUCLEASE OR GLYCOSYL HYDROLASE                                 | 10  | 2352998| 2353916|
| VIT_00s0194g00200 | FAR1-RELATED SEQUENCE 10 ISOFORM 1                                          | 10  | 2358663| 2360918|
| VIT_00s0194g00210 | OUTWARD RECTIFYING POTASSIUM CHANNEL PROTEIN                               | 10  | 2364396| 2365775|
| VIT_00s0194g00220 | PROTEASOME ACTIVATING PROTEIN 200                                           | 10  | 2366617| 2367043|
| VIT_00s0194g00230 | PROTEASOME SUBUNIT PAB1                                                      | 10  | 2369713| 2370453|
| VIT_00s0194g00240 | UB-LIKE PROTEASE 1A                                                          | 10  | 2395023| 2395911|
| VIT_02s0012g02920 | ACYL-COA OXIDASE 6                                                           | 2   | 1113398| 1113634|
| VIT_02s0012g02970 | UNCHARACTERIZED PROTEIN                                                      | 2   | 11183822| 11186832|
| VIT_02s0109g00080 | PHOSPHORIBULOKINASE                                                          | 2   | 12348163| 12351903|
| VIT_02s0109g00120 | UNCHARACTERIZED PROTEIN                                                      | 2   | 12422839| 12441134|
| VIT_02s0109g00160 | UNCHARACTERIZED PROTEIN                                                      | 2   | 12557348| 12557838|
| VIT_02s0109g00180 | AMINOPHOSPHOLIPID ATPASE 2                                                   | 2   | 12640920| 12641306|
| VIT_02s0109g00190 | ABC TRANSPORTER-LIKE PROTEIN                                                 | 2   | 12642481| 12643852|
| VIT_02s0109g00230 | EARLY-RESPONSIVE TO DEHYDRATION STRESS PROTEIN (ERD4)                       | 2   | 12786746| 12808921|
| VIT_06s0004g04370 | HISTONE H4                                                                   | 6   | 5352023| 5352367|
| VIT_06s0004g04380 | MITOCHONDRIAL F1F0-ATP SYNTHASE                                               | 6   | 5357240| 5358176|
| VIT_06s0004g04390 | UDP-XYL SYNTHASE 6                                                           | 6   | 5358853| 5364485|
| VIT_06s0004g04400 | P-HYDROXYBENZOIC ACID EFFLUX PUMP SUBUNIT                                   | 6   | 5365814| 5368974|
| VIT_06s0004g04410 | SEC14 CYTOSOLIC FACTOR FAMILY PROTEIN / PHOSPHOGLYCERIDE TRANSFER FAMILY PROTEIN | 6   | 5375905| 5383629|
| VIT_06s0004g04420 | METALLOTHIOL TRANSFERASE FOSB                                                | 6   | 5386184| 5387156|
| VIT_06s0004g04430 | UBQUITIN-PROTEIN LIGASE 7                                                    | 6   | 5387285| 5399913|
| VIT_06s0004g04440 | PATHOGENESIS-RELATED THAUMATIN SUPERFAMILY PROTEIN                           | 6   | 5401080| 5402253|
| VIT_06s0004g04450 | PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C4                               | 6   | 5406399| 5406925|
| VIT_06s0004g04460 | INNER MEMBRANE PROTEIN ALBINO3, CHLOROPLASTIC                               | 6   | 5407544| 5415216|
| VIT_06s0004g04470 | HEAT SHOCK COGNATE PROTEIN 70                                                | 6   | 5418191| 5420677|
| VIT_06s0004g04480 | HSP20-LIKE CHAPERONES SUPERFAMILY PROTEIN ISOFORM 1                          | 6   | 5422703| 5425089|
| VIT_06s0009g00070 | ALDEHYDE OXIDASE 4                                                           | 6   | 12734005| 12755925|
| VIT_06s0009g000780 | ABI3-INTERACTING PROTEIN 2                                                   | 6   | 12797100| 12798180|
| VIT_06s0009g000790 | SHUGOSHIN 2                                                                  | 6   | 12798181| 12799747|
| VIT_06s0080g00370 | F-BOX/LRR PLANT PROTEIN                                                      | 6   | 21456653| 21457918|
Other Supplementary Materials

**Data S1.** Annotation of genes identified in introgression tracts of the CwIB2 study group, following analysis of $f_{st}$ statistics.

**Data S2.** List of genes identified as belonging to selective sweeps (95$^{th}$ percentile cut-off), after DCMS analysis of positive selection regions for pairwise comparison of $W_{EAST}$ against remaining study groups.

**Data S3.** Annotation of genes identified in selective sweeps (95$^{th}$ percentile cut-off), after DCMS analysis of positive selection regions in the pairwise comparison between wild groups $W_{IBERIA}$ vs $W_{EAST}$. 
REFERENCES AND NOTES

1. P. This, T. Lacombe, M. R. Thomas, Historical origins and genetic diversity of wine grapes. *Trends Genet.* **22**, 511–519 (2006).

2. Z.-Y. Ma, J. Wen, S. M. Ickert-Bond, Z.-L. Nie, L.-Q. Chen, X.-Q. Liu, Phylogenomics, biogeography, and adaptive radiation of grapes. *Mol. Phylogenet. Evol.* **129**, 258–267 (2018).

3. D. Zohary, M. Hopf, E. Weiss, *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin* (Oxford Univ. Press on Demand, 2012).

4. R. Arroyo-García, L. Ruiz-García, L. Bolling, R. Ocete, M. A. López, C. Arnold, A. Ergul, G. Söylemezoğlu, H. I. Uzun, F. Cabello, J. Ibáñez, M. K. Aradhya, A. Atanassov, I. Atanassov, S. Balint, J. L. Cenis, L. Costantini, S. Goris-Lavets, M. S. Grando, B. Y. Klein, P. E. McGovern, D. Merdinoglu, I. Pejic, F. Pelsy, N. Primikirios, V. Rissovannaya, K. A. Roubelakis-Angelakis, H. Snoussi, P. Sotiri, S. Tamhankar, P. This, L. Troshin, J. M. Malpica, F. Lefort, J. M. Martínez-Zapater, Multiple origins of cultivated grapevine (*Vitis vinifera* L. *ssp. sativa*) based on chloroplast DNA polymorphisms. *Mol. Ecol.* **15**, 3707–3714 (2006).

5. F. Grassi, M. Labra, S. Imazio, R. Ocete Rubio, O. Failla, A. Scienza, F. Sala, Phylogeographical structure and conservation genetics of wild grapevine. *Conserv. Genet.* **7**, 837–845 (2006).

6. S. Riaz, G. De Lorenzis, D. Velasco, A. Koehmstedt, D. Maghradze, Z. Bobokashvili, M. Musayev, G. Zdunic, V. Laucou, M. Andrew Walker, O. Failla, J. E. Preece, M. Aradhya, R. Arroyo-Garcia, Genetic diversity analysis of cultivated and wild grapevine (*Vitis vinifera* L.) accessions around the Mediterranean basin and Central Asia. *BMC Plant Biol.* **18**, 137 (2018).

7. P. McGovern, M. Jalabadze, S. Batiuk, M. P. Callahan, K. E. Smith, G. R. Hall, E. Kvavadze, D. Maghradze, N. Rusishvili, L. Bouby, O. Failla, G. Cola, L. Mariani, E. Boaretto, R. Bacilieri, P. This, N. Wales, D. Lordkipanidze, Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10309-E10318 (2017).

8. Z. Liang, S. Duan, J. Sheng, S. Zhu, X. Ni, J. Shao, C. Liu, P. Nick, F. Du, P. Fan, R. Mao, Y. Zhu, W. Deng, M. Yang, H. Huang, Y. Liu, Y. Ding, X. Liu, J. Jiang, Y. Zhu, S. Li, X. He, W. Chen, Y. Dong, Whole-genome resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history analyses. *Nat. Commun.* **10**, 1190 (2019).

9. Y. Zhou, M. Massonnet, J. S. Sanjak, D. Cantu, B. S. Gaut, Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11715–11720 (2017).

10. A. G. Reynolds, The grapevine, viticulture, and winemaking: A brief introduction, in *Grapevine Viruses: Molecular Biology, Diagnostics and Management* (Springer, 2017), pp. 3–29.
11. Y. Zhou, A. Muyle, B. S. Gaut, Evolutionary genomics and the domestication of grapes, in *The Grape Genome* (Springer, 2019), pp. 39–55.

12. H. P. Olmo, Origin and distribution of grapes, in *Evolution of Crop Plants*, N. W. Simmonds, Ed. (Longman, London and New York, 1976), pp. 294–298.

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

14. S. Myles, A. R. Boyko, C. L. Owens, P. J. Brown, F. Grassi, M. K. Aradhya, B. Prins, A. Reynolds, J.-M. Chia, D. Ware, C. D. Bustamante, E. S. Buckler, Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3530–3535 (2011).

15. V. Laucou, A. Launay, R. Bacilieri, T. Lacombe, A.-F. Adam-Blondon, A. Bérard, A. Chauveau, M. T. de Andrés, L. Hausmann, J. Ibáñez, M.-C. Le Paslier, D. Maghradze, J. M. Martinez-Zapater, E. Maul, M. Ponnaiah, R. Töpfer, J.-P. Péros, J.-M. Boursiquot, Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLOS ONE* **13**, e0192540 (2018).

16. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGS: Analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

17. S. Y. Kim, K. E. Lohmueller, A. Albrechtsen, Y. Li, T. Korneliussen, G. Tian, N. Grarup, T. Jiang, G. Andersen, D. Witte, T. Jorgensen, T. Hansen, O. Pedersen, J. Wang, R. Nielsen, Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics* **12**, 231 (2011).

18. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

19. Y. Zhou, A. Minio, M. Massonnet, E. Solares, Y. Lv, T. Beridze, D. Cantu, B. S. Gaut, The population genetics of structural variants in grapevine domestication. *Nat. Plants* **5**, 965–979 (2019).

20. M. A. Gazda, P. Andrade, S. Afonso, J. Dilyte, J. P. Archer, R. J. Lopes, R. Faria, M. Carneiro, Signatures of selection on standing genetic variation underlie athletic and navigational performance in racing pigeons. *Mol. Biol. Evol.* **35**, 1176–1189 (2018).

21. K. P. Oh, C. L. Aldridge, J. S. Forbey, C. Y. Dadabay, S. J. Oyler-McCance, Conservation genomics in the Sagebrush Sea: Population divergence, demographic history, and local adaptation in sage-grouse (*Centrocercus* spp.). *Genome Biol. Evol.* **11**, 2023–2034 (2019).

22. Z. Migicovsky, J. Sawler, K. M. Gardner, M. K. Aradhya, B. H. Prins, H. R. Schwaninger, C. D. Bustamante, E. S. Buckler, G.-Y. Zhong, P. J. Brown, S. Myles, Patterns of genomic and phenomic diversity in wine and table grapes. *Hortic. Res.* **4**, 17035 (2017).
23. R. E. Green, J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H.-Y. Fritz, N. F. Hansen, E. Y. Durand, A.-S. Malaspina, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušić, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. de la Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, S. Pääbo, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).

24. S. Wright, The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).

25. S. H. Martin, J. W. Davey, C. D. Jiggins, Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).

26. T. Kuromori, M. Seo, K. Shinozaki, ABA transport and plant water stress responses. *Trends Plant Sci.* **23**, 513–522 (2018).

27. J. Santiago, F. Dupeux, A. Round, R. Antoni, S.-Y. Park, M. Jamin, S. R. Cutler, P. L. Rodriguez, J. A. Márquez, The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* **462**, 665–668 (2009).

28. H. Fernandes, K. Michalska, M. Sikorski, M. Jaskolski, Structural and functional aspects of PR-10 proteins. *FEBS J.* **280**, 1169–1199 (2013).

29. Y. Ma, X. Ding, S. Qanbari, S. Weigend, Q. Zhang, H. Simianer, Properties of different selection signature statistics and a new strategy for combining them. *Heredity* **115**, 426–436 (2015).

30. C. Burgarella, A. Barnaud, N. A. Kane, F. Jankowski, N. Scarcelli, C. Billot, Y. Vigouroux, C. Berthouly-Salazar, Adaptive introgression: An untapped evolutionary mechanism for crop adaptation. *Front. Plant Sci.* **10**, 4 (2019).

31. G. M. Janzen, L. Wang, M. B. Hufford, The extent of adaptive wild introgression in crops. *New Phytol.* **221**, 1279–1288 (2019).

32. A. Sivan, O. Rahimi, B. Lavi, M. Salmon-Divon, E. Weiss, E. Drori, S. Hübner, Genomic evidence supports an independent history of Levantine and Eurasian grapevines. *Plants People Planet* **3**, 414–427 (2021).

33. M. S. Lopes, D. Mendonça, M. Rodrigues dos Santos, J. E. Eiras-Dias, A. da Câmara Machado, New insights on the genetic basis of Portuguese grapevine and on grapevine domestication. *Genome* **52**, 790–800 (2009).

34. G. De Lorenzis, F. Mercati, C. Bergamini, M. F. Cardone, A. Lupini, A. Mauceri, A. R. Caputo, L. Abbate, M. G. Barbagallo, D. Antonacci, F. Sunseri, L. Brancadoro, SNP genotyping elucidates the genetic diversity of *Magna Graecia* grapevine germplasm and its historical origin and dissemination. *BMC Plant Biol.* **19**, 7 (2019).
35. G. Zdunić, J. E. Preece, M. Aradhya, D. Velasco, A. Koehmstedt, G. S. Dangl, Genetic diversity and differentiation within and between cultivated (Vitis vinifera L. ssp. sativa) and wild (Vitis vinifera L. ssp. sylvestris) grapes. Vitis 52, 29–32 (2013).

36. J. Cunha, M. Teixeira Santos, L. Carneiro, P. Fevereiro, J. E. Eiras-Dias, Portuguese traditional grapevine cultivars and wild vines (Vitis vinifera L.) share morphological and genetic traits. Genet. Resour. Crop Evol. 56, 975–989 (2009).

37. M. T. De Andrés, A. Benito, G. Pérez-Rivera, R. Ocete, M. A. Lopez, L. Gaforio, G. Muñoz, F. Cabello, J. M. Martínez Zapater, R. Arroyo-Garcia, Genetic diversity of wild grapevine populations in Spain and their genetic relationships with cultivated grapevines. Mol. Ecol. 21, 800–816 (2012).

38. R. De Michele, F. La Bella, A. S. Gristina, I. Fontana, D. Pacifico, G. Garfi, A. Motisi, D. Crucitti, L. Abbate, F. Carimi, Phylogenetic relationship among wild and cultivated grapevine in Sicily: A hotspot in the middle of the Mediterranean Basin. Front. Plant Sci. 10, 1506 (2019).

39. J. Cunha, J. Ibáñez, M. Teixeira-Santos, J. Brazão, P. Fevereiro, J. M. Martínez-Zapater, J. E. Eiras-Dias, Genetic relationships among Portuguese cultivated and wild Vitis vinifera L. Germplasm. Front. Plant Sci. 11, 127 (2020).

40. V. Maraš, J. Tello, A. Gazivoda, M. Mugoša, M. Perišić, J. Raičević, N. Štajner, R. Ocete, V. Božović, T. Popović, E. García-Escudero, M. Grbić, J. M. Martínez-Zapater, J. Ibáñez, Population genetic analysis in old Montenegrin vineyards reveals ancient ways currently active to generate diversity in Vitis vinifera. Sci. Rep. 10, 15000 (2020).

41. J. Ramos-Madrigal, A. K. W. Runge, L. Bouby, T. Lacombe, J. A. S. Castruita, A.-F. Adam-Blondon, I. Figueiral, C. Hallavant, J. M. Martínez-Zapater, C. Schaal, R. Töpfer, B. Petersen, T. Sicheritz-Pontén, P. This, R. Bacilieri, M. T. P. Gilbert, N. Wales, Palaeogenomic insights into the origins of French grapevine diversity. Nat. Plants 5, 595–603 (2019).

42. J. Cunha, J. Ibáñez, M. Teixeira-Santos, J. Brazão, P. Fevereiro, J. M. Martínez-Zapater, J. E. Eiras-Dias, Characterisation of the Portuguese grapevine germplasm with 48 single-nucleotide polymorphisms. Aust. J. Grape Wine Res. 22, 504–516 (2016).

43. M. J. Iriarte-Chiapusso, C. A. Ocete-Pérez, B. Hernández-Beloqui, R. Ocete-Rubio, Vitis vinifera in the Iberian Peninsula: A review. Plant Biosyst. 151, 245–257 (2017).

44. L. Bouby, I. Figueiral, A. Bouchette, N. Rovira, S. Ivorra, T. Lacombe, T. Pastor, S. Picq, P. Marinval, J.-F. Terral, Bioarchaeological insights into the process of domestication of grapevine (Vitis vinifera L.) during Roman times in Southern France. PLOS ONE 8, e63195 (2013).

45. M. K. Aradhya, G. S. Dangl, B. H. Prins, J.-M. Boursiquot, M. A. Walker, C. P. Meredith, C. J. Simon, Genetic structure and differentiation in cultivated grape Vitis vinifera L. Genet. Res. 81, 179–192 (2003).

46. C. D’Onofrio, Introggression among cultivated and wild grapevine in Tuscany. Front. Plant Sci. 11, 202 (2020).
47. W. De Coster, M. H. Weissensteiner, F. J. Sedlazeck, Towards population-scale long-read sequencing. *Nat. Rev. Genet.* **9**, 572–587 (2021).

48. H. Badouin, A. Velt, F. Gindraud, T. Flutre, V. Dumas, S. Vautrin, W. Marande, J. Corbi, E. Sallet, J. Ganofsky, S. Santoni, D. Guyot, E. Ricciardelli, K. Jepsen, J. Kafer, H. Berges, E. Duchene, F. Picard, P. Hugueney, R. Tavares, R. Bacilieri, C. Rustenholz, G. A. B. Marais, The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome Biol.* **21**, 223 (2020).

49. A. Yu, Y. Xie, X. Pan, H. Zhang, P. Cao, X. Su, W. Chang, M. Li, Photosynthetic phosphoribulokinase structures: Enzymatic mechanisms and the redox regulation of the Calvin-Benson-Bassham cycle. *Plant Cell* **32**, 1556–1573 (2020).

50. M. B. Hufford, P. Lubinksy, T. Pyhajarvi, M. T. Devengenzo, N. C. Ellstrand, J. Ross-Ibarra, The genomic signature of crop-wild introgression in maize. *PLOS Genet.* **9**, e1003477 (2013).

51. M. A. Hardigan, F. P. E. Laimbeer, L. Newton, E. Crisovan, J. P. Hamilton, B. Vaillancourt, K. Wieght-Rininger, J. C. Wood, D. S. Douches, E. M. Farré, R. E. Veilleux, C. R. Buell, Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9999-E10008 (2017).

52. S. Rivas-Martínez, Á. Penas, S. del Río, T. E. D. González, S. Rivas-Sáenz, Bioclimatology of the Iberian Peninsula and the Balearic Islands, in *The Vegetation of the Iberian Peninsula* (Springer, 2017), pp. 29–80.

53. Q. Lin, J. Yang, Q. Wang, H. Zhu, Z. Chen, Y. Dao, K. Wang, Overexpression of the trehalose-6-phosphate phosphatase family gene *AtTPPF* improves the drought tolerance of *Arabidopsis thaliana*. *BMC Plant Biol.* **19**, 381 (2019).

54. S. Rattanakon, R. Ghan, G. A. Gambetta, L. G. Deluc, K. A. Schlauch, G. R. Cramer, Abscisic acid transcriptomic signaling varies with grapevine organ. *BMC Plant Biol.* **16**, 72 (2016).

55. T. Kuromori, T. Miyaji, H. Yabuuchi, H. Shimizu, E. Sugimoto, A. Kamiya, Y. Moriyama, K. Shinozaki, ABC transporter AtABCG25 is involved in abscisic acid transport and responses. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2361–2366 (2010).

56. J. C. Isner, A. Begum, T. Nuehse, A. M. Hetherington, F. J. Maathuis, KIN7 kinase regulates the vacuolar TPK1 K⁺ channel during stomatal closure. *Curr. Biol.* **28**, 466–472.e4 (2018).

57. C. D. M. Castroverde, The AREB1-ADA2b-GCN5 complex regulates gene expression during drought stress. *Plant Cell* **31**, 559–560 (2019).

58. S. E. Murthy, A. E. Dubin, T. Whitwam, S. Jojoa-Cruz, S. M. Cahalan, S. A. R. Mousavi, A. B. Ward, A. Patapoutian, OSCA/TMEM63 are an evolutionarily conserved family of mechanically activated ion channels. *eLife* **7**, e41844 (2018).
59. K. Thor, S. Jiang, E. Michard, J. George, S. Scherzer, S. Huang, J. Dindas, P. Derbyshire, N. Leitão, T. A. DeFalco, P. Köster, K. Hunter, S. Kimura, J. Gronnier, L. Stransfeld, Y. Kadota, C. A. Bücherl, M. Charpentier, M. Wrzaczek, D. MacLean, G. E. D. Oldroyd, F. L. H. Menke, M. R. G. Roelfsema, R. Hedrich, J. Feijó, C. Zipfel, The calcium-permeable channel OSCA1.3 regulates plant stomatal immunity. *Nature* **585**, 569–573 (2020).

60. A. C. Vlot, J. H. Sales, M. Lenk, K. Bauer, A. Brambilla, A. Sommer, Y. Chen, M. Wenig, S. Nayem, Systemic propagation of immunity in plants. *New Phytol.* **229**, 1234–1250 (2021).

61. E. Sinapidou, K. Williams, L. Nott, S. Bahkt, M. Tör, I. Crute, P. Bittner-Eddy, J. Beynon, Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in Arabidopsis. *Plant J.* **38**, 898–909 (2004).

62. R. Ocete, P. Fevereiro, O. Failla, Proposal for the wild grapevine (*Vitis vinifera* L. subsp. *sylvestris* (Gmelin) Hegi) conservation in the European countries. *Vitis* **54**, 281–282 (2015).

63. E. Warschefsky, R. V. Penmetsa, D. R. Cook, E. J. B. von Wettberg, Back to the wilds: Tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *Am. J. Bot.* **101**, 1791–1800 (2014).

64. P. E. McGovern, *Ancient Wine: The Search for the Origins of Viticulture* (Princeton Univ. Press, Princeton, NJ, 2003).

65. J. E. Aura, Y. Carrión, E. Estrelles, G. P. Jordà, Plant economy of hunter-gatherer groups at the end of the last Ice Age: Plant macroremains from the cave of Santa Maira (Alacant, Spain) ca. 12000–9000 BP. *Veget. Hist. Archaeobot.* **14**, 542–550 (2005).

66. M. Lara, M. J. Iriarte-Chiapusso, M. Cantos, J. L. García Jiménez, R. Morales, C. A. Ocete, M. Á. López, J. A. Salinas, I. Rubio, J. Hidalgo, M. Íñiguez, A. Rodríguez, J. M. Valle, R. Arroyo-García, M. C. Ayala, I. Armendáriz, D. Maghradze, C. Arnold, R. Ocete, La vid silvestre. Un importante recurso fitogenético sin protección legal en España. *Rev. Iberoam. Vitic. Agroind. Y Rural.* **4**, 46–68 (2017).

67. G. Pérez-Jordà, L. Peña-Chocarro, S. Pardo-Gordó, Fruits arriving to the west. Introduction of cultivated fruits in the Iberian Peninsula. *J. Archaeol. Sci. Rep.* **35**, 102683 (2021).

68. S. Boso, P. Gago, J. L. Santiago, A. Teira-Brión, M. Martín-Seijo, J. Rey-Castiñeira, R. Ocete, C. Ocete, M. C. Martínez, Morphometric comparison of current, Roman-era and medieval *Vitis* seeds from the north-west of Spain. *Aust. J. Grape Wine Res.* **26**, 300–309 (2020).

69. R. Fernandes, A. J. M. Barros, *Descrição do terreno ao redor de Lamego duas léguas: 1531–1532* (Caleidoscópio — Edição e Artes Gráficas, Casal de Cambra, 2012).

70. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

71. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
72. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

73. F. G. Vieira, F. Lassalle, T. S. Korneliussen, M. Fumagalli, Improving the estimation of genetic distances from next-generation sequencing data. *Biol. J. Linn. Soc.* **117**, 139–149 (2016).

74. M. Fumagalli, F. G. Vieira, T. S. Korneliussen, T. Linderoth, E. Huerta-Sanchez, A. Albrechtsen, R. Nielsen, Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* **195**, 979–992 (2013).

75. M. Fumagalli, F. G. Vieira, T. Linderoth, R. Nielsen, ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487 (2014).

76. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).

77. L. Skotte, T. S. Korneliussen, A. Albrechtsen, Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).

78. R. Nielsen, T. Korneliussen, A. Albrechtsen, Y. Li, J. Wang, SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLOS ONE* **7**, e37558 (2012).

79. T. S. Korneliussen, I. Moltke, A. Albrechtsen, R. Nielsen, Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* **14**, 289 (2013).

80. G. A. Watterson, On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).

81. M. Nei, W. H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979).

82. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

83. J. C. Fay, C. I. Wu, Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).

84. J. M. Flowers, K. M. Hazzouri, M. Gros-Balthazard, Z. Mo, K. Koutroumpa, A. Perrakis, S. Ferrand, H. S. M. Khierallah, D. Q. Fuller, F. Aberlenc, C. Fournaraki, M. D. Purugganan, Cross-species hybridization and the origin of North African date palms. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1651–1658 (2019).

85. S. R. Grossman, I. Shlyakhter, E. K. Karlsson, E. H. Byrne, S. Morales, G. Frieden, E. Hostetter, E. Angelino, M. Garber, O. Zuk, E. S. Lander, S. F. Schaffner, P. C. Sabeti, A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883–886 (2010).
86. H. Innan, Y. Kim, Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* **179**, 1713–1720 (2008).

87. G. F. Templeton, A two-step approach for transforming continuous variables to normal: Implications and recommendations for IS research. *Comm Assoc Info Syst* **28**, 41–58 (2011).

88. H. Mi, A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, P. D. Thomas, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v. 14.0). *Nat. Protoc.* **14**, 703–721 (2019).

89. F. Cabanettes, C. Klopp, D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

90. M. F. Cardone, P. D'Addabbo, C. Alkan, C. Bergamini, C. R. Catachcio, F. Anaclerio, G. Chiatante, A. Marra, G. Giannuzzi, R. Perniola, M. Ventura, D. Antonacci, Inter-varietal structural variation in grapevine genomes. *Plant J.* **88**, 648–661 (2016).

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