

This is a repository copy of *Palaeogenomic insights into the origins of French grapevine diversity*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/147520/>

Version: Accepted Version

Article:

Ramos Madrigal, Jazmín, Runge, Anne Kathrine Wiborg, Bouby, Laurent et al. (14 more authors) (2019) Palaeogenomic insights into the origins of French grapevine diversity. Nature Plants. ISSN 2055-026X

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

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

1 **Palaeogenomic insights into the origins of French grapevine**

2 **diversity**

3 Jazmín Ramos-Madrigal¹, Anne Kathrine Wiborg Runge^{1,2}, Laurent Bouby³,
4 Thierry Lacombe⁴, José Alfredo Samaniego Castruita¹, Anne-Françoise Adam-
5 Blondon⁵, Isabel Figueiral⁶, Charlotte Hallavant⁷, José M. Martínez-Zapater⁸,
6 Caroline Schaal⁹, Reinhard Töpfer¹⁰, Bent Petersen^{1,11}, Thomas Sicheritz-
7 Pontén^{1,11}, Patrice This⁴, Roberto Bacilieri⁴, M. Thomas P. Gilbert^{1,12,*}, Nathan
8 Wales^{1,2,13,14*}

9
10 ¹ Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350
11 Copenhagen, Denmark. ² BioArCh, Department of Archaeology, University of York, Wentworth
12 Way, York YO10 5DD, UK. ³ ISEM - UMR 5554, CNRS-IRD-EPHE-Université Montpellier, Place
13 Eugène Bataillon, CC 065, 34095 Montpellier Cedex, France ⁴ UMR AGAP, Université Montpellier,
14 CIRAD, INRA, Montpellier SupAgro, 2 Place Viala, 34060 Montpellier, France. ⁵ URGI, Unité de
15 Recherche Génomique-Info, UR1164, INRA, Université Paris-Saclay, Route de Saint-Cyr 78026
16 Versailles, France. ⁶ Inrap, Méditerranée and ISEM - UMR 5554, CNRS-IRD-EPHE-Université
17 Montpellier, Place Eugène Bataillon, CC 065, 34095 Montpellier Cedex, France. ⁷ Bureau d'
18 études Hadès, laboratoire TRACES - UMR 5608 (pôle Terrae) - UT2J, 5 allées A. Machado, 31058
19 Toulouse Cedex 9, France. ⁸ Instituto de Ciencias de la Vid y del Vino (CSIC-UR-Gobierno de La
20 Rioja), Ctra. de Burgos km 6, 26007 Logroño Spain. ⁹ GéoArchEon Sarl, Laboratoire Chrono-
21 Environnement - UMR 6249, Université de Franche Comté, 16 route de Gray, 25000 Besançon,
22 France. ¹⁰ Julius Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen, Institut für
23 Rebenzüchtung, Geilweilerhof, D-76833 Siebeldingen, Germany. ¹¹ Centre of Excellence for
24 Omics-Driven Computational Biodiscovery, Faculty of Applied Sciences, AIMST University, Kedah,
25 Malaysia. ¹² NTNU University Museum, 7491 Trondheim, Norway. ¹³ Department of Plant and
26 Microbial Biology, University of California, 111 Koshland Hall, Berkeley, CA 94720, USA. ¹⁴
27 Laboratoire d'Anthropobiologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288,

28 Université Paul Sabatier, 31000 Toulouse, France. * e-mail: nathan.wales@york.ac.uk or
29 tgilbert@snm.ku.dk

30

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

47

48 Since its domestication in Southwestern Asia more than 6000 years ago ¹⁻³, the
49 Eurasian grapevine (*Vitis vinifera* L.) has become one of the world’s most widely
50 produced and economically valuable fruit crops. Although grapevine products
51 are widely consumed as table grapes, dried raisins, fruit preserves, and cooked
52 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, wherein favored varieties are maintained through horticultural techniques like 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 flavor 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 three to five 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 future development of breeding schemes (*e.g.* through better understanding 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 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 vineyards were first planted at the Greek colony of *Massalia*, present day Marseille, during the 6th century BCE ^{10,11}. Winemaking subsequently spread along the Mediterranean coast ¹², but it was not until 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 (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 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 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–225 CE) contained greater numbers of

103 morphologically wild seeds than the following period (225–600 CE), raising the
104 question of whether Romans collected and cultivated wild berries for
105 winemaking. Through this time series, seed shapes tended toward domesticated
106 morphotypes, a finding the authors hypothesize represents a combination of
107 continued selective pressures with a sporadic incorporation of native varieties
108 through sexual reproduction. While these interpretations are thought provoking,
109 the authors also recognize critiques that some wild and domesticated vines
110 produce morphologically indistinguishable seeds.

111

112 One of the most promising avenues of research for ancient viniculture is
113 palaeogenomic (or ancient DNA, aDNA) analysis of well-preserved
114 archaeological pips^{17–19}. For example, Wales *et al.*²⁰ demonstrated that many
115 waterlogged grape seeds contain high proportions of endogenous DNA that
116 could be interrogated with state-of-the-art, high-throughput aDNA sequencing.
117 With the establishment of genomic databases for hundreds of modern cultivars
118 and wild grapevines²¹, we sought to examine how DNA recovered from
119 archaeological samples could sidestep some of the challenges of conventional
120 archaeobotanical methods and reveal relationships between ancient samples
121 and modern varieties, thereby providing otherwise unachievable insights on past
122 implementation of vegetative propagation and the antiquity of some of the
123 world's most produced grapevine varieties.

124

125 **Results and discussion**

126

127 **Successful enrichment of SNP loci in archaeological pips**

128 We performed targeted enrichment and shotgun sequencing of 10,000 SNP loci
129 in 28 archaeological grape seeds. The pips were recently excavated from
130 waterlogged features (wells, latrines, ditches, and pits) at 9 French sites
131 (Supplementary Fig. 1), and based on archaeological context, date as early as the
132 Iron Age (510–475 BCE) and as late as the medieval period (1050–1200 CE)
133 (Table 1). SNP loci were selected from the GrapeReSeq panel, a DNA microarray
134 that was developed to authenticate varieties for breeding and germplasm
135 management ²¹. This reference panel provides data for 783 domesticated
136 varieties (*V. vinifera* subsp. *vinifera*), 112 wild (*V. vinifera* subsp. *sylvestris*)
137 accessions collected from Eurasia and North Africa, and 11 other *Vitis* species.
138 We obtained a 4- to 400-fold enrichment at the targeted SNP sites, leading to an
139 on-target depth of coverage of $1\text{--}25.7\times$ (Supplementary Table 1 and
140 Supplementary Fig. 2). Nucleotide misincorporation patterns observed in the
141 sequencing data and read length distributions were consistent with those
142 expected for degraded DNA ²² (Supplementary Figs. 3, 4 and 5a) .

143

144 **Archaeological seeds related to European winemaking lineages**

145 We employed multidimensional scaling (MDS) to investigate whether
146 archaeological samples were more closely related to wild accessions or
147 domesticated varieties. Samples were compared to the GrapeReSeq panel
148 following the random allele sampling strategy described in *bammds* ²³, to
149 account for varying depth of coverage in the archaeological samples.
150 Additionally, we expanded our reference dataset with publicly-available whole-
151 genome sequencing data from 27 wild and domesticated grape accessions ^{24–26}
152 (Supplementary Table 2). The MDS plots showed all 28 archaeological samples

153 fall within the variability of domesticated grapevines, suggesting none of the
154 seeds originated from truly wild vines (Fig. 1a). While it is plausible that samples
155 near the boundary of the domesticated and wild clusters could represent F₁
156 hybrids between domesticated varieties and wild grapevines (*e.g.* specimen R-
157 LLE_09), we find no evidence for large-scale collection of wild berries by Romans
158 or medieval people at these sites. Likewise, the oldest sample, from the Iron Age
159 site of La Cougourlude dating to 510–475 BCE, also falls within the MDS space
160 composed of cultivated grapevines. These findings support Bouby *et al.*'s¹⁰
161 hypothesis that even though many pips from Roman and medieval sites exhibit
162 wild morphotypes they in fact originate from domesticated varieties.

163

164 Once we determined that archaeological seeds likely originated from
165 domesticated grapevines, we repeated the MDS analysis without wild accessions
166 to achieve a more refined picture of the relationships to regional varieties and
167 types of berries (*i.e.*, predominantly used in winemaking or as table grapes). The
168 majority of the archaeological pips were most closely related to wine cultivars
169 from West and Central Europe (Fig. 1b), although the three Early Roman samples
170 from the Mas de Vignoles XIV site had a closer affinity to wine grapes from the
171 Balkans and the Iberian Peninsula. Overall, this analysis shows that the
172 archaeological seeds are predominantly related to Western European varieties
173 that are used for winemaking, and not grapevines that are today grown further
174 east for wine or table grapes. These data suggest that 2000 years ago cultivated
175 vines in the modern territory of France were distinct from their Near Eastern
176 ancestors and well on their way to founding the germplasm of modern varieties
177 used in Western European winemaking. We also verified that the patterns

178 observed in the MDS analysis using the GrapeReSeq panel were consistent with
179 those obtained from a whole-genome (WG) reference panel (Supplementary
180 Table 2 and Supplementary Fig. 6).

181

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

189

190 As there is evidence for gene flow with local wild grapevines in Western Europe
191 ¹, we explored the wild ancestry components identified through the clustering
192 analysis. Since these proportions are estimated on the GrapeReSeq SNPs they do
193 not necessarily represent whole-genome ancestry proportions. However, this
194 allowed us to: 1) compare the proportions between present-day varieties and
195 the archaeological seeds at these diagnostic sites, and 2) identify the potential
196 source of the wild grape ancestry in the archaeological seeds. Wild grapevines
197 carry four main ancestry components when assuming 8 clusters (Fig. 1c). While
198 American and Asian *Vitis* species (yellow) and Eurasian wild grapes from the
199 Caucasus and Turkey (light blue) separate into individual clusters that do not
200 contribute significant ancestry to any other group, wild grapes from the African
201 and Western European populations display two ancestry components (dark and
202 light green) that are found in some domesticated grapes. All archaeological

203 samples except for the most recent (M-LM_22) show evidence of genetic
204 contributions from wild grapevines (Fig. 1c and Supplementary Fig. 7), and these
205 wild ancestries are primarily associated with Western and Central European
206 vines. While these data provide the first clues on the timing of genetic
207 introgression from local vines into domesticated lineages, the amount of wild
208 ancestry does not follow a consistent pattern related to sample age. For example,
209 the oldest sample (La Cugourlud, 510–475 BCE) shows some of the highest
210 levels of wild ancestry (~45%), while other early samples from Mas de Vignoles
211 XIV (2nd–1st century BCE) have marginal amounts of wild ancestry (3.5–4.5%),
212 and five samples from La Lesse-Espagnac (175–225 CE) range from ~10–38%. In
213 fact, these proportions of wild ancestry are similar to those observed in modern
214 French varieties, suggesting that the admixture with wild grapevines took place
215 at the earliest stages of viniculture in France, and potentially before other
216 cultivated lineages were introduced to France (*i.e.*, from Greece or the Italian
217 Peninsula). Together, these results suggest that the local wild gene pool played
218 an early role for domesticated varieties, with the gene flow between wild
219 grapevines and domesticated cultivars occurring at least 2500 years ago.

220

221 **Ancient use of vegetative propagation**

222 The availability of genotype data for hundreds of cultivars in the GrapeReSeq
223 panel, allowed us to explore relationships between archaeological pips excavated
224 from individual sites and across different regions of France. We estimated
225 kinship coefficients among pairs of samples using *KING*²⁹ and *NgsRelate*³⁰. Pairs
226 of samples were classified based on the kinship coefficients and the proportion
227 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, since grape seeds that have been cross-
fertilized contain paternal derived DNA³¹ 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;
Supplementary Section 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 $\leq 10\%$ (Supplementary Figs. 8 to 11).
Moreover, we studied the dependence of the relatedness analyses on such
contribution and found that $\leq 10\%$ paternal DNA does not significantly affect the
results (Supplementary Fig. 12). Therefore, we consider that clonal and parent-
offspring relationships are not affected in most samples. On the other hand, full-
sibling relationships could derive from multiple scenarios if the samples
involved contain paternal DNA (Supplementary Fig. 12c), and 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
(*e.g.* in up to 30% of the cases for 'Chardonnay' variety ^{32,33}). Our results
suggest that the observed clonal clusters among archaeological samples

253 represent empty seeds with only maternal tissue, either produced by the same
254 plant, such as might occur at one archaeological site, or by one grapevine variety
255 spread through vegetative propagation (Fig. 2 and Supplementary Table 3). Five
256 of these clonal clusters consist of two or three seeds from a single stratigraphic
257 context: an Early Roman ditch at Mas de Vignoles XIV near Nîmes city (2nd–1st
258 century BCE), a Roman well at Mont Ferrier, Tourbes (1st century CE), a Roman
259 well at La Lesse-Espagnac (ca. 200 CE), a Roman well at Terrasses de Montfau,
260 Magalas (4th century CE), and an early medieval well at Mas de Vignoles XIV (ca.
261 800 CE). Given that bunches of grapes might have been pressed for juice and
262 discarded *en masse*, these genetically identical specimens may well represent
263 seeds from single plants. The other genetic cluster consists of three seeds from
264 Horbourg-Wihr in Alsace and one seed from La Lesse-Espagnac in
265 Mediterranean France (Fig 2b); while all four samples date to the 2nd century CE,
266 these genetic clones suggest that Romans transported grapevine across long
267 distances (>600 km), most likely as cuttings.

268 Five archaeological sites in Southern France demonstrated the presence of
269 multiple genotypes within a single temporal stratum, providing genetic evidence
270 that multiple lineages or varieties were maintained at individual vineyards. For
271 example, we identified six different genotypes at Mas de Vignoles XIV near
272 Nîmes, three of which shared first-degree relationships and three of which were
273 unrelated (Fig. 2b). Overall, these relationship data indicate that vegetative
274 propagation, long-distance transportation of varieties, and multivarietal
275 cultivation have been practiced in France since the Roman era, consistent with
276 historic accounts ⁴.

277

278 **The antiquity of modern French varieties**

279 We lastly investigated the relatedness between archaeological and modern
280 varieties, by computing kinship coefficients and the proportion of IBS0 sites
281 between pairs of archaeological samples and samples present in the GrapeReSeq
282 panel using *KING* ²⁹ (Fig. 3 and Supplementary Tables 4 and 5). Our results
283 confirm long-held beliefs that Roman and medieval viticulturists maintained
284 ancient lineages using vegetative propagation ¹³, and that modern French
285 viniculture is in large part a product of these traditions. One archaeological
286 sample from La Madeleine (Orléans), dating to 1050–1200 CE, was an identical
287 clone of ‘Savagnin Blanc’ (VIVC17636), a variety today cultivated for wine
288 production in Northeastern France and other countries from Central Europe
289 (*kinship coeff.* = 0.496; *IBS0* ~ 0.0001; Identity of 99.7% and 99.9% for the
290 GrapeReSeq and WG panels, respectively) (Supplementary Tables 4 and 5).
291 Several researchers previously identified ‘Savagnin Blanc’, also known as ‘
292 Traminer Weiss’, as a recurrent parent of many commercially important
293 European varieties ^{1,34,35}, and written accounts document the appellation as early
294 as 1539 CE ³⁶. Our findings extend the presence of this variety in France by
295 hundreds of years and furthermore, suggest that either ‘Savagnin Blanc’ or its
296 direct relatives have been cultivated in France since the 1st century CE, since
297 archaeological seeds from Mont Ferrier, Tourbes have a parent-offspring
298 relationship with ‘Savagnin Blanc’ (Figs. 2b and 3).

299

300 Several archaeological seeds were closely related to ‘Mondeuse Blanche’
301 (VIVC7919), a French variety characteristic of the Northern French Alps that has

302 been suggested to have acted as a key progenitor ^{35,37}. We found that four
303 genetically identical 2nd century CE seeds from Horbourg-Wihr and La Lesse-
304 Espagnac have a parent-offspring relationship with ‘Mondeuse Blanche’,
305 indicating that just one reproductive cycle has taken place in this lineage in the
306 past 1800 years (Fig. 3). This finding presents an exciting consilience of genetic
307 and archaeobotanical data; using morphometric analysis, Terral *et al.* ¹⁶ also
308 found evidence for ‘Mondeuse Blanche’ among 1st–2nd century CE pips from the
309 Rec-de-Ligno site, which lies less than 10 km from La Lesse-Espagnac. We also
310 observed that ‘Mondeuse Blanche’ is highly related (full-sibling or similar
311 relationship) to an archaeological seed from Colletiere, dating to circa 1000 CE,
312 close to the region where ‘Mondeuse Blanche’ is still grown today (Savoie, Ain)
313 (Fig. 3). Interestingly, the medieval seed is also highly related to ‘Tressot’
314 (12640) (cited since 1396 in France ³⁸) and ‘Servanin’ (VIVC11526), both French
315 varieties that are rarely cultivated today.

316

317 In addition to ‘Mondeuse Blanche’, four other Roman seeds from Southern
318 France provided parent-offspring relationships to modern Alpine varieties: three
319 1st century CE seeds from Mont Ferrier are highly related to ‘Arvine’
320 (VIVC664) and ‘Amigne’ (VIVC425) and one 1st–3rd century CE seed from
321 Roumeges is a first-degree relative to ‘Humagne Blanc’ (VIVC5450) (Fig. 3).
322 All three are Swiss varieties used for white wine, and the former two are
323 recorded in Switzerland by the 17th century CE ³⁹. Tradition holds that the
324 Romans brought ‘Amigne’ to Switzerland as a variety they referred to as
325 ‘*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 there indeed is a close genetic link between the varieties
grown by the Romans and some modern Swiss cultivars, including ‘Amigne’.
Moreover, these data suggest that modern Alpine varieties may have been
cultivated in a more widespread geographic 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: 2nd –1st 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, while 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

351 during this period, as well as replacement and incorporation of new varieties
352 through time.

353

354 **Concluding remarks**

355

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

371

372 **Materials and methods**

373

374 **Archaeological sample description**

375 Grape seeds were collected from nine archaeological sites in France during
376 excavations of wells, latrines, pits, and ditches (Supplementary Fig. 1; see
377 Supplementary Section 1 for a description of the archaeological sites). Sediment
378 samples were systematically collected and immediately isolated to prevent
379 contamination and stored in cool conditions (4° C). The sediment samples were
380 processed at the Institut des Sciences de l' Evolution (ISEM) in Montpellier,
381 France. To prevent contamination with modern material, seeds were isolated in
382 a clean room separate from the archaeobotanical laboratory. Additionally,
383 surfaces and tools were cleaned with bleach prior to handling. Most of the
384 samples included in this study were photographed inside the clean room, with
385 specific equipment in order to carry out morphological analyses. Archaeological
386 samples were dated either by association with archaeological artifacts found in
387 the same stratigraphic units, dendrochronology, or radiocarbon dating. The age
388 of the samples ranged from the Iron Age (510–475 BCE) to the medieval period
389 (1050–1200 CE) (Supplementary Fig. 1 and Table 1).

390

391 **Archaeological samples processing**

392 Archaeological samples were processed in dedicated aDNA facilities at the
393 University of Copenhagen following standard measures to prevent
394 contamination. Individual seeds were decontaminated with 10% bleach, rinsed
395 with molecular biology grade water, and pulverized. DNA was extracted from the
396 resulting powder following standard protocols standardized for
397 archaeobotanical remains ⁴⁶. DNA extracts were converted into double-stranded
398 Illumina libraries using the NEBnext DNA Library Prep Mast Mix Set 2 (E6070L,
399 New England BioLabs) with modifications described in Wales *et al.* ⁴⁷ (see

Supplementary Section 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 Section 5). This panel covers genomic sites known to be informative for identification of grape cultivars. Libraries were captured following the MYbaits protocol as described in Supplementary Section 6. Finally, pre- 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.

408

409 **Sequencing data processing**

AdapterRemoval2.0 ⁴⁸ was used to remove Illumina adapter sequences, low quality stretches and ambiguous bases from the read ends. Resulting reads ≥ 30 base pairs were mapped to the grape nuclear reference genome 12X.2 ⁴⁹, chloroplast ⁵⁰ and mitochondrial ⁵¹ genomes using *bwa aln* (0.7.5a) ⁵²; seeding was disabled (-l was set to 1000) to improve the mapping sensitivity of aDNA reads ⁵³. Reads with mapping quality below 30 or ambiguously mapping were discarded, PCR duplicates were removed using *MarkDuplicates* (<http://picard.sourceforge.net>), reads were realigned around indels using *GATK* ⁵⁴ and the MD-tag was recalculated using *samtools* 1.2 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 Section 12. To evaluate the genotyping pipeline, we generated sequencing data for two modern grape cultivars using the

425 SNP capture protocol. These two varieties are present in the GrapeReSeq panel,
426 thus provide a direct comparison between our method and the GrapeReSeq
427 microarray. We found a concordance of 99.4% and 99.3%, between the called
428 genotypes and their corresponding genotypes in the GrapeReSeq panel.

429

430 **Ancient DNA authentication**

431 The authenticity of the aDNA data was assessed on the basis of the length
432 distribution and the nucleotide misincorporation patterns observed in the
433 sequencing data. We used *bamdamage*²³ to estimate per base nucleotide
434 substitutions in the mapped reads. Reads with mapping quality lower than 30
435 and base quality lower than 20 were discarded. Archaeological samples
436 displayed increased C-to-T and G-to-A substitutions as well as short reads
437 (Supplementary Figs. 3 and 4), consistent with aDNA data²².

438

439 **Reference datasets**

440 We used two reference datasets to compare the archaeological grape seeds to
441 present-day grape varieties (see Supplementary Section 11 for a detailed
442 description of the reference panels used). 1) The GrapeReSeq panel consists of
443 783 modern grape cultivars (*V. vinifera* subsp. *vinifera*) and 112 wild grape
444 individuals (*V. vinifera* subsp. *sylvestris*) representative of the genetic diversity
445 found in Europe (81 accessions), as well as from North Africa (18 accessions)
446 and the Caucasus (13 accessions) genotyped for 10,000 diagnostic SNPs²¹. 2) We
447 assembled a whole-genome (WG) reference panel incorporating sequencing data
448 from 27 publicly available wild and domesticated grape accessions^{24–26}. Raw
449 reads were obtained from the NCBI SRA, 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.

453

454 **Genetic structure analyses**

455 We explored the genetic relationships between the archaeological grape seeds
456 and the samples in the two assembled reference panels using multidimensional
457 scaling as implemented in *bammds* ²³ (Figs. 1a and 1b, and Supplementary Fig.
458 6). Samples with an on-target depth of coverage $\geq 3\times$ were included to the
459 reference panel by sampling a random allele from the called genotypes; while the
460 six low coverage samples were incorporated from a majority count consensus
461 sequence (Supplementary Table 1). After filtering low-quality SNPs, the final
462 datasets consisted of 9,896 and 3,076,549 sites for the GrapeReSeq and WG
463 panels, respectively. Note that, for analyses using the GrapeReSeq panel we did
464 not exclude transition sites. However, data from the genotype calls and majority
465 count consensus sequences obtained for the archaeological samples showed
466 error rates comparable to those of modern grape samples (Supplementary Fig.
467 5), suggesting that the aDNA derived error is unlikely to have a substantial effect
468 in the analyses.

469 We used the model-based clustering approaches implemented in *fastNGSadmix*
470 ²⁸ and *ADMIXTURE* ²⁷ to estimate ancestry proportions in the archaeological
471 samples (Fig. 1c and Supplementary Fig. 7). First, ADMIXTURE was run on the
472 GrapeReSeq panel assuming 2-8 populations/clusters ($K=2-8$). We obtained
473 1,000 independent replicates for each value of K and kept the one with the best
474 likelihood. Then, we estimated genotype likelihoods for the archaeological

475 samples using the *samtools* model implemented in *ANGSD* v1.9 ⁵⁷ at the sites
476 included in the GrapeReSeq panel. Finally, we obtained maximum likelihood
477 estimates of the ancestry proportions for the archaeological samples using the
478 genotype likelihoods and the *ADMIXTURE*-inferred allele frequencies for each
479 value of *K* using *fastNGSadmix*. Figure 1c shows the results for the *K*=8, which
480 resulted among the lowest cross-validation errors (Supplementary Fig. 7b).

481

482 **Relatedness analyses**

483 To explore the relationships among pairs of archaeological samples and between
484 the archaeological samples and samples in the reference panels, we estimated
485 kinship coefficients using two approaches: the called genotype-based approach
486 implemented in *KING* ²⁹ and the genotype likelihood-based approach
487 implemented in *NgsRelate* ³⁰ (Supplementary Tables 3 to 5).

488 *KING* was run assuming non-homogeneous population structure for the two
489 reference panels and using called genotypes for the archaeological samples. Pairs
490 of samples were classified based on the kinship coefficients and the proportion
491 of sites with ‘zero alleles Identical by State’ (IBS0), as suggested in
492 Manichaikul *et al.* ²⁹, in the following categories: identical clones ($K \geq 0.49$ and
493 $IBS0 \leq 0.001$), parent-offspring ($0.177 < K < 0.354$ and $IBS0 \leq 0.001$), highly
494 related/sibling ($0.177 < K < 0.354$ and $IBS0 \leq 0.25$) or unrelated (Supplementary
495 Tables 3 to 5). These values have been shown to be reliable in discerning known
496 first-degree relationships among grape cultivars ²¹.

497 *NgsRelate* was used as a complementary method to validate the results obtained
498 using *KING* and to include low coverage samples for which it was not possible to
499 call genotypes. To run *NgsRelate*, we first estimated genotype likelihoods for the

archaeological samples using the *samtools* model (-gl 1) implemented in *ANGSD*
v1.9⁵⁷. Reads with mapping quality lower than 30 and bases with quality lower
than 20 were discarded. We then estimated allele frequencies for the two
reference panels using *PLINK 1.9*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 (IBD) between pairs of samples
(Supplementary Table 3). These results were evaluated together with the
obtained from the genotype-based approach to assign relationships between
pairs of archaeological samples.

In Supplementary Section 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 to 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 four seeds contain $\geq 10\%$ of paternal DNA (Supplementary Fig.
11). Additionally, we evaluate 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 from Supplementary Section 16.

524 We further explored the effect of sequencing depth and panel ascertainment in
 525 the robustness of the relatedness inferences (Supplementary Section 17). The
 526 results indicate that the metrics used to identify relationships between the
 527 samples are reliable for samples with an on-target depth of coverage of $\geq 2 \times$
 528 when using genotypes, and $1 \times$ when using genotype likelihoods
 529 (Supplementary Table 7). Additionally, we confirmed that samples identified as
 530 identical clones display an IBS distance < 0.0001 both in the sites overlapping the
 531 GrapeReSeq panel and in off-target sites (Supplementary Fig. 13; Supplementary
 532 Section 18).

533

534 **References**

- 535 1. Myles, S. *et al.* Genetic structure and domestication history of the grape. *Proc.*
 536 *Natl. Acad. Sci.* **108**, 3530–3535 (2011).
- 537 2. Olmo, H. P. Grapes: *Vitis*, Muscadinia (Vitaceae). in *In Evolution of Crop Plants.*
 538 *J. Smartt and N. W. Simmonds, eds.* 485–490 (Longman Scientific & Technical,
 539 1995).
- 540 3. Zohary, D., Maria Hopf & Ehud Weiss. *Domestication of Plants in the Old*
 541 *World: The origin and spread of domesticated plants in south-west Asia,*
 542 *Europe, and the Mediterranean Basin.* (Oxford University Press, 2012).
- 543 4. McGovern, P. E. *Ancient wine: the search for the origins of viniculture.*
 544 (Princeton University Press, 2003).
- 545 5. McGovern, P. *et al.* Early Neolithic wine of Georgia in the South Caucasus.
 546 *Proc. Natl. Acad. Sci.* **114**, E10309–E10318 (2017).

- 547 6. Goldschmidt, E. E. The Evolution of Fruit Tree Productivity: A Review. *Econ.*
548 *Bot.* **67**, 51–62 (2013).
- 549 7. Hartmann, H. T., Kester, D. E. & Davies, F. T. *Plant propagation: principles and*
550 *practices*. (Prentice-Hall, Upper Saddle River, NJ, 1997).
- 551 8. Janick, J. The Origins of Fruits, Fruit Growing, and Fruit Breeding. in *Plant*
552 *Breeding Reviews* (ed. Janick, J.) 255–321 (John Wiley & Sons, Inc., 2010).
553 doi:10.1002/9780470650301.ch8
- 554 9. This, P., Lacombe, T. & Thomas, M. Historical origins and genetic diversity of
555 wine grapes. *Trends Genet.* **22**, 511–519 (2006).
- 556 10. Bouby, L. *et al.* Bioarchaeological Insights into the Process of Domestication
557 of Grapevine (*Vitis vinifera* L.) during Roman Times in Southern France. *PLoS*
558 *ONE* **8**, e63195 (2013).
- 559 11. J. M. Renfrew. Archaeology and the origins of wine production. in *Wine: A*
560 *Scientific Exploration*. Sandler, M. (Ed.), Pinder, R. (Ed.). (2003). (CRC Press,
561 2013).
- 562 12. McGovern, P. E. *et al.* Beginning of viniculture in France. *Proc. Natl. Acad. Sci.*
563 **110**, 10147–10152 (2013).
- 564 13. Bostock, J. & Riley, H. T. The Natural History of Pliny. in (Taylor and Francis,
565 1855).
- 566 14. Royer, C. Mouvement historiques de la vigne dans le monde. in *La Vigne et le*
567 *Vin (La Manufacture et la Cité des sciences et de l'industrie, eds)* 15–25
568 (Graficas, 1988).
- 569 15. Figueiral, I., Bouby, L., Buffat, L., Petitot, H. & Terral, J.-F. Archaeobotany, vine
570 growing and wine producing in Roman Southern France: the site of
571 Gasquinoy (Béziers, Hérault). *J. Archaeol. Sci.* **37**, 139–149 (2010).

- 572 16. Terral, J.-F. *et al.* Evolution and history of grapevine (*Vitis vinifera*) under
573 domestication: new morphometric perspectives to understand seed
574 domestication syndrome and reveal origins of ancient European cultivars.
575 *Ann. Bot.* **105**, 443–455 (2010).
- 576 17. Bacilieri, R. *et al.* Potential of combining morphometry and ancient DNA
577 information to investigate grapevine domestication. *Veg. Hist. Archaeobotany*
578 (2016). doi:10.1007/s00334-016-0597-4
- 579 18. Cappellini, E. *et al.* A multidisciplinary study of archaeological grape seeds.
580 *Naturwissenschaften* **97**, 205–217 (2010).
- 581 19. Manen, J.-F. *et al.* Microsatellites from archaeological *Vitis vinifera* seeds
582 allow a tentative assignment of the geographical origin of ancient cultivars. *J.*
583 *Archaeol. Sci.* **30**, 721–729 (2003).
- 584 20. Wales, N. *et al.* The limits and potential of paleogenomic techniques for
585 reconstructing grapevine domestication. *J. Archaeol. Sci.* **72**, 57–70 (2016).
- 586 21. Laucou, V. *et al.* Extended diversity analysis of cultivated grapevine *Vitis*
587 *vinifera* with 10K genome-wide SNPs. *PloS ONE* **13**, e0192540 (2018).
- 588 22. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a
589 Neandertal. *Proc. Natl. Acad. Sci.* **104**, 14616–14621 (2007).
- 590 23. Malaspinas, A.-S. *et al.* bammds: a tool for assessing the ancestry of low-depth
591 whole-genome data using multidimensional scaling (MDS). *Bioinformatics*
592 **30**, 2962–2964 (2014).
- 593 24. Zhou, Y., Massonnet, M., Sanjak, J. S., Cantu, D. & Gaut, B. S. Evolutionary
594 genomics of grape (*Vitis vinifera* ssp. *vinifera*) domestication. *Proc. Natl. Acad.*
595 *Sci.* 201709257 (2017). doi:10.1073/pnas.1709257114

- 596 25. Di Genova, A. *et al.* Whole genome comparison between table and wine
597 grapes reveals a comprehensive catalog of structural variants. *BMC Plant*
598 *Biol.* **14**, 7 (2014).
- 599 26. Cardone, M. F. *et al.* Inter-varietal structural variation in grapevine genomes.
600 *Plant J.* **88**, 648–661 (2016).
- 601 27. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of
602 ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 603 28. Jørsboe, E., Hanghøj, K. & Albrechtsen, A. fastNGSadmix: admixture
604 proportions and principal component analysis of a single NGS sample.
605 *Bioinformatics* btx474 (2017).
- 606 29. Manichaikul, A. *et al.* Robust relationship inference in genome-wide
607 association studies. *Bioinformatics* **26**, 2867–2873 (2010).
- 608 30. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating
609 pairwise relatedness from next-generation sequencing data. *Bioinformatics*
610 btv509 (2015). doi:10.1093/bioinformatics/btv509
- 611 31. Bleckmann, A., Alter, S. & Dresselhaus, T. The beginning of a seed: regulatory
612 mechanisms of double fertilization. *Front. Plant Sci.* **5**, (2014).
- 613 32. Ebadi, A., Sedgley, M., May, P. & Coombe, B. G. Seed Development and
614 Abortion in *Vitis vinifera* L., cv. Chardonnay. *Int. J. Plant Sci.* **157**, 703–712
615 (1996).
- 616 33. Cadot, Y., Miñana-Castelló, M. T. & Chevalier, M. Anatomical, Histological, and
617 Histochemical Changes in Grape Seeds from *Vitis vinifera* L. cv Cabernet franc
618 during Fruit Development. *J. Agric. Food Chem.* **54**, 9206–9215 (2006).

- 619 34. Boursiquot, J. Le Savagnin blanc. in *Berthet-Bondet J, Roulière-Lambert M-J*
620 *(eds) Le Château-Chalon: un vin, son terroir et ses hommes. Méta Jura, Lons-le-*
621 *Saunier, France, pp 46-55 (2013).*
- 622 35. Lacombe, T. *et al.* Large-scale parentage analysis in an extended set of
623 grapevine cultivars (*Vitis vinifera* L.). *Theor. Appl. Genet.* **126**, 401–414
624 (2013).
- 625 36. R. Regner, A. Stadlhuber & H. Kaserer. Considerations about the evolution of
626 grapevine and the role of Traminer. *Acta Hortic.* 179–184 (2000).
- 627 37. Bowers, J. E., Siret, R., Meredith, C. P., This, P. & Boursiquot, J.-M. A single pair
628 of parents proposed for a group of grapevine varieties in northeastern
629 France. *Acta Hortic.* 129–132 (2000). doi:10.17660/ActaHortic.2000.528.15
- 630 38. Galet, P. *Dictionnaire encyclopédique des cépages et de leurs synonymes.* (Libre
631 et Solidaire, 2015).
- 632 39. Périsset, Z. *Histoire de la vigne et du vin en Valais: des origines à nos jours.*
633 (Infolio, 2010).
- 634 40. Robinson, J., Harding, J. & Vouillamoz, J. *Wine Grapes: A Complete Guide to*
635 *1,368 Vine Varieties, including their Origins and Flavours.* (Ecco (Harper
636 Collins), 2012).
- 637 41. Mascher, M. *et al.* Genomic analysis of 6,000-year-old cultivated grain
638 illuminates the domestication history of barley. *Nat. Genet.* (2016).
639 doi:10.1038/ng.3611
- 640 42. Ramos-Madrigal, J. *et al.* Genome Sequence of a 5,310-Year-Old Maize Cob
641 Provides Insights into the Early Stages of Maize Domestication. *Curr. Biol.* **26**,
642 3195–3201 (2016).

- 643 43. Vallebuena-Estrada, M. *et al.* The earliest maize from San Marcos Tehuacán is
644 a partial domesticate with genomic evidence of inbreeding. *Proc. Natl. Acad.*
645 *Sci.* **113**, 14151–14156 (2016).
- 646 44. Malenica, N. *et al.* Whole genome amplification and microsatellite genotyping
647 of herbarium DNA revealed the identity of an ancient grapevine cultivar.
648 *Naturwissenschaften* **98**, 763–772 (2011).
- 649 45. Fuller, D. Q. Long and attenuated: comparative trends in the domestication of
650 tree fruits. *Veg. Hist. Archaeobotany* (2017). doi:10.1007/s00334-017-0659-2
- 651 46. Wales, N., Andersen, K., Cappellini, E., Ávila-Arcos, M. C. & Gilbert, M. T. P.
652 Optimization of DNA Recovery and Amplification from Non-Carbonized
653 Archaeobotanical Remains. *PLoS ONE* **9**, e86827 (2014).
- 654 47. Wales, N. *et al.* New insights on single-stranded versus double-stranded DNA
655 library preparation for ancient DNA. *BioTechniques* **59**, (2015).
- 656 48. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter
657 trimming, identification, and read merging. *BMC Res. Notes* **9**,
658 10.1186/s13104-016-1900-2 (2016).
- 659 49. Canaguier, A. *et al.* A new version of the grapevine reference genome
660 assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* **14**, 56–62
661 (2017).
- 662 50. Jansen, R. K. *et al.* Phylogenetic analyses of *Vitis* (Vitaceae) based on complete
663 chloroplast genome sequences: effects of taxon sampling and phylogenetic
664 methods on resolving relationships among rosids. *BMC Evol. Biol.* **6**, 32
665 (2006).

- 666 51. Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of
667 *Vitis vinifera* and the Issue of Rampant Horizontal Gene Transfer. *Mol. Biol.*
668 *Evol.* **26**, 99–110 (2008).
- 669 52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-
670 Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- 671 53. Schubert, M. *et al.* Improving ancient DNA read mapping against modern
672 reference genomes. *BMC Genomics* **13**, 178 (2012).
- 673 54. DePristo, M. A. *et al.* A framework for variation discovery and genotyping
674 using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
- 675 55. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
676 *Bioinformatics* **25**, 2078–2079 (2009).
- 677 56. Maul. *et al* *Vitis* International Variety Catalogue - www.vivc.de. (Accessed
678 January 2019).
- 679 57. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
680 Generation Sequencing Data. *BMC Bioinformatics* **15**, 10.1186/s12859-014-
681 0356-4 (2014).
- 682 58. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger
683 and richer datasets. *GigaScience* **4**, (2015).

684

685 **Acknowledgments**

686 We thank the Danish National High-throughput Sequencing Centre for assistance
687 in generating the sequencing data. This project was funded by the Danish Council
688 for Independent Research (10-081390) and the Danish National Research
689 Foundation (DNRF94). L.B. and R.B. were supported by the French National
690 Agency of Research (VINICULTURE project - ANR-16-CE27-0013). We would like

691 to thank the following scientific and technical directors and corresponding
692 institutions for providing the archaeological material used in this project as well
693 as contextual information: P. Blanchard (Inrap, site: La Madeleine), E. Verdel
694 (Isère Patrimoine, site: Colletière), H. Pomarèdes (Inrap, sites: Mas de Vignoles
695 XIV and La Lesse-Espagnac), O. Ginouvez (Inrap, site: Terrasses de Montfau), R.
696 Bourgaut (Communauté d'Agglomération du Bassin de Thau, site: Roumèges), P.
697 Flotte (Archéologie Alsace, site: Horbourg-Wihr), M. Compan (Inrap, site : Mont
698 Ferrier), I. Daveau (Inrap, site: Cougourlude) and C. Tardy (Inrap). We are also
699 grateful to the GrapeReSeq consortium for early access to the genotype data.
700 Finally, we would like to thank J. Víctor Moreno-Mayar, Shyam Gopalakrishnan,
701 Filipe G. Vieira, David Maghradze, and Angela Schlumbaum for their helpful
702 discussion.

703

704 **Corresponding authors**

705 Correspondence to Nathan Wales and M. Thomas P. Gilbert.

706

707 **Author contributions**

708 The project was conceived by N.W., M.T.P.G., R.B., and L.B., and headed by N.W.
709 and M.T.P.G.; J.A.S.C., A.K.W.R, R.B. and N.W. designed experimental enrichment
710 methodology with input from J.M.M.Z., R.T. and A.F.A.B.; A.K.W.R. processed
711 ancient DNA with input from N.W.; J.R.M., A.K.W.R., R.B. and N.W. designed
712 analysis strategy; J.R.M. performed bioinformatic analysis with assistance from
713 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.,
714 L.B., T.L. and P.T. interpreted the results; L.B., I.F., C.S. and C.H., curated

715 archaeological material; J.R.M., N.W. and M.T.P.G. wrote the manuscript with
716 input from R.B., L.B. and T.L. and the other authors. All authors revised, edited
717 and accepted the manuscript. Primary funding acquired by M.T.P.G.

718

719 **Competing interest**

720 The authors declare no competing interests.

721

722 **Data availability**

723 Sequencing data produced in this study are available at the NCBI SRA under the
724 reference PRJNA489970. Genotype data are available in the *figshare* repository
725 under the following DOI: 10.6084/m9.figshare.7610987.

726

727 **Figure legends**

728 **Figure 1. Genetic affinities between archaeological grape seeds and**
729 **modern *Vitis vinifera* accessions.** a. Multidimensional scaling plot (MDS)
730 including archaeological samples, wild *V. vinifera subsp. sylvestris* accessions, and
731 domesticated varieties. b. MDS plot restricted to archaeological samples and
732 domesticated varieties. Colors correspond to the main ancestry clusters
733 identified in Laucou *et al.* ²¹. *Archaeological samples that were incorporated to
734 the dataset by sampling a random allele from a majority count consensus
735 sequence instead of called genotypes c. Model-based clustering analysis of the
736 GrapeReSeq panel assuming K=8 clusters. Vertical bars represent individual
737 accessions, colors represent the inferred ancestry components, and the fraction
738 of each color corresponds to the estimated ancestry proportion. Archaeological

739 samples are sorted by age, and by sample identification within a stratigraphic
740 context. Samples that were identified as identical clones are grouped with black
741 lines and capital letters (A-F) at the bottom.

742 **Figure 2. Geographic distribution and relationships between the distinct**
743 **genetic types of archaeological samples.** a. Relatedness among pairs of
744 archaeological samples. Kinship coefficients were estimated using *NgsRelate*
745 between pairs of samples for SNP loci present in the GrapeReSeq panel. Capital
746 letters (A-F) on the left indicate genetically identical clones, *i.e.*, putative ancient
747 and historical varieties. *Archaeological seeds that were found consistent with
748 carrying >10% paternal DNA. b. Map displaying the distribution of genetic types
749 (circles) in each archaeological site. Capital letters (A-F) on the circles indicate
750 clusters of genetically identical seeds represented by more than one seed.
751 Shading of the circles indicates sample age. In red is shown the genetic type that
752 was found in more than one archaeological site. Lines connect pairs of samples
753 that are related as parent-offspring (solid lines) or highly-related/full-sibling
754 (dotted lines). Note that, since in the presence of paternal DNA full-sibling
755 relationships could derive from multiple scenarios, we classified samples
756 consistent with full-sibling relationships as ‘highly-related’ (see Supplementary
757 Section 16).

758 **Figure 3. Genetic origins of ancient and historic French grapevine varieties.**
759 Relationships identified between archaeological samples and modern cultivars
760 included in the GrapeReSeq panel. Solid lines represent parent-offspring
761 relationships and dotted lines represent pairs of highly related (full-sibling or
762 similar) samples. Sibling relationships involving pairs of modern cultivar are not
763 displayed for simplicity. *Archaeological seeds that were found consistent with

764 carrying >10% paternal DNA. The VIVC (<http://www.vivc.de>) and GrapeReSeq

765 identifiers for the modern cultivars can be found in Supplementary Table 4.

766

767

768 **Table 1.** Description of the archaeological grape seeds used in the study.

769

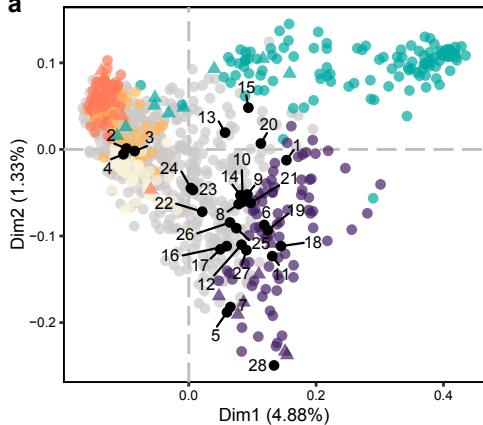
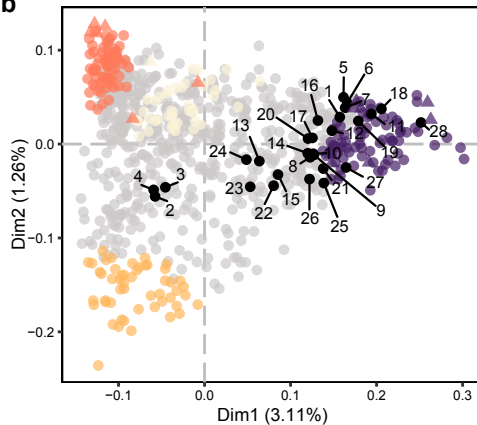
#	Sample ID	Geographic coordinates	Archaeological site	Stratigraphic unit	Structure	Age	Dating method	Period	GC †
1	IA-LC_01	43.573639, 3.914750	La Cougourlude, Lattes	US 31084	Ditch FO 30277	510-475 BCE/2480 ± 30 BP (769- 417 cal BCE)	Archaeological artifacts/C14	Iron Age	
2	R-MDV14_04	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	2nd-1st c BCE	Archaeological artifacts	Early Roman	A
3	R-MDV14_07	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	2nd-1st c BCE	Archaeological artifacts	Early Roman	A
4	R-MDV14_09	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 14152	Rural ditch FO 14194	2nd-1st c BCE	Archaeological artifacts	Early Roman	A
5	R-MF_21	43.432556, 3.394222	Mont Ferrier, Tourbes	US 2076	Well PT 2052	1st c CE	Archaeological artifacts	Roman	B
6	R-MF_23	43.432556, 3.394222	Mont Ferrier, Tourbes	US 2076	Well PT 2052	1st c CE	Archaeological artifacts	Roman	
7	R-MF_25	43.432556, 3.394222	Mont Ferrier, Tourbes	US 2076	Well PT 2052	1st c CE	Archaeological artifacts	Roman	B
8	R-HW70_18	48.080500, 7.399194	Horbours-Wihr	N.D.	Pit ST7054	2nd c CE	Dendrochronology / Archaeological artifacts	Roman	C
9	R-HW71_03	48.080500, 7.399194	Horbours-Wihr	N.D.	Pit ST7172	2nd c CE	Archaeological artifacts	Roman	C
10	R-HW71_17	48.080500, 7.399194	Horbours-Wihr	N.D.	Pit ST7172	2nd c CE	Archaeological artifacts	Roman	C
11	R-R_09	43.471306, 3.670139	Roumeges, Poussan	US 5007(12/13)	Well PT 5001	1st-3rd c CE	Archaeological artifacts	Roman	
12	R-R_14	43.471306, 3.670139	Roumeges, Poussan	US 5007(12/13)	Well PT 5001	1st-3rd c CE	Archaeological artifacts	Roman	
13	R-LLE_02	43.300806, 3.239917	La Lesse- Espagnac, Sauvian	US 3019	Well PT 3005	175-225 CE	Archaeological artifacts	Roman	
14	R-LLE_08	43.300806, 3.239917	La Lesse- Espagnac, Sauvian	US 3019	Well PT 3005	175-225 CE	Archaeological artifacts	Roman	C
15	R-LLE_09	43.300806, 3.239917	La Lesse- Espagnac, Sauvian	US 3019	Well PT 3005	175-225 CE	Archaeological artifacts	Roman	
16	R-LLE_13	43.300806, 3.239917	La Lesse- Espagnac, Sauvian	US 3019	Well PT 3005	175-225 CE	Archaeological artifacts	Roman	D
17	R-LLE_14	43.300806, 3.239917	La Lesse- Espagnac, Sauvian	US 3019	Well PT 3005	175-225 CE	Archaeological artifacts	Roman	D
18	R-TDM_06	43.472806, 3.223000	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	4th c CE	Archaeological artifacts	Roman	E
19	R-TDM_08	43.472806, 3.223000	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	4th c CE	Archaeological artifacts	Roman	E
20	R-TDM_10	43.472806, 3.223000	Terrasses de Montfau, Magalas	US 4015	Well PT 4000	4th c CE	Archaeological artifacts	Roman	
21	M-MDV13_07	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 13525	Well PT 13319	1605 ± 35 BP (417-515 CE)	C14	Late Roman/ Medieval	

22	M-MDV12_02	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1220 ± 30 BP (731-851 CE)	C14	Early Medieval	
23	M-MDV12_04	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1220 ± 30 BP (731-851 CE)	C14	Early Medieval	F
24	M-MDV12_05	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1220 ± 30 BP (731-851 CE)	C14	Early Medieval	
25	M-MDV12_07	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1220 ± 30 BP (731-851 CE)	C14	Early Medieval	
26	M-MDV12_09	43.808222, 4.368222	Mas de Vignoles XIV, Nîmes	US 12111	Well PT 12024	1220 ± 30 BP (731-851 CE)	C14	Early Medieval	F
27	M-C_27	45.436417, 5.520306	Colletiere, Charavines	N.D.	Cultural layer, rubbish deposits	1006-1040 CE	Dendrochronology	Medieval	
28	M-LM_22	47.900472, 1.884333	La Madeleine, Orléans	US 15126	Cesspit F 1517	1050-1200 CE	Archaeological artifacts	Medieval	

770 † Genetic clusters composed of identical clones. The genetic cluster was assigned
771 according to the relatedness analyses described in the results section.

772

773

a**b**

● Wine grapes – West and Central Europe
 ● Wine grapes – Iberian Peninsula and Northwest Africa
 ● Wine grapes – Balkans and East Europe
 ● Table grapes
 ● Other domesticated grapes
 ● *V. vinifera* subsp. *sylvestris*

● Grapereseq panel and ancient

▲ Whole-genome reference data (Table S2)

Iron Age:	9. R-HW71_03*	20. R-TDM_10
1. IA-LC_01*	10. R-HW71_17	
	11. R-R_09	Medieval Period:
Roman Period:	12. R-R_14*	21. M-MDV13_07
2. R-MDV14_04	13. R-LLE_02	22. M-MDV12_02
3. R-MDV14_07*	14. R-LLE_08	23. M-MDV12_04
4. R-MDV14_09	15. R-LLE_09	24. M-MDV12_05
5. R-MF_21	16. R-LLE_13	25. M-MDV12_07
6. R-MF_23	17. R-LLE_14	26. M-MDV12_09
7. R-MF_25	18. R-TDM_06*	27. M-C_27
8. R-HW70_18	19. R-TDM_08*	28. M-LM_22

c