nature plants

ARTICLES

https://doi.org/10.1038/s41477-019-0437-5

Palaeogenomic insights into the origins of French grapevine diversity

Jazmín Ramos-Madrigal[1], Anne Kathrine Wiborg Runge[1,2], Laurent Bouby[3], Thierry Lacombe[4], José Alfredo Samaniego Castruita[1], Anne-Françoise Adam-Blondon[5], Isabel Figueiral[6], Charlotte Hallavant[7], José M. Martínez-Zapater[8], Caroline Schaal[9], Reinhard Töpfer[10], Bent Petersen[1,11], Thomas Sicheritz-Pontén[1,11], Patrice This[4], Roberto Bacilieri[4], M. Thomas P. Gilbert[1,12*] and Nathan Wales[1,2,13,14*]

The Eurasian grapevine (*Vitis vinifera*) has long been important for wine production as well as being a food source. Despite being clonally propagated, modern cultivars exhibit great morphological and genetic diversity, with thousands of varieties described in historic and contemporaneous records. Through historical accounts, some varieties can be traced to the Middle Ages, but the genetic relationships between ancient and modern vines remain unknown. We present target-enriched genome-wide sequencing data from 28 archaeological grape seeds dating to the Iron Age, Roman era and medieval period. When compared with domesticated and wild accessions, we found that the archaeological samples were closely related to western European cultivars used for winemaking today. We identified seeds with identical genetic signatures present at different Roman sites, as well as seeds sharing parent–offspring relationships with varieties grown today. Furthermore, we discovered that one seed dated to ~1100 CE was a genetic match to 'Savagnin Blanc', providing evidence for 900 years of uninterrupted vegetative propagation.

Since its domestication in Southwestern Asia more than 6,000 years ago[1–3], the Eurasian grapevine (*Vitis vinifera* L.) has become one of the world's most widely produced and economically valuable fruit crops. Although grapevine products are widely consumed as table grapes, dried raisins, fruit preserves and cooked leaves, both archaeological and historical evidence indicates that wine has been its primary use[4,5]. A key unresolved question in ancient viniculture is the origin and proliferation of vegetative propagation[6]. Like many other fruit crops, grapevine is grown almost exclusively as clonal lineages, in which favoured varieties are maintained through horticultural techniques, such as grafting, layering and planting of shoots[7,8]. These methods take advantage of its natural ability to reproduce asexually under certain conditions and, ultimately, enable the establishment of genetic clones of valuable cultivars. With vegetative propagation, viniculturists can consistently harvest berries with a desired flavour profile and, with relatively limited effort, have the potential to expand cultivars to new vineyards and distant regions. The alternative approach of sowing seeds is unreliable because grapevine genomes are highly heterozygous and individuals grown from seed are highly diverse in quality, yield, phenotype and phenology[8]. Moreover, winemakers have to wait from 3 to 5 years until vines reach maturity[9], before it is possible to assess berry quality and yield. Thus, clonal lineages of high-quality vines have become indispensable in modern viniculture. Discovering the antiquity of vegetative propagation technologies and the unique histories of individual grapevine varieties will mark a major advancement in our understanding of ancient viniculture, provide a means to investigate longstanding local agricultural traditions and generate pertinent information for the future development of breeding schemes (for example, through a better understanding of why some varieties have been more successful than others, or adding historical value to present-day cultivars).

The history of winemaking in France provides a useful model to explore how vegetative propagation helped to establish ancient vineyards, and how those actions ultimately shaped the economy and landscape of one of the world's most esteemed winegrowing countries. Written sources and archaeological records indicate that vineyards were first planted at the Greek colony of *Massalia*, present day Marseille, during the sixth century before Common Era (BCE)[10,11]. Winemaking subsequently spread along the Mediterranean coast[12], but it was not until the end of the first century BCE that Romans greatly increased wine production across southern France[10]. Roman authors, including Pliny the Elder in the first century CE (see ref. [13], Book XIV), discussed grafting and grapevine varieties, thereby demonstrating their proficiency in vegetative propagation techniques. While Pliny describes 91 varieties, it is currently impossible to link Roman names to modern grapevines; however, it is frequently speculated that some living varieties were grown by the

[1]Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [2]BioArCh, Department of Archaeology, University of York, York, UK. [3]ISEM—UMR 5554, CNRS-IRD-EPHE-Université Montpellier, Montpellier, France. [4]UMR AGAP, Université Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France. [5]URGI, Unité de Recherche Génomique-Info, UR1164, INRA, Université Paris-Saclay, Versailles, France. [6]Inrap, Méditerranée and ISEM—UMR 5554, CNRS-IRD-EPHE-Université Montpellier, Montpellier, France. [7]Bureau d'études Hadès, laboratoire TRACES—UMR 5608 (pôle Terrae)—UT2J, Toulouse, France. [8]Instituto de Ciencias de la Vid y del Vino (CSIC-UR-Gobierno de La Rioja), Logroño, Spain. [9]GéoArchEon Sarl, Laboratoire Chrono-Environnement—UMR 6249, Université de Franche Comté, Besançon, France. [10]Julius Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen, Institut für Rebenzüchtung Geilweilerhof, Siebeldingen, Germany. [11]Centre of Excellence for Omics-Driven Computational Biodiscovery, Faculty of Applied Sciences, AIMST University, Kedah, Malaysia. [12]NTNU University Museum, Trondheim, Norway. [13]Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA. [14]Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université Paul Sabatier, Toulouse, France. *e-mail: tgilbert@bio.ku.dk; nathan.wales@york.ac.uk

Romans, and that those genetic clones have been maintained for two millennia[9]. After the fall of the Roman Empire, winemaking traditions continued in France, and by the Middle Ages, contemporary variety names appear in written records[14]. Even though historic names are still used today, it remains unknown whether the same genetic clone has been maintained, or if names have been assigned to other lineages.

Archaeobotanical remains, in particular, seeds, have the potential to shed new light on the legacy of French grapevine varieties and, more generally, on the history of viniculture. Using morphometric analyses of seed shape, researchers have shown that seeds from most domesticated grapevines (*V. vinifera* subsp. *vinifera*) can be distinguished from those produced by wild vines (*V. vinifera* subsp. *sylvestris*)[15,16]. With this approach, Bouby et al.[10] determined that early Roman sites in southern France (50 BCE to 225 CE) contained greater numbers of morphologically wild seeds than the following period (225–600 CE), raising the question of whether Romans collected and cultivated wild berries for winemaking. Through this time series, seed shapes tended towards domesticated morphotypes, a finding the authors hypothesize represents a combination of continued selective pressures with a sporadic incorporation of native varieties through sexual reproduction. While these interpretations are thought-provoking, the authors also recognize critiques that some wild and domesticated vines produce morphologically indistinguishable seeds.

One of the most promising avenues of research for ancient viniculture is palaeogenomic (or ancient DNA (aDNA)) analysis of well-preserved archaeological pips[17–19]. For example, Wales et al.[20] demonstrated that many waterlogged grape seeds contain high proportions of endogenous DNA that could be interrogated with state-of-the-art, high-throughput aDNA sequencing. With the establishment of genomic databases for hundreds of modern cultivars and wild grapevines[21], we sought to examine how DNA recovered from archaeological samples could sidestep some of the challenges of conventional archaeobotanical methods and reveal relationships between ancient samples and modern varieties, thereby providing otherwise unachievable insights on past implementation of vegetative propagation and the antiquity of some of the world's most produced grapevine varieties.

Results and discussion

Successful enrichment of single-nucleotide polymorphism loci in archaeological pips. We performed targeted enrichment and shotgun sequencing of 10,000 single-nucleotide polymorphism (SNP) loci in 28 archaeological grape seeds. The pips were recently excavated from waterlogged features (wells, latrines, ditches and pits) at nine French sites (Supplementary Fig. 1) and based on archaeological context, date as early as the Iron Age (510–475 BCE) and as late as the medieval period (1050–1200 CE) (Table 1). SNP loci were selected from the GrapeReSeq panel, a DNA microarray that was developed to authenticate varieties for breeding and germplasm management[21]. This reference panel provides data for 783 domesticated varieties (*V. vinifera* subsp. *vinifera*), 112 wild (*V. inifera* subsp. *sylvestris*) accessions collected from Eurasia and North Africa, and 11 other *Vitis* species. We obtained a 4–400-fold enrichment at the targeted SNP sites, leading to an on-target depth of coverage of 1–25.7× (Supplementary Table 1 and Supplementary Fig. 2). Nucleotide misincorporation patterns observed in the sequencing data and read length distributions were consistent with those expected for degraded DNA[22] (Supplementary Figs. 3, 4 and 5a).

Archaeological seeds related to European winemaking lineages. We employed multidimensional scaling (MDS) to investigate whether archaeological samples were more closely related to wild accessions or domesticated varieties. Samples were compared to the GrapeReSeq panel following the random allele sampling strategy described in bammds[23], to account for varying depth of coverage in the archaeological samples. Additionally, we expanded our reference data set with publicly available whole-genome sequencing data from 27 wild and domesticated grape accessions[24–26] (Supplementary Table 2). The MDS plots showed that all 28 archaeological samples fall within the variability of domesticated grapevines, suggesting that none of the seeds originated from truly wild vines (Fig. 1a). While it is plausible that samples near the boundary of the domesticated and wild clusters could represent F_1 hybrids between domesticated varieties and wild grapevines (for example, specimen R-LLE_09), we find no evidence for large-scale collection of wild berries by Romans or medieval people at these sites. Likewise, the oldest sample, from the Iron Age site of La Cougourlude dating to 510–475 BCE, also falls within the MDS space composed of cultivated grapevines. These findings support the hypothesis by Bouby et al.[10] that even though many pips from Roman and medieval sites exhibit wild morphotypes, they in fact originate from domesticated varieties.

Once we determined that archaeological seeds probably originated from domesticated grapevines, we repeated the MDS analysis without wild accessions to achieve a more refined picture of the relationships to regional varieties and types of berries (that is, predominantly used in winemaking or as table grapes). The majority of the archaeological pips were most closely related to wine cultivars from West and Central Europe (Fig. 1b), although the three early Roman samples from the Mas de Vignoles XIV site had a closer affinity to wine grapes from the Balkans and the Iberian Peninsula. Overall, this analysis shows that the archaeological seeds are predominantly related to western European varieties that are used for winemaking, and not grapevines that are today grown further east for wine or table grapes. These data suggest that, 2,000 years ago, cultivated vines in the modern territory of France were distinct from their near eastern ancestors and well on their way to founding the germplasm of modern varieties used in western European winemaking. We also verified that the patterns observed in the MDS analysis using the GrapeReSeq panel were consistent with those obtained from a whole-genome reference panel (Supplementary Table 2 and Supplementary Fig. 6).

We further explored the genetic structure of the archaeological seeds with a two-step model-based clustering analysis. First, ADMIXTURE[27] was used to infer the ancestry proportions within the samples in the reference panel, and then FastNGSAdmix[28] was used to estimate the ancestry proportions in the archaeological samples (Fig. 1c and Supplementary Fig. 7). The results were consistent with the MDS analysis, showing that most archaeological seeds were related to wine grapes from western Europe.

As there is evidence for gene flow with local wild grapevines in western Europe[1], we explored the wild ancestry components identified through the clustering analysis. As these proportions are estimated on the GrapeReSeq SNPs, they do not necessarily represent whole-genome ancestry proportions. However, this allowed us to: (1) compare the proportions between present-day varieties and the archaeological seeds at these diagnostic sites, and (2) identify the potential source of the wild grape ancestry in the archaeological seeds. Wild grapevines carry four main ancestry components when assuming eight clusters (Fig. 1c). While American and Asian *Vitis* species (yellow) and Eurasian wild grapes from the Caucasus and Turkey (light blue) separate into individual clusters that do not contribute significant ancestry to any other group, wild grapes from the African and western European populations display two ancestry components (dark and light green) that are found in some domesticated grapes. All archaeological samples except for the most recent (M-LM_22) show evidence of genetic contributions from wild grapevines (Fig. 1c and Supplementary Fig. 7), and these wild ancestries are primarily associated with western and Central European vines. While these data provide the first clues on the

Table 1 | Description of the archaeological grape seeds used in the study

No.	Sample ID	Geographical coordinates	Archaeological site (France)	Stratigraphic unit	Structure	Age	Dating method	Period	GC[a]
1	IA-LC_01	43.573639° N, 3.914750° E	La Cougourlude, Lattes	US 31084	Ditch FO 30277	510–475 BCE/2,480 ± 30 yr BP (769–417 calendar years BCE)	Archaeological artifacts or ^{14}C	Iron Age	
2	R-MDV14_04	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	Second to first century BCE	Archaeological artifacts	Early Roman	A
3	R-MDV14_07	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	Second to first century BCE	Archaeological artifacts	Early Roman	A
4	R-MDV14_09	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	Second to first century BCE	Archaeological artifacts	Early Roman	A
5	R-MF_21	43.432556° N, 3.394222° E	Mont Ferrier, Tourbes	US 2076	Well PT 2052	First century CE	Archaeological artifacts	Roman	B
6	R-MF_23	43.432556° N, 3.394222° E	Mont Ferrier, Tourbes	US 2076	Well PT 2052	First century CE	Archaeological artifacts	Roman	
7	R-MF_25	43.432556° N, 3.394222° E	Mont Ferrier, Tourbes	US 2076	Well PT 2052	First century CE	Archaeological artifacts	Roman	B
8	R-HW70_18	48.080500° N, 7.399194° E	Horbourg-Wihr	NA	Pit ST7054	Second century CE	Dendrochronology or archaeological artifacts	Roman	C
9	R-HW71_03	48.080500° N, 7.399194° E	Horbourg-Wihr	NA	Pit ST7172	Second century CE	Archaeological artifacts	Roman	C
10	R-HW71_17	48.080500° N, 7.399194° E	Horbourg-Wihr	NA	Pit ST7172	Second century CE	Archaeological artifacts	Roman	C
11	R-R_09	43.471306° N, 3.670139° E	Roumeges, Poussan	US 5007(12/13)	Well PT 5001	First to third century CE	Archaeological artifacts	Roman	
12	R-R_14	43.471306° N, 3.670139° E	Roumeges, Poussan	US 5007(12/13)	Well PT 5001	First to third century CE	Archaeological artifacts	Roman	
13	R-LLE_02	43.300806° N, 3.239917° E	La Lesse-Espagnac, Sauvian	US 3019	Well PT 3005	175–225 CE	Archaeological artifacts	Roman	
14	R-LLE_08	43.300806° N, 3.239917° E	La Lesse-Espagnac, Sauvian	US 3019	Well PT 3005	175–225 CE	Archaeological artifacts	Roman	C
15	R-LLE_09	43.300806° N, 3.239917° E	La Lesse-Espagnac, Sauvian	US 3019	Well PT 3005	175–225 CE	Archaeological artifacts	Roman	
16	R-LLE_13	43.300806° N, 3.239917° E	La Lesse-Espagnac, Sauvian	US 3019	Well PT 3005	175–225 CE	Archaeological artifacts	Roman	D
17	R-LLE_14	43.300806° N, 3.239917° E	La Lesse-Espagnac, Sauvian	US 3019	Well PT 3005	175–225 CE	Archaeological artifacts	Roman	D
18	R-TDM_06	43.472806° N, 3.223000° E	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	Fourth century CE	Archaeological artifacts	Roman	E
19	R-TDM_08	43.472806° N, 3.223000° E	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	Fourth century CE	Archaeological artifacts	Roman	E
20	R-TDM_10	43.472806° N, 3.223000° E	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	Fourth century CE	Archaeological artifacts	Roman	
21	M-MDV13_07	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 13525	Well PT 13319	1,605 ± 35 yr BP (417–515 CE)	^{14}C	Late Roman/ medieval	

Continued

Table 1 | Description of the archaeological grape seeds used in the study (Continued)

No.	Sample ID	Geographical coordinates	Archaeological site (France)	Stratigraphic unit	Structure	Age	Dating method	Period	GC[a]
22	M-MDV12_02	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1,220 ± 30 yr BP (731–851 CE)	^{14}C	Early medieval	
23	M-MDV12_04	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1,220 ± 30 yr BP (731–851 CE)	^{14}C	Early medieval	F
24	M-MDV12_05	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1,220 ± 30 yr BP (731–851 CE)	^{14}C	Early medieval	
25	M-MDV12_07	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1,220 ± 30 yr BP (731–851 CE)	^{14}C	Early medieval	
26	M-MDV12_09	43.808222° N, 4.368222° E	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1,220 ± 30 yr BP (731–851 CE)	^{14}C	Early medieval	F
27	M-C_27	45.436417° N, 5.520306° E	Colletiere, Charavines	NA	Cultural layer, rubbish deposits	1006–1040 CE	Dendrochronology	Medieval	
28	M-LM_22	47.900472° N, 1.884333° E	La Madeleine, Orléans	US 15126	Cesspit F 1517	1050–1200 CE	Archaeological artifacts	Medieval	

[a]Genetic clusters composed of identical clones; the genetic cluster was assigned according to the relatedness analyses described in the Results section. NA, not attributed.

timing of genetic introgression from local vines into domesticated lineages, the amount of wild ancestry does not follow a consistent pattern related to sample age. For example, the oldest sample (La Cougourlude, 510–475 BCE) shows some of the highest levels of wild ancestry (~45%), whereas other early samples from Mas de Vignoles XIV (second to first century BCE) have marginal amounts of wild ancestry (3.5–4.5%), and five samples from La Lesse-Espagnac (175–225 CE) range from ~10% to 38%. In fact, these proportions of wild ancestry are similar to those observed in modern French varieties, suggesting that the admixture with wild grapevines took place at the earliest stages of viniculture in France, and potentially before other cultivated lineages were introduced to France (that is, from Greece or the Italian Peninsula). Together, these results suggest that the local wild gene pool played an early role for domesticated varieties, with the gene flow between wild grapevines and domesticated cultivars occurring at least 2,500 years ago.

Ancient use of vegetative propagation. The availability of genotype data for hundreds of cultivars in the GrapeReSeq panel, allowed us to explore relationships between archaeological pips excavated from individual sites and across different regions of France. We estimated kinship coefficients among pairs of samples using KING[29] and NgsRelate[30]. Pairs of samples were classified based on the kinship coefficients and the proportion of sites with 'zero alleles identical by state' (IBS0)[29] into the following categories: identical clones, parent–offspring, highly related/full siblings or unrelated[21] (Supplementary Table 3). We found six instances of genetically identical pairs or groups of seeds (Fig. 2a). Additionally, we identified first-degree relationships (parent–offspring and highly related/full siblings) and unrelated varieties (Fig. 2b). However, as grape seeds that have been cross-fertilized contain paternal-derived DNA[31], which could affect the relatedness analyses, we explored whether the archaeological seeds contained maternal DNA only (as expected from empty seeds), or both paternal and maternal DNA. To do so, we generated sequencing data from three seeds and a wood sample of the same plant and conducted a simulation study, in an attempt to estimate the parental contribution in the archaeological samples (Supplementary Fig. 8 and Supplementary Methods 16). We found that data from all archaeological seeds, except R-TDM_06, R-TDM_08, R-HW71_03 and M-MDV12_09, were consistent with a paternal DNA contribution of ≤10% (Supplementary Figs. 8–11). Moreover, we studied the dependence of the relatedness analyses on such contribution and found that ≤10% paternal DNA does not significantly affect the results (Supplementary Fig. 12). Thus, we consider that clonal and parent–offspring relationships are not affected in most samples. Conversely, full-sibling relationships could derive from multiple scenarios if the samples involved contain paternal DNA (Supplementary Fig. 12c); thus, we classified pairs of samples with this type of relationship as 'highly related'.

Grape seeds have been found to follow a degradation process of the two tissues that contain paternal DNA, the endosperm and embryo, resulting in empty seeds (for example, in up to 30% of the cases for the 'Chardonnay' variety[32,33]). Our results suggest that the observed clonal clusters among archaeological samples represent empty seeds with only maternal tissue, either produced by the same plant, such as might occur at one archaeological site, or by one grapevine variety spread through vegetative propagation (Fig. 2 and Supplementary Table 3). Five of these clonal clusters consist of two or three seeds from a single stratigraphic context: an early Roman ditch at Mas de Vignoles XIV near Nîmes city (second to first century BCE), a Roman well at Mont Ferrier, Tourbes (first century CE), a Roman well at La Lesse-Espagnac (about 200 CE), a Roman well at Terrasses de Montfau, Magalas (fourth century CE) and an early medieval well at Mas de Vignoles XIV (about 800 CE). Given that bunches of grapes might have been pressed for juice and discarded en masse, these genetically identical specimens may well represent seeds from single plants. The other genetic cluster consists of three seeds from Horbourg-Wihr in Alsace and one seed from La Lesse-Espagnac in Mediterranean France (Fig. 2b); while all four samples date to the second century CE, these genetic clones suggest that Romans transported grapevine across long distances (>600 km), most likely as cuttings.

Five archaeological sites in southern France demonstrated the presence of multiple genotypes within a single temporal stratum,

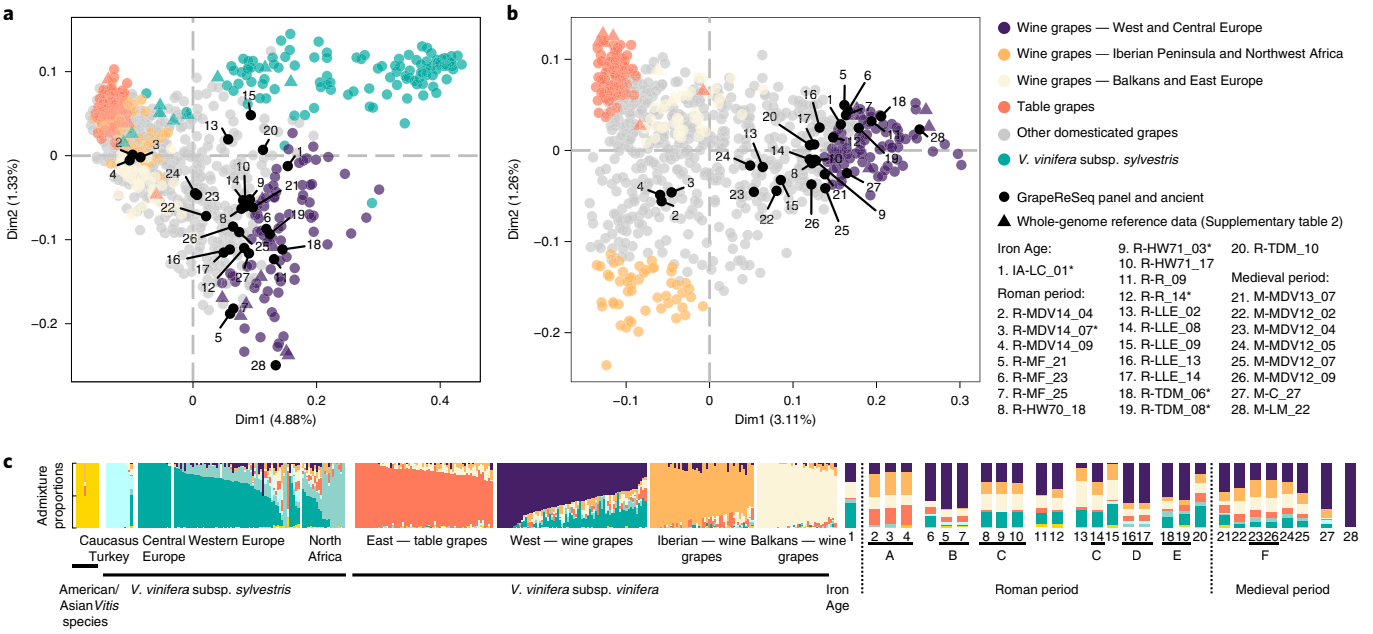

Fig. 1 | Genetic affinities between archaeological grape seeds and modern *V. vinifera* accessions. a, MDS plot including archaeological samples, wild *V. vinifera* subsp. *sylvestris* accessions and domesticated varieties. **b**, MDS plot restricted to archaeological samples and domesticated varieties. The colours correspond to the main ancestry clusters identified in Laucou et al.[21]. Asterisks indicate archaeological samples that were incorporated into the data set by sampling a random allele from a majority count consensus sequence instead of called genotypes. **c**, Model-based clustering analysis of the GrapeReSeq panel assuming $K = 8$ clusters. The vertical bars represent individual accessions, the colours represent the inferred ancestry components and the fraction of each colour corresponds to the estimated ancestry proportion. Archaeological samples are sorted by age and by sample identification within a stratigraphic context. Samples that were identified as identical clones are grouped with black lines and capital letters (A–F) at the bottom.

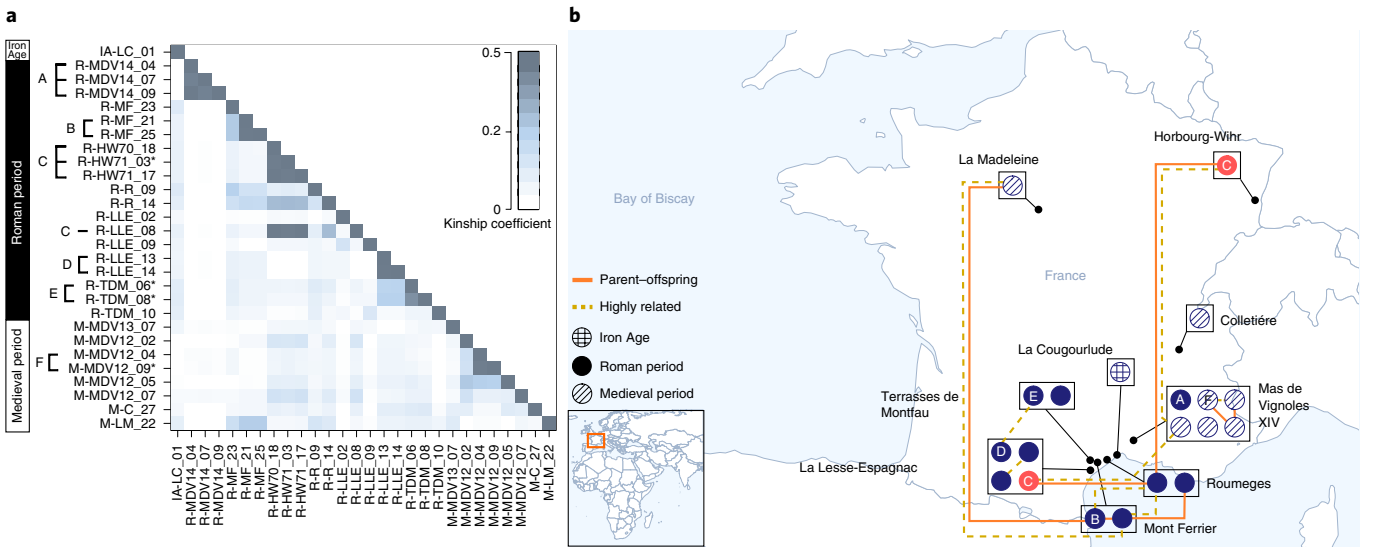

Fig. 2 | Geographical distribution and relationships between the distinct genetic types of archaeological samples. a, Relatedness among pairs of archaeological samples. Kinship coefficients were estimated using NgsRelate between pairs of samples for SNP loci present in the GrapeReSeq panel. The capital letters (A–F) on the left indicate genetically identical clones, that is, putative ancient and historical varieties. Asterisks indicate archaeological seeds that were found consistent with carrying >10% paternal DNA. **b**, Map displaying the distribution of genetic types (circles) in each archaeological site. The capital letters (A–F) on the circles indicate clusters of genetically identical seeds represented by more than one seed. The shading of the circles indicates sample age. In red is shown the genetic type that was found in more than one archaeological site. The lines connect pairs of samples that are related as parent–offspring or highly related/full sibling. Note that, as in the presence of paternal DNA full-sibling relationships could derive from multiple scenarios, we classified samples consistent with full-sibling relationships as 'highly related' (see Supplementary Methods 16).

providing genetic evidence that multiple lineages or varieties were maintained at individual vineyards. For example, we identified six different genotypes at Mas de Vignoles XIV near Nîmes, three of which shared first-degree relationships and three of which were unrelated (Fig. 2b). Overall, these relationship data indicate that vegetative propagation, long-distance transportation of varieties and multivarietal cultivation have been practiced in France since the Roman era, consistent with historic accounts[4].

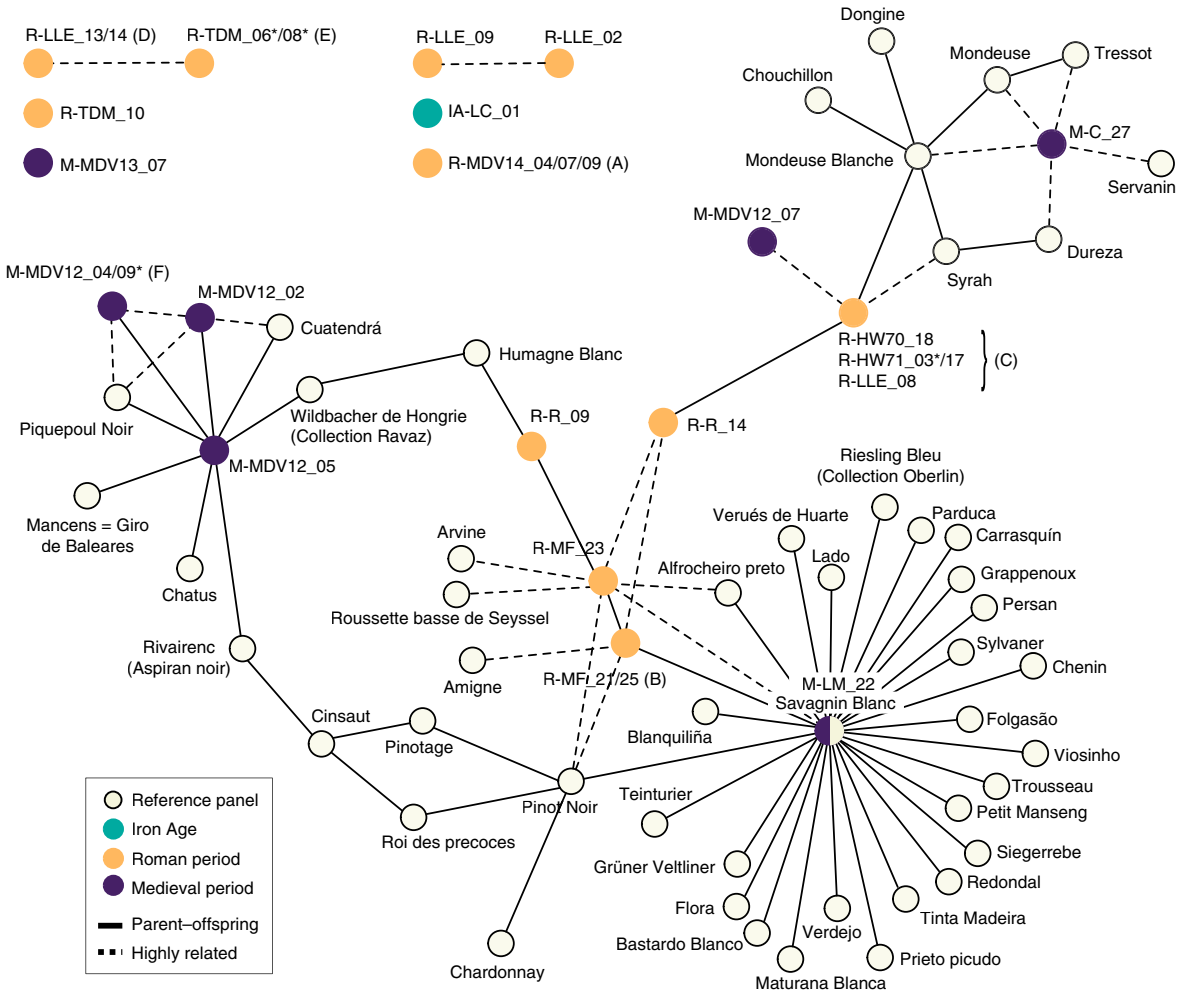

Fig. 3 | Genetic origins of ancient and historic French grapevine varieties. Relationships identified between archaeological samples and modern cultivars included in the GrapeReSeq panel. Highly related refers to full-sibling or similar samples. Sibling relationships involving pairs of modern cultivar are not displayed for simplicity. Asterisks indicate archaeological seeds that were found consistent with carrying >10% paternal DNA. The VIVC (http://www.vivc.de) and GrapeReSeq identifiers for the modern cultivars can be found in Supplementary Table 4.

The antiquity of modern French varieties. We finally investigated the relatedness between archaeological and modern varieties, by computing kinship coefficients and the proportion of IBS0 sites between pairs of archaeological samples and samples present in the GrapeReSeq panel using KING[29] (Fig. 3 and Supplementary Tables 4 and 5). Our results confirm long-held beliefs that Roman and medieval viticulturists maintained ancient lineages using vegetative propagation[13], and that modern French viniculture is in large part a product of these traditions. One archaeological sample from La Madeleine (Orléans), dating to 1050–1200 CE, was an identical clone of 'Savagnin Blanc' (VIVC17636), a variety today cultivated for wine production in northeastern France and other countries from Central Europe (kinship coefficient = 0.496; IBS0: ~0.0001; identity of 99.7% and 99.9% for the GrapeReSeq and whole-genome panels, respectively) (Supplementary Tables 4 and 5). Several researchers previously identified 'Savagnin Blanc', also known as 'Traminer Weiss', as a recurrent parent of many commercially important European varieties[1,34,35], and written accounts document the appellation as early as 1539 CE[36]. Our findings extend the presence of this variety in France by hundreds of years and also suggest that either 'Savagnin Blanc' or its direct relatives have been cultivated in France since the first century CE, as archaeological seeds from Mont Ferrier, Tourbes, have a parent–offspring relationship with 'Savagnin Blanc' (Figs. 2b and 3).

Several archaeological seeds were closely related to 'Mondeuse Blanche' (VIVC7919), a French variety characteristic of the northern French Alps that has been suggested to have acted as a key progenitor[35,37]. We found that four genetically identical second century CE seeds from Horbourg-Wihr and La Lesse-Espagnac have a parent–offspring relationship with 'Mondeuse Blanche', indicating that just one reproductive cycle has taken place in this lineage in the past 1,800 years (Fig. 3). This finding presents an exciting consilience of genetic and archaeobotanical data; using morphometric analysis, Terral et al.[16] also found evidence for 'Mondeuse Blanche' among first to second century CE pips from the Rec-de-Ligno site, which lies less than 10 km from La Lesse-Espagnac. We also observed that 'Mondeuse Blanche' is highly related (full-sibling or similar relationship) to an archaeological seed from Colletiere, dating to about 1000 CE, close to the region where 'Mondeuse Blanche' is still grown today (Savoie, Ain) (Fig. 3). Interestingly, the medieval seed is also highly related to 'Tressot' (12640) (cited since 1396 in France[38]) and 'Servanin' (VIVC11526), both French varieties that are rarely cultivated today.

In addition to 'Mondeuse Blanche', four other Roman seeds from southern France provided parent–offspring relationships to modern Alpine varieties: three first century CE seeds from Mont Ferrier are highly related to 'Arvine' (VIVC664) and 'Amigne' (VIVC425)

and one first to third century CE seed from Roumeges is a first-degree relative to 'Humagne Blanc' (VIVC5450) (Fig. 3). All three are Swiss varieties used for white wine, and the former two are recorded in Switzerland by the seventeenth century CE[39]. Tradition holds that the Romans brought 'Amigne' to Switzerland as a variety they referred to as '*Aminea*'; however some researchers have suggested the connection is primarily etymological, with the retained usage of the Latin word *amoenus* for 'delicious'[40]. Our findings suggest that there indeed is a close genetic link between the varieties grown by the Romans and some modern Swiss cultivars, including 'Amigne'. Moreover, these data indicate that modern Alpine varieties may have been cultivated in a more widespread geographical region during the Roman period, thus posing an important question on their origin and the adaptation of modern grapes. The approaches established here can be applied to other archaeological pip assemblages with the aim of detecting when regional and economically important lineages first appeared and how they were maintained.

Impact of cultural changes in the viniculture of France. Specimens from the Mas de Vignoles XIV site in Nîmes provide one final observation on the changing nature of viniculture in France. This site allowed us to investigate a transect of three time periods: second to first century BCE in the early Roman period, 417–515 CE in the late Roman period when viticulture was fully established in the region, and 731–851 CE in the early medieval period. While cultivars from the most recent period were found to share first-degree relationships with modern French varieties, no relationship was found between cultivars from the Roman period and the modern varieties (Fig. 3). Our results from Mas de Vignoles XIV suggest a change in grapevine diversity from Roman to medieval times. This transition can also be observed in the MDS analyses (Fig. 1b); the three seeds from the early Roman period (R-MDV14_04/07/09) are placed closer to East European and Iberian grape varieties, whereas late Roman and early medieval seeds are more similar to West Europe varieties. These results show the relatively high diversity of grapes cultivated in this region during this period, as well as replacement and incorporation of new varieties through time.

Concluding remarks

Palaeogenomic analysis of archaeobotanical remains has helped to reveal the evolutionary histories of annual crops, such as barley[41] and maize[42,43], but this project represents the first nuclear aDNA study of a vegetatively propagated fruit crop. Our results highlight the utility of state-of-the-art palaeogenomic methods in the study of ancient viniculture through space and time. While previous studies on ancient chloroplast DNA[20], microsatellites[18,19,44] and proteins[18] have provided insights into the history of grapevine cultivation, their resolution is limited. With the availability of a nuclear DNA diversity panel, we interrogated genome-wide data from archaeological grape seeds, identified relationships between ancient pips and modern varieties, observed connections between distant sites and traced the history of vegetative propagation in France. Future palaeogenomic research on archaeological grape seeds holds great potential in identifying the links between past and present grape varieties, and especially for refining our knowledge of the pace of domestication and improvement under vegetative propagation[45].

Methods

Archaeological sample description. Grape seeds were collected from nine archaeological sites in France during excavations of wells, latrines, pits and ditches (Supplementary Fig. 1; see Supplementary Methods 1 for a description of the archaeological sites). Sediment samples were systematically collected and immediately isolated to prevent contamination and stored in cool conditions (4 °C). The sediment samples were processed at the Institut des Sciences de l'Evolution (ISEM) in Montpellier, France. To prevent contamination with modern material, seeds were isolated in a clean room separate from the archaeobotanical laboratory. Additionally, surfaces and tools were cleaned with bleach before handling. Most of the samples included in this study were photographed inside the clean room, with specific equipment to carry out morphological analyses. Archaeological samples were dated either by association with archaeological artifacts found in the same stratigraphic units, dendrochronology or radiocarbon dating. The age of the samples ranged from the Iron Age (510–475 BCE) to the medieval period (1050–1200 CE) (Table 1 and Supplementary Fig. 1).

Archaeological samples processing. Archaeological samples were processed in dedicated aDNA facilities at the University of Copenhagen (Copenhagen, Denmark) following standard measures to prevent contamination. Individual seeds were decontaminated with 10% bleach, rinsed with molecular biology grade water and pulverized. DNA was extracted from the resulting powder following standard protocols standardized for archaeobotanical remains[46]. DNA extracts were converted into double-stranded Illumina libraries using the NEBNext DNA Library Prep Mast Mix Set 2 (E6070L, New England Biolabs) with modifications described in Wales et al.[47] (see Supplementary Methods 4 for a description of the protocol). Resulting Illumina libraries were enriched for a set of genomic loci present in the GrapeReSeq reference panel[21] (Supplementary Methods 5). This panel covers genomic sites known to be informative for the identification of grape cultivars. Libraries were captured following the MYbaits protocol as described in Supplementary Methods 6. Finally, pre-capture and post-capture libraries were sequenced on an Illumina 2500 HiSeq platform in SR100 mode. Sequencing reads obtained from the pre-captured libraries were used to assess the capture efficiency only.

Sequencing data processing. AdapterRemoval2.0 (ref. [48]) was used to remove Illumina adapter sequences, low-quality stretches and ambiguous bases from the read ends. Resulting reads of ≥30 base pairs were mapped to the grape nuclear reference genome 12×0.2 (ref. [49]), chloroplast[50] and mitochondrial[51] genomes using bwa aln (0.7.5a)[52]; seeding was disabled (-l was set to 1,000) to improve the mapping sensitivity of aDNA reads[53]. Reads with a mapping quality below 30 or ambiguously mapping were discarded, PCR duplicates were removed using MarkDuplicates (http://picard.sourceforge.net), reads were realigned around indels (insertions or deletions) using GATK[54] and the MD-tag was recalculated using SAMtools 1.2 (ref. [55]). Finally, we excluded 5 bases from the 5′ and 3′ ends of each read from subsequent analyses to reduce the proportion of bases with aDNA damage. Genotype calling was performed in the resulting alignments using a combination of the HaplotypeCaller and UnifiedGenotyper algorithms from GATK[54] on sites with a minimum coverage of 10× as described in Supplementary Methods 12. To evaluate the genotyping pipeline, we generated sequencing data for two modern grape cultivars using the SNP capture protocol. These two varieties are present in the GrapeReSeq panel, and thus provide a direct comparison between our method and the GrapeReSeq microarray. We found a concordance of 99.4% and 99.3% between the called genotypes and their corresponding genotypes in the GrapeReSeq panel.

Ancient DNA authentication. The authenticity of the aDNA data was assessed on the basis of the length distribution and the nucleotide misincorporation patterns observed in the sequencing data. We used bamdamage[23] to estimate per base nucleotide substitutions in the mapped reads. Reads with a mapping quality lower than 30 and a base quality lower than 20 were discarded. Archaeological samples displayed increased C-to-T and G-to-A substitutions as well as short reads (Supplementary Figs. 3 and 4), consistent with aDNA data[22].

Reference data sets. We used two reference data sets to compare the archaeological grape seeds to present-day grape varieties (see Supplementary Methods 11 for a detailed description of the reference panels used). (1) The GrapeReSeq panel consists of 783 modern grape cultivars (*V. vinifera* subsp. *vinifera*) and 112 wild grape individuals (*V. vinifera* subsp. *sylvestris*) representative of the genetic diversity found in Europe (81 accessions), as well as from North Africa (18 accessions) and the Caucasus (13 accessions) genotyped for 10,000 diagnostic SNPs[21]. (2) We assembled a whole-genome reference panel incorporating sequencing data from 27 publicly available wild and domesticated grape accessions[24–26]. Raw reads were obtained from the NCBI Sequence Read Archive, mapped and processed using similar parameters as the archaeological samples. To avoid ambiguities due to synonymy, the VIVC number[56] is assigned to cultivars as indicated in Supplementary Table 4.

Genetic structure analyses. We explored the genetic relationships between the archaeological grape seeds and the samples in the two assembled reference panels using multidimensional scaling as implemented in bammds[23] (Fig. 1a,b and Supplementary Fig. 6). Samples with an on-target depth of coverage of ≥3× were included to the reference panel by sampling a random allele from the called genotypes, whereas the six low-coverage samples were incorporated from a majority count consensus sequence (Supplementary Table 1). After filtering low-quality SNPs, the final data sets consisted of 9,896 and 3,076,549 sites for the GrapeReSeq and whole-genome panels, respectively. Note that, for analyses using the GrapeReSeq panel, we did not exclude transition sites. However, data from the genotype calls and majority count consensus sequences obtained for the archaeological samples showed error rates comparable to those of modern grape

samples (Supplementary Fig. 5), suggesting that the aDNA-derived error is unlikely to have a substantial effect in the analyses.

We used the model-based clustering approaches implemented in fastNGSadmix[28] and ADMIXTURE[27] to estimate ancestry proportions in the archaeological samples (Fig. 1c and Supplementary Fig. 7). First, ADMIXTURE was run on the GrapeReSeq panel assuming 2–8 populations per clusters ($K=2$–8). We obtained 1,000 independent replicates for each value of K and kept the one with the best likelihood. Then, we estimated genotype likelihoods for the archaeological samples using the SAMtools model implemented in ANGSD v1.9 (ref. [57]) at the sites included in the GrapeReSeq panel. Finally, we obtained maximum likelihood estimates of the ancestry proportions for the archaeological samples using the genotype likelihoods and the ADMIXTURE-inferred allele frequencies for each value of K using fastNGSadmix. Figure 1c shows the results for the $K=8$, which resulted among the lowest cross-validation errors (Supplementary Fig. 7b).

Relatedness analyses. To explore the relationships among pairs of archaeological samples and between the archaeological samples and samples in the reference panels, we estimated kinship coefficients using two approaches: the called genotype-based approach implemented in KING[29] and the genotype likelihood-based approach implemented in NgsRelate[30] (Supplementary Tables 3–5).

KING was run assuming a non-homogeneous population structure for the two reference panels and using called genotypes for the archaeological samples. Pairs of samples were classified based on the kinship coefficients and the proportion of sites with IBS0, as suggested by Manichaikul et al.[29], in the following categories: identical clones ($K \geq 0.49$ and IBS0 ≤ 0.001), parent–offspring ($0.177 < K < 0.354$ and IBS0 ≤ 0.001), highly related/sibling ($0.177 < K < 0.354$ and IBS0 ≤ 0.25) or unrelated (Supplementary Tables 3–5). These values have been shown to be reliable in discerning known first-degree relationships among grape cultivars[21].

NgsRelate was used as a complementary method to validate the results obtained using KING and to include low-coverage samples for which it was not possible to call genotypes. To run NgsRelate, we first estimated genotype likelihoods for the archaeological samples using the SAMtools model (-gl 1) implemented in ANGSD v1.9 (ref. [57]). Reads with a mapping quality lower than 30 and bases with a quality lower than 20 were discarded. We then estimated allele frequencies for the two reference panels using PLINK 1.9 (ref. [58]). These frequencies, together with genotype likelihoods, were used to obtain kinship coefficients and the proportion of sites sharing 0, 1 or 2 alleles identical by descent between pairs of samples (Supplementary Table 3). These results were evaluated together with those results obtained from the genotype-based approach to assign relationships between pairs of archaeological samples.

In Supplementary Methods 16, we explore the possibility of paternal DNA present in the archaeological seeds through a simulation study and comparing the archaeological seeds data with that obtained from fresh seeds (Supplementary Figs. 8–11 and Supplementary Table 6). While most of the archaeological seeds were found to be consistent with data derived from a single individual, our analyses indicate that four seeds contain ≥10% of paternal DNA (Supplementary Fig. 11). Additionally, we evaluate the potential effects of paternal DNA in the relatedness analyses and found that: (1) clonal relationships can only be detected from true identical individuals even in the presence of paternal DNA, (2) parent–offspring relationships are only ambiguous when the sample contains >10% paternal DNA, and (3) apparent full-sibling relationships can result from multiple scenarios; thus, pairs of samples with this type of relationship were classified as 'highly related pairs' (Supplementary Fig. 12). Relationships between archaeological seeds and modern grapes were evaluated based on the conclusions in Supplementary Methods 16.

We further explored the effect of sequencing depth and panel ascertainment in the robustness of the relatedness inferences (Supplementary Methods 17). The results indicate that the metrics used to identify relationships between the samples are reliable for samples with an on-target depth of coverage of ≥2× when using genotypes, and 1× when using genotype likelihoods (Supplementary Table 7). Additionally, we confirmed that the samples identified as identical clones display an IBS distance of <0.0001 both in the sites overlapping the GrapeReSeq panel and in off-target sites (Supplementary Fig. 13 and Supplementary Methods 18).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data produced in this study are available at the NCBI Sequence Read Archive under the reference PRJNA489970. Genotype data are available in the figshare repository at: https://doi.org/10.6084/m9.figshare.7610987.

Received: 3 September 2018; Accepted: 30 April 2019;
Published online: 10 June 2019

References

1. Myles, S. et al. Genetic structure and domestication history of the grape. *Proc. Natl Acad. Sci. USA* **108**, 3530–3535 (2011).
2. Olmo, H. P. in *Evolution of Crop Plants* (eds Smartt, J. & Simmonds, N. W.) 485–490 (Longman Scientific & Technical, 1995).
3. Zohary, D., Hopf, M. & Weiss, E. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in South-west Asia, Europe, and the Mediterranean Basin* (Oxford Univ. Press, 2012).
4. McGovern, P. E. *Ancient Wine: the Search for the Origins of Viniculture* (Princeton Univ. Press, 2003).
5. McGovern, P. et al. Early Neolithic wine of Georgia in the South Caucasus. *Proc. Natl Acad. Sci. USA* **114**, E10309–E10318 (2017).
6. Goldschmidt, E. E. The evolution of fruit tree productivity: a review. *Econ. Bot.* **67**, 51–62 (2013).
7. Hartmann, H. T., Kester, D. E. & Davies, F. T. *Plant Propagation: Principles and Practices* (Prentice-Hall, 1997).
8. Janick, J. in *Plant Breeding Reviews* (ed. Janick, J.) 255–321 (John Wiley & Sons, 2010).
9. This, P., Lacombe, T. & Thomas, M. Historical origins and genetic diversity of wine grapes. *Trends Genet.* **22**, 511–519 (2006).
10. Bouby, L. et al. Bioarchaeological insights into the process of domestication of grapevine (*Vitis vinifera* L.) during Roman times in southern France. *PLoS ONE* **8**, e63195 (2013).
11. Renfrew, J. M. in *Wine: A Scientific Exploration* (eds Sandler, M. & Pinder, R.) 56–69 (CRC Press, 2003).
12. McGovern, P. E. et al. Beginning of viniculture in France. *Proc. Natl Acad. Sci. USA* **110**, 10147–10152 (2013).
13. Bostock, J. & Riley, H. T. *The Natural History of Pliny* (Taylor and Francis, 1855).
14. Royer, C. in *La Vigne et le Vin* (ed. La Manufacture et la Cite des Sciences et de l'industrie) 15–25 (Graficas, 1988).
15. Figueiral, I., Bouby, L., Buffat, L., Petitot, H. & Terral, J.-F. Archaeobotany, vine growing and wine producing in Roman southern France: the site of Gasquinoy (Béziers, Hérault). *J. Archaeol. Sci.* **37**, 139–149 (2010).
16. Terral, J.-F. et al. Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Ann. Bot.* **105**, 443–455 (2010).
17. Bacilieri, R. et al. Potential of combining morphometry and ancient DNA information to investigate grapevine domestication. *Veg. Hist. Archaeobot.* **26**, 345–356 (2016).
18. Cappellini, E. et al. A multidisciplinary study of archaeological grape seeds. *Naturwissenschaften* **97**, 205–217 (2010).
19. Manen, J.-F. et al. Microsatellites from archaeological *Vitis vinifera* seeds allow a tentative assignment of the geographical origin of ancient cultivars. *J. Archaeol. Sci.* **30**, 721–729 (2003).
20. Wales, N. et al. The limits and potential of paleogenomic techniques for reconstructing grapevine domestication. *J. Archaeol. Sci.* **72**, 57–70 (2016).
21. Laucou, V. et al. Extended diversity analysis of cultivated grapevine *Vitis vinifera* with 10K genome-wide SNPs. *PLoS ONE* **13**, e0192540 (2018).
22. Briggs, A. W. et al. Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
23. Malaspinas, A.-S. et al. bammds: a tool for assessing the ancestry of low-depth whole-genome data using multidimensional scaling (MDS). *Bioinformatics* **30**, 2962–2964 (2014).
24. Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D. & Gaut, B. S. Evolutionary genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl Acad. Sci. USA* **114**, 11715–11720 (2017).
25. Di Genova, A. et al. Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. *BMC Plant Biol.* **14**, 7 (2014).
26. Cardone, M. F. et al. Inter-varietal structural variation in grapevine genomes. *Plant J.* **88**, 648–661 (2016).
27. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
28. Jørsboe, E., Hanghøj, K. & Albrechtsen, A. fastNGSadmix: admixture proportions and principal component analysis of a single NGS sample. *Bioinformatics* **33**, 3148–3150 (2017).
29. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
30. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* **31**, 4009–4011 (2015).
31. Bleckmann, A., Alter, S. & Dresselhaus, T. The beginning of a seed: regulatory mechanisms of double fertilization. *Front. Plant Sci.* **5**, 452 (2014).
32. Ebadi, A., Sedgley, M., May, P. & Coombe, B. G. Seed development and abortion in *Vitis vinifera* L., cv. Chardonnay. *Int. J. Plant Sci.* **157**, 703–712 (1996).
33. Cadot, Y., Miñana-Castelló, M. T. & Chevalier, M. Anatomical, histological, and histochemical changes in grape seeds from *Vitis vinifera* L. cv Cabernet franc during fruit development. *J. Agric. Food Chem.* **54**, 9206–9215 (2006).

34. Boursiquot, J. in *Le Château-Chalon: un Vin, son Terroir et ses Hommes* (eds Berthet-Bondet J. & Roulière-Lambert M.-J.) 46–55 (Mêta Jura, 2013).
35. Lacombe, T. et al. Large-scale parentage analysis in an extended set of grapevine cultivars (*Vitis vinifera* L.). *Theor. Appl. Genet.* **126**, 401–414 (2013).
36. Regner, R., Stadlhuber, A. & Kaserer, H. Considerations about the evolution of grapevine and the role of Traminer. *Acta Hortic.* **528**, 179–184 (2000).
37. Bowers, J. E., Siret, R., Meredith, C. P., This, P. & Boursiquot, J.-M. A single pair of parents proposed for a group of grapevine varieties in northeastern France. *Acta Hortic.* **5281**, 129–132 (2000).
38. Galet, P. *Dictionnaire Encylcopédique des Cépages et de leurs Synonymes* (Libre et Solidaire, 2015).
39. Périsset, Z. *Histoire de la Vigne et du Vin en Valais: des Origines à nos Jours* (Infolio, 2010).
40. Robinson, J., Harding, J. & Vouillamoz, J. *Wine Grapes: A Complete Guide to 1,368 Vine Varieties, including their Origins and Flavours* (Ecco, 2012).
41. Mascher, M. et al. Genomic analysis of 6,000-year-old cultivated grain illuminates the domestication history of barley. *Nat. Genet.* **48**, 1089–1093 (2016).
42. Ramos-Madrigal, J. et al. Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr. Biol.* **26**, 3195–3201 (2016).
43. Vallebueno-Estrada, M. et al. The earliest maize from San Marcos Tehuacán is a partial domesticate with genomic evidence of inbreeding. *Proc. Natl Acad. Sci. USA* **113**, 14151–14156 (2016).
44. Malenica, N. et al. Whole genome amplification and microsatellite genotyping of herbarium DNA revealed the identity of an ancient grapevine cultivar. *Naturwissenschaften* **98**, 763–772 (2011).
45. Fuller, D. Q. Long and attenuated: comparative trends in the domestication of tree fruits. *Veg. Hist. Archaeobot.* **27**, 165–176 (2018).
46. Wales, N., Andersen, K., Cappellini, E., Ávila-Arcos, M. C. & Gilbert, M. T. P. Optimization of DNA recovery and amplification from non-carbonized archaeobotanical remains. *PLoS ONE* **9**, e86827 (2014).
47. Wales, N. et al. New insights on single-stranded versus double-stranded DNA library preparation for ancient DNA. *Biotechniques* **59**, 368–371 (2015).
48. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
49. Canaguier, A. et al. A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genom. Data* **14**, 56–62 (2017).
50. Jansen, R. K. et al. Phylogenetic analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. *BMC Evol. Biol.* **6**, 32 (2006).
51. Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* **26**, 99–110 (2008).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Schubert, M. et al. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
54. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
55. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
56. Vitis *International Variety Catalogue* (JKI, accessed 7 January 2019); www.vivc.de
57. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
58. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

Acknowledgements

We thank the Danish National High-throughput Sequencing Centre for assistance in generating the sequencing data. This project was funded by the Danish Council for Independent Research (10–081390) and the Danish National Research Foundation (DNRF94). L.B. and R.B. were supported by the French National Agency of Research (VINICULTURE project—ANR-16-CE27–0013). We thank the following scientific and technical directors and corresponding institutions for providing the archaeological material used in this project as well as contextual information: P. Blanchard (Inrap, site: La Madeleine), E. Verdel (Isère Patrimoine, site: Colletière), H. Pomarèdes (Inrap, sites: Mas de Vignoles XIV and La Lesse-Espagnac), O. Ginouvez (Inrap, site: Terrasses de Montfau), R. Bourgaut (Communauté d'Agglomération du Bassin de Thau, site: Roumèges), P. Flotte (Archéologie Alsace, site: Horbourg-Wihr), M. Compan (Inrap, site: Mont Ferrier), I. Daveau (Inrap, site: Cougourlude) and C. Tardy (Inrap). We are also grateful to the GrapeReSeq consortium for early access to the genotype data. Finally, we thank J. V. Moreno-Mayar, S. Gopalakrishnan, F. G. Vieira, D. Maghradze and A. Schlumbaum for their helpful discussion.

Author contributions

The project was conceived by N.W., M.T.P.G., R.B. and L.B., and headed by N.W. and M.T.P.G. J.A.S.C., A.K.W.R, R.B. and N.W. designed the experimental enrichment methodology with input from J.M.M.-Z., R.T. and A.-F.A.-B. A.K.W.R. processed the aDNA with input from N.W. J.R.-M., A.K.W.R., R.B. and N.W. designed the analysis strategy. J.R.-M. performed the bioinformatic analysis with assistance from B.P. and T.S.-P. and input from N.W., M.T.P.G. and R.B. J.R.-M., N.W., M.T.P.G., R.B., L.B., T.L. and P.T. interpreted the results. L.B., I.F., C.S. and C.H. curated the archaeological material. J.R.-M., N.W. and M.T.P.G. wrote the manuscript with input from R.B., L.B. and T.L. and the other authors. All authors revised, edited and accepted the manuscript. Primary funding was acquired by M.T.P.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41477-019-0437-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.T.P.G. or N.W.

Peer Review Information: *Nature Plants* thanks David Caramelli, Elizabeth Zimmer and other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Nathan Wales

Last updated by author(s): Jan 24, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We did not used specific software for the collection of the data used in this work.
Data analysis	The software used in this work is publicly available. We cite the corresponding publications in the main text and supplementary texts. The list of software used and their respective versions: CASAVA v1.8.2 AdapterRemoval v2.1.3 bwa v0.7.15 samtools v 1.2 picard-tools v1.130 GATK v3.4.0 bammds angsd v0.915 ADMIXTURE v1.3 FastNGSadmix KING v2.0 ngsRelate PLINK v1.90b3.42 R v3.4.1 perl5 v24

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data used in this study will be made available upon publication at the NCBI SRA under the reference PRJNA489970. Genotype data will be made available in the figshare repository under the following DOI: 10.6084/m9.figshare.7610987.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not rely on statistical methods to predetermine the sample sizes. The number of samples (archaeological seeds) used in the study was determined based on the preservation of the archaeological material and the amount of endogenous DNA.
Data exclusions	We did not exclude any of the 28 samples sequenced in this study. However for the clustering analysis we excluded admixed individuals from the reference panel as it is described in the supplementary information.
Replication	We did not attempted to replicate any of the experimental procedures.
Randomization	We did not performed any measurement on specific experimental groups or effect sizes, thus no randomization strategy was followed.
Blinding	We did not applied any blinding techniques, since the type of analysis performed in this work does not require it.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☐	☒ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Palaeontology

Specimen provenance	Archaeological seeds were obtained during excavations in France. The following scientific and technical directors and corresponding institutions provided permits for the archaeological material used in this project: P . Blanchard (Inrap, archaeological site: La Madeleine), E. V erdel (Isère Patrimoine, archaeological site: Colletière), H. Pomarèdes (Inrap, archaeological site: Mas de Vignoles XIV and La Lesse-Espagnac), O. Ginouvez (Inrap, archaeological site: Terrasses de Montfau), R. Bourgaut (Communauté d'Agglomération du Bassin de Thau, archaeological site: Roumèges), P. Flotte (Archéologie Alsace, archaeological site: Horbourg-Wihr), M. Compan (Inrap, archaeological site: Mont Ferrier), I. Daveau (Inrap, archaeological site: Cougourlude).
Specimen deposition	Samples were completely pulverized during the DNA extraction.
Dating methods	Archaeological seeds were dated based on their association with archaeological artifacts found in the same stratigraphic unit,

Dating methods	dendrochronology o radiocarbon dating.

☒ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	The study did not involve animals.
Wild animals	The study did not involve wild animals.
Field-collected samples	We use fresh plant tissue for DNA extraction and sequencing. Grapevine leaves, seeds and wood samples were obtained from the INRA (https://www6.montpellier.inra.fr/vassal) grapevine collections. We provide accession numbers for all the grape varieties used in the study in the corresponding supplementary tables.
Ethics oversight	The plant material used in this study did not require ethical approval.

Note that full information on the approval of the study protocol must also be provided in the manuscript.