



**HAL**  
open science

## Ancient genomes from present-day France unveil 7,000 years of its demographic history

Samantha Brunel, E. Andrew Andrew Bennett, Laurent Cardin, Damien Garraud, H el ene Barrand Emam, Alexandre Beylier, Bruno Boulestin, Fanny Chenal, Elsa Ciesielski, Fabien Convertini, et al.

► **To cite this version:**

Samantha Brunel, E. Andrew Andrew Bennett, Laurent Cardin, Damien Garraud, H el ene Barrand Emam, et al.. Ancient genomes from present-day France unveil 7,000 years of its demographic history. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117 (23), pp.12791-12798. 10.1073/pnas.1918034117. hal-03039135

**HAL Id: hal-03039135**

**<https://hal.science/hal-03039135v1>**

Submitted on 3 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

# Ancient genomes from present-day France unveil 7,000 years of its demographic history.

Samantha Brunel<sup>1</sup>, E. Andrew Bennett<sup>1</sup>, Laurent Cardin<sup>1,2</sup>, Damien Garraud<sup>1,2</sup>, H el ene Barrant Emam<sup>3,4</sup>, Alexandre Beylier<sup>5,6</sup>, Bruno Boulestin<sup>7</sup>, Fanny Chenal<sup>4,8</sup>, Elsa Ciesielski<sup>6</sup>, Fabien Convertini<sup>6,8</sup>, Bernard Dedet<sup>6</sup>, St ephanie Desbrosse-Degobertiere<sup>8</sup>, Sophie Desenne<sup>8,9</sup>, Jer ome Dubouloz<sup>9</sup>, Henri Duday<sup>7</sup>, Gilles Escalon<sup>6,8</sup>, V eronique Fabre<sup>6,8</sup>, Eric Gailledrat<sup>6</sup>, Muriel Gandelin<sup>8,10</sup>, Yves Gleize<sup>7,8</sup>, S ebastien Goepfert<sup>3</sup>, Jean Guilaine<sup>11,12</sup>, Lamys Hachem<sup>8,9</sup>, Michael Ilett<sup>9</sup>, Fran ois Lambach<sup>7</sup>, Florent Maziere<sup>6,8</sup>, Bertrand Perrin<sup>3,4</sup>, Suzanne Plouin<sup>4</sup>, Estelle Pinard<sup>8,9</sup>, Ivan Praud<sup>8,9</sup>, Isabelle Richard<sup>8,12</sup>, Vincent Riquier<sup>8,9</sup>, R ejane Roure<sup>5</sup>, Benoit Sendra<sup>6,8</sup>, Corinne Thevenet<sup>8,9</sup>, Sandrine Thiol<sup>8</sup>, Elisabeth Vauquelin<sup>8</sup>, Luc Vergnaud<sup>3,4</sup>, Thierry Grange<sup>1,\*</sup>, Eva-Maria Geigl<sup>1,\*</sup>, Melanie Pruvost<sup>1,7,\*</sup>

<sup>1</sup> Institut Jacques Monod, UMR 7592 CNRS, Universit  Paris Diderot, Paris, France

<sup>2</sup> Institut de Recherche Criminelle de la Gendarmerie Nationale (IRCGN), 95000 Pontoise

<sup>3</sup> ANTEA-Arch ologie, Habsheim, France

<sup>4</sup> UMR7044 Archim de, CNRS Universit  de Strasbourg et Universit  de Haute-Alsace, Strasbourg et Mulhouse, France

<sup>5</sup> Service Arch ologie S te agglom re m diterran e, 34110 Frontignan

<sup>6</sup> UMR 5140 ASM, CNRS, Universit  Paul Val ry Montpellier 3, 34199 Montpellier, France

<sup>7</sup> UMR 5199 PACEA, CNRS, Universit  de Bordeaux, 33615 Pessac, France

<sup>8</sup> INRAP, Institut National de Recherches Arch ologiques Pr ventives, 75685 Paris cedex 14, France

<sup>9</sup> UMR 8215 Trajectoires, CNRS, Universit  Paris 1 Pantheon Sorbonne, 92023 Nanterre, France

<sup>10</sup> UMR 5608 TRACES, CNRS, Universit  de Toulouse II31058 Toulouse, France

<sup>11</sup> Coll ge de France, 75231 Paris cedex 05

<sup>12</sup> UMR 7264 CEPAM, CNRS Universit  Nice Sophia Antipolis, 06357 Nice cedex 4, France

\* Corresponding authors: Melanie Pruvost, UMR 5199 PACEA, all e Geoffroy Saint-Hilaire, Universit  de Bordeaux, 33615 Pessac, France, [melanie.pruvost@u-bordeaux.fr](mailto:melanie.pruvost@u-bordeaux.fr)  
Eva-Maria Geigl, [eva-maria.geigl@ijm.fr](mailto:eva-maria.geigl@ijm.fr) and Thierry Grange, [thierry.grange@ijm.fr](mailto:thierry.grange@ijm.fr)

Competing interest statement: The authors declare no competing interests.

## Abstract

Genomic studies conducted on ancient individuals across Europe have revealed how migrations have contributed to its present genetic landscape, but the territory of present-day France has yet to be connected to the broader European picture. We generated a large dataset comprising the

complete mitochondrial genomes, Y chromosome markers and genotypes on a number of nuclear loci of interest of 243 individuals sampled across present-day France over a period spanning 7,000 years, complemented with a partially overlapping dataset of 58 low-coverage genomes. This panel provides a high-resolution transect of the dynamics of maternal and paternal lineages in France as well as of autosomal genotypes. Parental lineages and genomic data both revealed demographic patterns in France for the Neolithic and Bronze Age transitions consistent with neighboring regions, first with a migration wave of Anatolian farmers followed by varying degrees of admixture with autochthonous hunter-gatherers, then substantial gene flow from individuals deriving part of their ancestry from the Pontic Steppe at the onset of the Bronze Age. Our data have also highlighted the persistence of the Magdalenian-associated ancestry in hunter-gatherer populations outside of Spain and thus provide arguments for an expansion of these populations at the end of the Paleolithic period more northerly than what has been described so far. Finally, no major demographic changes were detected during the transition between the Bronze and Iron Ages.

Key words: paleogenomics, migration, Neolithic, Mesolithic, Protohistory, population genomics

## Significance Statement

Using genomic data as well as paternal and maternal lineages from more than 200 individuals, including 58 low-coverage ancient genomes, we show the population structure from the Mesolithic to the Iron Age in France, and trace the changing frequency of genotypes associated with phenotypic traits. Importantly, we also report the late persistence of Magdalenian-associated ancestry in hunter-gatherer populations, showing the presence of this ancestry beyond the Iberian Peninsula in the Late Paleolithic. This study complements the genomic history of Western Europe for this broad period by supplying a large genetic transect of three regions of France.

## Introduction

Over the last 10,000 years, populations in Western Eurasia have undergone two major cultural shifts: from a hunter-gatherer lifestyle transitioning to a lifestyle based on food production during the Neolithic<sup>1</sup>, and the development and improvement of metallurgy over the course of the third and second millennia BCE, giving rise to the Bronze Age that in turn evolved into the Iron Age during the last seven centuries BCE<sup>2</sup>.

Ancient genomes have been instrumental in characterizing past populations<sup>3-8</sup>. Diachronic series have investigated temporal genome-wide dynamics within defined European regions such as Central Europe<sup>9</sup>, Great Britain<sup>10</sup> or Iberia<sup>11,12</sup>, revealing that these cultural changes were the result of demographic changes that deeply altered the genetic makeup of Western Eurasian populations<sup>5,9,10,13</sup>.

In France, the demographic processes underlying these transitions are not yet explored at a territory-wide scale. Only a handful of studies restricted to individual archeological sites have been

performed, relying on partial mitochondrial sequence information<sup>14-17</sup> or on partial Y chromosome sequences<sup>15</sup>. Due to its geographic location, France occupies a strategic position in the understanding of population migrations in Western Europe<sup>18-20</sup>. Both the Central European and Mediterranean currents of neolithization (the Linearbandkeramik (LBK) complex in the North and the Impressa and Cardial complexes in the South) became established there, although the extent of their interactions remains an open question<sup>18-20</sup>. Moreover, the consequences on the gene pool of the region during the development and spread of metallurgy across Europe, at the beginning of the Bronze and Iron Ages, are still unknown.

To investigate the demographic dynamics over the 7,000-year-transect from the Mesolithic period, before the onset of agriculture, to the Iron Age, we genetically analyzed 243 unique individuals sampled from 54 different archeological sites, enriching for both complete mitochondrial genomes and a panel of 120 nuclear single nucleotide polymorphisms (SNPs) including Y chromosome SNPs, as well as a partially overlapping dataset of 58 low coverage genomes. Our data from ancient Western Europe fill a geographic gap that was so far missing, allowing a more global view of past population dynamics in Europe.

## Results and Discussion

All analyzed individuals belong to well-defined archaeological contexts from the Mesolithic period to the Iron Age and 40 were directly radiocarbon dated for this study (see SI Appendix, Table S1-2). We built sequencing libraries for 243 individuals and enriched them for mitochondrial genomes and a panel of 120 selected SNPs across the nuclear genome (see SI Appendix, Table S1-1). Along with complete mitochondrial genomes of 223 individuals and Y chromosome haplogroups of 62 male individuals, we report 58 newly generated ~0.05X to 0.5X genomes sequenced from UDG-treated libraries. We combined our genotype data with either complete genomes or 1240k genome-wide polymorphisms in ancient individuals, and genotypes of modern individuals from Europe, Caucasus and the Near East genotyped on the Affymetrix Human Origins array<sup>3,5,9,10,12,13,21,22</sup>. Complete mitochondrial genomes were retrieved from this ancient dataset and were grouped based on temporal and geographic proximity, often translating into cultural classifications (see SI Appendix S3). Contamination estimates derived from X chromosome polymorphisms in male individuals were consistently low (between 0.68%±1.26% and 3.55%±0.62% depending on the method considered), with only one individual displaying more than 5% X chromosome contamination (PIR3037AB) (see SI Appendix Table S1-1).

### The French Mesolithic substrate

We generated genomic data for five Mesolithic individuals from Les Perrats cave (Agris, Charente) dated from 7177-7057 cal. BCE including low coverage shotgun sequencing for three of

them (0.111-0.134X). All possess haplogroup U5b, demonstrating in the territory of present-day France the presence of a Mesolithic substrate akin to that described elsewhere where it was characterized by haplogroups U5b (85%) and, to a lesser extent, U5a (15%)<sup>3,23</sup>.

Mesolithic hunter-gatherers discovered in France are located near the edge of a PCA plot obtained from genome-wide genotypes, falling near the Mesolithic individuals from Western Europe<sup>5,6,9</sup> (Figure 1B and Figure S4-3). While the ancestry from a ~14,000-year-old individual from Villabruna in Italy was found to be dominant in Holocene hunter-gatherers across Western and Central Europe<sup>3</sup>, recently published studies conducted on ancient Iberians highlighted the survival of Magdalenian-associated ancestry<sup>11</sup>. This dual ancestry is found in all Iberian hunter-gatherers (HGs; ranging from 23.7% to 75.3%), while in most other regions only the Villabruna-associated ancestry remained. In order to investigate the proportions of ancestry in French HGs derived from these two Late Pleistocene lineages, we used qpAdm to model it as a mixture of two sources: ~15,000-year-old GoyetQ2 from Belgium, the least admixed individual associated with the Magdalenian complex in Europe sequenced so far<sup>11</sup>, and Villabruna. We found that French HGs from Les Perrats harbored relatively high proportions of GoyetQ2 ancestry ranging from 31.3% to 45.6% (Figure 1C), comparable to proportions described in the La Braña or Canes1 Mesolithic individuals from Spain, suggesting a late survival of Magdalenian-associated ancestry in HGs outside the Iberian Peninsula.

### **Successive waves of migration and admixture over the course of the Neolithic in France**

The arrival of an Anatolian Neolithic-associated ancestry component at the onset of the Neolithic is clearly visible in our dataset from both uniparentally-inherited markers and at the genome-wide level. Maternal lineages from Neolithic farmers in France (from ~5300 BCE onwards) are more diverse than those from HGs, and display variable frequencies across the time transect (see SI Appendix, Figure S3-2). Early and Middle Neolithic individuals share more affinity with present-day Southern Europeans (see SI Appendix, Figure S4-2), and their genetic variation is encompassed within that of contemporaneous European populations (Figure 1B). Three of our individuals associated with the LBK culture, which stems from the Danubian Neolithic current, cluster genetically with other Central European Early Neolithic (EN) individuals and share more drift with other early farmers associated with the LBK cultural complex (see SI Appendix, Figure S4-6). Values of *D* (Mbuti, France\_EN; Western Hunter-Gatherers (WHG), Anatolia\_Neolithic) are consistent with those observed in LBK-associated individuals, with the exception of Mor6 (~7,100 calBP) in north-eastern France (Figure 2B). Despite a consistent dating and cultural assignment to the LBK, this individual falls within the genetic diversity of Iberian\_EN individuals and harbors the highest proportion of shared alleles with WHG. After testing the potential origin of this ancestry with qpAdm, we found the best fit to be with a two-source ancestry model between Anatolia\_Neolithic and GoyetQ2 (p-value: 0,0911155), a situation so far only described in Early Neolithic individuals from the Iberian Peninsula. This is the

northern-most description of such ancestry. To date, the distribution of the Magdalenian-associated ancestry in the hunter-gatherer population in Western Europe that admixed with the Neolithic farmers in Europe is not clear because of the scarcity of genetic data from Mesolithic HGs. The observation of this ancestry in the Iberian Peninsula has been interpreted as a feature distinguishing the Cardial from the LBK Neolithic migrants<sup>12</sup>.

The observation that the ancestry of GoyetQ2 was found alongside Villabruna ancestry in HGs from Les Perrats in western France (~9,100 calBP), as well as in HGs from Rigney1 (central-eastern France; ~15,500 cal. BP), and from two caves in south-western Germany, Hohlefels (~15,000 calBP) and Burkhardtshöhle (~14,600 calBP), rather indicates that this mixed ancestry is characteristic of Western European HGs<sup>11</sup>. Thus, the admixture between HGs with GoyetQ2 ancestry and Neolithic migrants is not necessarily characteristic of the Cardial Neolithic migration wave as it seems to have occurred as well in more northern locations in Western Europe.

While the only male Mesolithic individual from our dataset was assigned to Y haplogroup I, LBK individuals from France showed a relative diversity, belonging to Y haplogroups C1a2, G2a and H2. G2a and H2 have been described in various European Early Neolithic contexts in both Central Europe and Iberia<sup>3,6,24</sup>, haplogroups I, whereas I2 are likely to have been introduced into the Neolithic pool through admixture with hunter-gatherers.

In eastern France, the transition into the Middle Neolithic is represented by individuals from the Grossgartach culture, that derives from the Danubian sphere (~4,700-4,500 BCE, Figure S1) Their proximity to Central European contemporaneous individuals on PCA plots generated from both mitochondrial haplogroup frequencies and genome-wide markers suggests genetic homogeneity within the Danubian current (Figure 1B, SI Appendix Figure S3-3). Interestingly, the BUCH2 individual harbors mixed GoyetQ2 and Villabruna HG ancestry (Figure 1C) suggesting again that the GoyetQ2 ancestry is not a specific signature of south-western Europe. However, we can observe a subtle shift away from LBK individuals and towards WHG, reflecting an increase in shared alleles with the latter. This echoes observations from other ancient populations across Europe (Figure 1C and 2A), yet with a higher proportion. This phenomenon seems to appear in France very early and becomes stronger as the Neolithic progresses, since individuals from French Middle Neolithic sites share a higher proportion of ancestry with HGs (Figure 2A) and display Mesolithic signature haplogroups (e.g. U5b) at variable frequencies across the territory (see SI Appendix, Figure S3-2). Predominant in the Champagne region of northeastern France, these haplogroups suggest variable proportions of admixture between autochthonous hunter-gatherers and farmers. During the second half of the Middle Neolithic (MN2, ~4500 BCE), individuals from northern and southern France are poorly distinguished on the PCA plot (Figure 1B).

Unlike in Iberia, the migration wave from which the Neolithic ancestry in ancient populations from France originated could not clearly be established using D statistics comparing the ancestry proportion between Iberia\_EN and LBK\_EN, as most individuals display non significant values (see

SI Appendix, Figure S4-6)<sup>12</sup>. Using qpAdm, we investigated the ancestral source of the HG ancestry in individuals from French MN2 sites. Only individuals from southern France harbor varying levels of GoyetQ2-like ancestry, suggesting admixture with different populations of hunter-gatherers (Figure 1C).

### **The Bell Beaker period in France**

At the end of the Neolithic, the Bell Beaker complex (named after the particular shape of its ceramics) appears and becomes widespread across Europe. The Beaker phenomenon is peculiar because of its large geographic distribution and because of its coexistence with local Late Neolithic and Copper Age cultures.

We report two Bell Beaker-associated individuals (CBV95 and PEI2), that we co-analyzed with previously reported contemporaneous individuals from Europe, including France<sup>10</sup>. French Beaker-associated individuals display a wide range of Steppe ancestry proportions (Figure 1C, 2D). CBV95 in northern France derives the highest proportion of alleles from the Yamnaya in our dataset, and belongs to Y chromosome haplogroup R1b (Figure 1C, 2D, SI Appendix Figure S4-5), providing the earliest clear evidence of the presence of this haplogroup in France around 2500 BCE (see SI Appendix, Table S5-1). This lineage was associated with the arrival of migrants from the Steppe in Central Europe during the Late Neolithic, and was described in other parts of Europe and in Bell Beaker associated individuals from southern France, while being almost absent in Iberia prior to the Bronze Age<sup>10,13</sup>. PEI2, a male unearthed from a collective burial site near Carcassonne in southwestern France with artifacts of the Bell Beaker complex, falls within the genetic diversity of Neolithic individuals on the PCA. Modeling admixture proportions between three source populations: Anatolia\_Neolithic, Villabruna and Yamnaya\_Samara, we could, however, detect 28.3% of Steppe ancestry in PEI2 (Figure 1C). These observations are consistent with previous findings and confirm that Steppe ancestry appeared later and with a lower impact in Southwestern Europe than in other parts of the continent<sup>10,25</sup>. Surprisingly, the admixture model did not include Villabruna as an ancestry source for either CBV95 and PEI2, which differs from previously known Late Neolithic individuals.

### **Relative continuity between the Bronze Age and the Iron Age**

Bronze Age individuals in France display new mitochondrial haplogroups such as U2, U4 and I, although at a low frequency. First described in Eastern European populations and in Late Neolithic Central Europe<sup>13</sup>, these haplogroups suggest gene flow with a population deriving maternal ancestry from the Pontic Steppe herders at the onset of the Bronze Age. As foretold by its early occurrence in Beaker-associated CBV95, a drastic Y chromosome turnover occurs during the Bronze Age, where R1b replaces the preexisting diversity of Neolithic lineages in our sampling (Bronze Age: 11/13; Iron Age: 7/10 individuals having R1b, Table S5-1 and S5-2 and SI Appendix Figure S5-1).

Bronze and Iron Age France share a common space on the PCA plot, both shifted towards modern Central Europe and falling within the genetic diversity of Bronze Age Britain and Central Europe (Figure 1B), with a homogenization of the Steppe Component (Figure 1C and 2C). In contrast to what was described for Central Europe<sup>9</sup>, there is no further shift toward Eastern Eurasian genotypes during the Iron Age<sup>9</sup>. Instead, the Steppe component, heterogeneously distributed between individuals during the Bronze Age (ranging from 30% to 70%), becomes homogeneous (Figure 1C and 2C), and individuals from the Hallstatt and La Tène culture on the French territory display similar affinities toward both modern and ancient populations. This could indicate that the transition from the Bronze Age to the Iron Age in France was mostly driven by cultural diffusion, without major gene-flow from an external population. This would be consistent with an archeological and linguistic hypothesis proposing that the Celts from the second Iron Age descended from populations already established in Western Europe, within the boundaries of the Bell Beaker cultural complex<sup>26–29</sup>. It is important to mention, however, that due to the relative genetic homogeneity among European populations by the Bronze Age, subsequent migrations between different parts of Europe could easily remain unnoticed at this level of coverage<sup>10,25,30</sup>.

### **Analysis of genetic markers associated to phenotypes**

To follow the evolutionary timelines of genetic adaptation to changing lifestyles and environments, we analyzed 73 autosomal loci, associated with both physical and physiological traits, genotyped across 149 individuals (Figure 3, see SI Appendix Table S2-1 for details on the targeted positions). Among the phenotypic traits of interest, we focused on genetic variants involved in eye and skin pigmentation (such as SLC24A5, SLC45A2, GRM5, HERC2, IRF4, TYR). It has been previously shown that Mesolithic HGs and Anatolian Neolithic farmers differ in their pigmentation alleles, with the latter carrying derived alleles responsible for a light skin that are near fixation in Europe today<sup>22</sup>, while HGs were described as having a rather dark skin and light eye color<sup>6,31</sup>. The Mesolithic individuals from whom we retrieved reliable genotype information after stringent filtering (1 to 3 individuals) possessed the ancestral pigmentation variant for loci SLC45A2 and GRM5, which had been associated with darker skin. In contrast, the derived alleles of these two loci are found at significantly higher frequency in the French Neolithic population (SLC45A2: 24.9%, GRM5: 38.4%), although they remain far from the frequency reached by these alleles in present-day Europe, where they are near fixation in the population (93.8 and 68.9%, respectively). SLC24A5, described as the principal mutation associated with light skin pigmentation, is found in 96% in our Neolithic population, close to the 99% found in present-day Europe. These results suggest different evolutionary timelines for these mutations in France, where a similar phenotype (light skin pigmentation) possibly resulted from several selection events, although the genetic determinism of these traits is likely more complex and not encoded by these three loci alone<sup>32</sup>. The Mesolithic individual that displayed coverage of variants involved in the pigmentation of the eye color (HERC2 and IRF4), carried the



alleles associated with the blue eye color. At the time of the Neolithic, the frequency of these two variants reaches 29.6% and 24.4%, respectively. Interestingly, these two variants seemingly have different trajectories after the Neolithic. While *HERC2* rises in frequency again from 37.5% during the Bronze Age to finally reach more than 63% in present-day Europe, *IRF4* further decreases from 23.81% during the Bronze Age to reach a frequency of 11% at present. None of the genotyped Neolithic individuals carried the mutation responsible for the persistence of lactase in Europeans, consistent with a later origin for this mutation<sup>13,30</sup>. We also analyzed several innate immunity-related variants that show signs of recent positive selection in modern European populations<sup>6,33–35</sup>. For a number of genes involved in the immune response (*NOS2A*, *TOLLIP*, *CCL18*, *STAT6*, *IL3*), the frequency of derived alleles was comparable to that found in present-day Europeans, indicating that selection on these loci predates the Neolithic, whereas frequencies of variants in the *TLR1-6-10* locus, the Toll-Like receptor family involved in the innate immune response, are much lower than those found in present-day Europeans<sup>36</sup>. The *TLR1-6-10* gene cluster is possibly associated with resistance to leprosy, tuberculosis or other mycobacteria<sup>36</sup>. Our data suggest that the onset of agriculture and the proximity with domesticated animals did not exert a selective pressure on this locus. Likewise, variants associated with coeliac disease in *SH2B3* or *SLC22A4* that could have been the target for positive selection owing to the role they play in inflammations and the elimination of a wide array of environmental toxins<sup>33</sup>, did not reach yet the frequency of modern populations.

## Conclusion

The study of 7000 years of demographic history through the study of ancient genomes throughout France reveals successive migration events mostly accompanied by major cultural changes and admixture events whose traces are still found in current European populations.

Our genomic data obtained for three Mesolithic individuals from Western France show that the late survival of the Upper Paleolithic Magdalenian-associated ancestry (represented by GoyetQ2) was not restricted to the Iberian Peninsula. This finding expands the region where this ancestry is found and raises the question of where this admixture occurred. More Mesolithic individuals from Western Europe need to be studied to better characterize the history of this ancestry.

While the first Neolithic people who inhabited the territory of present-day northern France descended from early farmers in Anatolia, therefore forming a group distinct from the local hunter-gatherers, we could detect later admixture with hunter-gatherers supporting their progressive integration within farming communities. Although widespread, the process of admixture with hunter-gatherers shows regional variability and origins. However, the scarcity of the data available for the South of France does not allow us to conclude whether contact occurred between the LBK and Cardial complexes, nor to identify any legacy that may be present in the Neolithic cultures that followed. Genomic data from other French regions in the North and the South will be necessary to confirm the

hypotheses on the origin and the relations of the different cultural groups of the Middle and Late Neolithic.

The study of a selection of nuclear markers revealed differences between Neolithic people and present-day Europeans. Derived allele frequencies indicate selection on loci involved in pigmentation, diet and immunity that are consistent with adaptation to high latitudes and changes in diet. Finally, neither Mesolithic HGs nor Neolithic farmers on the territory of present-day France carried the allele allowing for lactase persistence during adulthood, in agreement with studies of Neolithic populations further East that reported the presence of this allele only by the end of the Bronze Age<sup>13,30</sup>.

An important outcome of our study is the genetic continuity between Bronze and Iron Age individuals of France but with less heterogeneity between individuals from Iron Age. In such context, tracing migrations from one part of Europe to the other becomes more challenging and will require a deeper temporal and geographical resolution.

## **Material and methods**

### **Processing of archaeological samples**

A total of 243 individuals from 54 archeological sites across France passed screening for the preservation of ancient DNA and were processed for sequencing (see SI Appendix Table S1, Figure S1-1). These individuals covered a time period ranging from the Mesolithic to the Iron Age, with the majority of samples covering the Neolithic period (~5500-3500 BCE). As direct dates were not available for all samples, they were labeled and grouped for analyses based on geographical proximity and cultural assignment (see SI Appendix Figure S1-1).

### **Targeted DNA enrichment and sequencing**

DNA was extracted from petrous bones and teeth and used to build double-stranded Illumina libraries with thorough decontamination procedures and optimized protocols in a dedicated facility at the Institut Jacques Monod (see SI Appendix SI.2)<sup>37</sup>. Sequence enrichment for both the whole mitochondrial genome and nuclear regions was performed through a capture approach using biotinylated RNA baits obtained through in vitro transcription of PCR amplified target regions (<sup>37</sup>, see SI Appendix SI.2 for details on the procedure). Both enriched and whole genome shotgun libraries were sequenced on an Illumina MiSeq or Illumina NextSeq at the Institut Jacques Monod in Paris, and the Institut de Recherche Biomédicale des Armées in Brétigny-sur-Orge, respectively.

### **Data processing and authentication**

Adapters were trimmed and overlapping paired-end reads were merged using LeeHom<sup>38</sup>. Merged reads were then quality trimmed with cutadapt and only reads longer than 30bp were retained.

Mapping was performed with BWA version 1.2.3<sup>11</sup> against the Cambridge reference sequence (rCRS)<sup>39</sup> for human mitochondrial DNA with a duplication of its first 100 bases at the end to ensure mapping of the reads overlapping the junction resulting from the virtual linearization of the circular mitogenome, and to the human reference genome hs37d5. In both cases, PCR duplicates were removed using Picard's MarkDuplicate (<https://broadinstitute.github.io/picard/>). Ancient DNA authentication was assessed using MapDamage<sup>40</sup> and contamination was estimated using both the mitochondrial genome and X chromosome<sup>41</sup>.

### **Uniparental haplogroups**

Consensus sequences of the mitochondrial genome were obtained using the software ANGSD v0.910, relying only on reads with base quality above 20, mapping quality above 30, and with more than 3X coverage. Mitochondrial haplotypes were determined based on the PhyloTree phylogenetic tree<sup>42</sup> build 17 and by using the HAPLOFIND web application<sup>43</sup> and Phy-Mer<sup>44</sup>. Unexpected or missing mutations were visually inspected in Geneious version R6<sup>45</sup> to check if they could be the results of misincorporations in low coverage regions. Sequences with more than 5% undetermined sites were removed. Y chromosome haplogroups were retrieved using the Yleaf software<sup>46</sup>.

### **Nuclear markers**

Target regions were summarized in a file, allowing their extraction using the mpileup tool from the SAMTools toolkit<sup>47</sup>. Proper variant calling files were obtained using bcftools call and filtered for base quality 20 and a minimum coverage of 5 reads using the SelectVariants tool from the Genome Analysis Toolkit<sup>48</sup>. Allelic count and frequency estimation was carried out with plink<sup>49</sup>.

### **Reference datasets**

We merged our dataset with published ancient individuals from Western Eurasia as well as genotyping data from the modern populations of the Human Origins panel<sup>50</sup>. We only used published data obtained from UDG treated libraries. Genotype was called for the 1.2 million SNPs broadly used in aDNA in-solution capture procedures<sup>22</sup>. A single allele was drawn at random for each position (minimum mapping and base quality of 30) using PileupCaller (<https://github.com/stschiff/sequenceTools>), rendering the individuals from the dataset homozygous for each locus. For genotyping data from modern individuals of the Human Origins panel, heterozygous sites in the eigenstrat file were randomly recoded as homozygous for the reference or the alternative allele using an awk script. The two files were then merged using plink v1.9<sup>49</sup>.

## Population Genetic Analyses

We conducted PCA on ancient individuals, projecting them onto the genetic variation of present-day Eurasians using the `lsqproject` option from `smartpca`<sup>51</sup>. We computed  $D$  and  $f$  statistics using `Admixtools` to estimate gene flow and shared drift between populations and to test tree topologies<sup>52</sup>. We used `qpAdm` to estimate the ancestry proportions attributable to early Anatolian farmers, hunter-gatherers, and steppe herders in the individuals of this study<sup>13</sup>.

### Data Availability:

Sequencing data can be found on the European Nucleotide Archive, ENA: ERP119731

### Funding:

This work was supported by the Agence National de la Recherche (grant: ANR15-CE27-0001), the Institut National de Recherche en Archéologie préventive (INRAP) (PAS ANCESTRA) and the CNRS. The ARTEMIS program (French Ministry of Culture) funded the C14 datations. The paleogenomic facility of the Institut Jacques Monod (IJM) obtained support from the University Paris Diderot within the program “Actions de recherches structurantes”. The sequencing facility of the IJM is supported by grants from the University Paris Diderot, the Fondation pour la Recherche Médicale (DGE20111123014), and the Région Ile-de-France (11015901).

### Acknowledgments:

We would like to thank the Service Régional de l’Archéologie (SRA) from Occitanie, Grand Est, Hauts-de-France, ANTEA and INRAP for providing bone materials. We are grateful to Olivier Gorgé for making shotgun sequencing possible at the Institut de Recherche Biomédicale des Armées (IRBA) in the Division defense NRBC, département des services, Unité de Biologie moléculaire.

Contributions: S.B., T.G., E.-M.G. and M.P. performed or supervised laboratory work with input from E.A.B., L.C., D.G.; archaeological materials were collected and information about archaeological contexts were provided by H.B.E., A.B., B.B., F.Ch., E.C., F.Co., B.D., S. D.-D., S.D., J.D., H.D., G.E., V.F., E.G., M.G., Y.G., S.G., J.G., L.H., M.I., F.L., F.M., B.P., S.P., E.P., I.P., I.R., V.R., R.R., B.S., C.T., S.T., E.V. and L.V.; S.B. and M.P. analyzed data with input from E.A.B., T.G., E.-M.G.; S.B. and M.P. wrote the manuscript with the contribution of E.A.B., T.G., E.-M.G., J.D., B.B., M.G., J.G., R.R. and all other co-authors.

1. Bellwood, P. *et al.* First Farmers: the Origins of Agricultural Societies, by Peter Bellwood. Malden (MA): Blackwell, 2005; *Camb. Archaeol. J.* **17**, 87 (2007).
2. Kristiansen, K. & Larsson, T. B. *The Rise of Bronze Age Society: Travels, Transmissions and Transformations.* (Cambridge University Press, 2005).
3. Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
4. Keller, A. *et al.* New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat. Commun.* **3**, 698 (2012).
5. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
6. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
7. Sánchez-Quinto, F. *et al.* Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. *Curr. Biol. CB* **22**, 1494–1499 (2012).
8. Skoglund, P. *et al.* Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* **336**, 466–469 (2012).
9. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5**, 5257 (2014).
10. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
11. Villalba-Mouco, V. *et al.* Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula. *Curr. Biol. CB* **29**, 1169-1177.e7 (2019).
12. Valdiosera, C. *et al.* Four millennia of Iberian biomolecular prehistory illustrate the impact of prehistoric migrations at the far end of Eurasia. *Proc. Natl. Acad. Sci.* 201717762 (2018) doi:10.1073/pnas.1717762115.

13. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* (2015).
14. Beau, A. *et al.* Multi-scale ancient DNA analyses confirm the western origin of Michelsberg farmers and document probable practices of human sacrifice. *PLoS ONE* **12**, (2017).
15. Lacan, M. *et al.* Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18255–18259 (2011).
16. Rivollat, M. *et al.* When the Waves of European Neolithization Met: First Paleogenetic Evidence from Early Farmers in the Southern Paris Basin. *PLOS ONE* **10**, e0125521 (2015).
17. Rivollat, M. *et al.* Ancient mitochondrial DNA from the middle neolithic necropolis of Obernai extends the genetic influence of the LBK to west of the Rhine. *Am. J. Phys. Anthropol.* **161**, 522–529 (2016).
18. *La protohistoire de la France*. [The Protohistory of France] dir: Guilaine J. & Garcia D.(Hermann, 2018). *French*
19. Demoule, J.-P. *La révolution néolithique en France*. [The neolithic revolution in France] (La Découverte, 2007). *French*
20. Constantin, C. & Vachard, D. Anneaux d'origine méridionale dans le Rubané récent du Bassin parisien. [Rings of southern origin in the Recent Rubane of the Paris Basin.] *Bull. Société Préhistorique Fr.* **101**, 75–83 (2004). *French*
21. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372 (2017).
22. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).

23. Posth, C. *et al.* Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Curr. Biol.* **26**, 827–833 (2016).
24. Jones, E. R. *et al.* Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.* **6**, 8912 (2015).
25. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019)
26. Brun, P. L'origine des Celtes. Communautés linguistiques et réseaux sociaux. [The origin of the Celts. Linguistic communities and social networks] (2008).in Danièle Vitali (Ed) Celtes et Gaulois. L'archéologie face à l'histoire. La préhistoire des Celtes. Glux-en-Glenne, Bibracte p29-44. French
27. Demoule, J.-P. Les indo-européens, un mythe sur mesure. [Indo-europeans, a tailor-made myth] *La Recherche* 40–47 (1998). French
28. Demoule, J.-P. *Mais où sont passés les Indo-Européens?: le mythe d'origine de l'Occident.* [But where have the Indo--europeans gone: the origin myth of the West] (Éditions du Seuil, 2014). French
29. Hawkes, C. F. C. & Hawkes, S. C. *Greeks, Celts, and Romans: studies in venture and resistance.* (Rowman and Littlefield, 1973).
30. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172 (2015).
31. Brace, S. *et al.* Population Replacement in Early Neolithic Britain. *bioRxiv* 267443 (2018) doi:10.1101/267443.
32. Hysi, P. G. *et al.* Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* **50**, 652–656 (2018).

33. Hunt, K. A. *et al.* Newly identified genetic risk variants for celiac disease related to the immune response. *Nat. Genet.* **40**, 395–402 (2008).
34. Yasuda, H. *et al.* Association of single nucleotide polymorphisms in endothelin family genes with the progression of atherosclerosis in patients with essential hypertension. *J. Hum. Hypertens.* **21**, 883–892 (2007).
35. Zhou, L.-T., Qin, L., Zheng, D.-C., Song, Z.-K. & Ye, L. Meta-analysis of genetic association of chromosome 9p21 with early-onset coronary artery disease. *Gene* **510**, 185–188 (2012).
36. Barreiro, L. B. *et al.* Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* **5**, e1000562 (2009).
37. Massilani, D. *et al.* Past climate changes, population dynamics and the origin of Bison in Europe. *BMC Biol.* **14**, 93 (2016).
38. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Res.* **42**, e141 (2014).
39. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
40. Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E. & Orlando, L. mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinforma. Oxf. Engl.* **27**, 2153–2155 (2011).
41. Renaud, G., Slon, V., Duggan, A. T. & Kelso, J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* **16**, 224 (2015).
42. van Oven Mannis & Kayser Manfred. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2008).



43. Vianello, D. *et al.* HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
44. Navarro-Gomez, D. *et al.* Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. *Bioinforma. Oxf. Engl.* **31**, 1310–1312 (2015).
45. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
46. Ralf, A., Montiel González, D., Zhong, K., Kayser, M. & Su, B. Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Mol. Biol. Evol.* **35**, 1291–1294 (2018).
47. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
48. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
49. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
50. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
51. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
52. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065–1093 (2012).

## Legends

### **Figure 1: Overview of the ancient French dataset**

Filled symbols are used for ancient individuals from France whereas open symbols are used for other ancient Western Eurasian. The shape of the symbol indicates the geographic origin and the color, the time period.

A. Location of the samples included in the study.

B. Principal component analysis of ancient Western Eurasians projected onto the variation of present-day genotypes, restricted to Europe.

C. Ancestry proportion for new and previously reported French individuals ranging from the Mesolithic to the Iron Age established using qpAdm. Each bar represents one individual with associated mtDNA haplogroup and Y chromosome haplogroup on the right side (error bars indicated  $S \pm 1SE$ ). Symbol used in Figure 1A and B are indicated for each individual.

### **Figure 2: Genetic affinity between ancient French and either Western hunter-gatherers or Steppe herders**

A. Boxplot of the D-statistics values of the form  $D(\text{Mbuti}, \text{Test: WHG}, \text{Anatolia\_Neolithic})$  for Early and Middle Neolithic European population with WHG consisting in Loschbour, ElMiron, GoyetQ2, Villabruna, KO1 and LaBrana to represent better the GoyetQ2 and Villabruna ancestry.

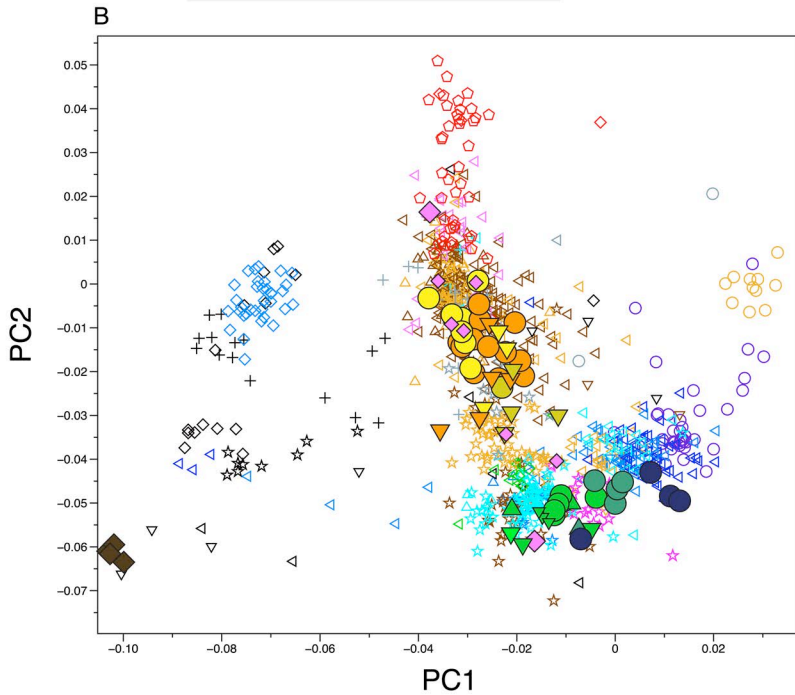
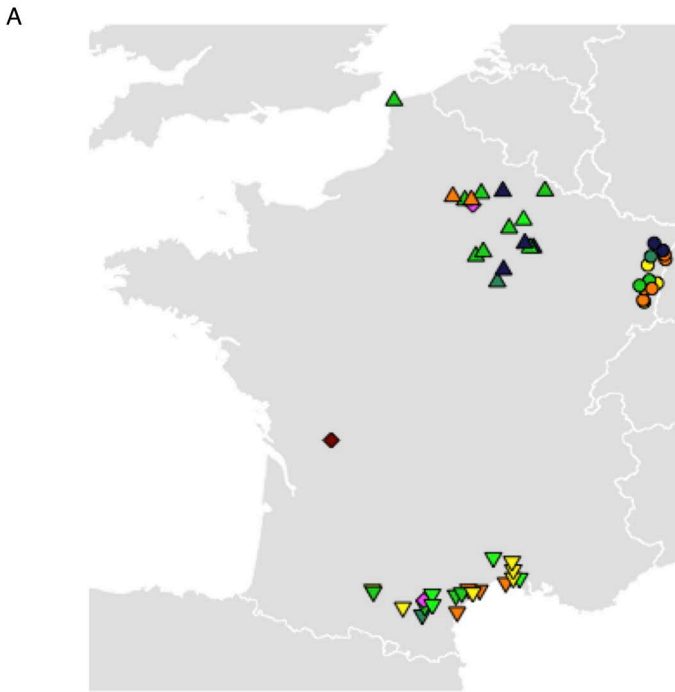
B.  $-D$ -statistics of the form  $D(\text{Mbuti}, \text{Test: WHG}, \text{Anatolia\_Neolithic})$  for Early Neolithic individuals of present-day Germany, France and Iberia. Error bars represent one standard error.

C. Boxplot of the D-statistics values of the form  $D(\text{Mbuti}, \text{Test; Yamnaya\_Samara}, \text{Anatolia\_Neolithic})$  for Late Neolithic, Bronze and Iron Age populations.

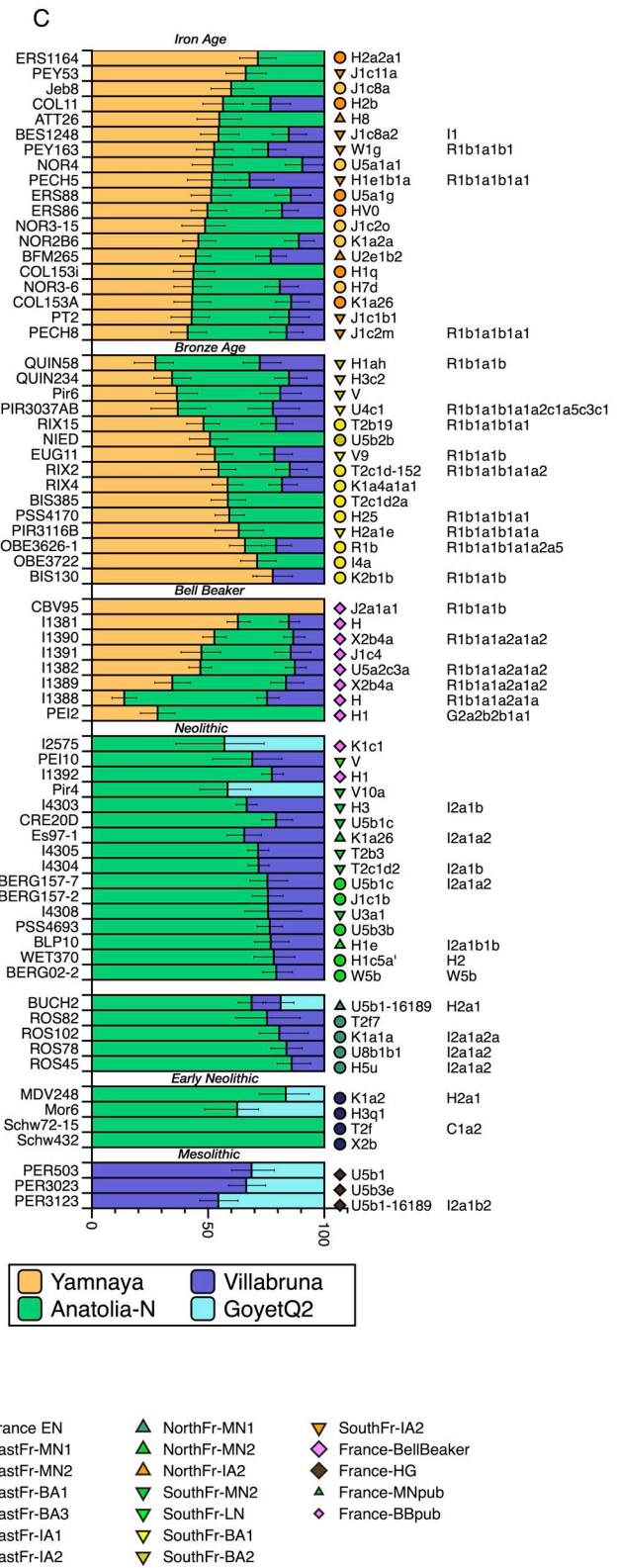
D. D-statistics of the form  $D(\text{Mbuti}, \text{Test; Yamnaya\_Samara}, \text{Anatolia\_Neolithic})$  for individuals associated with the Bell Beaker cultural complex. Error bars represent one standard error.

### **Figure 3: Frequency of the derived allele for several types of genetic markers in Neolithic French and present-day Europeans.**

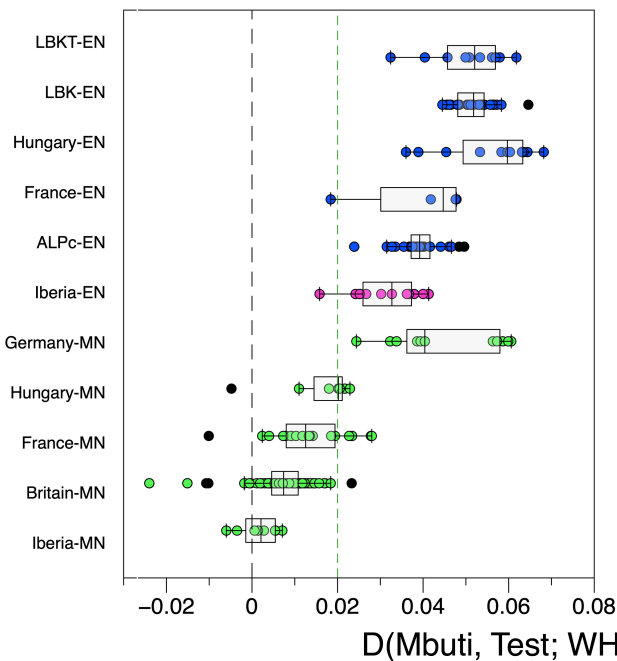
Numbers indicated above the bar chart correspond to the total number of observations of the derived allele for each variant in the Neolithic population. The color code reflects the function associated with the considered locus. The frequency of the alleles in present-day European populations is indicated in grey.



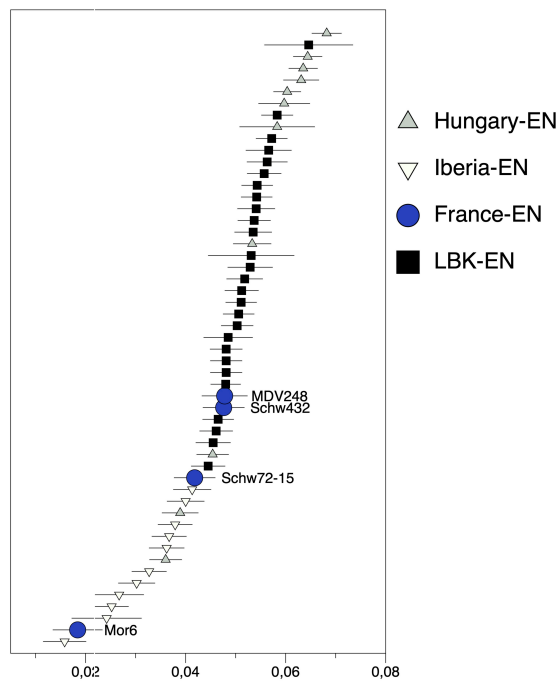
- Anatolia
- ◁ Central Europe
- △ Great Britain
- ☆ Iberia
- ▽ Italy
- ◇ Steppe
- ◇ East Europe
- + Northern Europe
- Mesolithic
- LBK
- Middle Neo
- Late Neo
- Corded ware
- Globular Amphora
- Bell Beaker
- Bronze Age
- Iron Age
- Yamnaya



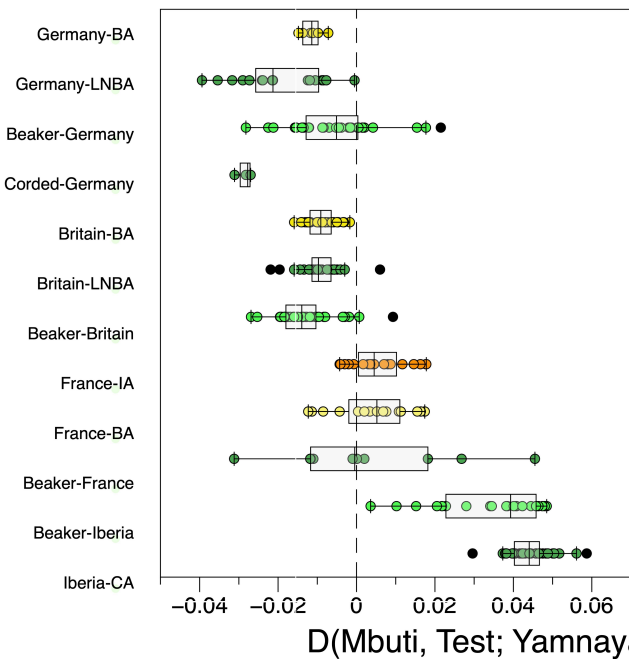
A



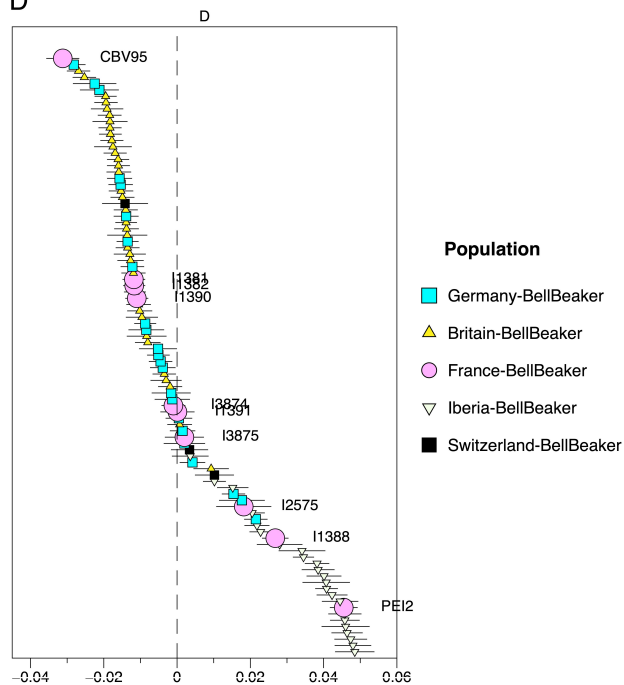
B



C



D



Frequency

