Genomic evidence supports an independent history of Levantine and Eurasian grapevines

Aviad Sivan1 | Oshrit Rahimi2 | Bar Lavi1 | Mali Salmon-Divon3 | Ehud Weiss4 | Elyashiv Drori2,5 | Sariel Hübner1

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors, Plants, People, Planet © New Phytologist Foundation
INTRODUCTION

Since ancient times, domesticated grapevine (Vitis vinifera ssp. sativa) has had an inspiring role in the culture of human societies. Among grapevine products, wine is the most popular and influential; thus, many legends and beliefs are tied with its consumption in the Mediterranean region and the Near East (This et al., 2006; Zohary et al., 2012). Today, grapevine is among the most valuable horticulture crops in the world, cultivated on more than 7 million ha. around the globe, mostly for wine production (www.fao.org/faostat).

According to archaeological evidence, grapevine was domesticated 8000–10,000 years ago in the Taurus, Caucasus, and Zagros Mountains (McGovern, 2019). However, molecular evidence from chloroplast DNA suggests at least one more domestication event outside the Near East (Arroyo-Garcia et al., 2006) albeit not conclusively (Aradhya et al., 2003). From the Caucasian region, domesticated grapevines have presumably distributed southwards to the Levant and later to Europe (McGovern, 2019; Zohary et al., 2012). Tracing the history of grapevines is challenging due to its spread as vegetative propagation material and mixtures between different genetic stocks (Myles et al., 2011; This et al., 2006). Thus, the complexity of grapevine domestication and distribution remains largely obscure (Zhou, Minio, et al., 2019).

Recently, whole-genome sequence data were generated for a few grapevine collections and allowed to study the domestication history of grapevine with higher confidence and to estimate the time since the split of the domesticated European varieties from wild grapevines to circa 22,000 years ago (Liang et al., 2019; Zhou et al., 2017). These estimates significantly predate archaeological evidence, presumably due to a missing link to the closest wild ancestral population. Thus, the exact population from which domesticated grapevine emerged is still a missing key in the domestication history of cultivated grapevines.

In the Levant region, winemaking was highly prevalent at least from the Bronze Age and contributed significantly to the culture and economy of ancient societies. Indeed, charred pips and occasionally also entire berries or raisins have been discovered in numerous pre-historic and historic sites in Europe and Southwest Asia throughout the ages (Zohary et al., 2012). Recently, Fuks et al., (2020) demonstrated by using archaeobotanical pip remains, compiled with ceramic evidence, the rise and fall of Negev Deserters’ viticulture. This important ancient industry flourished for many centuries until the collapse of the Byzantine Empire in the region during the 7th century CE (Frankel, 1997). For five centuries, the wine industry was suppressed but continued to exist until the Mamluk conquest of the Levant in the 13th century when winemaking, cultivation, and production became completely forbidden (Levanoni, 2005; Maier et al., 2003). Consequently, Levantine wine grapevine varieties which had an important role in ancient culture in this region for many centuries were abandoned and considered lost thereafter.

Recently, a comprehensive grapevine survey was conducted in an attempt to revive the ancient Levantine wine industry (Drori et al., 2015, 2017). However, it is unclear whether the collected varieties are authentic ancient Levantine varieties or rather the outcome of a recent introduction of European varieties during the 19th century.

Here, we present the analysis of whole-genome sequencing data obtained for 81 domesticated (sativa) and wild (sylvestris) accessions representing a diversity panel of Levantine and Eurasian grapevines. We provide new evidence for the history of grapevines in the Levant which support the authenticity of this material. Our results indicate that some of the grapevine varieties that were cultivated in the Levant in ancient times survived the suppression of the wine industry in the region commencing in the 7th century. In addition, genomic screening of the different populations provided evidence that the Levantine and Eurasian sativa lineages are distinguishable and that selective sweeps have affected known domestication syndrome traits in both groups in addition to group-specific loci.

MATERIALS AND METHODS

2.1 Whole-genome sequencing and variant calling

Samples were obtained from two separate sources (Tables S1 and S2). Levantine samples were selected from a recently established large Vitis collection comprising 372 accessions which were genotyped using 22 standard SSR markers (Drori et al., 2017). Based on these data, clone accessions were removed, and a representative diversity panel of 55 accessions was selected for our study including 46 accessions identified as sativa and 9 accessions identified as sylvestris. The wild grapevines, initially collected from typical habitats, were characterized by clear sylvestris phenotypes including small leaves, thin shoots, bushy architecture, and small thin clusters with small berries. The flowers of all Levantine sylvestris individuals were inspected and verified to be dioecious (Table S2). In addition, three Eurasian varieties were selected for sequencing using the same procedure for comparison. From each accession, a young leaf tissue (shoot tip) was sampled for DNA extraction using
the QIAamp DNA Micro cleanup kit (Qiagen, Valencia, USA), and a library was prepared using the NEBNext Ultra DNA library preparation kit (catalog number E7370L; NEB). Whole-genome sequencing of 150-bp paired-end reads was generated on Illumina HiSeq2500 machine to a target coverage of 20x per sample. Raw sequence data for 14 Eurasian *sativa* and 9 *sylvestris* samples were obtained from the short-read archive (SRA) project number PRJNA388292 (Zhou et al., 2017). The Eurasian samples were verified to represent their expected type (wild/domesticated); thus, suspected contaminated accessions from the Ketsch population or other mis-clustered and unclear assignments were not included (Badouin et al., 2020; Liang et al., 2019). Altogether, whole-genome sequence data of 81 accessions were obtained and analyzed.

The quality of raw sequence data for all Levantine and Eurasian samples was inspected using FastQC v0.11.8 (Simon, 2010). Low-quality reads were trimmed, and adapters were removed using Trimomatic v0.32 (Bolger et al., 2014) with default parameters. Following trimming, reads were inspected again with FastQC to guarantee only high-quality reads are included in the analysis. Clean reads from each sample were aligned to the Pinot Noir 40.024 reference genome (Jaillon et al., 2007) using BWA-MEM (Li, 2013) with default parameters. The obtained alignment files were further processed following the GATK best practice protocol (Vand der Auwera et al., 2013) and included marking duplicates with picard-tools v2.8.1 (github.com/broadinstitute/picard), realignment around indels and indexing using samtools v.1.3.1 (Li et al., 2009). Variant calling was conducted across all 81 accessions in one batch using the HaplotypeCaller program and variants were filtered using the variant quality score recalibration (VQSR) algorithm as implemented in GATK v3.6. Briefly, the VQSR uses a machine-learning algorithm to develop a model of true variants based on validated SNPs and allows the discrimination between true and false calls. We used the 20K Illumina SNP-chip data (Paslier et al., 2013) as a training set in the VQSR analysis and a minimum LOD score of 4 was set to guarantee that the highest confident SNPs are kept for downstream analyses. The obtained SNP set was further filtered to exclude sites with more than 20% missing data across all samples and a minimum minor allele frequency of 5%.

### 2.2 Population stratification analyses

For population structure analysis, we used both the FastStructure v1.0 (Raj et al., 2014) and ADMIXTURE v1.3.0 (Alexander et al., 2009) programs. For each analysis, the number of ancestral populations (K) tested ranged from 2 to 10 with 20 replicates for each K. The cross-validation procedure implemented in ADMIXTURE and the “chooseK” tool implemented in FastStructure were used to select the most likely number of clusters explaining the population structure among *Vitis* accessions.

A neighbor-joining (NJ) network was constructed with SpitsTrees4 (Huson & Bryant, 2006) using all SNPs that passed the filtering procedure. In addition, a principal component analysis (PCA) was conducted using the smartPCA program as implemented in EIGENSOFT v6.1.4 package (Price et al., 2006).

### 2.3 Demographic analyses

To infer the historical relationship including events of splits and migrations among populations, we used the TreeMix v1.13 program (Pickrell & Pritchard, 2012). To avoid sample size bias on demographic inferences, the number of accessions in each group was set to nine individuals in accordance with the smallest groups (*sylvestris* groups). In each *sativa* group, the 9 accessions with the highest ancestry assignment, as obtained by FastStructure, were selected for downstream analyses. To avoid bias introduced by genomic regions affected by nonneutral processes, the SNP data set was restricted to intergenic regions. In addition, SNPs were called in a *Vitis rotundifolia* accession (Muscadine, SRA accession number: SRR5627788; Zhou et al., 2017) which was used in the model as an outgroup for rooting the tree. To filter nonindependent SNPs in the model, windows were set to the size of twice the calculated average number of SNPs per 20 Kbp (the evaluated extent of LD across populations). Additionally, zero to four migration events were tested in the model.

To further infer the demographic history and split time in the Levantine and Eurasian populations, we used the likelihood-free SMC++ program v1.15.2 (Terhorst et al., 2017). The SMC++ program allows to leverage information from multiple individuals of each population and infer changes in the effective population size also in recent history. A mutation rate of $\mu = 5.4 \times 10^{-9}$ mutations per base-pair per generation (Liang et al., 2019) and a generation time of 3 years were assumed in all models. The input data from each population were filtered for long stretches of homozygosity (>20 Kbp), and the cross-validation (CV) module was used to infer the effective population sizes with 10-fold CV steps. A polarization error rate was set to 0.5 to allow uncertainty on the identity of the ancestral allele. Finally, the “clean-split” model was used with default parameters to estimate split times between pairs of populations.

In addition, the MSMC v2 program (Schiffels & Durbin, 2014) was used to estimate changes in the effective population size ($N_e$) in each population over time. Four individuals were randomly chosen from each population after phasing and imputing the entire SNP data set using BEAGLE v5.1 (Browning & Browning, 2007). A mutation rate of $\mu = 5.4 \times 10^{-9}$ mutations per base-pair per generation and a generation time of three years were set in the model (Liang et al., 2019).

### 2.4 Pedigree network analysis

To explore the breeding history of grapevine vitreis in the Levant and their relationship to the Eurasian varieties, a pair-wise
identity by descent (IBD) was calculated across all domesticated *sativa* accessions (Eurasia and Levant). The refined-IBD program in BEAGLE v5.1 (Browning & Browning, 2013) was used to calculate the pair-wise IBD, and the results were converted to a kinship score using the “relatedness” tool as implemented in BEAGLE v5.1 (Browning & Browning, 2013). A relatedness threshold of 0.466 was set following a previous study (Myles et al., 2011) on the familial relationship among grapevine varieties based on IBD scores calculated in the program plink (Purcell et al., 2007). To adjust the relatedness scores calculated from the refined-IBD to the calculated threshold, a Spearman correlation analysis was conducted between the refined-IBD and the plink scores which were computed for each pair of accessions. The obtained pair-wise IBD matrix was visualized with the network analysis program Cytoscape v3.8.2 (Shannon et al., 2003).

### 2.5 | Population genomics statistics

Linkage disequilibrium (LD) was computed using the PopLDdecay software (Zhang, Dong, et al., 2019) across all accessions and for each population separately. LD was also measured within each population at 1 Mbp windows using plink (Purcell et al., 2007). Population diversity statistics were calculated in 1 Mbp windows for each group separately using the PopGenome package (Pfeifer et al., 2014) and included nucleotide diversity (π), Tajima’s D, and Watterson’s θ. Observed heterozygosity was obtained for each population using VCFtools v0.1.15 (Danecek et al., 2011). D-statistics were calculated using the admixr package (Petr et al., 2019) with the Levantine *sativa* group set as W and rotating all other groups to test all possible scenarios of allele sharing. The analysis was conducted for the selected 9 accessions from each group.

Genome scans for the footprints of selective sweeps were conducted within each domesticated *sativa* group. Within each group, the ω-statistic was calculated using the RAiSD software for 9 accessions from each group (Alachiotis & Pavlidis, 2018). The ω-statistic is a composite score of the changes in site frequency spectrum (SFS), linkage disequilibrium, and genetic diversity. Top-ranked windows (>99.95%) were considered outliers. Overlapping windows with a maximum gap of 20 Kbp were merged with BEDtools v2.26.0 (Quinlan & Hall, 2010) to allow a comparison of genomic regions between groups.

### 2.6 | Genetic load estimation

To estimate the genetic load in each accession, we first identified nonsynonymous mutations using the *Vitis vinifera* reference genome (NCBI_Assembly: GCF_000003745.3) and associated annotation files (annotations release 102, GCF_000003745.3_12X) to build a SIFT genomic database in SIFT4G (Vaser et al., 2016). The identified nonsynonymous mutations within coding regions were considered as potentially deleterious if the obtained SIFT score was lower than 0.05. To avoid a reference bias effect on the genetic load predictions, alleles identified also in the outgroup species *Vitis rotundifolia* were not considered deleterious. The genetic load was calculated by summing the number of deleterious alleles in each accession with a score of one for heterozygote and two for homozygote deleterious alleles.

### 3 | RESULTS

#### 3.1 | Population stratification in Eurasian and Levantine grapevines

A panel of 81 grapevine accessions was obtained from two independent sources (Tables S1 and S2). The Levantine panel included 46 and 9 putative domesticated (*sativa*) and wild (*sylvestris*) accessions, respectively (Figure 1a). These accessions were previously confirmed to be unique material based on SSR markers (Drori et al., 2017) and were whole-genome sequenced (WGS) to a coverage of 20x. In addition, we sequenced to the same coverage 3 Eurasian *sativa* accessions and obtained publicly available raw WGS data for additional 14 Eurasian *sativa* varieties and 9 Eurasian *sylvestris* accessions from the SRA repository. Altogether, 2.1 Tera base-pairs were obtained for 81 grapevine accessions representing domesticated *sativa* and wild *sylvestris* types of Levantine and Eurasian origin. High quality trimmed reads from all accessions were aligned to the Pinot Noir PN40024 reference genome followed by a variant calling procedure, yielding a total of 26,083,120 high quality (QUAL > 20) SNPs across all accessions. To further reduce the false positive rate among called variants, a machine-learning filtering approach was implemented, and a set of 1,824,029 robustly called SNPs was kept for downstream analyses.

To investigate the population stratification among Levantine and Eurasian accessions, model-based analyses were conducted using fastStructure and ADMIXTURE (Figure 1b; Figure S1 and S2). Considering both analyses, the optimal clustering was obtained at K = 4 in accordance with the geographic origin and type of accessions: Levantine *sativa*, Levantine *sylvestris*, Eurasian *sativa*, and Eurasian *sylvestris*. Signs of admixture were observed among all four groups with the highest rate of admixed individuals in the Eurasian *sylvestris* group (n = 9, 100%), and lowest admixture rate in the Levantine *sylvestris* (n = 9, 33%). The difference in the level of admixture detected in each of the two *sylvestris* groups conceivably reflects the distribution range and associated inbreeding in each group, that is, broad and narrow geographic range represented in the Eurasian and Levantine groups, respectively.

Overall, the assignment of accessions to clusters as identified in the population stratification analyses well supported previous characterization records with minor exceptions (Drori et al., 2017; Liang et al., 2019). Two accessions in the Eurasia *sativa* group (“Thompson-RLK” and “Thompson2A”) known as table-grapes varieties were assigned to a distinct cluster from the remaining wine-grapes type. The same pattern was also observed in three Levantine *sativa* accessions...
which were assigned to the same cluster as the two “Thompson” varieties based on the fastStructure analysis and are also considered table-grape varieties. These results support previous reports for both Eurasian (Zhou et al., 2017) and Levantine accessions (Drori et al., 2017) that table-grapes cluster separately from wine and wild grapevines. Strong signs of admixture

FIGURE 1 Population structure among all Eurasian and Levantine, domesticated sativa and wild sylvestris, grapevine accessions. (a) Geographic map and locations where Levantine sativa (red) and sylvestris (orange) accessions were collected; (b) population structure among the 81 Eurasian and Levantine accessions. Analysis was conducted with FastSTRUCTURE and the barplot represents K = 4. Accessions are sorted by their expected group of origin (top-bottom): domesticated Eurasian sativa, wild Eurasian sylvestris, domesticated Levantine sativa, and wild Levantine sylvestris; and (c) neighbor-joining network representing the resemblance between Eurasian sativa (blue), Eurasia sylvestris (purple), Levantine sativa (red), and Levantine sylvestris (orange) (“Tamar-H1,” “Tamar-H2,” and “Suka”) which were assigned to the same cluster as the two “Thompson” varieties based on the fastStructure analysis and are also considered table-grape varieties. These
were observed in specific accessions from both the Eurasian and Levantine *sativa* groups. For example, among the Levantine *sativa* accessions, a high rate of admixture with Levantine *sylvestris* (65%) was observed for the "Buffalo" accession, presumably due to recent hybridization with wild Levantine grapevine although historical admixture and maintenance through vegetative propagation is also plausible (Figure 1b). Among the Eurasian *sativa* accessions, high rates of admixture were observed in "Zinfandel"/"Primitivo" (56%), "Muscat of Alexandria," and "Carignan" (74%).

To further investigate the level of divergence between the four grapevine groups, a neighbor-joining network was built using the SNP dataset called across all 81 accessions (Figure 1c). The split into four groups in the network analysis supported the results of the model-based stratification analyses. Moreover, the obtained network clearly discriminated between the Eurasian and Levantine groups, implying that the Eurasian *sativa* group has branched from the Eurasian *sylvestris* group while the Levantine *sativa* group has branched from the Levantine *sylvestris* group. This pattern of divergence suggests that the Eurasian and Levantine *sativa* lineages do not share the same domestication history and may have developed in independent processes. To further confirm this observation, a PCA was conducted and designated the same pattern of divergence as obtained in the neighbor-joining network and model-based stratification analyses (Figure S3). In addition, the network analysis well supported the misassignment of accessions identified by the model-based stratification analyses. Interestingly, one wild white-berry accession collected near the Sea of Galilee (“Majrase”) was confirmed as a *sylvestris* type by all analyses. White-berry is a common phenotype among domesticated grapevines and considered a post-domestication characteristic (Migicovsky et al., 2017). A white-berry phenotype in wild grapevine is possibly the result of introgression from cultivated *sativa*; however, none of the analyses conducted supported signs of introgression from cultivated grapevines into this wild white-berry accession.

Next, the level of nucleotide diversity (θ), Watterson’s θ, Tajima’s D, observed heterozygosity, and linkage disequilibrium (LD) were investigated in each group after excluding five misassigned accessions (defined as less than 10% assignment to their expected cluster). Linkage disequilibrium analysis conducted across all *Vitis* accessions indicated that decay is reached at 20 Kbp (Figure S4; Table S3). Within groups, steep LD decay and high genetic diversity were observed among the Eurasian *sylvestris* (θ = 0.013 ± 0.002, LD = 8 Kbp) compared with the Eurasian *sativa* group (θ = 0.012 ± 0.0101, LD = 10 Kbp) which is characterized by a significantly lower diversity (θ = 3.84, p = 0.001) due to domestication bottleneck (Table S3). An opposite trend was observed among the Levantine groups, that is, steep LD decay and high genetic diversity within the Levantine *sativa* group (θ = 0.012 ± 0.001, LD = 6 Kbp) compared with Levantine *sylvestris* (θ = 0.011 ± 0.001, LD = 25 Kbp) which was also characterized by a significantly lower diversity (θ = 6.35, p = 5.92 × 10⁻³). This pattern among the Levantine groups is attributed to the constrained geographic range of wild grapevines at its southern distribution edge. In addition, both the population stratification and genetic diversity analyses imply that the Levantine *sativa* group is likely a mixture of two distinct subgroups (Figure 1b). Thus, the domesticated Levantine *sativa* group was further split into two subgroups (Table S3), a homogenous Levantine *sativa*, and a highly diverse group with signs of admixture with other genetic sources.

### 3.2 | The demographic history of Eurasian and Levantine grapevines

To test the hypothesis that Eurasian and Levantine *sativa* are distinct lineages that were developed in two independent processes, various complementary demographic analyses were conducted. To reduce the confounding effect of differences in sample size and level of admixture across groups, nine representative accessions were selected from each group to adjust the sample size to the smallest groups (nine accessions in Eurasia *sylvestris* and Levantine *sylvestris*). First, the D-statistic was used to test for signs of ancestry in the Levatine *sativa* group from each of the other three groups. No significant contribution of alleles (D = 0.025, Z = 1.85) was identified from the Eurasian *sylvestris* or Eurasian *sativa* groups while a significant gene-flow was identified between the Levantine *sylvestris* and the Levatine *sativa* group (D = 0.123, Z = 13.5). To further explore historical splits and gene-flow among the four groups, a graphical model-based analysis was conducted using TreeMix. The SNP dataset was restricted in this analysis to intergeneric regions to reduce the effect of selective sweeps on the demographic inferences; thus, a total of 1,055,512 SNPs were kept for the analysis. The model uses the allele frequencies in modern populations and a Gaussian approximation to infer historical demographic events. First, the model was carried out without migration (f_{index} = 0.996) and the obtained graph supported the hypothesis of independent demographic history (Figure 2a). Although the model does not allow to precisely estimate times of split events, it was noted that domestication in Eurasia predated the development of the Levantine *sativa*. Allowing one migration event in the model improved its likelihood (f_{index} = 0.999) and indicated a historical gene-flow from Levantine *sativa* into the Eurasian *sativa* group (Figure 2b). This inference supports previous reports on the possible exchange of germplasm by the Romans, Crusaders, or Islamic rulers (This et al., 2006). Allowing more migration events in the model did not improve its likelihood (Figure S5).

To further investigate the demographic history of grapevines in Eurasia and the Levant, a coalescence analysis was conducted using the SMC++ program which allows to incorporate information from all nine individuals representing each group. To infer the demographic changes on a timescale, a generation time of 3 years and a mutation rate of 5.4 × 10⁻⁹ per nucleotide per generation (Liang et al., 2019) were used in the model (Figure 2c).

The SMC++ analysis was conducted and inspected with and without the "clean-split" model to allow a better interpretation of the demographic process. In all models, the divergence between groups predated the "clean-split" model indicating that grapevine
domestication was prolonged over a period of time until the two groups were fixed as distinct lineages. In agreement with previous studies (Liang et al., 2019; Zhou et al., 2017), the SMC++ model denoted that the divergence between Eurasian *sylvestris* and *sativa* grapevines commenced circa 30,000 years ago and reached a clear differentiation circa 15,000 years ago based on the "clean-split" model (Figure 2c). Conducting the analysis for the Levantine *sativa* group indicated that divergence from Eurasian *sylvestris* and *sativa* predates the estimated domestication event in Eurasia (20,000 and 17,000 years ago, respectively). The same demographic analysis conducted for Levantine *sylvestris* and *sativa* groups indicated that divergence between them commenced approximately 15,000 years ago and reached a clear differentiation 10,000 years ago according to the "clean-split" model. Based on the demographic analyses, it is unclear which wild population is the direct progenitor of the Eurasian nor the Levantine *sativa*. Nevertheless, the demographic inferences for the Levantine *sativa* group provide evidence that these accessions are authentic Levantine varieties that sustained the suppression of grapevines cultivation and winemaking industry in this region. Arguably, these accessions are the descendants of the lost Levantine wine varieties that were cultivated in this region in ancient times.
To further validate these results, a second analysis was conducted with the MSMC software using four individuals that were randomly sampled from each population. The obtained trajectories in the effective population size of each group over time supported the observed split between the Eurasian *sylvestris* and *sativa* grapevines circa 20,000 years ago and a more recent split between the Levantine *sylvestris* and *sativa* grapevines circa 9,000 years ago. The MSMC inferences well support the results obtained by the SMC++ and TreeMix analyses (Figure S6).

3.3 Pedigree network in domesticated Eurasian and Levantine varieties

To better understand the recent history and pedigree relationship among grapevine varieties, a relatedness network was constructed for all 58 domesticated (*sativa*) accessions of both Levantine and Eurasian origins. The relatedness matrix was computed at 1-cM segments using the refined-IBD tool which allows to identify traces of shared ancestry signatures. This approach has the advantage of distinguishing between “old” ancestry (short segments) and recent ancestry (long segments) signatures. The minimum threshold to delineate a parent-offspring or sibling relationship was determined based on a confirmed cutoff (IBD = 0.466) which was computed using a short segments ancestry detection approach (Myles et al., 2011). To link between this confirmed cutoff and the calculated refined-IBD score, a correlation analysis was conducted between overlapping pairs of accessions presented in both studies ($\rho = 0.62, p = 0.028$). To graphically visualize the pedigree network, a first-degree relationship graph was constructed for all accessions that had at least one pedigree link with other accessions (Figure 3). Overall, the Levantine *sativa* formed a distinguished cluster from the Eurasian *sativa* group which further support the independent history of the Levantine domesticated accessions. Clonal relationships were identified in both the Eurasian (“Zinfandel”/“Primitivo”) and Levantine (“Marawi”/“Marawi-GB”) groups. Interestingly, a pedigree relationship was observed between the European “Chardonnay” variety and a Levantine accession sampled at the Galilee (“Hadari”, IBD = 0.587).

3.4 Footprints of selection in cultivated verities

To investigate whether the independent history of the Levantine and Eurasian *sativa* resulted in distinct patterns of selective sweeps, we performed a genome scan analysis using the $\mu$-statistic score. To maintain a balanced comparison between groups, the analysis was performed for 9 accessions that were found to properly represent each group based on the population stratification analysis. Despite the small sample size, high SNP density can compensate at least some of the confounding effects (Willing et al., 2012). The $\mu$-statistic is a composite score combining the site frequency spectrum (SFS), linkage disequilibrium, and genetic diversity calculated at overlapping windows that are adaptively determined according to the calculated metrics. We considered a significant outlier window following the default cutoff of 99.95% (Figure 4). Altogether, 5,581 outlier windows were detected in the Eurasian *sativa* group and 6,333 outlier windows were detected in the Levantine *sativa*. The average $\mu$-statistic in the

---

**FIGURE 3** Relatedness among domesticated Eurasian (green) and Levantine (blue) *sativa* accessions. Relatedness was calculated from identity by descent (IBD) in 1cM fragments across all 58 *sativa* accessions. In the network, only links higher than 0.466 are indicated, where dashed lines represent weak links and solid lines represent an IBD values greater than 0.5. The thickness of the line corresponds to the calculated IBD value.
Levantine *sativa* group was significantly higher than in the Eurasian *sativa* group ($t = 10.6, p < 0.0001$). The overall stronger signal of selective sweep detected in the Levantine *sativa* group implies that this group has experienced stronger sweep by selection or that this group experienced a genetic bottleneck more recently than the Eurasian group. To identify footprints of a selective sweep as a result of domestication and breeding, overlapping windows were merged and extended to form swept regions (SRs). Merging was performed only...
for outlier windows found in distance shorter than 20 Kbp in accordance with LD decay in *Vitis*. A total of 222 and 260 outlier SRs were detected across all chromosomes in the Eurasian and Levantine *sativa* groups, respectively (Table S4). Among the identified SRs, 80 were found in both groups across all chromosomes except for chromosome 8 where no SR was detected in either group. Enrichment in the number of SRs was observed in the Eurasian *sativa* group on chromosomes 12 (26 SRs), 13 (17 SRs), and 19 (22 SRs), and a strong signal of selection \( (\mu > 100) \) was observed on chromosomes 7 \((\mu = 103.1)\) and 13 \((\mu = 145.5)\). In the Levantine *sativa* group, enrichment in the number of SRs was observed on chromosomes 4 (25 SRs), 5 (32 SRs), and 18 (21 SRs), and a strong signal of selection \( (\mu > 100) \) was observed on chromosomes 2 \((\mu = 110.8)\), 11 \((\mu = 751.5)\), 12 \((\mu = 106)\), 17 \((\mu = 409.8)\), and 18 \((\mu = 150.4)\).

Several candidate genes were found within outlier SRs in both Eurasian and Levantine *sativa*. For example, the genes resveratrol synthase and stilbene synthase that are involved in the response to biotic stress as well as in taste and aroma (Cantu & Walker, 2019) were found in an overlapping SR on chromosome 16 in both groups. Another example was observed on chromosome 2 in both *sativa* groups where bifunctional nitrilase/nitrile hydratase genes were found. These genes were recently targeted elsewhere for their potential role in grapevine domestication (Marrano et al., 2018; Zhou et al., 2017). Within the Levantine *sativa* group, a strong signal of selective sweep was observed on chromosome 18 where a cluster of phenylalanine ammonia-lyase (PAL) genes was identified. PAL genes contribute to anthocyanin concentration in the berry pericarp which affects berry color and wine quality in addition to enhancing resistance to biotic and abiotic stresses (Roubelakis-Angelakis & Kliwer, 1986). In the same SR located on chromosome 18, we also detected an RPW8 gene that confers a basal resistance to powdery mildew in *Arabidopsis* (Xiao et al., 2005). Other SRs that were detected exclusively in the Levantine *sativa* group were on chromosome 17 where several abiotic stress-responsive genes were identified including the basic helix-loop-helix transcription factor (Gao et al., 2019) and HVA22-like gene (Chen et al., 2002; Shen et al., 2001).

Outlier SRs that were detected exclusively in the Eurasian *sativa* group harbored the anthocyanin synthesis genes (MYBA1 and MYBA3) on chromosome 2 (Fischer et al., 2004; Salmaso et al., 2008), the pathogen response gene (ACD6) on chromosome 16 (Zhang et al., 2014), and the disease response genes (RPM1) on chromosome 7 which were described to be up-regulated in response to pathogen infection in grapevine (Zhang, Yan, et al., 2019).

### 3.5 Genetic load among cultivated grapevines

To evaluate the extent of genetic load due to the accumulation of deleterious mutations along the genome in each of the *sativa* groups, we used the SIFT4G software. A total of 92,936 nonsynonymous variants were detected of which 37,635 were predicted as deleterious (SIFT score < 0.05). To correct for potential bias introduced by the use of a reference genome, alleles identified also in the outgroup species *V. rotundifolia* were not considered deleterious. Altogether, 29,386 sites identified as deleterious across all accessions were used to calculate the genetic load in each group. The Eurasian *sylvestris* group represents a broad distribution range and a composite source of alleles, thus the variation observed among individuals was too substantial to reach conclusions regarding the genetic load in this group. Nevertheless, the results in the remaining groups were clearer and pointed the main differences between groups. Not surprisingly, the calculated genetic load in the Levantine *sativa* group \((7,142 \pm 101)\) was significantly higher \((t = 4.51, p = 1.70 \times 10^{-5})\) than in the Levantine *sylvestris* group \((6,228 \pm 598)\) due to the effect of domestication bottleneck. Moreover, significantly higher genetic load \((t = 14.70, p = 1.24 \times 10^{-8})\) was observed in the Levantine *sativa* group compared with the Eurasian *sativa* \((5,168 \pm 390)\). This observation supports the results obtained in the selective sweep analysis indicating that the Levantine group was under stronger selective pressure or was fixed as a distinct group more recently.

### 4 Discussion

Along the history of human societies, wine has provided a special cultural flavor. According to archaeological observations, the grapevine was domesticated and spread by ancient societies in the Near East (circa 10,000 BC), and later was introduced to East Asia, the Mediterranean basin and Europe (Liang et al., 2019; McGovern, 2019).

In the Levant region, grapevine cultivation has flourished for several millennia until the collapse of the Byzantine Empire during the 7th century CE (Frankel, 1997). Since then, wine production has declined and eventually was abandoned under the Mamluk Empire conquest, and the ancient grapevine varieties were considered lost (Levanoni, 2005; Maier et al., 2003). In this study, we provide new genomic evidence for the demographic history of grapevine varieties in the Levant, their origin, and the genomic landscape of their fixation as a distinct group.

#### 4.1 The origin of cultivated grapevines in the Levant

To study the history of Levantine varieties, we compared whole-genome sequence data from *sylvestris* and *sativa* types of Levantine and Eurasian origin. All population stratification analyses supported the deviation into four distinct groups by type (*sylvestris/sativa*), and geography (Levant/Eurasia). Moreover, the clustering pattern implied that the Eurasian *sativa* group has branched from the Eurasian *sylvestris* group while the Levantine *sativa* group branched from the Levantine *sylvestris* group, with few minor exceptions (Figure 1). In accordance with previous studies (Zhou et al., 2017), the Eurasian table-grapes varieties
“Thompson” and “Muscat of Alexandria” are distinguished from remaining varieties and a similar pattern was also observed among three Levantine *sativa* accessions which presumably represent introduced table-grapes varieties (“Tamar-H1,” “Tamar-H2,” “Suka”). Also, one Levantine accession was identified as a potential feral grapevine (“Buffalo”), and one white-berry accession (“Majrase”) was confirmed to be of wild origin. The latter could be an interesting example of a sporadic occurrence of white-berry mutation in the wild, as we failed to identify signs of admixture of this accession with other domesticated varieties.

The process of domestication can be generally divided into three phases which include management of wild material, selection of desirable basic domestication traits, and dispersal of the domesticated material. During the third stage of domestication, introgression from local wild populations can increase adaptation of the alien crop to the local environment (Purugganan, 2019; Zhou, Minio, et al., 2019). These introgressions may be spread across the genome yet the genetic background should reflect the origin of the crop. Once the genetic turnover is so profound that the origin is masked by introgression, the question of the origin of domesticates becomes quantitative. The demographic analyses conducted in this study supported a distinct origin of the Levantine *sativa* from the Eurasian *sativa*. In accordance with previous studies (Liang et al., 2019; Zhou et al., 2017), our analyses pointed that the divergence of Eurasian *sativa* from *sylvestris* commenced approximately 30,000 years ago; however, the “clean-split” model in SMC++ provided an improved estimate for the domestication of grapevine in Eurasia to approximately 15,000 years ago. It should be noted that these models are limited in their ability to estimate recent demographic events especially in the past few centuries (Terhorst et al., 2017). Therefore, it is difficult to infer robustly from these models whether the Levantine accessions are truly descendants of ancient varieties or the outcome of a recent introduction of grapevine varieties from Eurasia. The pedigree and genome scans analyses imply these are indeed authentic Levantine varieties. Nevertheless, to distinguish between incomplete lineage sorting and post-domestication admixture with high confidence it is necessary to obtain a collection that was sampled continuously throughout the distribution range of wild grapevine.

### 4.2 Pedigree relationship between Eurasian and Levantine varieties

To test for potential recent admixture between Levantine and Eurasian *sativa*, and the effect of vegetative propagation on the similarity between cultivars, a relatedness network was constructed based on pair-wise identity by descent (IBD) analysis. In accordance with previous studies (Myles et al., 2011) conducted for the Eurasian *sativa* group, clonal relationship was observed among the Eurasian accessions and also among the Levantine accessions but no pedigree links were observed between the two groups with one exception. A pedigree link between the European “Chardonnay” variety and a Levantine accession sampled at the Galilee region (“Hadari”) imply a potential Levantine ancestry for “Chardonnay.” Previously, ampelographers suggested that “Chardonnay” has ancestral roots in the Levant, although it was later contradicted by genetic analysis which indicated that “Chardonnay” was produced by a cross between “Pinot Noir” and “Gouais blanc” (Bowers et al., 1999; Hunt et al., 2010). While “Pinot Noir” is a confirmed French variety, “Gouais blanc” is considered an introduced variety from elsewhere. The results obtained from the pedigree analysis do not allow to track back the entire lineage of “Chardonnay” and how it is linked to the Levant, but to the best of our knowledge, this is the first genomic evidence for this ancestry relationship.

### 4.3 Development of grapevine varieties in Eurasia and the Levant

All population stratification, demography, and pedigree analyses conducted supported the hypothesis that Levantine *sativa* evolved in a distinct route from the Eurasian *sativa*. To test whether these independent histories are reflected in different genomic footprints of selective sweeps, genome scans were conducted for each *sativa* group. Although the small sample size used in the genome scan analyses may have limited the power of these statistics, footprints of selective sweeps were detected across all chromosomes in both groups with many overlapping SRs. However, several SRs were group-specific and included candidate genes that are involved in biotic and abiotic stress resistance. Environmental stress has always troubled farmers and breeders around the world, however, since the pathogen identity and type of stress varies between regions, so are the resistance genes selected to confront them.

This was further supported by the genetic load analysis which pointed to higher similarity among the Levantine *sylvestris* and *sativa* groups than to the Eurasian group. The high inbreeding in the Levantine *sylvestris* is reflected in an elevated genetic load which is maintained also in the Levantine *sativa* population as expected between a progenitor and a domesticate (Gaut et al., 2018). Interestingly, the Eurasia *sativa* has lower genetic load compared with the Levantine populations despite the domestication bottleneck. These results support previous indications that the transition to clonality may compensate the genetic load associated with domestication (Gaut et al., 2018).

The revolution of genomics provides powerful tools to fill gaps in the evolutionary history of crops at multiple genomic levels (Zhou, Muylo, et al., 2019). However, the use of a single reference genome may introduce some bias towards the genetic cluster from which the reference individual originated. Thus, the genomic differences observed between groups may be increased if multiple reference genomes are used to study the history of domestication. In addition, although coalescence models can compensate to some extent for small population sample, inferences may be biased by sampling especially when populations are restricted in
distribution (e.g., Levantine sylvestris group). The small sample size of the Levantine sylvestris population also reflects the impact of anthropogenic activity on the distribution of wild grapevines in the Levant. Despite their importance in human culture and economy, many ancient Levantine grapevine varieties were considered lost for many centuries. Here, we provide evidence that some of the Levantine ancient grapevine varieties survived, in some cases under harsh conditions, despite the cultural turnovers in this region throughout history. Nevertheless, to link between the discovered Levantine accessions and ancient varieties we need to obtain high-quality sequence data from archaeological samples which may be in reach for grapevines.

ACKNOWLEDGMENTS

This work was supported by the JNF grant # 90-23-020-12 and the Israeli Ministry of Science and Technology (MOST).

AUTHOR CONTRIBUTIONS

ED, EW, and SH designed this study; AS performed the research; AS, OR, BL, and ED analyzed the data; and AS, ED, EW, and SH wrote the manuscript.

DATA AVAILABILITY STATEMENT

All sequence data are available through the SRA repository under the manuscript.

ORCID

the bioproject accession PRJNA647155.

REFERENCES

Alachiotis, N., & Pavlidis, P. (2018). RAISD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. Communications Biology, 1, 1. https://doi.org/10.1038/s42003-018-0085-8

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 1655–1664. https://doi.org/10.1101/gr.094052.109

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

Arroyo-Garcia, R., Ruiz-Garcia, L., Bolling, L., Ocete, R., Lopez, M. A., Arnold, C., Ergul, A., Söylemez, L. G., Uzun, H. I., Cabello, F., & Ibanez, J. (2006). Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. sativa) based on chloroplast DNA polymorphisms. Molecular Ecology, 15, 3707–3714.

Badouin, H., Velt, A., Gindraud, F., Flutre, T., Dumas, V., Vautrin, S., Marande, W., Corbi, J., Sallet, E., Ganofsky, J., Santoni, S., Guyot, D., Ricciardelli, E., Jepsen, K., Käfer, J., Berges, H., Duchêne, E., Picard, F., Hugueney, P., ... Marais, G. A. B. (2020). The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. Genome Biology, 21, 1–24. https://doi.org/10.1186/s13059-020-02131-y

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bowers, J., Boursiquot, J. M., This, P., Chu, K., Johansson, H., & Meredith, C. (1999). Historical genetics: The parentage of Chardonnay, Gamay, and other wine grapes of northeastern France. Science, 285, 1562–1565. https://doi.org/10.1126/science.285.5433.1562

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics, 194, 459–471. https://doi.org/10.1534/genetics.113.150029

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics, 81, 1084–1097. https://doi.org/10.1086/521987

Cantu, D., & Walker, M. A. (2019). The grape genome. Springer.

Chen, C. N., Chu, C. C., Zentella, R., Pan, S. M., & Ho, T. H. (2002). ATHVA22 gene family in Arabidopsis: Phylogenetic relationship, ABA and stress regulation, and tissue-specific expression. Plant Molecular Biology, 49, 631–642.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., & McVean, G. (2011). The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.

Drori, E., Rahimi, O., Henig, Y., Lorenzi, S., Brauner, H., Marrano, A., Amar, Z., Netzer, Y., Failla, O., & Grando, M. S. (2015). Ampelographic and genetic characterization of an initial Israeli grapevine germplasm collection. Vitis., 54, 107–110.

Drori, E., Rahimi, O., Marrano, A., Henig, Y., Brauner, H., Salmon-Divon, M., Netzer, Y., Prazzoli, M. L., Stanevsky, M., Failla, O., & Weiss, E. (2017). Collection and characterization of grapevine genetic resources (Vitis vinifera) in the Holy Land, towards the renewal of ancient winemaking practices. Scientific Reports, 7, 44463.

Fischer, B. M., Salakhutdinov, I., Akkurt, M., Eibach, R., Edwards, K. J., Töpfer, R., & Zyprian, E. M. (2004). Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. Theoretical and Applied Genetics., 108, 501–515. https://doi.org/10.1007/s00122-003-1445-3

Frankel, R. (1997). Presses for oil and wine in the southern Levant in the Byzantine period. Dumbarton Oaks Papers, 51, 73–84. https://doi.org/10.2307/1291762

Fuchs, D., Bar-Oz, G., Tepper, Y., Erickson-Gini, T., Langgut, D., Weissbrod, L., & Weiss, E. (2020). The rise and fall of viticulture in the Late Antique Negev Highlands reconstructed from archaeobotanical and ceramic data. Proceedings of the National Academy of Sciences, 117, 19780–19791. https://doi.org/10.1073/pnas.1922200117

Gao, M., Zhu, Y., Yang, J., Zhang, H., Cheng, C., Zhang, Y., Wan, R., Fei, Z., & Wang, X. (2019). Identification of the grape basic helix-loop-helix transcription factor family and characterization of expression patterns in response to different stresses. Plant Growth Regulation, 88, 19–39. https://doi.org/10.1007/s10725-019-00485-3

Gaut, B. S., Seymour, D. K., Liu, Q., & Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. Nature Plants, 4, 512–520. https://doi.org/10.1038/s41477-018-0210-1

Hunt, H. V., Lawes, M. C., Bower, M. A., Haeger, J. W., & Howe, C. J. (2010). A banned variety was the mother of several major wine grapes. Biology Letters, 6, 367–369. https://doi.org/10.1098/rspb.2009.0810

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution, 23, 254–267. https://doi.org/10.1093/molbev/msj030

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., & Vezzi, A. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449, 463.

Le Paslier, M. C., Choisne, N., Scalabrini, S., Bacilieri, R., Berard, A., Bounon, R., Boursiquot, J. M., Bras, M., Brunel, D., Chauveau, A., & D’i Gaspero, G. (2013). The GRAPERESEQ 18K Vitis genotyping chip. In IX International Symposium on Grapevine Physiology & Biotechnology. Retrieved from https://hal.inrae.fr/view/index/identifiant-hal-02811055
in grapevine domestication. *Nature Plants*, 5, 965–979. https://doi.org/10.1038/s41477-019-0507-8

Zhou, Y., Muyle, A., & Gaut, B. S. (2019). Evolutionary genomics and the domestication of grapes. In D. Cantu & M. Walker (Eds.), *The grape genome. Compendium of Plant Genomes*. (pp. 39–55). Springer. https://doi.org/10.1007/978-3-030-18601-2_3

Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of plants in the old world: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford University Press. Retrieved from https://oxford.universitypressscholarship.com/view/10.1093/acprof:osobl/9780199549061.001.0001/acprof-9780199549061

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Sivan A, Rahimi O, Lavi B, et al. Genomic evidence supports an independent history of Levantine and Eurasian grapevines. *Plants, People, Planet*. 2021;3:414–427. https://doi.org/10.1002/ppp3.10197