RESEARCH ARTICLES

PLANT GENETICS

Dual domestications and origin of traits in grapevine evolution

Yang Dong[1,2]†, Shengchang Duan[1,2]†, Qiuju Xia[3]†, Zhenchang Liang[4]†, Xiao Dong[1,2]†§, Kristine Margaryan[5,6]‡, Mirza Musayev[7]‡, Svitlana Goryslavets[8]‡, Goran Zdunić[9]‡, Pierre-François Bert[10]‡, Thierry Lacombe[11]‡, Erika Maul[12]‡, Peter Nick[13]‡, Kakha Bitskinashvili[14]‡, György Dénes Bisztray[15]‡, Elyashiv Drori[16,17]‡, Gabriella De Lorenzis[18]‡, Jorge Cunha[19,20]‡, Carmen Florentina Popescu[21]‡, Rosa Arroyo-Garcia[22]‡, Claire Arnold[23]‡, Ali Ergül[24]‡, Yifan Zhu[1]‡, Chao Ma[25]‡, Shufen Wang[1,2], Siqi Liu[1,2], Liu Tang[1,2], Chunping Wang[1,2], Dawei Li[1,2], Yunbing Pan[1,2], Jingxian Li[1,2], Ling Yang[1,2], Xuzhen Li[1,2], Guisheng Xiang[1,2], Zijiang Yang[1,2], Baozheng Chen[1,2], Zhanwu Dai[4], Yi Wang[4], Arsen Arakelyan[5,26,27], Varis Kuliyev[28], Gennady Spotar[8], Nabil Girollet[10], Serge Delrot[10], Nathalie Ollat[10], Patrice This[11], Cécile Marchal[29], Gautier Sarah[11], Valérie Laucou[11], Roberto Bacilieri[11], Franco Röckel[12], Pingyin Guan[13], Andreas Jung[30], Michael Riemann[13], Levan Ujmajuridze[14], Tekle Zakalashvili[14], David Maghradze[14], Maria Höhn[15], Gizella Jahnke[15], Erzsébet Kiss[15], Tamás Deák[15], Oshrit Rahimi[16], Sariel Hübner[31], Fabrizio Grassi[32,33], Francesco Mercati[34], Francesco Sunseri[35], José Eiras-Dias[19,20], Anamaria Mirabela Dumitru[21], David Carrasco[22], Alberto Rodriguez-Izquierdo[22], Gregorio Muñoz[36], Tamer Uysal[37], Cengiz Özer[37], Kemal Kazan[38], Meilong Xu[39], Yunyue Wang[1], Shusheng Zhu[1], Jiang Lu[40], Maoxiang Zhao[25], Lei Wang[25], Songtao Jiu[25], Ying Zhang[41], Lei Sun[41], Huanming Yang[42], Ehud Weiss[43], Shiping Wang[25], Youyong Zhu[1], Shaohua Li[4]*, Jun Sheng[1,2]*, Wei Chen[1,2]*

We elucidate grapevine evolution and domestication histories with 3525 cultivated and wild accessions worldwide. In the Pleistocene, harsh climate drove the separation of wild grape ecotypes caused by continuous habitat fragmentation. Then, domestication occurred concurrently about 11,000 years ago in Western Asia and the Caucasus to yield table and wine grapevines. The Western Asia domesticates dispersed into Europe with early farmers, introgressed with ancient wild western ecotypes, and subsequently diversified along human migration trails into muscat and unique western wine grape ancestries by the late Neolithic. Analyses of domestication traits also reveal new insights into selection for berry palatability, hermaphroditism, muscat flavor, and berry skin color. These data demonstrate the role of the grapevines in the early inception of agriculture across Eurasia.

The cultivated grapevine (*Vitis vinifera* ssp. *vinifera*, hereafter *V. vinifera*) shares a close relationship with humans (*1*). With unmatched cultivar diversity, this food source (table and raisin grapes) and winemaking ingredient (wine grapes) became an emblem of cultural identity in major Eurasian civilizations (*1*–*3*), leading to intensive research in ampelography, archaeobotany, and historical records to reveal its history (*4*). Early work asserted that *V. vinifera* originated from its wild progenitor *Vitis vinifera* ssp. *sylvestris* (hereafter *V. sylvestris*) ~8000 years ago during the Neolithic agricultural revolution in the Western Asia (*5*, *6*). In recent years, various genetic studies explored this proposition (*6*–*13*), but the critical details of grapevine domestication were often inconsistent. Studies argued for the existence of domestication centers in the western Mediterranean (*13*), Caucasus (*12*, *14*), and Central Asia (*12*), which in turn cast doubt on the popular notion of a single past domestication event (*10*, *11*). Three demographic inferences yielded population split times between *V. vinifera* and *V. sylvestris* to dates between 15,000 and 400,000 years ago, predating the historical consensus on domestication time (*7*–*9*). Because early domesticates spread to other parts of Eurasia through poorly defined migration routes in the ensuing millennia (*5*), the single-origin theory also confounds the origin order between table and wine grapevines. One view proposes a wine grapevine–first model, with the two types diverging ~2500 years ago (*7*, *10*, *11*). Hybridization with local *V. sylvestris* was common in creating extant European wine grapes (*10*, *11*), but when these introgression events occurred is unknown. Several studies suggest that the earliest cultivation of European wine grapes in France and Iberia postdates 3000 years ago (*10*, *15*). These discrepancies primarily result from the inadequate sampling of grapevine accessions and the limited resolution of genetic data in previous analyses. Therefore, we report the genomic variation dataset from a global cohort to systematically delineate the structure of grapevine genetic diversity, explore the origin of *V. vinifera*, deduce a putative dispersal history, and investigate key domestication traits and diversification signatures.

Results

We constructed a chromosome-level reference *V. sylvestris* genome assembly (VS-1 from Tunisia) to attain genomic variations, which shows a higher percentage of anchored chromosomal lengths than PN40024 (fig. S1 and tables S1 to S9) (*16*). From the 3304 assembled accessions from a dozen Eurasian germplasm and private collections, we obtained good-quality Illumina paired-end sequencing data to an average 20× coverage for 3186 grapevine accessions (2237 *V. vinifera* and 949 *V. sylvestris*; tables S10 to S13). The sample selection preferentially included old, autochthonous, and economically important varieties to maximize the spectrum of genetic diversity. We also included genomic data for 339 previously sequenced accessions (266 *V. vinifera* and 73 *V. sylvestris*; table S14) in the analyses (*7*, *8*, *17*), producing the final cohort of 3525 grapevine accessions (2503 *V. vinifera* and 1022 *V. sylvestris*). The alignment of the Illumina reads to the VS-1 reference genome identifies 45,624,306 biallelic single-nucleotide polymorphisms (SNPs) and 7,314,397 biallelic short Indels [≤40 base pairs (bp); 73.2% shorter than 5 bp] (*16*), among which rare alleles (minor allele frequency ≤1%) accounted for the majority (fig. S2 and tables S15 to S22).

Core accessions differentiate by eight distinct genetic ancestries

Clones, mutants, synonyms, and homonyms are common phenomena in grapevine germplasm and collections (*18*). Using the identity-by-state sharing pattern estimators, we found 1534 accessions sharing the genetic profile with at least one other in the cohort, totaling 498 distinct genotypes (fig. S3 and table S23) (*16*). We kept one accession for each distinct genotype, corrected misidentified accessions, and excluded interspecific hybrids for a core cohort of 2448 grapevines (1604 *V. vinifera* and 844 *V. sylvestris*; fig. S3), which remain representative of the major viticultural regions (*19*) in the world (Fig. 1A and fig. S3).

Principal component analysis (PCA) showed that *V. sylvestris* and *V. vinifera* separately spread out along the first two axes (total variance explained: PC1 7.56% and PC2 1.71%), with both displaying a crude Western Asia to Western Europe gradient (Fig. 1B and figs. S4 and S5). The PC3 axis (1.26% variance) separates *V. vinifera* individuals according to their utilization, agreeing with the main table and wine grapevine clades in the maximum likelihood phylogenetic tree and reticulate phylogenetic network (figs. S6 and S7). The *V. vinifera* accessions show a weak isolation-by-distance correlation (Fig. 1C), suggesting a disconnection between the viticultural geographic pattern and the genetic structures in the grapevine

(*20*). This observation could be due to the extensive exchange of superior cultivars across regions and the subsequent interbreeding throughout history.

Given the poor resolution of viticultural regions in defining grapevine diversity, we leveraged genetic ancestry information from an unsupervised ADMIXTURE analysis to categorize core accessions (Fig. 1D and fig. S8) (*16*). At K = 2, all *V. vinifera* accessions contain a majority east (red) ancestry that matches the ancestry of the *V. sylvestris* accessions in the East Mediterranean region. At K = 8, hierarchical clustering of ancestry components identifies four *V. sylvestris* groups from distinct geographic regions: Western Asia (Syl-E1, 84.3% *K*2), the Caucasus (Syl-E2, 72.7% *K*6), Central Europe (Syl-W1, 94.7% *K*1), and the Iberian Peninsula (Syl-W2, 69.8% *K*8; Fig. 1, D to F). *V. sylvestris* accessions collected from other regions show admixed genetic structures (*16*). For cultivated grapevines (CGs), six genetic ancestries could designate six distinctive groups (CG1 to CG6), all covering a broad range of viticultural regions (Fig. 1, D to F) (*16*). Accessions with pure or close to pure ancestries (fig. S9) (*16*) helped to ascribe names to these groups as Western Asian table grapevines (CG1, 73.9% *K*2), Caucasian wine grapevines (CG2, 66.4% *K*6), muscat grapevines (CG3, 87.7% *K*5), Balkan wine grapevines (CG4, 69.9% *K*4), Iberian wine grapevines (CG5, 68.8% *K*7), and Western European wine grapevines (CG6, 68.4% *K*3). The admixed *V. vinifera* accessions showed different combinations of genetic ancestries (fig. S9). The four *V. sylvestris* and six *V. vinifera* groups, supported by archetypal analysis at K = 8 (fig. S10), formed identifiable clusters in the PCA plots (Fig. 1G and fig. S4) and were thus suitable for population genomic investigations.

Separation of V. sylvestris *ecotypes in Pleistocene*

According to the genetic ancestries and the occupied ecological niches in the western Eurasia continent, we designate *V. sylvestris* accessions in Western Asia and the Caucasus as the eastern ecotype (*V. sylvestris* eastern ecotype, hereafter Syl-E) and accessions in Central Europe and the Iberian Peninsula as the western ecotype (*V. sylvestris* western ecotype, hereafter Syl-W) (Fig. 2A). The large between-ecotype fixation index values [e.g., Syl-E1 versus Syl-W1, pairwise population fixation index (F_{ST}) = 0.340] and the small within-ecotype fixation index values (Syl-E1 versus Syl-E2, F_{ST} = 0.101; Syl-W1 versus Syl-W2, F_{ST} = 0.072; fig. S11 and table S26) support this designation. Both nucleotide diversity (π) and individual heterozygosity show that the western ecotype (especially Syl-W1) has significantly reduced variation compared with its eastern counterpart (fig. S11). Furthermore, the linkage disequilibrium decay (LD, r^2) was much slower in Syl-W (1.0 to 1.6 Kb at half of maximum r^2) than in Syl-E (400 to 600 bp at half of maximum r^2; fig. S12). These data demonstrate that the eastern ecotype retains more genetic diversity.

Demographic inference with folded SNP frequency spectra reveals an ancient population bottleneck in Syl-E ~400,000 to 800,000 years ago and in Syl-W ~150,000 to 400,000 years ago (Fig. 2B and fig. S13). This Pleistocene period, characterized by changing climate cycles (*21*, *22*), also witnessed the deduced population split (median time ~200,000 to 400,000 years ago) between the two ecotypes (Fig. 2C). The slow descent of the split line suggests that the geographic isolation process was gradual (fig. S13). At ~56,000 years ago, the population split between Syl-E1 and Syl-E2 occurred during the last glacial cycle (11,700 to 115,000 years ago), when the global climate trended toward dryer and colder conditions (*23*). Close to the time of the Last Glacial Maximum (LGM; ~21,000 years ago), *V. sylvestris* subgroups experienced a second population bottleneck (~40,000 years ago), with effective population sizes (N_e) reaching a minimum of 10,000 to 40,000 (Fig. 2B and fig. S13). After this result, ecological niche modeling predicts that the areas with suitable environmental conditions for Syl-E and Syl-W (suitability > 0.75) remained connected at the Pleistocene Last Interglacial (~130,000 years ago) (fig. S14) but became entirely separated at the LGM (Fig. 2D). The post-bottleneck N_e rebound was steeper in the Syl-W accessions, but the numbers decreased to lower levels in recent times (Fig. 2B and fig. S13). This result agrees with the reduced genetic diversity in Syl-W and the abrupt population split between Syl-W1 and Syl-W2 at ~2500 years ago.

Dual origin of V. vinifera *at the advent of agriculture*

The wet climate in the Early Holocene (11,700 to 8300 years ago) (*24*) facilitated the expansion of suitable habitats for Syl-E, resulting in a large geographic span from Central Asia to the Iberian Peninsula (Fig. 2D). This expansion supports the eastern origin and subsequent continental dispersal of *V. vinifera*. Because CG1 shares the main ancestral component with Syl-E1 and CG2 with Syl-E2 (Fig. 1, D and F), the possibility of two domestication events becomes evident. Indeed, both CG1 and CG2 maintain the highest genetic diversity and manifest the quickest LD decay among all CG groups (figs. S11 and S12). Furthermore, they are less differentiated from their corresponding wild ecotypes (Fig. 3A and fig. S11). The Akaike information criterion (AIC)–based phylogenetic selection also prefers a dual origin tree model (fig. S15), which agrees with the outgroup f_3 statistics biplots that CG1 and CG2 are genetically

[1]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China. [2]Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201, China. [3]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China. [4]Beijing Key Laboratory of Grape Science and Oenology and Key Laboratory of Plant Resources, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China. [5]Institute of Molecular Biology, NAS RA, 0014 Yerevan, Armenia. [6]Yerevan State University, 0014 Yerevan, Armenia. [7]Genetic Resources Institute, Azerbaijan National Academy of Sciences, AZ1106 Baku, Azerbaijan. [8]National Institute of Viticulture and Winemaking Magarach, Yalta 298600, Crimea. [9]Institute for Adriatic Crops and Karst Reclamation, 21000 Split, Croatia. [10]Bordeaux University, Bordeaux Sciences Agro, INRAE, UMR EGFV, ISVV, 33882 Villenave d'Ornon, France. [11]AGAP Institut, University of Montpellier, CIRAD, INRAE, Institut Agro Montpellier, 34398 Montpellier, France. [12]Julius Kühn Institute (JKI) – Federal Research Center for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, 76833 Siebeldingen, Germany. [13]Botanical Institute, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany. [14]LEPL Scientific Research Center of Agriculture, 0159 Tbilisi, Georgia. [15]Hungarian University of Agriculture and Life Sciences (MATE), 1118 Budapest, Hungary. [16]Department of Chemical Engineering, Ariel University, 40700 Ariel, Israel. [17]Eastern Regional R&D Center, 40700 Ariel, Israel. [18]Department of Agricultural and Environmental Sciences, University of Milano, 20133 Milano, Italy. [19]Instituto Nacional de Investigação Agrária e Veterinária, I.P./INIAV-Dois Portos, 2565-191 Torres Vedras, Portugal. [20]Green-it Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2780-157 Oeiras, Portugal. [21]National Research and Development Institute for Biotechnology in Horticulture, Stefanesti, 117715 Arges, Romania. [22]Center for Plant Biotechnology and Genomics, UPM-INIA/CSIC, Pozuelo de Alarcon, 28223 Madrid, Spain. [23]University of Lausanne, 1015 Lausanne, Switzerland. [24]Biotechnology Institute, Ankara University, 06135 Ankara, Turkey. [25]Department of Plant Science, School of Agriculture and Biology, Shanghai JiaoTong University, Shanghai 200240, China. [26]Armenian Bioinformatics Institute, 0014 Yerevan, Armenia. [27]Biomedicine and Pharmacy, RAU, 0051 Yerevan, Armenia. [28]Institute of Bioresources, Nakhchivan Branch of the Azerbaijan National Academy of Sciences, AZ7000 Nakhchivan, Azerbaijan. [29]Vassal-Montpellier Grapevine Biological Resources Center, INRAE, 34340 Marseillan-Plage, France. [30]Historische Rebsorten-Sammlung, Rebschule (K39), 67599 Gundheim, Germany. [31]Galilee Research Institute (Migal), Tel-Hai Academic College, 12210 Upper Galilee, Israel. [32]Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milano, Italy. [33]NBFC, National Biodiversity Future Center, 90133 Palermo, Italy. [34]Institute of Biosciences and Bioresources, National Research Council, 90129 Palermo, Italy. [35]Department AGRARIA, University Mediterranea of Reggio Calabria, Reggio 89122 Calabria, Italy. [36]IMIDRA, Alcalá de Henares, 28805 Madrid, Spain. [37]Viticulture Research Institute, Ministry of Agriculture and Forestry, 59200 Tekirdağ, Turkey. [38]Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St. Lucia, Queensland 4072, Australia. [39]Institute of Horticulture, Ningxia Academy of Agricultural and Forestry Sciences, Yinchuan 750002, China. [40]Center for Viticulture and Oenology, School of Agriculture and Biology, Shanghai JiaoTong University, Shanghai 200240, China. [41]Zhengzhou Fruit Research Institutes, CAAS, Zhengzhou 450009, China. [42]BGI-Shenzhen, Shenzhen 518083, China. [43]The Martin (Szusz) Department of Land of Israel Studies and Archaeology, Bar-Ilan University, 5290002 Ramat-Gan, Israel.

***Corresponding author. Email: wchenntr@gmail.com (W.C.); shengjun@dongyang-lab.org (J.S.); shhli@ibcas.ac.cn (S.L.)**

†These authors contributed equally to this work. ‡Institution contacts for biological samples. Ordered by country names. §Present address: Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany.

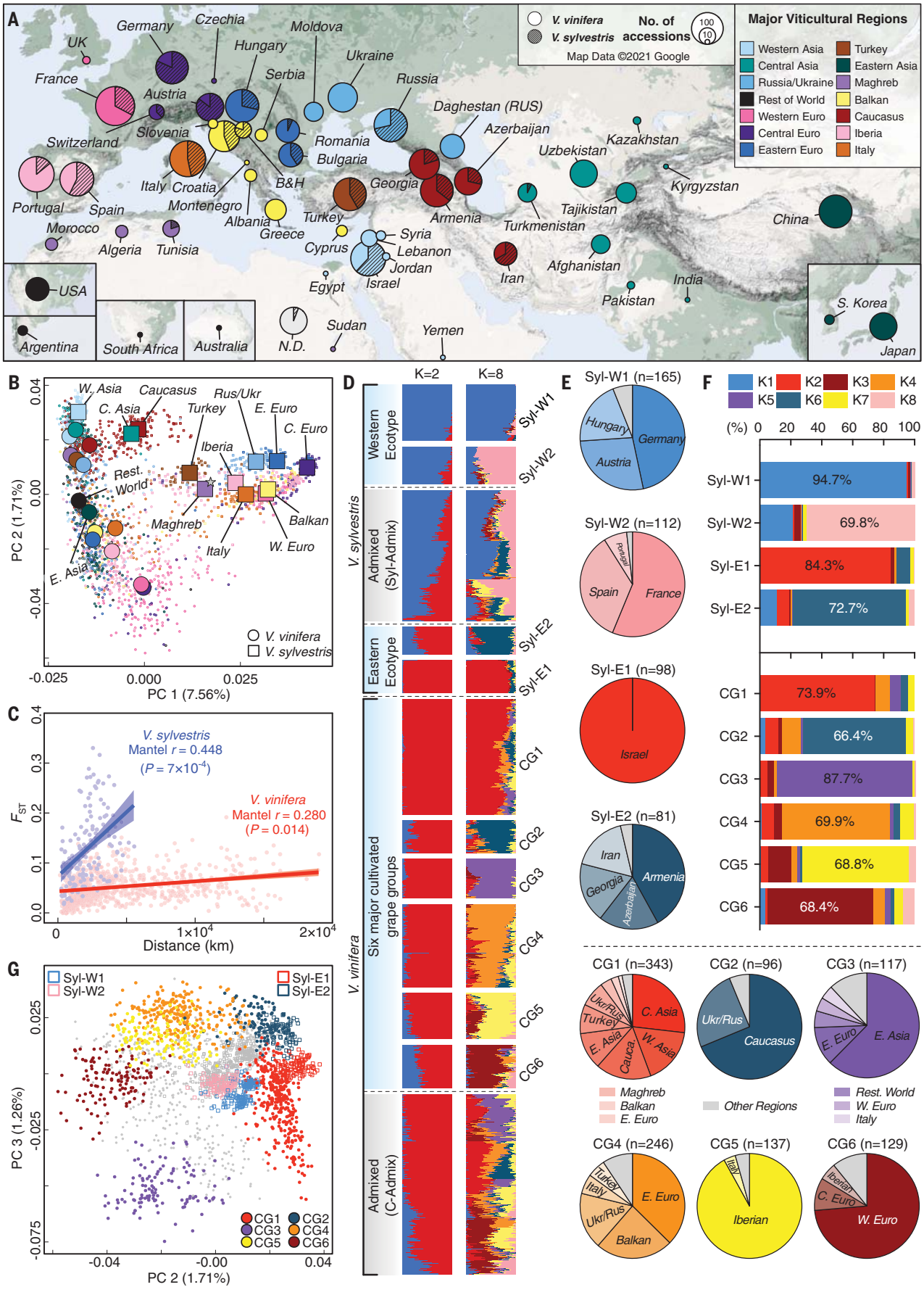

Fig. 1. Genetic diversity of global core *V. sylvestris* and *V. vinifera* accessions. (**A**) Geographical locations of the 2448 core grapevine accessions. (**B**) PCA according to major viticultural regions. Large square/circle highlights median position. Star shows VS-1 position. (**C**) Isolation-by-distance test of *V. sylvestris* and *V. vinifera* accessions. Linear regression with 95% confidence interval is shown. (**D**) ADMIXTURE clustering of the accessions. (**E**) Geographic locations of the accessions in each group. Gray represents minor locations. (**F**) Average proportion of major genetic ancestries in grapevine groups. (**G**) PC2 versus PC3 projection according to grapevine group.

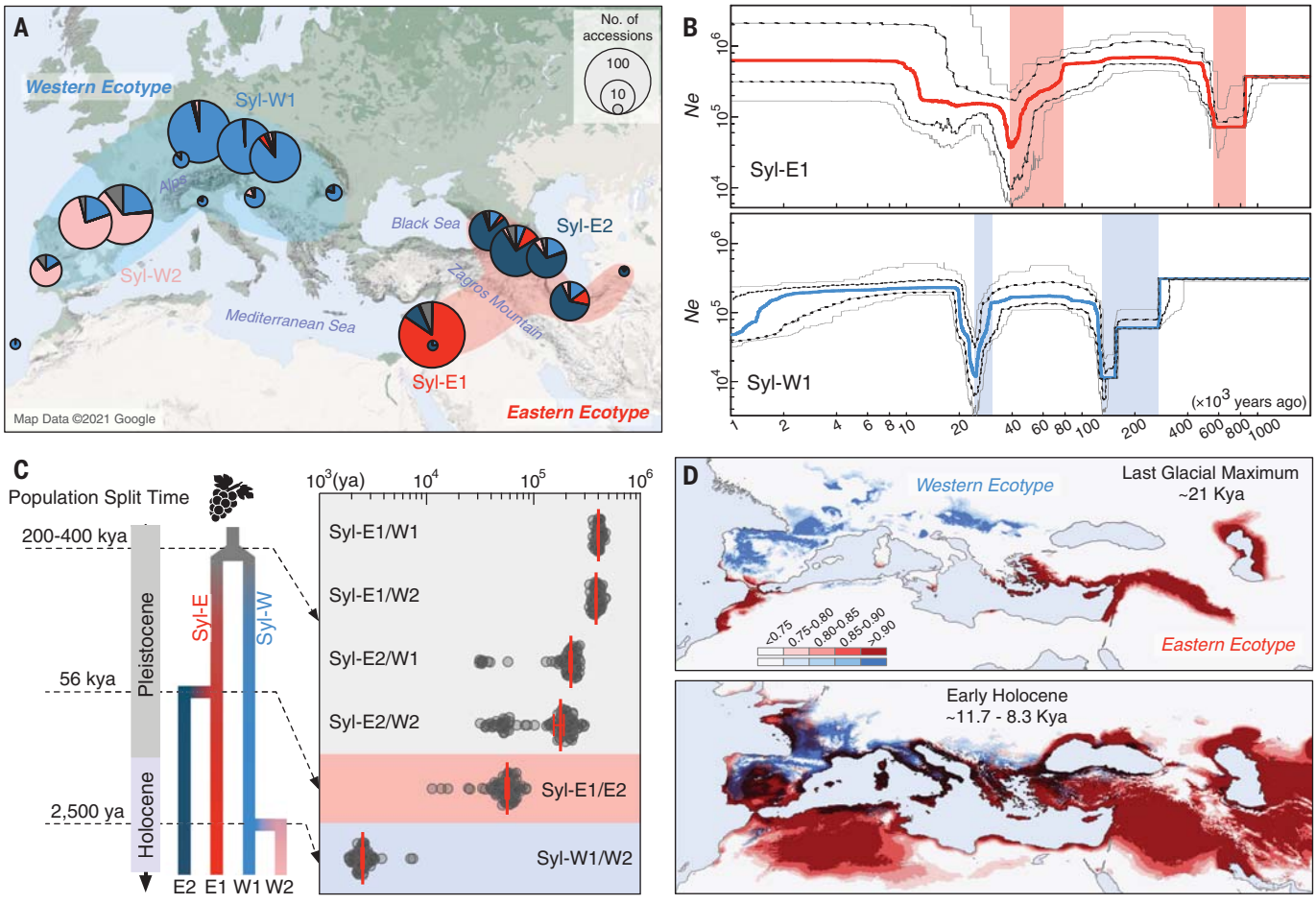

Fig. 2. Population history of *V. sylvestris* ecotypes. (**A**) Geographic isolation and population separation of *V. sylvestris* ecotypes. Pie charts show mean ancestry proportion at $K = 8$. Same color scheme as in Fig. 1B is used. (**B**) Demographic histories of *V. sylvestris* populations deduced from Stairway Plot 2. Lines indicate medians with 75% and 95% confidence intervals. (**C**) Population split times among ecotypes with MSMC2. Red bars indicate medians with 95% confidence intervals. (**D**) Ecological niche modeling of the suitable habitats for *V. sylvestris* ecotypes. The color scale shows suitability score.

closer to Syl-E1 and Syl-E2, respectively (Fig. 3B, fig. S15, and table S27). The population split lines of CG1/Syl-E2 and CG2/Syl-E1 pairs resemble that of Syl-E1/Syl-E2 and differ from those of CG1/Syl-E1 and CG2/Syl-E2 pairs (Fig. 3C and fig. S16). These data collectively support a dual origin of *V. vinifera* and reject the popular theory of a single primary domestication center (*10*, *11*). Both CG1/Syl-E1 and CG2/Syl-E2 population pairs separated quickly (Fig. 3C), which is compatible with a clean-split scenario. We estimate the median population split time to be ~11,000 years ago (95% confidence interval: ~10,500 to 12,500 years ago) for both pairs, suggesting that the domestication events took place concurrently around the advent of agriculture. Because CG1 and CG2 separately represent table and wine grapevine ancient genetic backgrounds (*K2* and *K6*; fig. S9), the dual origin rejects the assumption that wine grapevines predate table grapevines (*7*, *10*, *11*).

Dispersal of grapevine domesticates along human migration routes

The geographic distributions of CG1 and CG2 cultivars across Eurasia and North Africa correspond to vastly different human migration routes for the two grapevine groups (Fig. 3D). The CG2 cultivars were mainly confined to both sides of the Caucasus Mountains, with a limited dispersal into the Carpathian Basin by the northern Black Sea. This result contrasts with previous models implying that CG2 played a central role in the formation of wine grapevines in Europe (*3*). Instead, CG2 represents a local domestication effort that had a minor impact on grapevine diversification. By comparison, the dispersal of CG1 in four directions spanned Eurasia and North Africa. First, the eastward expansion through Central Asia into India and China follows the Inner Asian

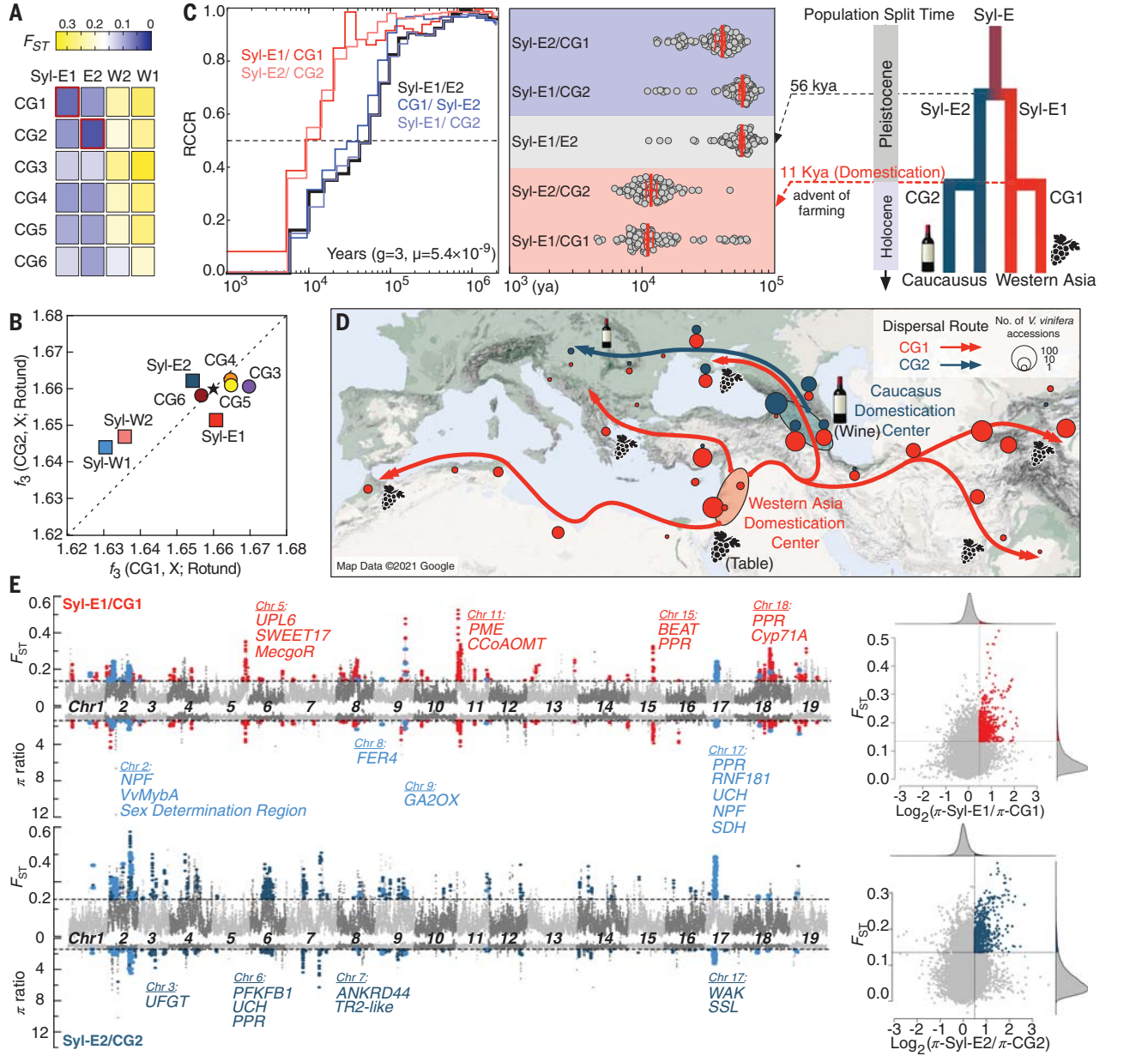

Fig. 3. Dual domestications of *V. vinifera* in Western Asia and the Caucasus. (**A**) Pairwise fixation index of the major grapevine groups. (**B**) Outgroup f_3 statistics biplot measuring genetic similarity. Rotund, *Muscadinia rotundifolia*. Stars mark the f_3 statistics for CG1/CG2. (**C**) Estimated split times among Syl-E1/2 and CG1/2 with MSMC2 (left). Red bars indicate medians with 95% confidence intervals. (**D**) Geographic distribution of CG1 and CG2 in relation to the domestication centers. Human dispersal routes are shown. (**E**) Shared (sky blue) and unique domestication selective sweep regions (red and dark teal) in *V. vinifera*.

Mountain Corridor, a path that also witnessed the exchange of other crops (i.e., wheat, barley, and millet) between the West and the East (*25*). Second, the northbound expansion could mirror the early cultural contact of Western Asia over the Zagros mountains with the Caucasus (*26*, *27*). Third, the northwest expansion through Anatolia into the Balkans bespeaks the spread of farming into Europe (*28*, *29*). Finally, a westward expansion moved across the North African coastline to reach Morocco (*30*). Even though grapevine domestications followed the trails of past human migration, the timing and dispersal details require paleogenomic data for delineation.

Shared and unique domestication signatures in CG1 and CG2 grapevines

Given the dual origin scenario, we investigated domestication signatures in both Syl-E1/CG1 and Syl-E2/CG2 group pairs by selecting genomic regions that display increased nucleotide diversity differences and population differentiation (both top 5%; Fig. 3D). This method yields 1140 domestication selective sweep genes in 132 regions for CG1 and 887 genes in 137 regions for CG2 (table S28), among which only 189 genes in 31 regions exist in both groups (table S29). Most shared signals are on chromosomes 2 and 17, confirming previous findings that the selection on flower sexual morphs (sex determination region, SDR), berry skin color (*VvMybA* gene cluster), and berry development (*SDH* gene cluster) were of great importance during grapevine domestication (*8*, *11*). In addition, our analysis identifies

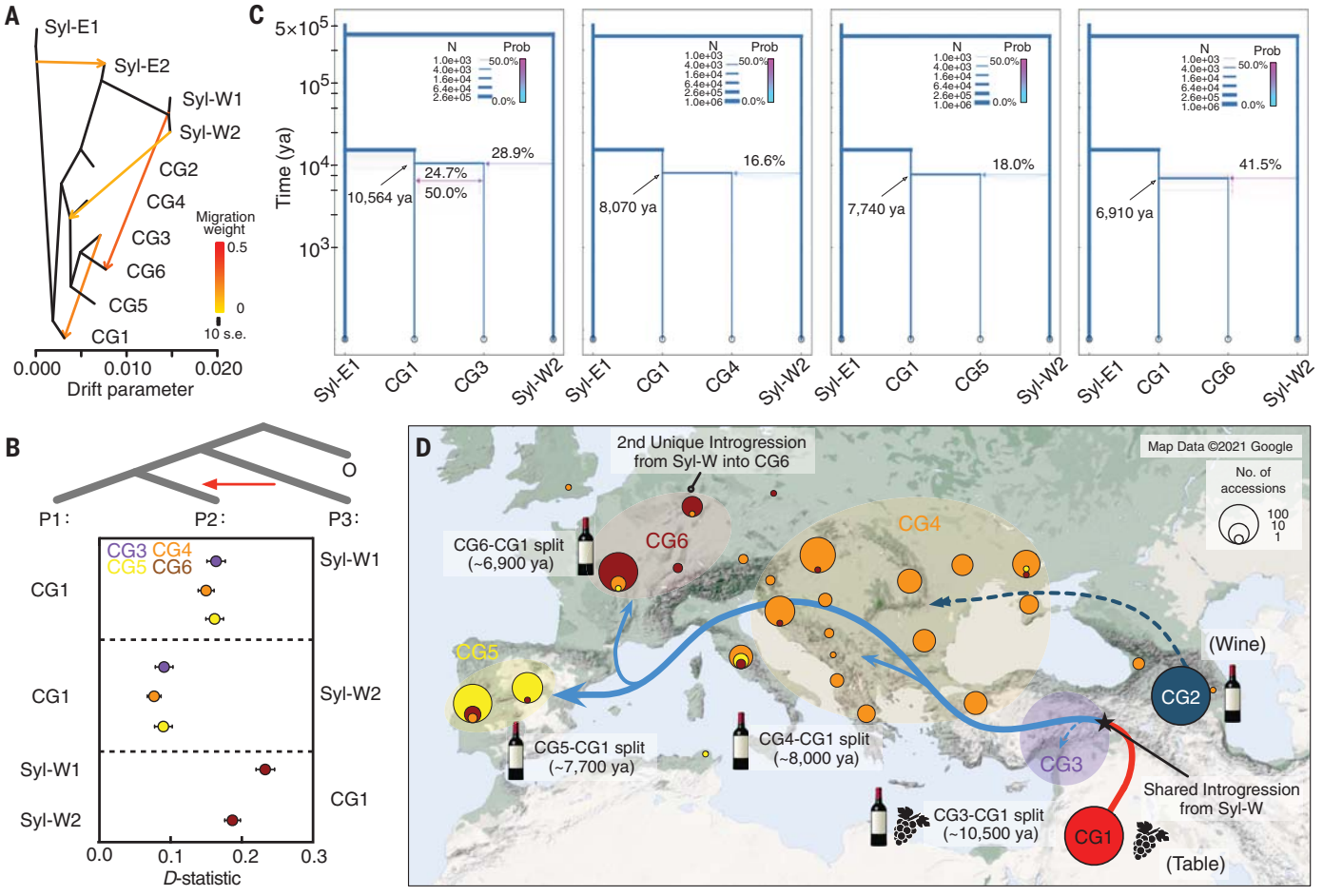

Fig. 4. Stepwise diversification of *V. vinifera* in Europe. (**A** and **B**) Introgression from Syl-W into European *V. vinifera* groups revealed by TreeMix (A) and confirmed by D-statistic (B). (**C**) Four population simulation of split times and genetic introgression using Momi2. Median numbers from 100 bootstrap runs are shown. (**D**) Origination of *V. vinifera* groups (CG3 to CG6) by the end of the Neolithic. Geographic distributions of CG groups are shown by colored circles. See fig. S24 for details on CG3.

shared domestication genes that possibly underlie grapevine growth (e.g., *NPF*), physiology (e.g., *FER4*), fruit set (e.g., the *GA2OX* gene cluster), and resistance to biotic/abiotic stress (e.g., *FER4*, the *PPR* gene cluster, and the *RNF181* gene cluster) [see (*16*) for gene descriptions].

As expected for dual domestications, most selective sweep signatures in CG1 and CG2 are unique and target distinctive chromosomal regions (Fig. 3E). Even though CG1 and CG2 correspondingly represent table and wine grapevines, many unique signatures seem to suggest a convergent selection mechanism targeting different aspects of common domestication traits. An obvious example is the improvement of berry palatability through the reduction of alkaloid biosynthesis (the *MecgoR* gene cluster in CG1 and the *TR2* and *SSL* gene clusters in CG2) and the enhancement of carbohydrate metabolism (*SWEET17* in CG1 and *PFKFB1* in CG2). Other examples include perceived berry desirability (the *BEAT* gene cluster for floral scent in CG1 and the *UFGT* gene cluster for berry color in CG2) and response to environmental stresses (*UPL6* in CG1 and *WAK* in CG2). These findings suggest that the initial cultivation of CG1 and CG2 may have been to serve early humans' caloric and micronutrient needs. The selection of genetic features suitable for winemaking in CG2 could have been serendipitous, and the practice of winemaking with CG2 (e.g., 8000 years ago) (*14*) possibly postdates grapevine domestication. Because gene annotation depends on homology-based inference, it should be noted that many genes mentioned here need further verification in grapevines.

Wine grapevine diversification in Europe

Because the CG1 early domesticates dispersed into Europe through Anatolia, a crucial question concerns the diversification history of European wine grapevines in the ensuing millennia. In particular, the shared areas of suitable habitats for Syl-E and Syl-W in the early Holocene (black area in Fig. 2D) formed an ecological foundation for the genetic exchange between CG1 and local refugia Syl-W accessions in the coastal regions of the northern Mediterranean Sea and the southern Black Sea, the Iberian Peninsula, and an area corresponding to present-day western France. It is therefore important to examine where and how distinct grapevine genetic ancestries (CG3 to CG6) formed with relevance to Syl-W introgression (*10*, *11*). We have chosen cultivars in each group with at least 75% major ancestry (and with an average Syl-W ancestry in each *V. vinifera* group <3%) to perform population analyses. This selection rules out many old varieties (i.e., 'Lambrusco' cultivars deriving about half of their ancestries from Syl-W; fig. S9), which likely showcase secondary diversification efforts after the distinct ancestries had been established. The TreeMix analysis finds one migration edge that points from Syl-W to a population ancestral to CG3 to CG6 (estimated weight, 0.114; Fig. 4A and fig. S17), suggesting an ancient introgression event occurred before the diversification of all European grapevines. An additional migration

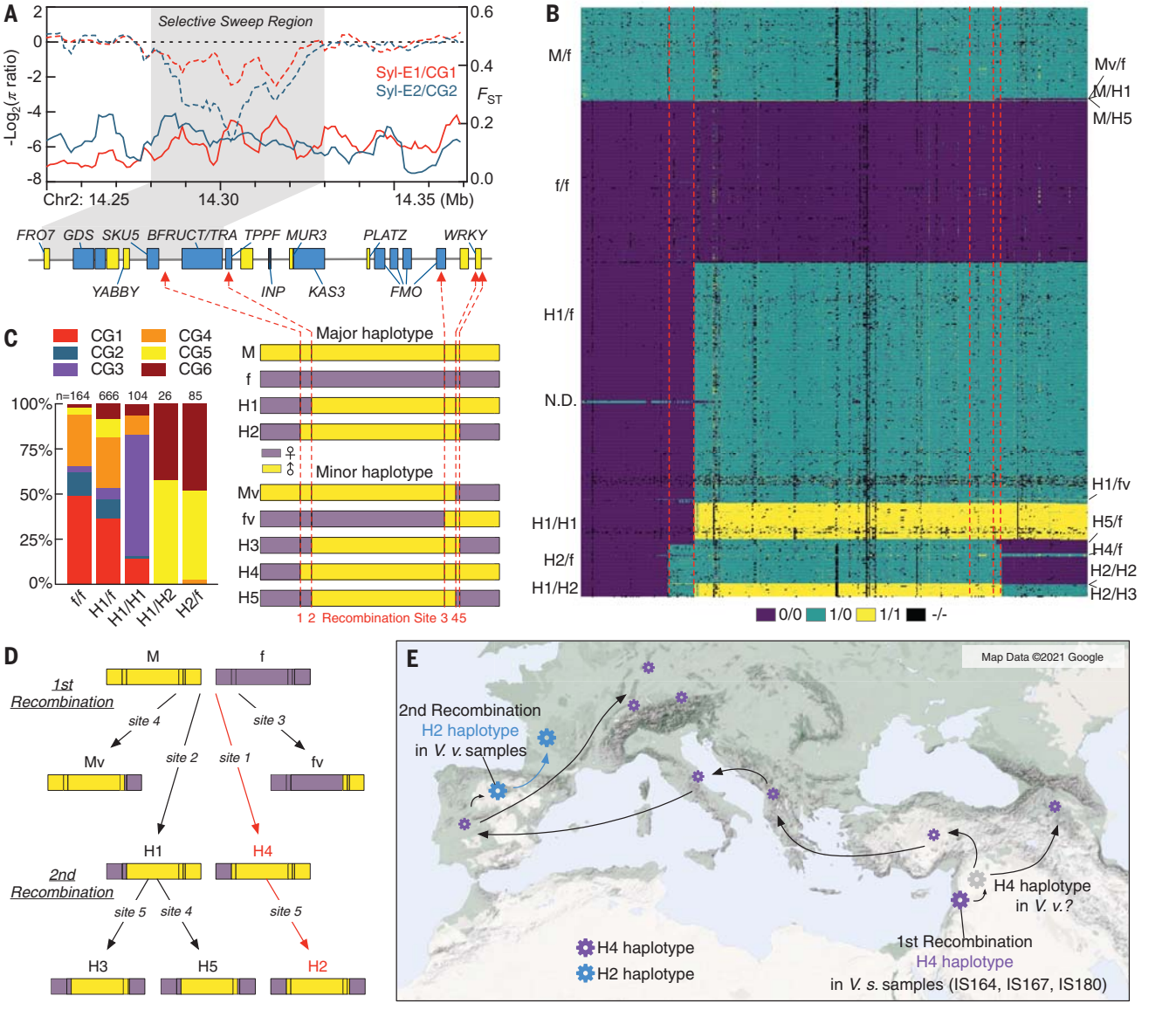

Fig. 5. Selection and evolution of the SDR in the core grapevine accessions. (**A**) The SDR in VS-1. Red arrows indicate identified recombination sites. (**B**) SDR genotypes from associated SNPs reveal five recombination sites (dashed lines) and genotype diversity (right). Major and minor haplotypes are shown on the left. (**C**) Distribution of SDR genotypes in the six major grapevine groups. (**D**) Recombination history of all SDR haplotypes. (**E**) Putative dispersal route of the H4 haplotype and the origination of H2 haplotype.

edge also points from Syl-W to CG6 (estimated weight, 0.292), which implies an independent introgression event unique to Western European wine grapevines in the past. Various combinations of D-statistics testing the gene flow from Syl-W into CG groups (Z score > 3.0, adjusted $P < 4.17 \times 10^{-5}$; Fig. 4B and table S31) support this introgression history. Additionally, gene flow from Syl-W into CG3 to CG6 inferred from Momi2 align with their corresponding divergence from CG1, further supporting the introgression history (Fig. 4C). The estimated median divergence times date the creation of muscat grapes (CG3) to 10,500 years ago, Balkan wine grapes (CG4) to 8070 years ago, Iberian wine grapes (CG5) to 7740 years ago, and Western European wine grapes to 6910 years ago (Fig. 4D). These stepwise diversification times agree with the historical migration of Anatolian farmers into Europe (*26*, *29*, *31*, *32*), substantiating the role of viticulture in forming Neolithic agricultural societies.

The migration edge weights, f_4 ratio, and Momi2 estimates collectively show that ancient introgression from Syl-W accounts for ~11.4 to 18.0% of the CG3 to CG6 genomes (Fig. 4 and table S30). In addition, at least one other independent introgression event contributed ~25.0 to 30.0% additional Syl-W to the CG6 ancestry. We have screened the introgression tracts in CG3 to CG6 by choosing the genomic windows with the top 1% d_f and f_{dM} values (fig. S18). Ten shared regions among the CG3 to CG6 groups contain genes that are putatively involved in plant immunity (e.g., *CYSK*), abiotic stress response (e.g., *GBA*), and carbohydrate metabolism (e.g., *TPS/TPP*) (table S31). This result agrees with the proposal that introgression helps grapevines adapt to new environments and become more suitable for winemaking (*10*, *11*).

Genetic analyses of domestication and diversification traits

Hermaphroditism: origin of H2 haplotype

The transition from dioecy in *V. sylvestris* (male, M/f; female, f/f) to hermaphroditism in *V. vinifera*

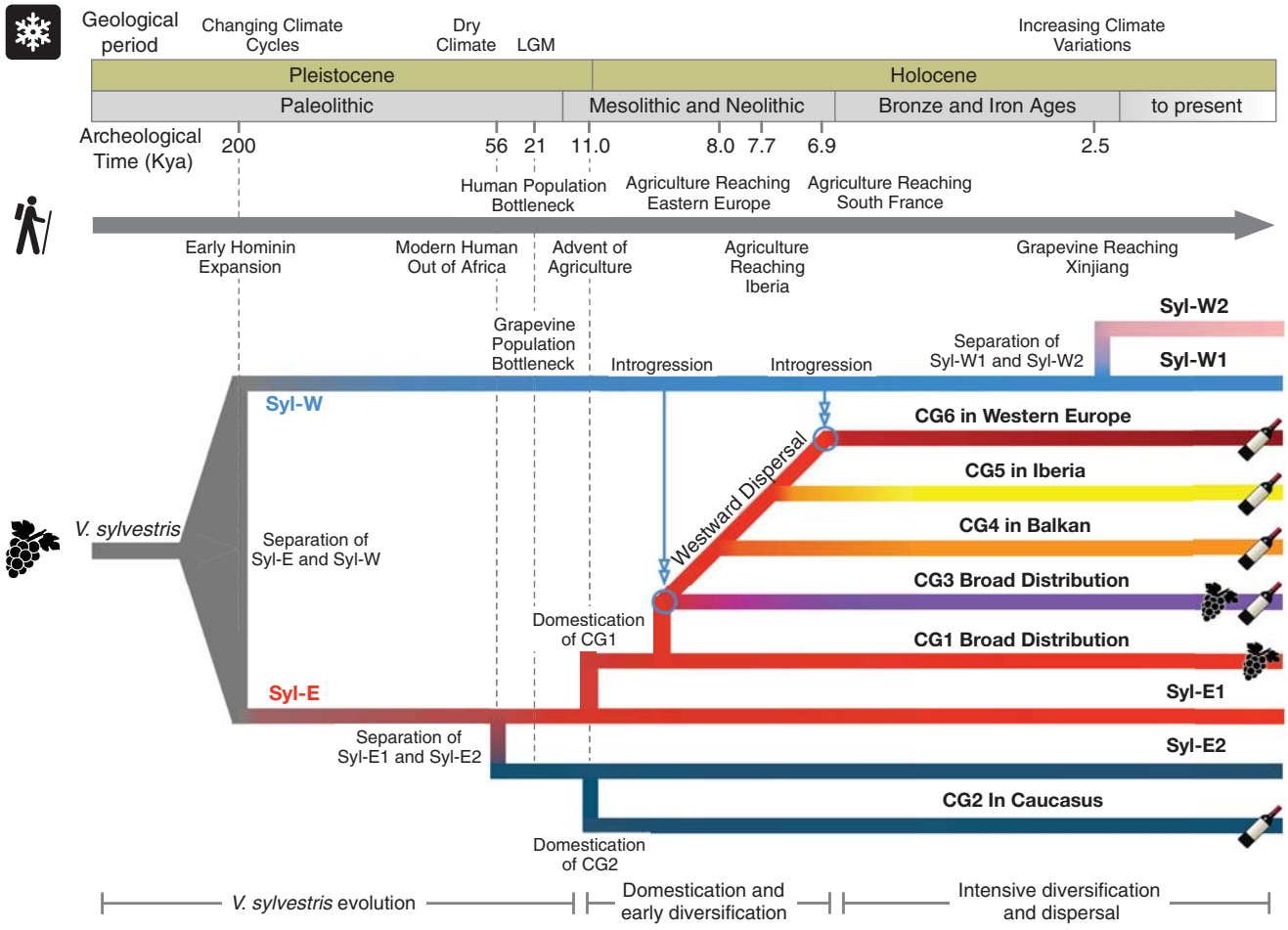

Fig. 6. Schematic graph of grapevine evolutionary history. Key events in the evolutionary history of grapevines are shown alongside major events in global climate change and human migration.

is the most prominent phenotypic change during domestication (*33*). It involves recombination events between M and f around a selective sweep region on chromosome 2 known as the SDR (Fig. 5A). Previous studies have identified two major hermaphroditic haplotypes (H1 and H2) and four hermaphroditic genotypes (H1/f, H2/f, H1/H1, and H1/H2) from select cultivars (*33*), but the recombination history remains unclear. The analysis of our grapevine cohort reveals five recombination sites in the SDR (Fig. 5B), which not only confirms known genotypes but also identifies new minor haplotypes (male variant Mv, female variant fv, H3, H4, and H5) and genotypes (Mv/f, M/H1, M/H5, H1/fv, H5/f, H4/f, H2/H2, and H2/H3) in both wild and cultivated grapevines (Fig. 5B and table S32). Among all SDR haplotypes, M and H1 manifest the highest subtype diversity (figs. S19 to S22). Furthermore, the SDR genotype statistics reveal a distribution bias of the H2-containing SDRs in the Iberian (CG5) and Western European (CG6) grapevines (Fig. 5C and fig. S23). To investigate this observation, we constructed a putative recombination history for all known SDR haplotypes (Fig. 5D), which showed that a first recombination event between the parental M and f haplotypes created Mv (site 4), fv (site 3), H1 (site 2), and H4 (site 1). On this basis, H1 experienced a second recombination event with f to produce H3 (site 5) and H5 (site 4), whereas H4 recombined again with f at site 5 to bring about H2. Because three Syl-E *V. sylvestris* (IS164, IS167, and IS180) and 11 *V. vinifera* accessions in the cohort contain H4 (Fig. 4G), a likely scenario supports a westward dispersal of H4 after human selection to reach the Iberian Peninsula [e.g., in extant old Iberian cultivar 'Malvasia Fina' (PO153)], where H2 originated from H4 through secondary recombination and later became dominant during the diversification of Iberian and Western European cultivars.

Muscat flavor: Trait selection may reduce grapevine fitness

Muscat grapevine is unique for its floral aromas, which result from a hard-to-define concoction of monoterpenoids in the fruit (*34*). Given the broad geographic distribution (fig. S24) and ancient history of muscat grapevines, it is not easy to pinpoint the center of origin. However, Momi2 estimate predicts a population split from CG1 at ~10,564 years ago (Fig. 4C), suggesting an origination site within the boundary of Western Asia. This scenario agrees with the relatively low F_{ST} values and sizeable gene flow with CG1 (Fig. 4 and fig. S11). The CG3 group also shows low genetic diversity and high LD extent compared with the others (figs. S11 and S12). One possible reason is the gradual loss of ancient CG3 cultivars in Anatolia and the surrounding regions throughout history (fig. S24). Even though the muscat aroma is a complex trait, genome-wide association analysis based on a binary differentiation reveals 18 SNP signatures on chromosomes 5 and 18 (fig. S24 and table S33). This set includes a nonsynonymous SNP Chr5:19419686 in the *VvDXS* gene linked to the trait (*34*). Examination of the genotype at this locus shows that 108 of the 134 muscat grapevines (including 'Muscat Hamburg,' 'Königin der Weingärten,' and 'Muscat of Alexandria,' which are commonly

used as parental cultivars) are heterozygous (G/T), and only eight individuals are homozygous (T/T) for the alternative SNP (exact test for Hardy-Weinberg equilibrium, D = 20.68, $P = 2.01 \times 10^{-13}$). Additionally, most grapevines without muscat aroma are homozygous for the reference SNP (G/G; 1451 of 1468; exact test for Hardy-Weinberg equilibrium, D = 0.049, $P = 1.00$). This result suggests that selection on this allele might have put a constraint on grapevine fecundity, thereby preventing the alternative SNP from reaching fixation.

Berry skin color: New genes associated with white grapes

The emergence of white grapes from their red-berried congeners is an essential domestication episode in viticulture history. The color change results from a reduction of anthocyanin synthesis in berry skin cells, where the expression of proposed master regulators such as *VvMybA* decreased significantly in select cultivars because of a *Gret1* retrotransposon (*35*), nonconservative exonic mutations (*36*), or large deletions in the locus (*37*). We performed genome-wide association analysis on this large grapevine cohort (fig. S25, A and B) and identified multiple significant SNPs across the genome (fig. S25C). The most prominent peak spans a broad genomic region from 3.51 to 16.05 Mb on chromosome 2, overlapping the *VvMybA* locus. Among all significant exonic SNPs in this region (table S34), nonsynonymous SNPs with the smallest *P* values localize to two uncharacterized genes outside the *VvMybA* locus (fig. S25D), the putative protein functions of which are acylaminoacyl-peptidase (*Vvsyl02G000229*) and lysine-specific demethylase (*Vvsyl02G001064*). These SNPs are overwhelmingly homozygous for the reference allele in white grapes and are heterozygous in red grapes (fig. S25E). We validated the SNPs in red-berried *V. sylvestris* accessions to account for possible false positives and confirmed their genotypes as being predominantly heterozygous (fig. S25E and table S34). By comparison, significant exonic SNPs in *VvMybA* genes [including Chr2:5116947 G/T reported previously in (*36*)] show shared genotypes between white grapes and the *V. sylvestris* accessions (fig. S25E). It is unclear how *Vvsyl02G000229* and *Vvsyl02G001064* might regulate anthocyanin synthesis, but these results demonstrate that exonic mutations in the two genes are better predictors of berry skin colors. Furthermore, the heterozygous SNP states in *V. sylvestris* accessions suggest that the white berry alleles existed in natural wild populations before grapevine domestication.

Discussion

Our systematic genomic survey of *V. sylvestris* and *V. vinifera* accessions paints a defined picture of grapevine evolutionary history, which echoes key events in the history of world climate change and human migration (Fig. 6). The Pleistocene era witnessed the continuous fragmentation of habitats, the decline of effective population size, and the separation of ecotypes for *V. sylvestris*. It is highly likely that modern humans extensively used grapevines as an energy source from the late Pleistocene, but the harsh climate was not suited for agriculture (*38*). As the climatic conditions ameliorated at the Pleistocene-Holocene transition, the grapevine, with its relatively stable perennial yield, unsurprisingly became one of the earliest candidates for domestication. The dual events underpin the model that plant domestication occurs in large, culturally connected areas over a long time (*39*), but the domestication time gap remains between genomic inference and archaeological evidence (table S35 and figs. S26 and S27) (*16*). The diverse SDR haplotypes suggest that an early goal could be the conscious selection (*40*) and propagation of rare, naturally occurring hermaphroditic individuals from the *V. sylvestris* population because they allow mass plantation without male plants. The selection on phenotype, but not on genotype, also implies that the different hermaphroditic haplotypes were subject to strong genetic drift, which is supported by the high frequency of H1 and the almost extinct H4 in extant cultivars. The Mesolithic and Neolithic periods also saw the early dispersal and diversification of grapevines such that unique ancestries emerged in the Balkans, Iberia, and Western Europe with the help of *V. sylvestris* introgression into CG1. This event mirrors early farmer migration in Europe, consolidating the role of viticulture in forming sedentary societies. A higher level of cultural exchange characterizes the last stage since the Bronze Age and the trading of superior grapevine cultivars along trade routes. This is especially evident in the plethora of Italian cultivars with three or more genetic ancestries, but unfortunately poses a challenge to disentangle the genealogical history of each grapevine cultivar (*20*). Finally, genetically reliable wild grapevines from Central Asia, a region battered by climate change and social instability for the past few millennia, are no longer available to test Vavilov's theory for a diversity center or a hypothetical turnover of grapevine types caused by Islam conversion in the region. Paleogenomic data may help to resolve these questions in the future.

REFERENCES AND NOTES

1. P. E. McGovern, U. Hartung, V. R. Badler, D. L. Glusker, L. J. Exner, *Expedition* **39**, 3–21 (1997).
2. P. This, T. Lacombe, M. R. Thomas, *Trends Genet.* **22**, 511–519 (2006).
3. F. Grassi, G. De Lorenzis, *Int. J. Mol. Sci.* **22**, 4518 (2021).
4. D. Cantu, M. A. Walker, *The Grape Genome* (Springer Nature, 2019).
5. 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, 2012).
6. S. Myles *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3530–3535 (2011).
7. Y. Zhou, M. Massonnet, J. S. Sanjak, D. Cantu, B. S. Gaut, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11715–11720 (2017).
8. Z. Liang *et al.*, *Nat. Commun.* **10**, 1190 (2019).
9. A. Sivan *et al.*, *Plants People Planet* **3**, 414–427 (2021).
10. S. Freitas *et al.*, *Sci. Adv.* **7**, eabi8584 (2021).
11. G. Magris *et al.*, *Nat. Commun.* **12**, 7240 (2021).
12. S. Riaz *et al.*, *BMC Plant Biol.* **18**, 137 (2018).
13. R. Arroyo-García *et al.*, *Mol. Ecol.* **15**, 3707–3714 (2006).
14. P. McGovern *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10309–E10318 (2017).
15. J. Ramos-Madrigal *et al.*, *Nat. Plants* **5**, 595–603 (2019).
16. See the supplementary materials.
17. M. J. Roach *et al.*, *PLOS Genet.* **14**, e1007807 (2018).
18. T. Lacombe *et al.*, *Theor. Appl. Genet.* **126**, 401–414 (2013).
19. R. Bacilieri *et al.*, *BMC Plant Biol.* **13**, 25–25 (2013).
20. F. Mercati *et al.*, *Front. Plant Sci.* **12**, 692661 (2021).
21. R. Hosfield, J. Cole, *Quat. Sci. Rev.* **190**, 148–160 (2018).
22. A. Timmermann *et al.*, *Nature* **604**, 495–501 (2022).
23. E. C. Corrick *et al.*, *Science* **369**, 963–969 (2020).
24. M. Engel *et al.*, *Quat. Int.* **266**, 131–141 (2012).
25. C. J. Stevens *et al.*, *Holocene* **26**, 1541–1555 (2016).
26. I. Lazaridis *et al.*, *Nature* **536**, 419–424 (2016).
27. C.-C. Wang *et al.*, *Nat. Commun.* **10**, 590 (2019).
28. R. Pinhasi, J. Fort, A. J. Ammerman, *PLOS Biol.* **3**, e410 (2005).
29. I. Mathieson *et al.*, *Nature* **555**, 197–203 (2018).
30. R. Fregel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6774–6779 (2018).
31. I. Olalde *et al.*, *Science* **363**, 1230–1234 (2019).
32. S. Brunel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 12791–12798 (2020).
33. C. Zou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023548118 (2021).
34. F. Emanuelli *et al.*, *BMC Plant Biol.* **10**, 241–241 (2010).
35. S. Kobayashi, N. Goto-Yamamoto, H. Hirochika, *Science* **304**, 982–982 (2004).
36. A. R. Walker *et al.*, *Plant J.* **49**, 772–785 (2007).
37. A. R. Walker, E. Lee, S. P. Robinson, *Plant Mol. Biol.* **62**, 623–635 (2006).
38. P. J. Richerson, R. Boyd, R. L. Bettinger, *Am. Antiq.* **66**, 387–411 (2001).
39. R. G. Allaby, C. J. Stevens, L. Kistler, D. Q. Fuller, *Trends Ecol. Evol.* **37**, 268–279 (2022).
40. R. S. Meyer, M. D. Purugganan, *Nat. Rev. Genet.* **14**, 840–852 (2013).
41. Code for: Y. Dong *et al.*, Dual domestications and origin of traits in grapevine evolution, Zenodo (2023); https://doi.org/10.5281/zenodo.7523647.

ACKNOWLEDGMENTS

We thank F. Pelsy, L. Garmendia Auckenthaler, A.-F. Adam-Blondon, C. Cornier, P. Kozma, O. Bachmann, F. Gillet, J.-M. Gobat, S. Dedet, J. Daumann, K. Huber, V. Risovannaya, A. Polulyah, B. Louis, M. Lafargue, G. Jean-Pascal, G. Melyan, D. I. Sumedrea, A. Naqinezhad, M. Filipova, technical staff from EGFV and UEVB, and the Danube-Auen National Park for assistance in the sample collection and laboratory work and P. Kupfer, E. D. O. Roberson, and D. Petkova for comments on the manuscript. **Funding:** This work was supported by the Natural Science Foundation of China (grant 32070599 to W.C.); Yunnan Agricultural University (Research Fund A2032002519 to W.C.); China Agriculture Research System of MOF and MARA CARS-29 (S.W.); the Science Committee at the Ministry of SCS (RA 20APP-4E007 to K.M.); Alliance of International Science Organization (ANSO-CR-PP-2020-04-A to K.M.); Ministerio de Ciencia, Innovación y Universidades and Agencia Estatal de Investigación of Spain (RTI2018-094470-R-C21 to R.A.G.); Predoctoral Fellowship PRE2019-088446 (A.R.I.); Israel Ministry of Science and Technology (90-23-020-12 to E.D.); Fondation Giacomi and Swiss National Science Foundation (SNSF 43307 to C.A.); European Regional Fund (KK.05.1.1.02.0010 to G.Z.); Georgian state budget (L.U., K.B., and T.Z.); TUBITAK and Ministry of Agriculture and Forestry of Republic of Türkiye (grant 105G078 to A.E.); and the Israel Science Foundation (551/18 to E.W.). **Author contributions:** Conceptualization:

Y.D., Z.L., S.W., J.S., W.C.; Formal analysis: S.D., Q.X., X.D.; Funding acquisition: W.C.; Investigation: Y.Z., C.M., S.W., S.L., L.T., C.W., D.L., Y.P., J.L., L.Y., X.L., G.X., Z.Y., B.C., Y.W., P.G., M.R., O.R., A.R.I., Y.W., S.Z.; Resources: Z.L., K.M., M.M., S.G., G.Z., P.F.B., T.L., F.R., P.N., K.B., G.D.B., E.D., G.D.L., J.C., C.F.P., R.A.G., C.A., A.E., Z.D., V.K., G.S., N.G., S.D., N.O., P.T., C.M., V.L., A.J., L.U., T.Z., D.M., M.H., G.J., E.K., T.D., F.G., F.M., F.S., J.E.D., A.M.D., D.C., G.M., T.U., C.Ö., K.K., M.X., J.L., M.Z., L.W., S.J., Y.Z., L.S., S.L.; Supervision: Y.D., H.Y., Y.Z., S.W., J.S., W.C.; Validation: All authors participated in the interpretation of the data; Visualization: S.D., Q.X., X.D.; Writing - original draft: Y.D., S.D., Q.X., X.D., W.C.; Writing - review & editing: W.C. with input from all coauthors. **Competing interests:** A.J. is the founder and owner of Historische Rebsorten vineyard. The remaining authors declare no competing interests. **Data and materials availability:** The VS-1 genome assembly is available at the Genome Warehouse in the National Genomics Data Center, China National Center for Bioinformation, under accession numbers CRA006898 and GWHBQCW00000000. The raw resequencing data are available at the Genome Warehouse in the National Genomics Data Center, China National Center for Bioinformation, under accession number CRA006917. The code for the work can be accessed at Zenodo (*41*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add8655
Materials and Methods
Supplementary Text
Figs. S1 to S27
Tables S1 to S35
References (*42*–*153*)
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 11 July 2022; accepted 23 January 2023
10.1126/science.add8655

HEMATOPOIESIS

USB1 is a miRNA deadenylase that regulates hematopoietic development

Ho-Chang Jeong[1,2]†, Siddharth Shukla[3,4]†‡, Wilson Chun Fok[1,2], Thao Ngoc Huynh[3,4], Luis Francisco Zirnberger Batista[1,2]*, Roy Parker[3,4]*

Mutations in the 3′ to 5′ RNA exonuclease USB1 cause hematopoietic failure in poikiloderma with neutropenia (PN). Although USB1 is known to regulate U6 small nuclear RNA maturation, the molecular mechanism underlying PN remains undetermined, as pre-mRNA splicing is unaffected in patients. We generated human embryonic stem cells harboring the PN-associated mutation c.531_delA in USB1 and show that this mutation impairs human hematopoiesis. Dysregulated microRNA (miRNA) levels in USB1 mutants during blood development contribute to hematopoietic failure, because of a failure to remove 3′-end adenylated tails added by PAPD5/7. Modulation of miRNA 3′-end adenylation through genetic or chemical inhibition of PAPD5/7 rescues hematopoiesis in USB1 mutants. This work shows that USB1 acts as a miRNA deadenylase and suggests PAPD5/7 inhibition as a potential therapy for PN.

Poikiloderma with neutropenia (PN) is an autosomal-recessive bone marrow failure (BMF) syndrome with marked clinical overlap with dyskeratosis congenita (DC) (*1*). However, unlike patients with DC, telomeres are not shortened in patients suffering from PN, providing a distinguishable feature for the correct diagnosis of PN (*2*). PN patients harbor homozygous or compound heterozygous mutations in the human gene *C16orf57*, which encodes the conserved 3′ to 5′ RNA exonuclease U6 biogenesis 1 (USB1) (*2*–*5*). USB1 is required for the processing of U6 and U6atac small nuclear RNAs (snRNAs), and some splicing defects are observed when using yeast and zebrafish models of USB1 deficiency (*6*–*10*). However, lymphoblastoid cells from PN patients do not exhibit reduced U6 snRNA levels and have normal pre-mRNA splicing (*8*). These results establish USB1-mediated PN as a singular BMF syndrome, in which the underlying genetic cause has been identified but the molecular mechanisms leading to tissue failure are unknown.

[1]Division of Hematology, Department of Medicine, Washington University in St. Louis, St. Louis, MO 63110, USA. [2]Center for Genome Integrity, Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110, USA. [3]Department of Biochemistry, University of Colorado, Boulder, CO 80303, USA. [4]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.
***Corresponding author. Email: lbatista@wustl.edu (L.F.Z.B.); roy.parker@colorado.edu (R.P.)**
†These authors contributed equally to this work.
‡Present address: Alltrna Inc., Cambridge, MA 02139, USA.

USB1 mutant hESCs have impaired hematopoietic development

To investigate the role of USB1 in a physiological context, we utilized CRISPR-Cas9 to create human embryonic stem cells (hESCs) containing a frequently occurring c.531_del_A loss-of-function mutation in the USB1 gene (hereafter referred to as USB1 mutant) (fig. S1, A and B). These USB1 mutant hESCs have normal karyotype (fig. S1C), have normal growth rate (fig. S1D), are pluripotent (fig. S1E), and display normal telomere length (fig. S1F), indicating that a clinically relevant USB1 mutation is not deleterious in undifferentiated hESCs.

To elucidate the role of USB1 during hematopoiesis, we performed serum-free hematopoietic differentiations (*11*–*15*) to derive hematopoietic progenitor cells from hESCs (Fig. 1A). Gene expression analysis confirmed the efficiency of this protocol, with silencing of pluripotency markers and efficient formation of hematopoietic lineages at the end (day 30) of differentiation (fig. S2A). USB1 mutant cells did not show any impairment during early stages of hematopoietic differentiation, including the formation of mesoderm (day 3; fig. S2B), and $CD34^+/CD43^-$ hemogenic endothelium (HE) populations (day 8; fig. S2C). However, the formation of $CD45^+$ hematopoietic progenitors (day 16) was decreased in USB1 mutant cells compared to wild-type (WT) cells (fig. S2D), and hematopoietic colony potential analysis showed compromised colony formation in USB1 mutant cells (Fig. 1B). Consistent with a role of USB1 in regulating hematopoiesis, USB1 mRNA levels increased about threefold in mature blood cells compared to undifferentiated hESCs (Fig. 1C). These observations indicate that loss-of-function mutations in USB1 negatively influence hematopoiesis.

As PN is usually associated with severe noncyclic neutropenia (*1*), we specifically analyzed the potential of neutrophil formation in WT and USB1 mutant cells. USB1 mutants had reduced formation of $CD15^+/CD66b^+$ lineages, indicating abnormal neutrophil development (Fig. 1D and fig. S2E). The conditional expression of the WT USB1 protein in USB1 mutants with the use of a Dox-inducible system (fig. S2F) rescued the hematopoietic potential of these cells (Fig. 1E). These results recapitulate major clinical manifestations of USB1 deficiency and establish USB1 as an important regulator of hematopoiesis.

Mutations in USB1 lead to increased 3′ adenylation but do not affect levels of the U6 snRNA in human stem cells and hematopoietic progenitors

To determine the mechanism by which USB1 regulates hematopoiesis, we initially examined if the USB1 mutation affected U6 snRNA. Northern blot analysis of WT and USB1 mutant cells at undifferentiated (D0) and hematopoietic progenitor (D16) stages showed no reduction in the levels of U6 and U6atac snRNAs in USB1 mutants (Fig. 1F and fig. S2, G and H). However, we observed that U6 and U6atac snRNA from USB1 mutant cells were slightly longer compared to WT cells (Fig. 1F), indicating aberrant posttranscriptional processing of these snRNAs similar to what is observed in patient-derived cells (*8*).

Sequencing the 3′ end of U6 snRNA from WT and USB1 mutant cells revealed two changes.