Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France

Andaine Seguin-Orlando, Richard Donat, Clio Der Sarkissian, John Southon, Catherine Thèves, Claire Manen, Yaramila Tchérémissinoff, Eric Crubézy, Beth Shapiro, Jean-François Deleuze, et al.

Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell BeakerGenomes from Present-Day France

Andaine Seguin-Orlando, Richard Donat, Clio Der Sarkissian, John Southon, Catherine Thèves, Claire Manen, Yaramila Tchérémissinoff, Eric Crubézy, Beth Shapiro, Jean-François Deleuze, Love Dalén, Jean Guilaine, and Ludovic Orlando.

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Current Biology

CellPress

Report

Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France

Andaine Seguin-Orlando,[1,2,13,15,*] Richard Donat,[1,3] Clio Der Sarkissian,[1] John Southon,[4] Catherine Thèves,[1] Claire Manen,[5] Yaramila Tchérémissinoff,[3,6] Eric Crubézy,[1] Beth Shapiro,[7,8] Jean-François Deleuze,[9] Love Dalén,[10,11] Jean Guilaine,[12] and Ludovic Orlando[1,14,*]

[1]Centre d'Anthropobiologie et de Génomique de Toulouse CAGT, CNRS UMR 5288, Université Toulouse III Paul Sabatier, Faculté de Médecine Purpan, Bâtiment A, 37 allées Jules Guesde, 31000 Toulouse, France
[2]Institute for Advanced Study in Toulouse IAST, Université Toulouse I Capitole, Esplanade de l'Université, 31080 Toulouse Cedex 06, France
[3]Institut National de Recherches Archéologiques Préventives INRAP, 561 Rue Etienne Lenoir, 30900 Nîmes, France
[4]Earth System Science Department, B321 Croul Hall, University of California, Irvine, Irvine, CA 92697-3100, USA
[5]Laboratoire Travaux et Recherches Archéologiques sur les Cultures, les Espaces et les Sociétés TRACES, CNRS UMR 5608, Université Toulouse II Jean Jaurès, Maison de la Recherche, 5 allées A. Machado, 31058 Toulouse Cedex 9, France
[6]Laboratoire Méditerranéen de Préhistoire Europe Afrique LAMPEA, CNRS UMR 7269, Aix Marseille Université, 5 rue du Château de l'horloge, 13094 Aix-en-Provence, France
[7]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[8]Howard Hughes Medical Institute, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
[9]Centre National de Recherche en Génomique Humaine CNRGH, Institut de Biologie François Jacob, Université Paris Saclay, CEA, 2 rue Gaston Crémieux CP 5721, 91057 Evry, France
[10]Centre for Palaeogenetics, Svante Arrhenius väg 20C, 10691 Stockholm, Sweden
[11]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Box 50007, 10405 Stockholm, Sweden
[12]Collège de France, 11, place Marcelin-Berthelot, 75005 Paris, France
[13]Twitter: @AndaineO
[14]Twitter: @LudovicLorlando
[15]Lead Contact
*Correspondence: andaine.seguin@univ-tlse3.fr (A.S.-O.), ludovic.orlando@univ-tlse3.fr (L.O.)
https://doi.org/10.1016/j.cub.2020.12.015

SUMMARY

The transition from the Late Neolithic to the Bronze Age has witnessed important population and societal changes in western Europe.[1] These include massive genomic contributions of pastoralist herders originating from the Pontic-Caspian steppes[2,3] into local populations, resulting from complex interactions between collapsing hunter-gatherers and expanding farmers of Anatolian ancestry.[4–8] This transition is documented through extensive ancient genomic data from present-day Britain,[9,10] Ireland,[11,12] Iberia,[13] Mediterranean islands,[14,15] and Germany.[8] It remains, however, largely overlooked in France, where most focus has been on the Middle Neolithic (n = 63),[8,9,16] with the exception of one Late Neolithic genome sequenced at 0.05× coverage.[16] This leaves the key transitional period covering ~3,400–2,700 cal. years (calibrated years) BCE genetically unsampled and thus the exact time frame of hunter-gatherer persistence and arrival of steppe migrations unknown. To remediate this, we sequenced 24 ancient human genomes from France spanning ~3,400–1,600 cal. years BCE. This reveals Late Neolithic populations that are genetically diverse and include individuals with dark skin, hair, and eyes. We detect heterogeneous hunter-gatherer ancestries within Late Neolithic communities, reaching up to ~63.3% in some individuals, and variable genetic contributions of steppe herders in Bell Beaker populations. We provide an estimate as late as ~3,800 years BCE for the admixture between Neolithic and Mesolithic populations and as early as ~2,650 years BCE for the arrival of steppe-related ancestry. The genomic heterogeneity characterized underlines the complex history of human interactions even at the local scale.

RESULTS AND DISCUSSION

Archaeological Sites and Radiocarbon Dates

A total of 30 ancient human petrosal bones, 5 post-cranial bones, and 18 teeth were prepared in ancient DNA facilities for shotgun sequencing on Illumina platforms (STAR Methods). Following shallow sequencing on the MiniSeq instrument, 26 (~49%) DNA extracts showing at least 10% of human DNA were selected for further sequencing on the NovaSeq and HiSeqX instruments (Data S1A). These correspond to remains excavated from

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

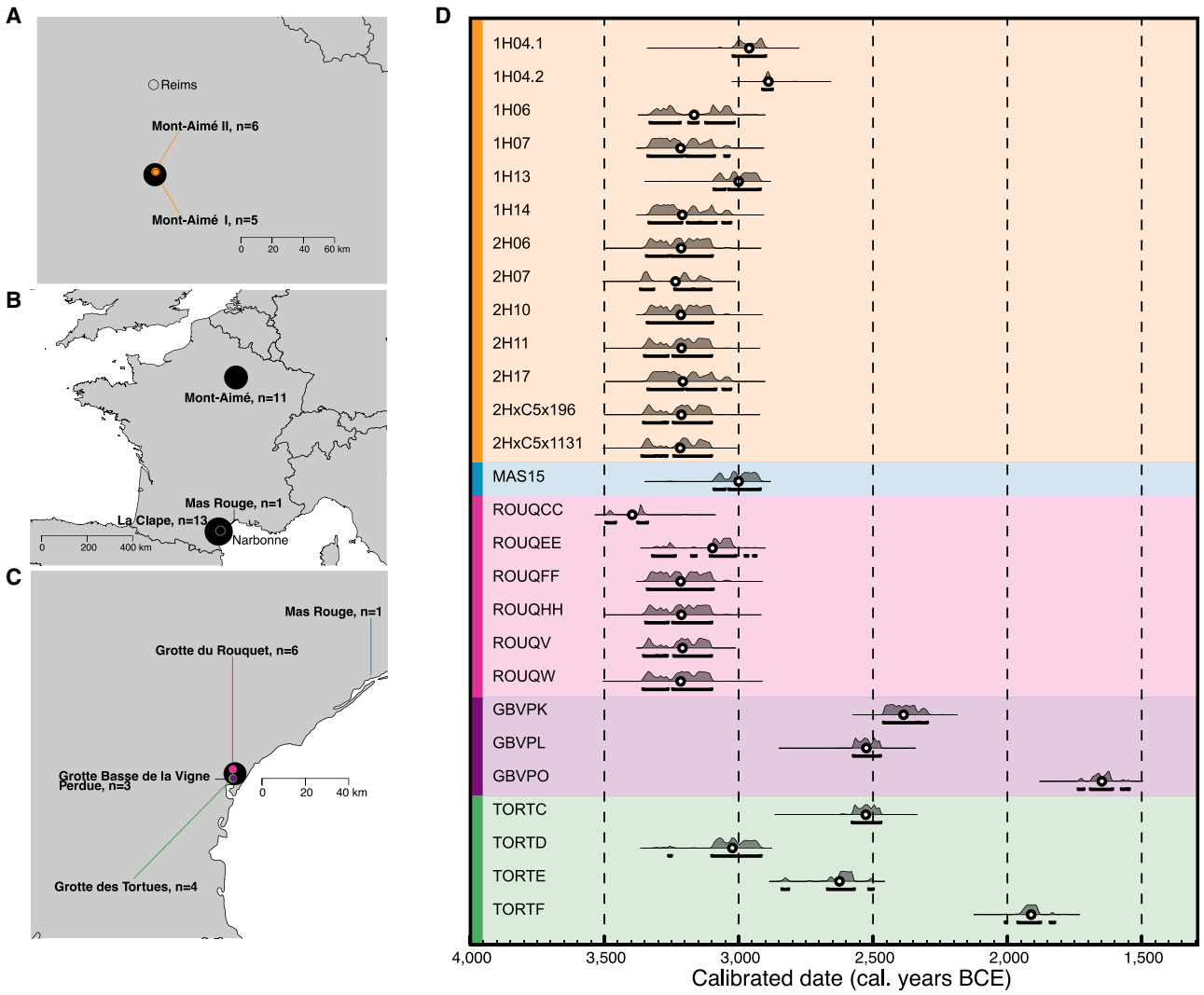

Figure 1. Sample Location and Radiocarbon Dates

(A) Geographic location of Mont-Aimé hypogea. The city of Reims, located north of Mont-Aimé, is shown.

(B) Geographic location of the main regions of present-day France investigated in this study. The city of Narbonne, located near the La Clape massif, is shown.

(C) Zoom in of the archaeological sites from southwestern France. The number of individuals analyzed genetically is indicated (n =).

(D) Bayesian posterior distributions of radiocarbon dates calibrated with OxCal online (https://c14.arch.ox.ac.uk/oxcal/OxCal.html) on IntCal20.[18]

See also Figure S1 and Data S1.

collective burials associated with Late Neolithic, Bell Beaker, and Early Bronze Age archaeological contexts and located in the Paris Basin (Mont-Aimé) and southwestern France (Mas Rouge, Grotte du Rouquet, Grotte Basse de la Vigne Perdue, and Grotte des Tortues; Figures 1A–1C; STAR Methods). Direct radiocarbon dating indicated that the two Mont-Aimé hypogea (i.e., human-made subterranean burials) were active between ~3,400 and 2,900 cal. years (calibrated years) BCE (Figure 1D). All six individuals from Grotte du Rouquet were virtually contemporary to those from Mont-Aimé. Grotte des Tortues included one Late Neolithic individual dated to ~3,100–2,900 cal. years BCE (TORTD), and three individuals spanning the mid-3rd (TORTC and TORTE) and the first half of the 2nd (TORTF) millennium BCE. This indicates that the Grotte des Tortues cave was used opportunistically as a collective burial site over a time period encompassing the Late Neolithic to the Early Bronze Age, in line with previous archaeological evidence.[17] The same was true for Grotte Basse de la Vigne Perdue, although only from the mid-3rd to the mid-2nd millennium BCE.

Genomic and Metagenomic Content

Metagenomic taxonomic profiles inferred from metaBIT[19] and the MetaPhLAn2 database[20] mainly contained contaminating bacterial species of unknown or environmental origin (Data S1B). Three out of 10 teeth from Mont-Aimé (2H10, 2H11, and 1H13), however, showed increasing fractions of human oral bacteria (4.8%–39.5%). These include the cariogenic *Streptococcus mutans*, as well as *Treponema denticola*, *Porphyromonas endodontalis*, and

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

CellPress

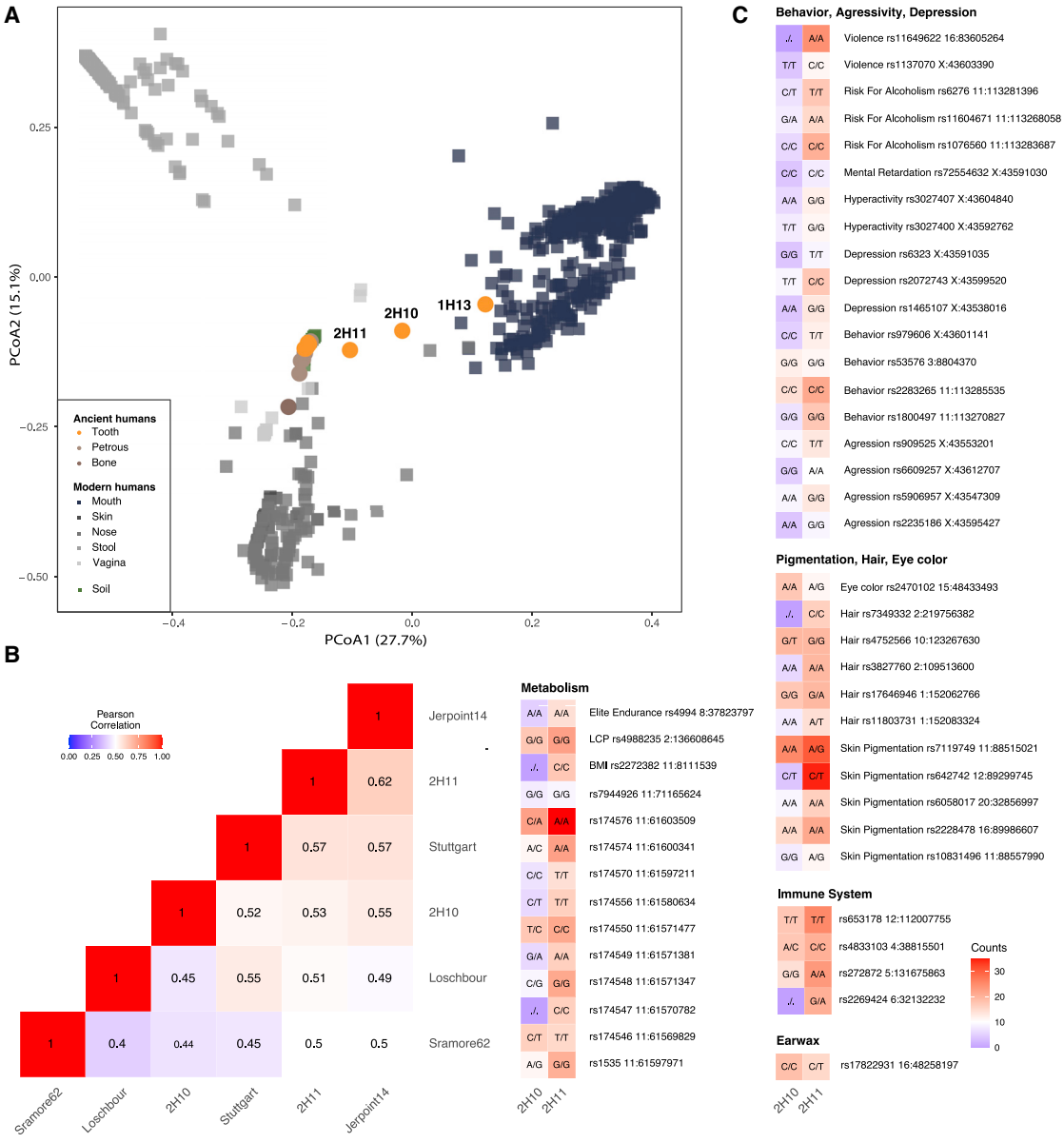

Figure 2. Microbial Profiling, DNA Methylation Analyses, and Phenotype Prediction

(A) Genus-level principal-coordinates analysis (PCoA) of bacterial profiles. Taxa supported by abundances strictly inferior to 1% were disregarded.

(B) Correlation matrix of DNA methylation levels at gene bodies.

(C) Genotypes observed for the two Mont-Aimé samples sequenced to high coverage at 49 loci underlying important morphological, metabolic, behavioral, and medical phenotypes. The number of independent sequence counts is provided for each locus considered. Undetermined genotypes are indicated as "./." See also Data S2 and S3.

Campylobacter rectus that are associated with peri- and endodontal diseases[21] but were also detected in healthy individuals today and in 120- to 5,700-year-old samples worldwide (Data S1D).[22–34] The presence of *S. mutans* as early as the Neolithic adds to the evidence[27,34,35] that this pathogen cannot provide a reliable marker for tooth decay following lifestyle changes in the Bronze Age and post-industrial era, despite earlier claims.[36] The proximity of oral metagenomic profiles measured in 1H13 and modern individuals (Figure 2A) is in line with the reduced fraction of unclassified and environmental microbes in this sample but may also indicate that microscopic calculus remained undetected and was co-extracted with the tooth cementum.

Mitochondrial sequence alignments indicated minimal human DNA contamination in all individuals (Data S1E), except two from Mont-Aimé; 2H06, which is characterized with minimal sequence data (and excluded from downstream analyses), and 1H14, which shows 10%–12% contamination. X-to-autosomal sequence coverage revealed 13 males, for whom X chromosome heterozygosity profiles confirmed minimal autosomal contamination ($\leq$1.01%; median = 0.12%; Data S1F).

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Genetically identified as female, individual 1H14 could thus not be tested with this approach. Sequence data displayed *post mortem* DNA degradation characteristic of DNA extracts partially treated with USER[37] (Figure S1A), and genetic affinities were in line with those Mont-Aimé males showing minimal contamination. As mtDNA data do not necessarily reflect autosomal contamination rates,[38] individual 1H14 was included in downstream analyses. Overall, we report a total of 24 ancient genomes at 0.20- to 23.88-fold average depth of coverage (median = 0.96-fold), including two at high depth (2H10 and 2H11, 13.91- and 23.88-fold, respectively). Supporting our data quality, error rates ranged between 0.057% and 0.083% substitutions per base (median = 0.069%), which is approximately half the average error rate estimated in ancient (high-coverage) genomes previously reported (Figures S1C–S1E).

Uniparental Markers and Kinship

No individuals from the same archaeological site carried identical mitochondrial haplotypes, indicating no maternal relatedness (Data S1E). In contrast, all males were assigned to the Y chromosome haplogroup I2a1, except one individual at Mont-Aimé belonging to haplogroup H2a1 and one ~4,400-year-old R1b1a1b1a1a2a1 individual from Grotte Basse de la Vigne Perdue (Data S1F). Close or incompletely covered but matching haplotypes were previously reported in contemporary Bell Beaker individuals (R1b1a1b for sample CBV95, La-Bouche-à-Vesle) and Bronze Age individuals from both northern (R1b1a1-b1a1a2 for sample RIX2, Rixheim-Zac du Petit-Prince; R1b1a1-b1a1a2a5 for sample OBE3626-1, Obernai PAEI) and southern France (R1b1a1b1a1a for sample PIR3116B Rec de Ligno).[16] This may reflect admixture from incoming steppe herders that, around that same time, have almost completely replaced other Y chromosome haplotypes in Iberian populations.[13] Haplogroup I2a1 was previously found among various hunter-gatherer groups from Europe.[4,6,12,13,39,40,41] This may indicate the persistence of western hunter-gatherer (WHG) ancestry in at least some Mont-Aimé individuals (2H11).

The apparently high frequency of I2a1 may reflect a general dominance of this haplogroup across Late Neolithic France. It may, however, also suggest communities organized patrilinearly, especially as individual sites showed no mitochondrial haplotypes shared between individuals. To explore this further, we assessed patterns of genetic relatedness at 1,240k autosomal SNP positions[40] using pseudo-diploid calls in READ[42] as well as genotype likelihood in lcMLkin[43] (Data S1G and S1H). This indicated that five individuals analyzed at Mont-Aimé were genetically related. In hypogeum 2, a father and his son (2H10 and 2H17) were related to a lesser degree to a third male (2HxC5x196x1131). They shared similar Y chromosomal haplotypes, confirming patrilinearity. However, we also identified a father-daughter relationship between two individuals anatomically determined as adults (2H11 and 1H06; Data S1A). That at least some women could remain in their native communities implies that no strict patrilocal rules were followed. Alternatively, this could also imply that at least a fraction of women were not involved in marital relationships or had their bodies repatriated to their father's community after death. Extending the analyses to more individuals and integrating isotopic signatures informing on birth location will help assess the extent to which this also applied to other women in the community. Nonetheless, the presence of first-degree relationships between Mont-Aimé I and Mont-Aimé II supports that both hypogea were used strictly contemporaneously and by the same group of people.

Inbreeding and Population Diversity

The fraction of the genome that is comprised within runs of homozygosity (ROHs) can indicate within-kin unions.[12] It can also result in under-estimating genome-wide levels of heterozygosity, which inform on the population effective size.[44] In order to both assess the presence of within-kin unions and estimate population effective sizes at Mont-Aimé, we applied ROHan[44] to the two individuals sequenced to high depth (2H10 and 2H11). The ROH genome fraction was, however, limited (1.5%–2.0%; Data S2A) and on par with that observed among other Neolithic individuals, except at the megalithic site of Newgrange, where a brother-sister incestuous union was previously reported.[12] This indicates limited, if any, within-kin unions at Mont-Aimé. Estimated heterozygosities outside ROHs were comparable to those found in other high-coverage Neolithic genomes, except Newgrange, but larger than those of west European Mesolithic hunter-gatherers (Loschbour and Sramore62; Data S2A). This is in line with Neolithic food production economies allowing population demographic expansion relative to hunting-gathering communities.[45]

DNA Methylation and Phenotypic Inference

Post mortem DNA decay, especially cytosine deamination, is known to have a differential impact on methylated and unmethylated CpGs.[46] This provides an opportunity to map DNA methylation marks on ancient genomes, assuming USER treatment of DNA extracts[47] and sufficient sequencing efforts.[48] We leveraged the two high-coverage genomes characterized here and four others previously reported[4,12] to identify genes that may have experienced important DNA methylation changes during the Mesolithic to Neolithic transition. Analyses were limited to teeth to avoid tissue-specific signatures. We found DNA methylation patterns within gene bodies generally more similar among pairs of Neolithic individuals than among Neolithic-Mesolithic pairs (Figure 2B). However, the Neolithic female (Stuttgart) was found closer to one Mesolithic male (Loschbour) prepared in the same laboratory than to the individual 2H10 prepared in our laboratory. This indicated technical batch effects impacting DNA methylation inference. To avoid these effects and sex-specific signatures, we further restricted the analyses to those male individuals characterized to highest coverage in different laboratories. We identified a total of 611 gene promoters in which DNA methylation shifted by at least 50% on average in both Neolithic individuals (Jerpoint14[12] and 2H11) relative to the Mesolithic individual (Loschbour[4]; Data S3A). Although not significantly enriched in particular functional categories, these included more instances undergoing hyper-methylation during the Neolithic than hypo-methylation ($p = 1.6 \times 10^{-15}$). Examples include C1Qb, the first chain of the humoral immune complement response,[49] and LCK, a key signaling molecule involved in T cell maturation.[50] Further work is required to assess whether such signatures reflect regulatory changes in dental or blood tissues.

We next attempted to predict the age at death of the two individuals sequenced to high coverage following a previously

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

reported approach,[51] leveraging DNA methylation levels inferred within 1 kb around 27,528 CpG positions forming the Illumina 27k methylation Bead Chip system. Although significantly correlated, the DNA methylation values returned by DamMet[48] were consistently lower than those measured on a normalization gold standard panel used for epigenetic age inference (Data S3B).[52] This precluded using the DNAmAge methylation clock predictor to precisely estimate age at death. The latter was instead estimated using DNA methylation levels at 11 clock-CpGs known to undergo DNA methylation changes in teeth with age.[53] Despite cross-validation rates of 0.897–0.939 on modern DNA data, this methodology showed limited precision with ancient DNA data and provided confidence intervals spanning up to 46 years (Data S3C). Pending necessary methodological improvements, this approach suggested that 2H10 potentially died younger than 2H11 (29 years old versus 42 years old).

Although precise phenotypic inference was not possible using DNA methylation patterns, we used the genotypes measured for the two high-coverage individuals at 90 loci to infer morphological, pigmentation, behavioral, and medical phenotypes (Figure 2C; Data S2B). Both individuals were lactose intolerant and were not carriers of the increased risk allele for the celiac disease at rs653178, for increased addiction behavior at rs1800497 and rs2283265, or for elite endurance performance at rs4994. They likely had brown eyes and dark hair (prediction probabilities = 0.844–0.997) but did not carry alleles showing additive effects on hair thickness (rs4752566 and rs3827760). Individual 2H11 was most likely dark skinned (prediction probability = 0.960; Data S2B), although 2H10 did not have a (very) pale but intermediate to dark skin. Interestingly, the two individuals carried different haplotypes at the mono-amine oxidase A locus, which encodes for an enzyme involved in the oxidation of dopamine, norepinephrine, and serotonin neurotransmitters.[54] These entail different combinations of alleles associated with higher (e.g., rs3027407 and rs909525) or lower (e.g., rs1137070, rs1465107, rs2072743, and rs6323) enzymatic production. This, and the existence of gene-environment interactions,[54] make the exact phenotypic consequences difficult to predict. However, the genetic differences observed at this locus suggest different risks of developing aggressive and impulsive behavior in these individuals.[55]

Heterogeneous Mesolithic Ancestry among Late Neolithic Individuals at Mont-Aimé

We next sought to identify genetic affinities between the 18 Late Neolithic individuals sequenced here and ~2,000 ancient individuals previously reported (STAR Methods). Projections of ancient individuals onto the principal components characterizing the genomic variation among 796 modern western Eurasians at 592,998 autosomal loci revealed that almost all Late Neolithic individuals from France clustered along an axis separating Anatolian Neolithic farmers (ANF) and western European hunter-gatherers (WHG), together with Early and Middle Neolithic individuals from France previously analyzed (Figures 3A and S2A). Segregating from the other Late Neolithic individuals along this axis, the father and daughter pair buried at Mont-Aimé (2H11 and 1H06) showed closer genetic affinities with WHG and PEN_B, an Early Neolithic individual from Pendimoun, southeastern France, known to have an excess of WHG ancestry.[8] Using f3-outgroup statistics and f4-statistics,[56] we also found that the ancient individuals genetically closest to 1H06 and 2H11 were hunter-gatherers, mostly from Ireland, Britain, France, Italy, and Luxembourg (Figures 3B–3F, S3A, and S3B). No such affinities were detected in any of the remaining 16 Late Neolithic individuals, for whom genetic affinities were more diffused at the western European scale (Figures 3F, S3A, and S3B). D-statistics[57] of the form (Mbuti, WHG; X, 1H06 or 2H11) also confirmed an excess of WHG ancestry in 1H06 and 2H11, more pronounced in the latter, relative to all other "X" individuals sequenced in this study (Figure S4). D-statistics of the form (Mbuti, ANF; X, 1H06 or 2H11) revealed a reciprocal lack of ANF ancestry in these individuals.

Supervised ADMIXTURE[58] further suggested 57.5% ± 3% and 65.7% ± 2.9% of WHG ancestry within 1H06 and 2H11, respectively (Figures 4B, S2B, and S2C). This contrasts with the 18.5% ± 1.6% to 28.8% ± 1.7% range estimated across the other Late Neolithic individuals reported here. Formal population modeling with qpADM,[56] in which Neolithic individuals are composed of WHG and ANF ancestries, also supported an excess of the former for both 1H06 and 2H11 (50.6% ± 5.1% and 63.3% ± 4.1%, respectively; Figure 4A; Data S4A). This excess shows no equivalent in other Late Neolithic individuals (9.9%–29.9%) and is exceptional among Neolithic individuals from France, except the Early Neolithic PEN_B individual (51.6% ± 4.8%; Figure 4A). For individuals MAS15, ROUQW, TORTC, TORTD, and TORTE, simpler models consisting of ANF ancestry only were supported, confirming WHG ancestry at best limited. However, more complex modeling did not support the presence of GoyetQ2-like ancestry as a second, previously described genetic group of hunter-gatherers present in western Europe (Data S4A).[8]

We next leveraged patterns of linkage disequilibrium (LD) decay in DATES[59] to assess whether the significant excess of WHG ancestry within individuals 2H11 and 1H06 reflects recent or ancient and single or multiple admixture pulses. Modeling all Mont-Aimé individuals, except 2H11 and 1H06, as a combination of ANF and WHG sources, we inferred admixture dates ~39.5 ± 8.5 generations prior to their age distribution (~4,300 years BCE; Figure 4E). Similar modeling indicated that the admixture occurred ~17.2 generations and ~6.2 generations later when considering individuals 1H06 and 2H11 (22.3 ± 4.6 and 33.3 ± 10.4 generations, respectively; i.e., ~4,100 and ~3,800 years BCE). This suggests multiple, sporadic contacts between Neolithic and Mesolithic groups during substantial time periods (at least 5 centuries) and up until ~3,800 BCE. Interestingly, the WHG ancestry estimated within the genomes of 1H06 and her father's (2H11) indicate that her mother had an average 37.9% WHG ancestry, which is larger than the proportion estimated in other Mont-Aimé individuals presently sequenced (average = 21.9%; max = 29.9%). This supports Late Neolithic communities, such as that of Mont-Aimé, as heterogeneous genetically, with WHG ancestors occupying various positions in their genealogies. Importantly, the most recent admixture date inferred here follows by an average of ~400 years the most recent admixture date previously estimated[8] within Middle Neolithic individuals from Esperstedt, Germany. There, archaeological evidence supports interactions between hunter-gatherers and Neolithic groups until 4,200 cal. BCE.[60] As no such evidence is available in northern France after 4,900 cal. BCE,[61] our results

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

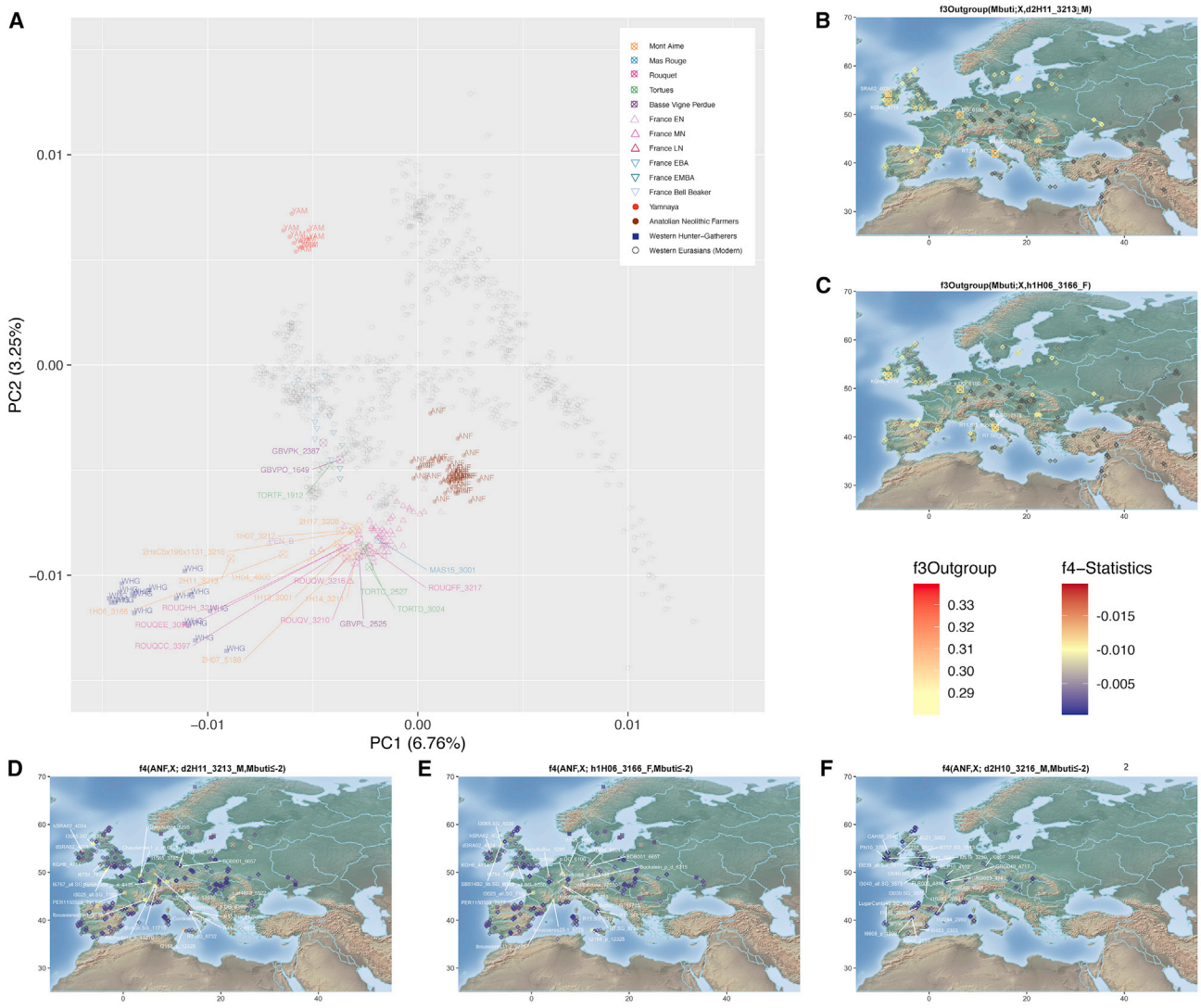

Figure 3. Population Genetic Affinities

(A) Principal-component analysis (PCA). Ancient individuals are projected onto the first two principal components summarizing the genetic variation present in 796 present-day Western Europeans. The fraction of the variance explained is indicated on both axes.

(B) Geographic map of the ancient individuals identified as most genetically related to the Late Neolithic sample 2H11 from Mont-Aimé on the basis of f3-outgroup statistics. Only individuals showing f3-outgroup statistics within the confidence interval of the top individual identified are shown. Crossed squares refer to Mesolithic hunter-gatherers. Other individuals are indicated with open squares if showing significant genetic affinities or with diamonds otherwise. The age of the closest samples is indicated in cal. years BCE.

(C) Same as (B) for the Late Neolithic 1H06_3166 sample from Mont-Aimé.

(D) Geographic map of the ancient individuals identified as most genetically related to the Late Neolithic sample 2H11_3213 from Mont-Aimé on the basis of f4 statistics. Only individuals showing f4 statistics in the form of (ANF,X; d2H11_3213,Mbuti) significantly negative (Z scores ≤ -2) are shown. Labels are only provided for the individuals showing the top 25 most negative f4 statistics for clarity.

(E) Same as (D) for the Late Neolithic h1H06_3166 sample from Mont-Aimé.

(F) Same as (D) for the Late Neolithic d2H10_3216 sample from Mont-Aimé. The prefixes h and d indicate whether the pseudo-diploid or diploid genotype calls were used in the analyses.

See also Figures S2–S4.

may indicate an acculturation process of hunter-gatherer groups following the arrival of farming in the region.

Bell Beaker and Bronze Age Population Changes

In addition to 18 Late Neolithic individuals, our dataset includes four Bell Beaker and two Early Bronze Age individuals from the Grotte des Tortues and Grotte Basse de la Vigne Perdue, both located in the La Clape massif, southern France (Figure 1C). Principal-component analysis (PCA) projections showed that three Bell Beaker individuals overlapped with other Late Neolithic individuals from France, except for the father-daughter pair (2H11 and 1H06; Figure 3A). One Bell Beaker individual (GBVPK) and

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

CellPress

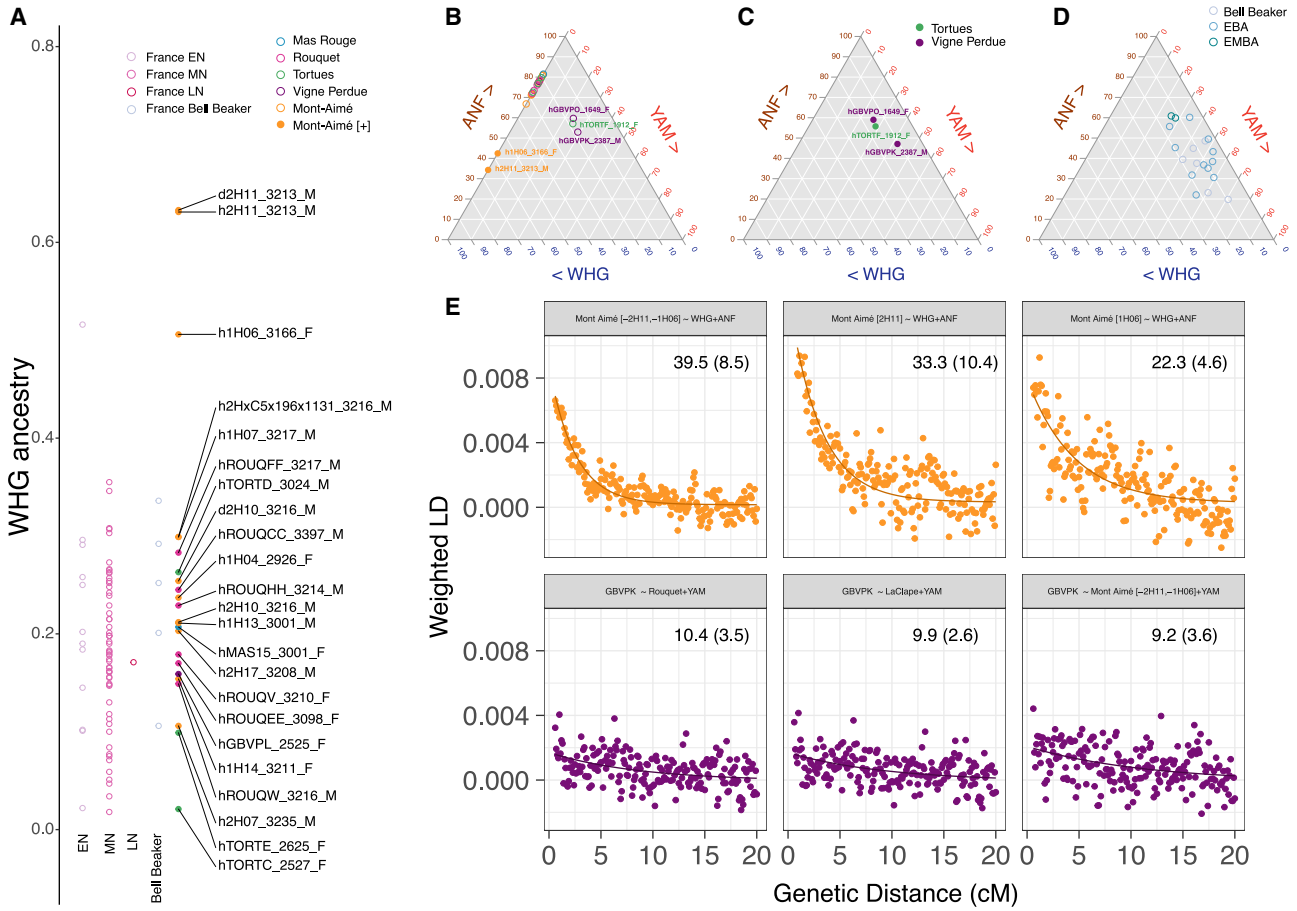

Figure 4. Genetic Ancestry and Admixture Time Estimates

(A) WHG and ANF ancestry proportions estimated in qpADM[56] among Neolithic and Bell Beaker genomes from present-day France. Colors are the same as in the PCA shown in Figure 3A. The prefixes h and d indicate whether the pseudo-diploid or diploid genotype calls were used in the analyses. Suffixes indicate the average radiocarbon date (cal. years BCE) and sex of each individual, respectively. EN, Early Neolithic; MN, Middle Neolithic; LN, Late Neolithic. Mont-Aimé [+] refers to the two individuals showing an excess of WHG ancestry (2H11 and 1H06).

(B) WHG, ANF, and Yamnaya_Samara (YAM) ancestry proportions estimated following supervised ADMIXTURE[58] among the individuals sequenced in this study.

(C) WHG, ANF, and YAM ancestry proportions estimated by qpAdm[56] for those individuals not correctly modeled when considering only WHG and ANF population sources.

(D) Same as (C) for those previously sequenced Neolithic and Bell Beaker genomes from present-day France.

(E) Admixture time estimates obtained using DATES[59]. Top: WHG and ANF admixture times are shown for (1) those individuals showing no excess of WHG ancestry at Mont-Aimé (i.e., all except 2H11 and 1H06; left), (2) the 2H11 individual, and (3) the 1H06 individual. Bottom: Yamnaya-related admixture time estimated using the GBVK genome and various proxies for Late Neolithic populations is shown (left, Grotte du Rouquet; center, sites from the La Clape Massif; right, all samples from Mont-Aimé, except 2H11 and 1H06).

See also Data S4.

the two Early Bronze Age individuals (GBVPO and TORTF), however, clustered with other previously reported Early to Middle Bronze Age individuals from France[13,16] but were displaced along an axis showing closer affinity with Yamnaya steppe herders from Samara.[2,40,59] D-statistics of the form (Mbuti, Yamnaya; X, GBVPK, GBVPO, or TORTF) also suggested a statistically significant excess of Yamnaya ancestry in these three individuals relative to some of the others from both sites (Figure S4). Supervised ADMIXTURE and formal three-way population modeling with qpADM confirmed the presence of ~23.6% ± 1.4% to 42.1% ± 4.7% of Yamnaya-related ancestry among these three genomes, in addition to the WHG and ANF ancestries (Figure 4C; Data S4B). This is in line with the ~25.9%–54.8% Yamnaya ancestry inferred in the genomes of Bell Beaker and Early-to-Middle Bronze Age individuals from France (Figure 4D), including individuals excavated from sites located within a 100-km distance from the La Clape massif (Quinquiris and RecDeLigno).[16]

Patterns of LD decay provided an estimate for the time when steppe-derived genetic ancestry entered southern France at around ~9.2 ± 3.6 to 10.4 ± 3.5 generations prior to the age of the earliest individual showing Yamnaya ancestry (i.e., GBVPK; Figure 4E). This corresponds to ~2,650 years BCE, which is earlier than previous reports showing significant steppe-derived ancestry in Britain ~2,400 years BCE[9] and in Iberia ~2,500–2,000 years BCE.[13] Our estimate was, however, found

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

consistent, regardless of the Late Neolithic proxy considered in population modeling, including the Mont-Aimé individuals showing least WHG admixture, the individuals from Grotte du Rouquet, or those from the La Clape massif (Figure 4E). Our data also reveal the co-existence of individuals with no to significant Yamnaya-related ancestry within Grotte Basse de la Vigne Perdue. This and the absence of Yamnaya ancestry inferred in one contemporary Bell Beaker individual from the nearby site from Dolmen des Peirières (individual PEI2)[16] support a complex genetic composition of Bell Beaker groups in southern France involving non-ubiquitous admixture with individuals related to steppe herders. This is in line with previous reports in Central Europe but contrasts to Britain, where ~90% of the local gene pool was replaced by steppe-related ancestry by the second half of the 3rd millennium BCE.[9]

In this study, we recovered ancient genomes from 24 ancient human individuals in burial sites from present-day France to document genomic changes in Late Neolithic to Early Bronze Age populations. Bell Beaker groups from southwestern France included some individuals, but not all, showing steppe-derived genetic ancestry, with an earliest appearance of Yamnaya-related steppe ancestry around ~2,650 years BCE. Late Neolithic genomes from Mont-Aimé provide evidence of multiple, sporadic contacts between Neolithic and Mesolithic groups at a time when no Mesolithic cultural evidence are available. This may indicate that Mesolithic groups survived at least until the first quarter of the 4th millennium BCE in the region but followed an acculturation process.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Archaeological and anthropological information
  - Mont-Aimé I and II sites
  - Mas Rouge site
  - La Clape sites
  - Grotte des Tortues
  - Grotte Basse de la Vigne Perdue
  - Grotte du Rouquet
- METHOD DETAILS
  - Radiocarbon dating
  - DNA extraction and sequencing
  - Read processing and alignment
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Post-mortem damage
  - Error rates
  - Uniparental markers and Contamination estimates
  - Phenotype prediction
  - Kinship analyses
  - Inbreeding and diversity estimates
  - Principal Component Analysis
  - f3-Outgroup and f4-statistics
  - Admixture
  - D-statistics
  - Population modeling
  - Admixture dating
  - Microbial profiling
  - DNA methylation
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.12.015.

ACKNOWLEDGMENTS

We thank Loreleï Chauvey, Stéphanie Schiavinato, and Laure Tonasso-Calvière for managing the ancient DNA lab facilities in Toulouse and running the MiniSeq sequencing instrument. We thank Prof. José Braga, Nancy Saenz-Oyhéréguy, and all members of the AGES research group for discussions. The authors are grateful to Flore Colette (curator of the Palais-Musée des Archevêques, Narbonne), as well as the Service Régional de l'Archéologie Champagne-Ardennes, the Direction Régionale des Affaires Culturelles du Grand Est, and the Institut National de Recherches Archéologiques Préventives for providing access to archaeological material. The authors acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure (NGI) in Sweden, the Knut and Alice Wallenberg Foundation, and UPPMAX for providing assistance in massively parallel DNA sequencing and computational infrastructure. This project is part of a "Collaboration CEA/JACOB/CNRGH - Unité Mixte de Recherches CNRS 5288." This project has received funding from the European Union's Horizon 2020 research and Innovation programme under the Marie Skłodowska-Curie grant agreement nos. 795916-NEO and 748122-ELITE; the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) programme, grant ANR-17-EURE-0010; the L'Oréal-UNESCO for Women in Science Young Talent France 2019 programme; the Villum Fonden miGENEPI research project; the ANR LifeChange; the CNRS MITI "Défi Ecologie de la Santé 2020" programme; the Simone & Cino Del Duca Foundation (Subventions scientifiques 2020, HealthTimeTravel); and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 681605).

AUTHOR CONTRIBUTIONS

A.S.-O., J.G., and L.O. conceived the project and designed research. R.D., C.M., Y.T., E.C., and J.G. provided samples and information about archaeological context. A.S.-O. carried out ancient DNA laboratory work, with input from C.D.S. and C.T. L.O. carried out computational analyses, with input from A.S.-O. and C.D.S. J.S. carried out radiocarbon dating, with input from B.S. J.-F.D., B.S., L.D., and L.O. provided reagents and material. A.S.-O. and L.O. wrote the Supplemental Information and STAR Methods, with input from C.D.S. L.O. wrote the paper, with input from A.S.-O., C.D.S., and all coauthors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 14, 2020
Revised: December 11, 2020
Accepted: December 11, 2020
Published: January 11, 2021

REFERENCES

1. Lazaridis, I. (2018). The evolutionary history of human populations in Europe. Curr. Opin. Genet. Dev. *53*, 21–27.

2. Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

3. Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. Nature *522*, 167–172.

4. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature *513*, 409–413.

5. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N., Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal complex genetic history of early European farmers. Nature *551*, 368–372.

6. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history of southeastern Europe. Nature *555*, 197–203.

7. Mittnik, A., Wang, C.C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R., Khartanovich, V., Moiseyev, V., Tõrv, M., et al. (2018). The genetic prehistory of the Baltic Sea region. Nat. Commun. *9*, 442.

8. Rivollat, M., Jeong, C., Schiffels, S., Küçükkalıpçı, İ., Pemonge, M.H., Rohrlach, A.B., Alt, K.W., Binder, D., Friederich, S., Ghesquière, E., et al. (2020). Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. Sci. Adv. *6*, eaaz5344.

9. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. Nature *555*, 190–196.

10. Brace, S., Diekmann, Y., Booth, T.J., van Dorp, L., Faltyskova, Z., Rohland, N., Mallick, S., Olalde, I., Ferry, M., Michel, M., et al. (2019). Ancient genomes indicate population replacement in Early Neolithic Britain. Nat. Ecol. Evol. *3*, 765–771.

11. Cassidy, L.M., Martiniano, R., Murphy, E.M., Teasdale, M.D., Mallory, J., Hartwell, B., and Bradley, D.G. (2016). Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc. Natl. Acad. Sci. USA *113*, 368–373.

12. Cassidy, L.M., Maoldúin, R.Ó., Kador, T., Lynch, A., Jones, C., Woodman, P.C., Murphy, E., Ramsey, G., Dowd, M., Noonan, A., et al. (2020). A dynastic elite in monumental Neolithic society. Nature *582*, 384–388.

13. Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C.J., Gandini, F., Pala, M., et al. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. Science *363*, 1230–1234.

14. Fernandes, D.M., Mittnik, A., Olalde, I., Lazaridis, I., Cheronet, O., Rohland, N., Mallick, S., Bernardos, R., Broomandkhoshbacht, N., Carlsson, J., et al. (2020). The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. Nat. Ecol. Evol. *4*, 334–345.

15. Marcus, J.H., Posth, C., Ringbauer, H., Lai, L., Skeates, R., Sidore, C., Beckett, J., Furtwängler, A., Olivieri, A., Chiang, C.W.K., et al. (2020). Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. Nat. Commun. *11*, 939.

16. Brunel, S., Bennett, E.A., Cardin, L., Garraud, D., Barrand Emam, H., Beylier, A., Boulestin, B., Chenal, F., Ciesielski, E., Convertini, F., et al. (2020). Ancient genomes from present-day France unveil 7,000 years of its demographic history. Proc. Natl. Acad. Sci. USA *117*, 12791–12798.

17. Guilaine, J. (1976-1977). Le Néolithique, le Chalcolithique et l'Âge du Bronze. Cah. Ligures Préhist. Archéol. *25-26*, 285–288.

18. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Chang, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon *62*, 725–757.

19. Louvel, G., Der Sarkissian, C., Hanghøj, K., and Orlando, L. (2016). metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data. Mol. Ecol. Resour. *16*, 1415–1427.

20. Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., and Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods *12*, 902–903.

21. Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., Gabbard, J.L., et al. (2017). Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. *45* (D1), D535–D542.

22. Warinner, C., Rodrigues, J.F., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R.Y., Fiddyment, S., et al. (2014). Pathogens and host immunity in the ancient human oral cavity. Nat. Genet. *46*, 336–344.

23. Eerkens, J.W., Nichols, R.V., Murray, G.G.R., Perez, K., Murga, E., Kaijankoski, P., Rosenthal, J.S., Engbring, L., and Shapiro, B. (2018). A probable prehistoric case of meningococcal disease from San Francisco Bay: Next generation sequencing of Neisseria meningitidis from dental calculus and osteological evidence. Int. J. Paleopathol. *22*, 173–180.

24. Mann, A.E., Sabin, S., Ziesemer, K., Vågene, A.J., Schroeder, H., Ozga, A.T., Sankaranarayanan, K., Hofman, C.A., Fellows Yates, J.A., Salazar-García, D.C., et al. (2018). Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. Sci. Rep. *8*, 9822.

25. Willmann, C., Mata, X., Hanghoej, K., Tonasso, L., Tisseyre, L., Jeziorski, C., Cabot, E., Chevet, P., Crubézy, E., Orlando, L., et al. (2018). Oral health status in historic population: macroscopic and metagenomic evidence. PLoS ONE *13*, e0196482.

26. Jensen, T.Z.T., Niemann, J., Iversen, K.H., Fotakis, A.K., Gopalakrishnan, S., Vågene, A.J., Pedersen, M.W., Sinding, M.S., Ellegaard, M.R., Allentoft, M.E., et al. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. Nat. Commun. *10*, 5520.

27. Velsko, I.M., Fellows Yates, J.A., Aron, F., Hagan, R.W., Frantz, L.A.F., Loe, L., Martinez, J.B.R., Chaves, E., Gosden, C., Larson, G., and Warinner, C. (2019). Microbial differences between dental plaque and historic dental calculus are related to oral biofilm maturation stage. Microbiome *7*, 102.

28. Bravo-Lopez, M., Villa-Islas, V., Rocha Arriaga, C., Villaseñor-Altamirano, A.B., Guzmán-Solís, A., Sandoval-Velasco, M., Wesp, J.K., Alcantara, K., López-Corral, A., Gómez-Valdés, J., et al. (2020). Paleogenomic insights into the red complex bacteria *Tannerella forsythia* in Pre-Hispanic and Colonial individuals from Mexico. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190580.

29. Eisenhofer, R., Kanzawa-Kiriyama, H., Shinoda, K.I., and Weyrich, L.S. (2020). Investigating the demographic history of Japan using ancient oral microbiota. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190578.

30. Fagernäs, Z., García-Collado, M.I., Hendy, J., Hofman, C.A., Speller, C., Velsko, I., and Warinner, C. (2020). A unified protocol for simultaneous extraction of DNA and proteins from archaeological dental calculus. J. Arch. Sci. *118*, 105135.

31. Fotakis, A.K., Denham, S.D., Mackie, M., Orbegozo, M.I., Mylopotamitaki, D., Gopalakrishnan, S., Sicheritz-Pontén, T., Olsen, J.V., Cappellini, E., Zhang, G., et al. (2020). Multi-omic detection of *Mycobacterium leprae* in archaeological human dental calculus. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190584.

32. Jacobson, D.K., Honap, T.P., Monroe, C., Lund, J., Houk, B.A., Novotny, A.C., Robin, C., Marini, E., and Lewis, C.M., Jr. (2020). Functional diversity of microbial ecologies estimated from ancient human coprolites and dental calculus. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190586.

33. Modi, A., Pisaneschi, L., Zaro, V., Vai, S., Vergata, C., Casalone, E., Caramelli, D., Moggi-Cecchi, J., Mariotti Lippi, M., and Lari, M. (2020). Combined methodologies for gaining much information from ancient dental calculus: testing experimental strategies for simultaneously analysing DNA and food residues. Archaeol. Anthropol. Sci. *12*, 10.

34. Neukamm, J., Pfrengle, S., Molak, M., Seitz, A., Francken, M., Eppenberger, P., Avanzi, C., Reiter, E., Urban, C., Welte, B., et al. (2020). 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals. BMC Biol. *18*, 108.

35. Achtman, M., and Zhou, Z. (2020). Metagenomics of the modern and historical human oral microbiome with phylogenetic studies on *Streptococcus mutans* and *Streptococcus sobrinus*. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190573.

36. Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., Bradshaw, C.J.A., Townsend, G., Sołtysiak, A., Alt, K.W., et al. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat. Genet. *45*, 450–455, 455e1.

37. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20130624.

38. Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H.A., Siebauer, M., Lachmann, M., and Pääbo, S. (2009). The Neandertal genome and ancient DNA authenticity. EMBO J. *28*, 2494–2502.

39. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nat. Commun. *5*, 5257.

40. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499–503.

41. Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınç, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: investigating early postglacial migration routes and high-latitude adaptation. PLoS Biol. *16*, e2003703.

42. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS ONE *13*, e0195491.

43. Lipatov, M., Sanjeev, K., Patro, R., and Veeramah, K. (2015). Maximum likelihood estimation of biological relatedness from low coverage sequencing data. bioRxiv. https://doi.org/10.1101/023374.

44. Renaud, G., Hanghøj, K., Korneliussen, T.S., Willerslev, E., and Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. Genetics *212*, 587–614.

45. Bocquet-Appel, J.-P. (2011). When the world's population took off: the springboard of the Neolithic demographic transition. Science *333*, 560–561.

46. Pedersen, J.S., Valen, E., Velazquez, A.M., Parker, B.J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D.J., Kelly, T.K., Vang, S., et al. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Res. *24*, 454–466.

47. Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S., Meshorer, E., and Carmel, L. (2014). Reconstructing the DNA methylation maps of the Neandertal and the Denisovan. Science *344*, 523–527.

48. Hanghøj, K., Renaud, G., Albrechtsen, A., and Orlando, L. (2019). DamMet: ancient methylome mapping accounting for errors, true variants, and post-mortem DNA damage. Gigascience *8*, giz025.

49. Kishore, U., and Reid, K.B. (2000). C1q: structure, function, and receptors. Immunopharmacology *49*, 159–170.

50. Gaud, G., Lesourne, R., and Love, P.E. (2018). Regulatory mechanisms in T cell receptor signalling. Nat. Rev. Immunol. *18*, 485–497.

51. Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J.S., Willerslev, E., and Orlando, L. (2016). Fast, accurate and automatic ancient nucleosome and methylation maps with epiPALEOMIX. Mol. Biol. Evol. *33*, 3284–3298.

52. Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biol. *14*, R115.

53. Bekaert, B., Kamalandua, A., Zapico, S.C., Van de Voorde, W., and Decorte, R. (2015). Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. Epigenetics *10*, 922–930.

54. Nilsson, K.W., Åslund, C., Comasco, E., and Oreland, L. (2018). Gene-environment interaction of monoamine oxidase A in relation to antisocial behaviour: current and future directions. J. Neural Transm. (Vienna) *125*, 1601–1626.

55. Dorfman, H.M., Meyer-Lindenberg, A., and Buckholtz, J.W. (2014). Neurobiological mechanisms for impulsive-aggression: the role of MAOA. Curr. Top. Behav. Neurosci. *17*, 297–313.

56. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

57. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. *28*, 2239–2252.

58. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

59. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.

60. Rowley-Conwy, P. (2011). Westward ho!: the spread of agriculture from Central Europe to the Atlantic. Curr. Anthropol. *52*, S431–S451.

61. Marchand, G., and Perrin, T. (2017). Why this revolution? Explaining the major technical shift in Southwestern Europe during the 7th millennium cal. BC. Quat. Int. *428*, 73–85.

62. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.

63. Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat. Protoc. *9*, 1056–1082.

64. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.

65. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.

66. Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., and Kronenberg, F. (2011). HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum. Mutat. *32*, 25–32.

67. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

68. Ralf, A., Montiel González, D., Zhong, K., and Kayser, M. (2018). Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. Mol. Biol. Evol. *35*, 1291–1294.

69. Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., Sijen, T., Knijff, P., Liu, F., Branicki, W., et al. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. Forensic Sci. Int. Genet. *35*, 123–135.

70. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.

71. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.

72. Gokhman, D., Kelman, G., Amartely, A., Gershon, G., Tsur, S., and Carmel, L. (2017). Gene ORGANizer: linking genes to the organs they affect. Nucleic Acids Res. *45* (W1), W138–W145.

73. Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *41*, D377–D386.

74. Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

75. Chertier, B. (1983). Informations archéologiques "Champagne-Ardenne". Gallia Préhistoire *26*, 388–389. https://www.persee.fr/doc/galip_0016-4127_1983_num_26_2_1725.

76. Donat, R., Mokrane, F.-Z., Rousseau, H., Dedouit, F., Telmon, N., and Crubézy, É. (2019). The antiquity of the spondyloarthritides: presentation of one of the oldest Neolithic cases in Western Europe. Int. J. Paleopathol. *24*, 229–235.

77. Langry-François, F. (2004). Le mobilier lithique des sépultures en hypogée du département de la Marne. Anthropol. et Praehist. *115*, 91–102.

78. Crubézy, E., and Mazière, G. (1990). L'hypogée II du Mont-Aimé à Val-Des-Marais (Marne). Premiers résultats. Bulletin de la Société d'Archéologie Champenoise *83*, 65–78.

79. Donat, R. (2014). Le mobilier de l'hypogée 2 du Mont-Aimé au Val-des-Marais (Marne) dans son cadre régional: nouvelles données. In La Fin du IVe Millénaire dans le Bassin Parisien. Le Néolithique Récent entre Seine, Oise et Marne (3500-2900 avant notre ère), R. Cottiaux, and L. Salanova, eds. (Revue Archéologique de l'Est), pp. 389–410.

80. Augereau, A., Brunet, P., Costa, L., Cottiaux, R., Hamon, T., Ihuel, E., Langry-Francois, F., Magne, P., Maingaud, A., Mallet, N., et al. (2007). Le Néolithique recent dans le Centre-Nord de la France (3400/3300-2800/2700 av. J.-C.): l'avenir du Seine-Oise-Marne en question. In Un Siècle de Construction du Discours Scientifique en Préhistoire. Congrès du Centenaire de la Société Préhistorique Française, J. Evin, ed. (Bull. SPF), pp. 165–184.

81. Renard, C., Polloni, A., Sohn, M., Ihuel, E., Langry-François, F., Magne, P., Maingaud, A., Martineau, R., and Salanova, L. (2014). La collection du baron Joseph de Baye au musée d'archéologie nationale. In La Fin du IVe Millénaire dans le Bassin Parisien. Le Néolithique Récent entre Seine, Oise et Marne (3500-2900 avant notre ère), R. Cottiaux, and L. Salanova, eds. (Revue Archéologique de l'Est), pp. 313–371.

82. Schmidt, C. (2016). Estimating age, sex, and individual ID from teeth. In A Companion to Dental Anthropology, J. Irish, D. Scott, and G. Richard, eds. (Wiley Blackwell), pp. 362–376.

83. Crubézy, E., Braga, J., and Larrouy, G. (2008). Abrégé d'Anthropobiologie et d'Évolution Humaine (Masson).

84. Kim, Y.K., Kho, H.S., and Lee, K.H. (2000). Age estimation by occlusal tooth wear. J. Forensic Sci. *45*, 303–309.

85. Brabant, H., and Sahly, A. (1962). La paléostomatologie en Belgique et en France. Acta Stomatol. Belg. *59*, 355–385.

86. Duchesne, S., and Crubézy, E. (2015). Les Cimetières du Haut Moyen Âge en Languedoc (Presses Universitaires de Perpignan).

87. Esclassan, R., Grimoud, A.M., Ruas, M.P., Donat, R., Sevin, A., Astie, F., Lucas, S., and Crubézy, E. (2009). Dental caries, tooth wear and diet in an adult medieval (12^{th}-14^{th} century) population from mediterranean France. Arch. Oral Biol. *54*, 287–297.

88. Becker, M.J. (1998). A Roman "implant" reconsidered. Nature *394*, 534.

89. Braga, J., Heuze, Y., Chabadel, O., Sonan, N.K., and Gueramy, A. (2005). Non-adult dental age assessment: correspondence analysis and linear regression versus Bayesian predictions. Int. J. Legal Med. *119*, 260–274.

90. Tchérémissinoff, Y., Leal, E., Breuil, J.-Y., and Martinez, C. (2018). Montpellier – Mas Rouge – La Cavalade. Une sépulture collective du Néolithique en architecture semi-enterrée. Inrap Méditerranée Rapport d'opération 9294.

91. Tchérémissinoff, Y., and Seguin, M. (2019). Mise en place d'un protocole photogrammétrie et SIG dès la fouille préventive: la sépulture collective de Mas Rouge à Montpellier (Hérault). Préhist. Médit. *7*, 17.

92. Héléna, P. (1925). Les grottes sépulcrales des Monges à Narbonne. Bulletin de la Commission Archéologique de Narbonne *16*, 66–73, 193–240.

93. Riquet, R. (1962). Les crânes préhistoriques de la collection Héléna (Narbonne). BMSAP *11*, 480–522.

94. Ramsey, C.B. (2009). Bayesian analysis of radiocarbon dates. Radiocarbon *51*, 337–360.

95. Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Sci. Rep. *5*, 11184.

96. Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., et al. (2019). Tracking five millennia of horse management with extensive ancient genome time series. Cell *177*, 1419–1435.e31.

97. Boessenkool, S., Hanghøj, K., Nistelberger, H.M., Der Sarkissian, C., Gondek, A.T., Orlando, L., Barrett, J.H., and Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. Mol. Ecol. Resour. *17*, 742–751.

98. Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nat. Protoc. *13*, 2447–2461.

99. Gaunitz, C., Fages, A., Hanghøj, K., Albrechtsen, A., Khan, N., Schubert, M., Seguin-Orlando, A., Owens, I.J., Felkel, S., Bignon-Lau, O., et al. (2018). Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science *360*, 111–114.

100. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

101. Poullet, M., and Orlando, L. (2020). Assessing DNA sequence alignment methods for characterizing ancient genomes and methylomes. Front. Ecol. Evol. *8*, 105.

102. Picard Toolkit (2019). Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/.

103. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

104. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature *499*, 74–78.

105. Meyer, M., Kircher, M., Gansauge, M.T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.

106. Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-Del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. Science *353*, 499–503.

107. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe directly descended from Neolithic Aegeans. Proc. Natl. Acad. Sci. USA *113*, 6886–6891.

108. Scheib, C.L., Hui, R., D'Atanasio, E., Wohns, A.W., Inskip, S.A., Rose, A., Cessford, C., O'Connell, T.C., Robb, J.E., Evans, C., et al. (2019). East Anglian early Neolithic monument burial linked to contemporary Megaliths. Ann. Hum. Biol. *46*, 145–149.

109. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. Science *334*, 94–98.

110. Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The demographic development of the first farmers in Anatolia. Curr. Biol. *26*, 2659–2666.

111. Feldman, M., Master, D.M., Bianco, R.A., Burri, M., Stockhammer, P.W., Mittnik, A., Aja, A.J., Jeong, C., and Krause, J. (2019). Ancient DNA sheds light on the genetic origins of early Iron Age Philistines. Sci. Adv. *5*, eaax0061.

112. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature *538*, 201–206.

113. Mondal, M., Casals, F., Xu, T., Dall'Olio, G.M., Pybus, M., Netea, M.G., Comas, D., Laayouni, H., Li, Q., Majumder, P.P., and Bertranpetit, J. (2016). Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. Nat. Genet. *48*, 1066–1070.

114. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

115. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229.

116. Gallego Llorente, M., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., et al. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science *350*, 820–822.

117. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

118. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

119. Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

120. Villalba-Mouco, V., van de Loosdrecht, M.S., Posth, C., Mora, R., Martínez-Moreno, J., Rojo-Guerra, M., Salazar-García, D.C., Royo-Guillén, J.I., Kunst, M., Rougier, H., et al. (2019). Survival of Late Pleistocene hunter-gatherer ancestry in the Iberian Peninsula. Curr. Biol. *29*, 1169–1177.e7.

121. Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., and Caporaso, J.G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. USA *109*, 21390–21395.

122. Human Microbiome Project Consortium (2012). A framework for human microbiome research. Nature *486*, 215–221.

123. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., and Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. USA *104*, 14616–14621.

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
osteological remain	this study	1H04
osteological remain	this study	1H06
osteological remain	this study	1H07
osteological remain	this study	1H13
osteological remain	this study	1H14
osteological remain	this study	2H06
osteological remain	this study	2H07
osteological remain	this study	2H10
osteological remain	this study	2H11
osteological remain	this study	2H17
osteological remain	this study	2HxC5x196
osteological remain	this study	2HxC5x1131
osteological remain	this study	GBVPK
osteological remain	this study	GBVPL
osteological remain	this study	GBVPO
osteological remain	this study	MAS15
osteological remain	this study	ROUQCC
osteological remain	this study	ROUQEE
osteological remain	this study	ROUQFF
osteological remain	this study	ROUQHH
osteological remain	this study	ROUQV
osteological remain	this study	ROUQW
osteological remain	this study	TORTC
osteological remain	this study	TORTD
osteological remain	this study	TORTE_
osteological remain	this study	TORTF
Chemicals, Peptides, and Recombinant Proteins		
N-Lauroylsarcosine solution 30% 500ml	Dutscher	Cat# 348533
Proteinase K 100MG	Thermo Fisher Scientific	Cat# 10103533
H2O, Molecular Biology Grade, Fisher BioReagents	Thermo Fisher Scientific	Cat# 10490025
Tween 20 100ML	Thermo Fisher Scientific	Cat# 10113103
Ethanol, Absolute, Mol Biology Grade	Thermo Fisher Scientific	Cat# 10644795
5M Sodium Chloride 100ML	Thermo Fisher Scientific	Cat# 10609823
guanidine hydrochloride	Sigma-Aldrich	Cat# G3272
2-Propanol	Merck	Cat# 109634
Silica magnetic beads	G-Biosciences	Cat# 786-915
USER Enzyme	New England Biolabs	Cat# M5505L
NEBNext End Repair Module	New England Biolabs	Cat# E6050L
Bst DNA Polymerase	New England Biolabs	Cat# M0275L
NEBNext Quick Ligation Module	New England Biolabs	Cat# E6056L
BSA Molecular Biology Grade	New England Biolabs	Cat# B9000S
ACCUPRIME PFX DNA POLYMERASE 100mL	Thermo Fisher Scientific	Cat# 10472482
Agencourt AMPure XP - 60ml	Beckman Coulter	Cat# A63881
Buffer PE	QIAGEN	Cat# 19065

(*Continued on next page*)

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Buffer PB	QIAGEN	Cat# 19066
Buffer EB	QIAGEN	Cat# 19086
DMSO molecular biol 250ML	Thermo Fisher Scientific	Cat# 10397841
dNTP Set 100mM 100mL	Thermo Fisher Scientific	Cat# 10336653
EDTA 0.5M pH 8.0 Fisher Bioreagents 500ML	Thermo Fisher Scientific	Cat# 10182903
Tris HCl, 1M, pH 8.0, 100ML	Thermo Fisher Scientific	Cat# 10336763
Critical Commercial Assays		
MinElute PCR Purification kit	QIAGEN	Cat# 28006
Tapestation screenTape D1000 HS	Agilent	Cat# 5067-5584
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat# Q32854
MiniSeq High Output Reagent Kit (150-cycles)	Illumina	Cat# FC-420-1002
Deposited Data		
Raw and analyzed data	this study	ENA: PRJEB41240
Software and Algorithms		
metaBIT	[19]	https://bitbucket.org/Glouvel/metabit/src/master/
READ	[42]	https://bitbucket.org/tguenther/read/src/master/
lcMLkin	[43]	https://github.com/COMBINE-lab/maximum-likelihood-relatedness-estimation
ROHan	[44]	https://github.com/grenaud/rohan
DamMet	[48]	https://gitlab.com/KHanghoj/DamMet
Admixtools	[56]	https://github.com/DReichLab/AdmixTools
ADMIXTURE	[58]	http://dalexander.github.io/admixture/index.html
DATES	[59]	https://github.com/priyamoorjani/DATES
AdapterRemoval2	[62]	https://github.com/MikkelSchubert/adapterremoval
Paleomix	[63]	https://github.com/MikkelSchubert/paleomix
mapDamage2	[64]	https://ginolhac.github.io/mapDamage
ANGSD	[65]	https://github.com/ANGSD/angsd
haplogrep	[66]	https://github.com/seppinho/haplogrep-cmd
Schmutzi	[67]	https://github.com/grenaud/schmutzi
Yleaf	[68]	https://github.com/genid/Yleaf
HIrisPlex-S	[69]	https://hirisplex.erasmusmc.nl/
smartPCA	[70]	https://github.com/DReichLab/EIG/tree/master/POPGEN
bamUtil	[71]	https://github.com/statgen/bamUtil
Gene Organizer	[72]	http://geneorganizer.huji.ac.il/
Panther	[73]	http://www.pantherdb.org/
David	[74]	https://david.ncifcrf.gov/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources, material and reagents should be addressed and will be fulfilled by the lead contact, Andaine Seguin-Orlando (andaine.seguin@univ-tlse3.fr).

Materials Availability

Raw sequence data and alignments are available at the European Nucleotide Archive (ENA) under accession number ENA: PRJEB41240.

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Data and Code Availability

The accession number for the genomic data reported in this paper is ENA: PRJEB41240. All other previously published genomic data used in this study is available at the sources referenced in the QUANTIFICATION AND STATISTICAL ANALYSIS section. The code used in this study is listed in the Key Resources Table.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Archaeological and anthropological information

Excavation inventory IDs are indicated in Data S1I for each sample analyzed in this study. The 27 remains that were not sequenced further but only screened for assessing the proportion of human DNA present in DNA extracts are not presented. They include three postcranial bones and eight teeth from the Mont-Aimé sites (endogenous content = 0.07%–8.27%), four petrosal bones from Grotte Basse de la Vigne Perdue (endogenous content = 0.01%–9.72%), three petrosal bones from Grotte des Tortues (endogenous content = 0.15%–6.52%), four petrosal bones from Grotte du Rouquet (endogenous content = 0.47%–5.24%) and five petrosal bones from Mas Rouge (endogenous content = 0.14%–3.26%).

Mont-Aimé I and II sites

The two Mont-Aimé hypogea are part of a wide complex of collective burials identified in the Marne region, France and including more than 160 artificial cavities, all organized in necropolises. They remain among the very few that were not emptied during early excavations by the end of the 19th century. Dug into Late Cretaceous chalk, the Mont-Aimé hypogea I and II are located only 30 m from one another, on the Southern side of the butte of Mont-Aimé. Both burials are oriented North-to-South. They open to the South, and show a very similar architecture consisting in an access corridor, an antechamber and a main funerary chamber of 16 to 17 m^2, split in two rooms by a lateral pillar. Within both burials, the skeletons were completely dislocated and dispersed.

The Mont-Aimé hypogeum I (lat. 48.858874, long. 3.996466) was discovered and excavated in 1982 (Champagne-Ardenne Regional Archaeological Service;[75]) and contained the skeletal remains of at least 57 individuals (Minimum Number of Individuals MNI, 25 adults, 32 children and adolescents[76]), as well as 272 artifacts consisting predominantly in shell ornaments, but also local flint arrowheads, bone and antlers artifacts and limestone beads[77].

The Mont-Aimé hypogeum II (lat. 48.858472, long. 3.995887) was excavated in 1988 and 1989[78] and included the remains of at least 57 individuals (41 adults and adolescents over 15 years old, 16 children,[79]). Many of the craniofacial blocks are missing and were most likely removed from the burial after decomposition, as both mandibles and atlas were well represented. To limit sampling bias between hypogea, we sampled teeth included in mandibles. In addition, two rib fragments 2HxC5x196 and 2HxC5x1131 were selected for DNA analysis as they showed cortical lesions compatible with tuberculosis infection. Grave goods were abundant (253 items) and included two axes in exogenous material, local flint arrowheads, bone and antler artifacts, as well as shell ornaments[79].

Published radiocarbon dates suggested that the communities using both hypogea as burials could be at least partially contemporaneous, and confirmed that the Mont-Aimé hypogea are among the earliest collective burials in the region, marking the transition between Middle and Late Neolithic in the Paris Basin. Artifacts found in both hypogea showed similar typology and technological characteristics, indicative of early and intermediate phases of the Late Neolithic period in the region[79–81]. In particular, the ornaments presented similarities with those from the Horgen culture, observed in French Jura and in Switzerland[79–81], suggesting exchanges between the two localities.

Estimating an age-at-death on the basis of tooth wear in archaeological individuals[82] is challenging due to the large variability in tooth wear, which increases with age[83] and is influenced by sex[84]. Difficulties also arise in absence of reference population (of known age and degree of wear), as is the case of European Neolithic populations. We therefore used a simple method[85], which is widely applied to populations from continental Europe[86,87], and minimizes intra- and inter-observer variation[88]. Teeth were classified according to five increasing tooth wear stages from 0 to 4. For infants and sub-adults, age estimation was performed according to teeth eruption[89]). Determined dental ages are summarized in Data S1A.

Mas Rouge site

The Neolithic collective burial known as "Mas Rouge" (lat. 43.594016, long. 3.91929), located near Montpellier, Hérault, France, was excavated in 2013-2014 by Y. Tchérémissinoff and colleagues[90]. The excavation was part of a preventive archaeology program linked to railway and highway construction works and benefitted from extensive Geographic Information System documentation[91]. The burial, on the edge of the contemporaneous La Cavalade village (Late Neolithic, Ferrières culture) is integrated into a fairly dense network of Neolithic occupations. The geographical context is that of a coastal plain in the proximity of a system of salt ponds. The burial consisted in a domestic cellar (6 m x 4 m x 1.7 m) whose elaborate internal architecture has been modified to host the remains of at least 189 individuals (91 adults and adolescents and 98 immature individuals, of all age classes). The funerary architectural rebuilding involved the implementation of complex infra- and superstructures (coated wooden floors, elevations in mud-bricks), which then formed a semi-subterranean burial, the emerged part of which constitutes a small, low-rise house with a light roof. The associated grave goods are not particularly prestigious, and mainly consist of necklace beads and everyday tools, as expected for the period. In addition to the radiocarbon date provided here (see below), a total of 14 radiocarbon dates were published in the excavation report, indicating that Mas Rouge was used as a cellar during the 3,295-2,945 cal. years BCE interval and as a collective burial between 3,035 and 2,890 cal. years BCE[90]. The sex and age-at-death anthropological determination, combined with the

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

elevated MNI, the narrow period of burial activity and the close proximity of the La Cavalade habitat are compatible with the hypothesis of Mas Rouge being used as a community burial for the vast majority of the villagers over four to five generations.

La Clape sites

The limestone La Clape massif, located between Narbonne and the Mediterranean Sea (Aude, France), hosts more than a dozen of natural caves used as collective burials during the Late Neolithic and the Early Bronze Age, many of which were excavated during the first half of the 20th century by Théophile and Philippe Héléna[92,93]. The remains are now part of the so-called 'Héléna' collection at the Archaeological Museum of Narbonne, France. Individuals from three of these caves (Grotte des Tortues, Grotte du Rouquet and Grotte Basse de la Vigne Perdue) were analyzed in this study (Data S1F).

Grotte des Tortues

The Grotte des Tortues site (lat. 43.149777, long. 3.060539) is located near Les Monges vineyard, between Narbonne and Gruissan (Aude, France) and was excavated around 1920 by Théophile and Philippe Héléna[92,93]. The 0.45 m by 0.75 m entrance of the cave was initially closed by a rock wall, blocking the access to a single 2.5 m wide chamber. Human skulls were arranged inside a recess and a short dead-end gallery, while dispersed bones were found in the main chamber, under large blocks. The total number of individuals was estimated to 22. The archaeological material is dated to the Late Neolithic to the Early Bronze Age period[17]. Considering the radiocarbon dates obtained in this study, the individual TORTE (2,839-2,500 cal. years BCE) belongs to the classical phase of the Veraza culture (final Neolithic), predating the Bell Beaker phenomenon while the individual TORTC (2,578-2,472 cal. years BCE), also attributed to the classical phase of the Veraza culture, could correspond to the timing of the first intrusions of early Bell Beaker individuals (maritime/international style). The individual TORTF (2,008-1,825 cal. years BCE) corresponds to the post-Bell Beaker Early Bronze Age period. The individual TORTD was dated 3,261-2,918 cal. years BCE, which corresponds to a middle phase of the final Neolithic, as supported by the presence of limestone '*à ailettes*' beads.

Grotte Basse de la Vigne Perdue

Also known as Grotte du Ruisseau (lat. 43.149150, long. 3.060582), the Grotte Basse de la Vigne Perdue cave is located on the right bank of an intermittent torrent ('Rec de la Vigne Perdue'). This 4.10-m-long cave is subdivided in two chambers of 2.25 and 2.85 m wide. The entrance was obstructed by limestone blocks[92]. During the excavation in 1920, Théophile and Philippe Héléna estimated the minimum number of individuals at 24 (20 adults and 4 children)[93]. In the second chamber, bones appeared grouped in several heaps, each covered by a stone mound. Archaeological material is characteristic of the Late Neolithic and the Early-Middle Bronze Age[17]. In particular hemispherical V-perforated buttons could be attributed to early Bell Beakers (Iberic influence). Two pottery fragments of the later Pyrenean style, as well as turtle-shaped V-perforated buttons were also present. Two double-ended copper awls of square cross-section and one barbed-and-tanged arrowhead could be attributed to any of these two early Bell Beaker phases.

Based on the radiocarbon dates presented here, individual GBVPL (2,574-2,473 cal. years BCE) corresponds to the classical phase of the Veraza culture, and could be contemporaneous with the first intrusions of early Bell Beakers. The individual GBVPK (2,461-2,299 cal. years BCE) is included in the late Bell Beaker timing period (Pyrenean). The individual GBVPO was dated to the middle Bronze Age (1,737-1,547 cal. years BCE), consistent with the presence in the burial of several bell-shaped cups, including one with an '*ad ascia*' handle.

Grotte du Rouquet

Also known as ''Grotte Vallée du Rouquet'' (lat. 43.166302, long. 3.060582), this site belongs to a series of natural caves in the Combe-Longue valley of the La Clape massif. Théophile and Philippe Héléna who excavated the site, also referred to this site as 'Grotte de la Palette' or 'Grotte du Triton' as they found among the grave goods a schist palette as well as a perforated triton's shell that could have been used as a horn[92]. This cavity consisted of two galleries, one in the axis of the main entrance obstructed by rock blocks and the other, on the right hand-side, giving access to a lateral chamber whose entrance was closed by a dry-stone wall. This chamber hosted nine well-preserved skulls. The axial gallery contained several bone and skull fragments, covered by flat stones. Grave goods included flint blades, a foliate point, a bone punch tool and numerous limestone and steatite beads. These artifacts could be dated as Languedocian Late Neolithic I-II (Early-Middle Veraza). More recent occupations (amber bead, pottery decorated with cord impressions) are attributed to the Early-Middle Bronze Age.

METHOD DETAILS

Radiocarbon dating

AMS-radiocarbon dating was performed at the Keck Carbon Cycle Accelerator Mass Spectrometry Facility, Earth System Science Department, UC Irvine, California. Bone pieces between 0.5 and 2.15 g were sampled in the bone laboratory in the ancient DNA facilities of the CAGT laboratory and sent for subsequent isotopic measurements on ultrafiltered collagen. For the Mont-Aimé remains, we sampled a fraction of the mandible bone holding the tooth analyzed for DNA to ensure that genomic and radiocarbon data correspond to the same individual. For individual 1H04, two aliquots were prepared and measured due to low collagen yields. Sample preparation backgrounds were estimated, and subtracted, based on measurements of ^{14}C-free mammoth and whale bone. Calibration was carried out using OxCal online (OxCal 4.4[94]) and the IntCal20 calibration curve. Values are provided in Data S1I.

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

DNA extraction and sequencing

Drilling, DNA extraction and sequencing library building (pre-PCR amplification steps) were performed in the ancient DNA facilities of the CAGT laboratory (Toulouse, France). DNA was extracted from 80-430 mg of petrosal bone (samples from Mas Rouge, Grotte des Tortues, Grotte du Rouquet and Grotte Basse de la Vigne Perdue), rib bones (samples 2HxC5x196 and 2HxC5x1131 from Mont-Aimé hypogeum II) or tooth cementum (all of the remaining Mont-Aimé samples). In short, after gentle surface abrasion using a drill, samples were pulverized using a manual mortar (for samples 2HxC5x196 and 2HxC5x1131 only) or a ball mill (Retsch MM 200). For teeth samples, the cementum was isolated using a procedure described in[95]. DNA of the Mont-Aimé samples was extracted following the method described in[96], using 4 mL of lysis buffer (0.45 M EDTA, 0.25 mg/mL proteinase K and 0,5% N-lauryl Sarcosyl) for a pre-digestion step of 1 hour at 37°C, before fully digesting the pellet overnight at 42°C in 4 mL of fresh lysis buffer. For samples 2HxC5x196 and 2HxC5x1131, both pre-digest and overnight digestion supernatants were further purified and built into sequencing libraries. For all the other samples, only the overnight digestion supernatants were further processed. Samples from Grotte des Tortues, Grotte du Rouquet and Grotte Basse de la Vigne Perdue were processed following a similar procedure, except that the lysis buffer volume was limited to 1 mL and that 250 μL of the supernatant obtained after overnight digestion was directly purified on a MinElute column (QIAGEN ©) and eluted in 23 μL of elution buffer (EB + 0,05% Tween 20). Of note, as the endogenous content of the first sequencing libraries built for samples TORTC and TORTD was relatively low (TORTC_AMIS_2_02252_U_i8_lr8 and TORTD_AMIS_2_02253_U_i10_lr10), a second bone powder aliquot was subjected to bleach pretreatment as described in[97] before proceeding with DNA extraction. After digestion in 1 mL lysis buffer as described above, DNA from the MAS15 sample was extracted using the combination of Buffer G and magnetic silica beads described in[98].

A 22.8 μL aliquot of each DNA extract was subjected to Uracil-Specific Excision Reagent treatment by incubation at 37°C for 3 hours with 7 μL of USER enzyme (NEB®), to limit the impact of nucleotide mis-incorporations in downstream analysis. Illumina sequencing libraries were built following a protocol first described in[37] and adapted in[99] introducing a unique index of seven nucleotides within adaptor P5 and within adaptor P7 before their ligation to the end-repaired sample's DNA. Libraries were enriched and indexed by performing 11-12 PCR cycles in 25 μL reaction volumes using 1 unit of AccuPrime™ Pfx DNA polymerase, 4-6 μL of DNA library and with an overall concentration of 200 nM of each primer, including the InPE1.0 primer and one custom PCR primer. The latter includes a 6 nucleotides sequence tag index used for sequence demultiplexing. After purification using Agencourt Ampure XP beads (1.4:1 beads:DNA ratio) and elution in 20 μL EB+0.05% tween, library concentration and size were checked on a Tapestation 4200 instrument (Agilent Technologies) and on a QuBit HS dsDNA assay (Invitrogen). Up to 13 amplified libraries were obtained per sample.

Amplified libraries were pooled with other indexed libraries and sequenced on Illumina platforms using the Paired-End mode. Screening and library validation were carried out on a MiniSeq instrument (2x80 bp reads) at the CAGT laboratory. Libraries showing endogenous human DNA content higher than 10% were selected for deep-sequencing and sequenced on a HiSeqX (2x150 bp reads) at the CEA/CNGM (Evry, France) or on a NovaSeq S4 (2x150 bp reads) at SciLifeLab (Stockholm, Sweden). These represented 5/12 and 7/11 remains for Mont-Aimé hypogea I and II respectively, 1/6 samples for Mas Rouge, 4/7 samples for Grotte des Tortues, 3/7 samples for Grotte Basse de la Vigne Perdue and 6/10 samples for Grotte du Rouquet.

Read processing and alignment

Illumina paired-end reads were demultiplexed on the basis of the seven-bp index sequences present at the starts of both read pairs[37], using a maximum edit distance of 1 mismatch for each index. Demultiplexed reads were then trimmed for adaptor sequences (–mm 5) and poor-quality ends (Phred quality scores $\leq$ 2), and collapsed into single reads using AdapterRemoval2[62]. Collapsed, collapsed truncated and non-collapsed pairs were mapped against the human reference genome (GRCh37, hg19) using Bowtie2 (version 2.3.5.1,[100]), and local sensitive mapping parameters, following the recommendations from[101]. Aligned reads shorter than 25 bp and alignments showing mapping quality strictly inferior to 30 were disregarded. PCR duplicates were removed using MarkDuplicates from Picard Tools (version 2.18.0,[102]). Local re-alignment around indels was carried out using GATK (version 3.8.1,[103]). Read collapsing, trimming and mapping, as well as PCR duplicate removal and local realignment were carried out using the Paleomix automated computational pipeline (version 1.2.13.*[63]). Sequencing statistics, including numbers of sequencing reads, endogenous DNA content, clonality, and average sequence length and coverage are provided in Data S1A.

QUANTIFICATION AND STATISTICAL ANALYSIS

Post-mortem damage

The Paleomix pipeline automatically produced mapDamage2[64] profiles, which provide an assessment of post-mortem DNA damage through the analysis of nucleotide mis-incorporations and base composition within reads (and on flanking genomic regions for the latter). Nucleotide mis-incorporation profiles for each individual DNA library looked similar and as expected following partial USER enzymatic treatment. Illustrative examples showing a limited excess of C→T transitions at alignment starts, mirrored by an equivalent excess of G→A transitions at alignment ends, are provided in Figures S1A and S1B. The base composition within reads was uniform, while the genomic position preceding alignments was significantly enriched in C, due to the cleavage of ancient DNA templates at deaminated non-methylated Cytosines[46].

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Error rates

Error rates were calculated using ANGSD (version 0.933-86-g3fefdc4, with htslib: 1.10.2-106-g9c35744[65]) and the methodology fully described in previous work[104] using 3-way alignments including an outgroup genome (here, the ancestral human genome), a high-quality genome (here, the genome from a modern Sardinian individual, sequenced to 39.4-fold average depth-of-coverage by[105]; accession number SS6004474) and each ancient genome as the test genome. The relative difference between the branch length leading to the ancient genome and the high-quality modern genome provides a direct measurement of relative error rates in the ancient genome, both globally and for each individual substitution class (Figure S1C). The analysis was repeated on previously published genomes for comparison (Figures S1D and S1E). Overall, the global error rates of each individual genome characterized in this study ranged between 0.000566 and 0.000848 substitutions per base on average, which is lower than the error rates presented in a random selection of comparative studies[4,16,106–108].

Uniparental markers and Contamination estimates

Individual sex was inferred on the basis of X-to-autosomal sequence coverage, resulting in the identification of 12 women and 13 men (Data S1A). Mitochondrial haplotypes were called using haplogrep (version 2.2[66]), and minimal mapping and base quality thresholds of 30. Briefly, read alignments against the rCRS reference mitogenome (GenBank Accession Number NC_120920.1) were converted into a VCF file using bcftools (version 0.1.19-44428cd) mpileup, call (–ploidy 1) and filter modules, disregarding sites covered by less than 5 independent reads. The resulting VCF file was then directly processed through haplogrep (version 2.1.25), calculating the best 100 hits. Contamination rates based on mitochondrial data were estimated using Schmutzi[67] and an identical base quality threshold. Contamination estimates were assessed to be within a 0%–2% confidence range for all individuals processed (Data S1E), except for 2H06 for whom extremely limited amounts of data were generated (mtDNA average coverage-of-depth = 1.25-fold) and 1H14 (confidence range 1%–12%). The former sample was disregarded in all other analyses. Sample 1H14 was, however, included and projected together with the majority of the other samples characterized at the Mont-Aimé site in PCA analyses. This sample also showed similar genetic profiles (e.g., f3-Outgroup statistics, f4- and D-statistics, ADMIXTURE, and PCA), supporting minimal nuclear contamination. Contamination rates were also estimated for males using heterozygosity measurements at polymorphic sites present on the X chromosome, following the methodology from[109] and implemented in ANGSD. Transition substitutions, and sites covered once or more than 200-times were disregarded. Contamination estimates were found to be extremely limited (median = 0.12%) and not greater than 1.01% (Data S1F). Y chromosome haplotypes were called using the Yleaf statistical package[68]. All the individuals analyzed belonged to haplogroup I2a1 (Data S1F), except individuals 2H07 and GBVPK who carried haplogroups H2a1a and R1b1a1b1a1a2a1, respectively.

Phenotype prediction

We assessed the genotypes of the two individuals sequenced to high depth-of-coverage, i.e., d2H11 and d2H10, d indicating that the high-coverage genome data (throughout the manuscript, pseudo-diploid genotypes called after random sampling one individual read per position are indicated using h as a prefix), for a total of 90 SNPs causative for or associated with a range of morphological, coloration, behavioral and health-related phenotypes (Figure 2C). Skin, eye and hair color was inferred using the 41 loci that are part of the HIrisPlex-S, HIrisPlex and IrisPlex sytems[39,69], available online at https://hirisplex.erasmusmc.nl/ (Data S2B). Individual SNP labels and their coordinates on the GRCh37 (hg19) reference genome are provided. Minimal mapping and base quality thresholds of 30 were required. Sites for which sequence coverage was above the 99.5% coverage quantile were disregarded to avoid any impact of mis-identified copy number variants. Bcftools mpileup, call and filter modules were assuming diploidy for autosomal SNPs and haploidy for those located on the X chromosome as both individuals were males.

Kinship analyses

Relatedness between individuals was assessed using a combination of two complementary methodologies, READ[42] and lcMLkin[43], respectively. Only those autosomal positions overlapping the 1240K dataset (see below) were considered. Prior to running lcMLkin, we filtered the *vcf.* matrix resulting from the SNPbam2*vcf.*py script using VCFtools (version 0.1.17) and the–thin 100000 and–maf 0.05 parameters in order to condition the analyses on those sites located at last 100 kb apart and present at minimal 5% allelic frequencies. In these analyses, the two genomes characterized to high-coverage were processed as the other genomes and were thus pseudo-diploidized, random sampling one read per position using ANGSD (-doHaploCall 1 -doCounts 1); they are, thus, hereafter labeled h2H10 and h2H11instead of d2H10 and d2H11 (hereafter, h stands of haploid, and d for diploid). Since the default pairwise distance thresholds provided in READ have shown proficiency when at least four individuals are considered, we carried out analyses grouping together individuals per archaeological site, providing four (Grotte Basse de la Vigne Perdue, and Grotte des Tortues) and eleven (Mont-Aimé) individuals per group considered. As only one individual was genome-sequenced at Mas Rouge, and three at Grotte Basse de la Vigne Perdue, we carried out an additional analysis including all individuals (for a total of 24 individuals). These analyses showed no relatedness except at the Mont-Aimé site (Data S1G), where two 1st degree relationships between two independent pairs of individuals were inferred (h1H06 and h2H11 on the one hand, and h2H10 and h2H17, on the other hand). For female and male h1H06 and h2H11, consistent results were obtained using READ (first-degree relationship) and lcMLkin (parent-offspring relationship). As these individuals carry different mitochondrial haplotypes, h2H11 can be identified as the father and h1H06 as his daughter. For the males h2H10 and h2H17, both READ and lcMLkin indicated a parent-offspring relationship without identifying the father and the son. They shared the I2a1b1b1 Y chromosome haplotype with h2HxC5x196x1131, and although READ and

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

lcMLkin suggested that those three individuals were related, they did not allow robust determination of their degrees of relatedness (Data S1G and S1H). In line with kinship analyses, related individuals showed overlapping radiocarbon date distributions.

Inbreeding and diversity estimates

Levels of inbreeding in the two individuals sequenced to high coverage were assessed using autosomal data in ROHan (–rohmu 2.0e-05[44]). The program returned no evidence of inbreeding in any of the two individuals and returned largely overlapping diversity theta estimates for both individuals. Inbreeding and theta outside run-of-homozygosity were estimated using the same procedure on a comparative panel of five Neolithic and three Mesolithic high-coverage genomes previously reported (Stuttgart and Loschbour[4]; NE1[39]; Bar8[107]; WC1[106]; sf12[41], and; Jerpoint14 and Sramore62[12]). The corresponding results are provided in Data S2A. The comparative panel was directly downloaded from public repositories as BAM files, except for individuals Bar8 and WC1 that were subject to the same read processing, mapping and filtering procedure as individuals d2H10 and d2H11. For samples Stuttgart and Loschbour, we noticed an excess of nucleotide mis-incorporations at the first and last sequence position of read alignment. These positions were trimmed using BAMtrim from the bamUtil repository (https://github.com/statgen/bamUtil) before running ROHan to limit the possible impact of sequencing errors.

Principal Component Analysis

Principal Component Analysis (PCA) was carried out using the Human Origin reference panel for 592,998 autosomal genotypes in 796 west Eurasian modern individuals, as reported by Patterson and colleagues[56] and Lazaridis and colleagues[4]. Genotypes were downloaded from David Reich's website (dataset v42.4.1240K_HumanOrigins, available at https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V42/V42.4/SHARE/public.dir/v42.4.1240K_HO.tar). We also included those individuals recently reported in three recent studies[8,12,16] and covering the Mesolithic, Neolithic and Bronze Age time periods. The individuals present in these publications were subjected to the same procedure as the individuals reported in this study, including read processing, mapping for those described by Brunel and colleagues[16] and Cassidy and colleagues[12], or starting from individual BAM files available from public repositories for those reported by Rivollat and colleagues[8]. The analysis was based on pseudo-diploid genotype calls for all the individuals presented in this study, except 2H06 for whom insufficient data could be collected. PCA was carried out using smartPCA from EIGENSOFT version 7.2.1[70], projecting ancient individuals onto the principal components (PCs) obtained from modern reference individuals (the lsqproject option was turned on). Projections on the first two PCs are provided in Figure 3A and projections PC2 and PC3 are provided in Figure S2A. In addition to pseudo-diploid genotype calls, the analysis was also carried out on truly diploid genotype calls for the two individuals sequenced to high-coverage in order to confirm individual projections.

f3-Outgroup and f4-statistics

f3-Outgroup statistics were carried out using q3Pop from Admixtools (version 5.0[56]) to identify those ancient individuals previously characterized showing highest genetic affinities with each individual sequenced in this study, except 2H06 for whom insufficient data could be collected. Here, we used all 1,150,639 autosomal positions overlapping the 1240K panel and the comparative genotype panel provided on David Reich's website (dataset v42.4.1240K, available at https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V42/V42.4/SHARE/public.dir/v42.4.1240K.tar). We also included all those individuals recently reported in[8,12,16], following the same procedure as that used in PCA. For those two individuals sequenced to high depth, we considered full genotype calls (d2H10 and d2H11), as well as their pseudo-diploid version based on random sampling of one read per position (h2H10 and h2H11). In addition to those individuals reported here for the first time, the final analyses included a total of four modern Mbuti individuals used as the outgroup (S_Mbuti-3.DG, B_Mbuti-4.DG, S_Mbuti-2.DG, S_Mbuti-1.DG) as well as 2,052 additional unrelated ancient individuals showing at least 10,000 positions covered and no evidence of contamination. In case several ancient individuals were annotated as related, the one showing the highest number of positions covered was selected. This dataset is referred to as the 1240K dataset hereafter.

Those individuals showing f3-Outgroups comprised within the confidence range of the maximal estimate are located on geographic maps in Figure S3A, highlighting a number of characteristic examples in Figures 3B and 3C. The ancient samples showing greatest genetic affinities to those presented in this study are indicated with their original label and radiocarbon date (cal. years BCE).

The genetic affinities between the ancient individuals characterized in this study and those previously reported were also investigated using f4-statistics. These statistics were calculated using qpDstat in Admixtools (version 5.0[56]) in the form of (ANF, X; Y, Mbuti), where ANF stand for Anatolian Neolithic Farmers, and X and Y refer to those individuals previously reported and those reported in this study, respectively. These statistics were designed in order to mitigate the various ANF ancestry proportions identified in each ancient individual, with negative values indicating proximity between individuals X and Y. Those f4-statistics identified as significantly deviating from zero through Z-scores inferior or equal to −2 are shown in Figure S3B, highlighting a number of characteristic examples in Figures 3D–3F. Only the top-25 X samples are annotated for clarity. A total of 36 ANF were selected following previous studies (ANF[40,107,110,111] = Bon001.SG, Bon002.SG, Bon004.SG, Bon005.SG, Bar31.SG, Bar8.SG, I1100, I1102, I1099, I1103, I1101, I1097, I0744, I1096, I1098, I0708, I0745, I0746, I0707, I0709, I0725, I0736, I0726, ZHAG_BON004.A0101_Luk10, ZHAJ_BON034.A0101_Luk9, ZHJ_BON024.A0101_Luk84, ZKO_BON001.A0101_Luk7, ZMOJ_BON014.A0101_Luk21, I1579_published, I1581_published, I1580_published, I1585_published, I1583_published, I0727_published, I0724_published, I0723_published). Whenever present, the word 'published' is replaced by 'p' in the sample labels reported on the figures.

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

Admixture

Supervised ADMIXTURE (version 1.3.0[58]) analyses were carried out to estimate different ancestry proportions present in each individual sequenced in this study, except 2H06 for whom not enough data could be collected. Here, we followed a previously published procedure[7] and used all autosomal positions overlapping those from the 1240K panel available for download from David Reich's website (dataset v42.4.1240K, available at https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V42/V42.4/SHARE/public.dir/v42.4.1240K.tar). The same 36 ANF and 15 WHG individuals as described above were selected together with ten ancient Yamnaya_Samara individuals, following previous work[40,59] (Yamnaya_Samara = I0370, I0441, I0444, I0439, I0357, I0429, I0438, I0443, I7489, I0231_published) that established that the genetic landscape of modern European individuals can be explained by various contributions of those three ancestries. Only sites that were covered in at least 50% of all individuals were considered. Termination criteria for convergence were based on a log-likelihood increase by a maximum of 0.0001 between iterations. Confidence intervals were estimated from 100 bootstrap pseudo-replicates. Analyses were repeated 10 times using 10 random seeds in order to assess cross-validate convergence. Individual profiles are provided in Figures S2B and S2C, averaging out the point estimates and confidence ranges obtained across the 10 analytical replicates. A ternary plot of the averaged point estimates are provided in Figure 4B.

D-statistics

Different combinations of D-statistics were calculated using qpDstat in Admixtools (version 5.0[56]) in order to estimate whether pairs of ancient individuals were symmetrically related to ANF, WHG and/or Yamnaya_Samara individuals previously reported (and defined as indicated in the previous section). Calculations were carried out on the 1240K dataset and using four Mbuti individuals as the outgroup (S_Mbuti-3.DG, B_Mbuti-4.DG, S_Mbuti-2.DG, S_Mbuti-1.DG). The topologies investigated were in the form of (Outgroup, ANF or WHG or Yamnaya_Samara; Ancient_Individual1, Ancient_Individual2), respectively. Only those combinations represented by a minimal number of 10,000 covered sites were considered. The returned values of the different D-statistics calculations are provided in Figure S4 for each individual reported in this study, except 2H06 for whom not enough data could be collected. Negative values indicate closer genetic affinities between H3 (i.e., WHG, ANG and Yamnaya_Samara) and H2, while positive values indicate closer genetic proximity to H1.

Population modeling

Affinities between various population groups were assessed using the 1240K dataset and qpAdm from Admixtools (version 5.0[56]). Analyses were aimed at assessing the respective proportions of ANF and WHG ancestries in the individual genomes presented in this study as well as in the genome-scale data previously reported for other contemporary ancient individuals from France[8,9,16]. This represented a total of 121 ancient individuals from France (I1381, I1382, I1388, I1390, I1391published, I1392, I2575published, I3874, I3875, I4303, I4304, I4305, I4308, d2H10, d2H11, h1H04, h1H06, h1H07, h1H13, h1H14, h2H07, h2H10, h2H11, h2H17, h2HxC5x196x1131, hBERG02-2, hBERG157-2, hBERG157-7, hBIS130, hBIS385, hBLP10, hBUCH2, hCBV95, hCRE20D, hEUG11, hEs97-1, hFLR001, hFLR002, hFLR003, hFLR004, hFLR005, hFLR007, hFLR010, hFLR013, hFLR014, hGBVPK, hGBVPL, hGBVPO, hGRG003, hGRG008, hGRG015, hGRG016, hGRG019, hGRG022, hGRG023, hGRG025, hGRG027, hGRG028, hGRG032, hGRG035, hGRG043, hGRG047, hGRG049, hGRG050, hGRG052, hGRG056, hLBR001, hLBR002, hLBR003, hLBR004, hLBR005, hMAS15, hMDV248, hMor6, hOBE3626-1, hOBE3722, hOBN001, hOBN002, hOBN003, hOBN004, hOBN005, hOBN006, hOBN007, hOBN008, hOBN009, hOBN010, hOBN011, hPEI10, hPEI2, hPEN001, hPEN001, hPEN003, hPIR3037AB, hPIR3116B, hPRI001, hPRI005, hPRI006, hPSS4170, hPSS4693, hQUIN234, hQUIN58, hRIX15, hRIX2, hRIX4, hROS102, hROS45, hROS78, hROS82, hROUQCC, hROUQEE, hROUQFF, hROUQHH, hROUQV, hROUQW, hSch72-15, hSchw432, hTORTC, hTORTD, hTORTE, hTORTF, hWET370). In all models tested, we followed Haak and colleagues[2], Olalde and colleagues[9] and Brunel and colleagues[16] and used Mbuti (n = 4; S_Mbuti-3.DG, B_Mbuti-4.DG, S_Mbuti-2.DG, S_Mbuti-1.DG[112]), Papuans (n = 15; S_Papuan-1.DG, S_Papuan-5.DG, S_Papuan-13.DG, S_Papuan-10.DG, B_Papuan-15.DG, S_Papuan-12.DG, S_Papuan-2.DG, S_Papuan-9.DG, S_Papuan-7.DG, S_Papuan-8.DG, S_Papuan-3.DG, S_Papuan-14.DG, S_Papuan-4.DG, S_Papuan-6.DG, S_Papuan-11.DG), Onge (n = 6; mondal_ONG-1.SG, mondal_ONG-12.SG, mondal_ONG-14.SG, mondal_ONG-4.SG, mondal_ONG-8.SG, mondal_ONG-9.SG[113]), Han (n = 106; S_Han-1 to S_Han-3, NA18525.SG, NA18526.SG, NA18528.SG, NA18530.SG-NA18539.SG, NA18541.SG-NA18550.SG, NA18552.SG, NA18553.SG, NA18555.SG, NA18557.SG-NA18567.SG, NA18570.SG-NA18574.SG, NA18577.SG, NA18579.SG, NA18582.SG, NA18591.SG-NA18593.SG, NA18595.SG-NA18597.SG, NA18599.SG, NA18602.SG, NA18603.SG, NA18605.SG, NA18606.SG, NA18608.SG-NA18648.SG, NA18740.SG, NA18745.SG, NA18747.SG-NA18749.SG, NA18757.SG), Karitiana (n = 4; B_Karitiana-3.DG, S_Karitiana-1.DG, S_Karitiana-2.DG, BI16.SG[112,114,115]) modern individuals as outgroup ('right') populations together with Mota (n = 1; mota.SG[116]), Ust-Ishim (n = 1; Ust_Ishim_published.DG[117]), Malta Ma-1 (n = 1; MA1.SG[118]), Koros (n = 1; I1507[5,109]), Vestonice (n = 1; Vestonice16), Natufian (n = 1, I1072[119]) and Goyet-116 (n = 1; GoyetQ116-1_published[117]). This selection was limited to those individuals showing sequence data for a minimum of 700,000 SNPs. The same coverage filter was applied to the list of ANF, WHG and Yamnaya_Samara individuals described above, leaving a final number of 18, 5 and 10 individuals, respectively.

In a first model, each ancient individual from France was grouped together with WHG[117] and ANF as 'left' populations. This provided WHG and ANF ancestry estimates in each ancient individual considered and assessed the validity of the model relative to models in which only one of the two possible ancestries modeled was considered (Data S4A).

A second model was aimed at assessing the possibly different contributions of WHG by adding the Goyet-Q2 individual[120] to the list of 'left' populations (Data S4A), as previous work identified the presence of a cline of ancestry related to this individual in some

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

hunter-gatherer populations from Western Europe[8]. This provided the opportunity to assess whether an additional and Goyet-Q2-related ancestry was necessary to explain the genetic profiles of the different ancient individuals from France analyzed at the genetic level. This analysis was complemented with a third model, which corresponded to the second model, except that the individual Goyet-Q2 was now considered as an additional outgroup (i.e., among the so-called 'right' populations; Data S4A). Since this modeling largely supported the conclusions obtained the previous two models, this allowed us to confirm the absence of Goyet-Q2 ancestry in the individuals analyzed.

Finally, a fourth model was assessed in order to estimate the proportion of Yamnaya_Samara ancestry in each ancient individual considered. This was done using a list of 'left' populations including each ancient individual from France, as well as Yamnaya_Samara individuals, WHG and ANF (Data S4B).

Admixture dating

The different analyses presented supported the presence of an excess of WHG ancestry in two individuals at Mont-Aimé (2H10 and 1H06) and an excess of Yamnaya_Samara ancestry in one individual at Grotte Basse de la Vigne Perdue (GBVPK). We used LD-decay profiles and DATES ([59], https://github.com/priyamoorjani/DATES) to estimate the timing of underlying admixture events. For timing the dates of admixture between WHG and ANF ancestors in the genealogy of 2H11and 1H06 individuals, DATES was run on the 1240K dataset, restricting the analyses to those two individuals and the 36 ANF and 15 WHG individuals defined in the f3Outgroup, f4-statistics and ADMIXTURE sections. Specifically, each of these individuals was modeled as a combination of the two ANF and WHG possible sources. The same analyses were repeated in a third run where all the other Mont-Aimé individuals for which at least 10,000 SNPs were covered and no particular excess of WHG ancestry was identified (n = 8, h1H04, h1H07, h1H13, h1H14, h2H07, h2H10, h2H17, h2HxC5x196x1131). Similarly, the timing of admixture with Yamnaya-related individuals in the genealogy of the GBVPK individual was assessed modeling this individual as a combination of two possible sources: on the one hand, a Late Neolithic population from France, and on the other hand, the ten Yamnaya_Samara individuals defined above. In these analyses, three proxies were used as Late Neolithic groups from France: (1) the individuals from Grotte du Rouquet, (2) the Late Neolithic individuals from the La Clape massif sequenced in this study, and including Grotte du Rouquet as well as MAS15 and TORTD, and; (3) those individuals from Mont-Aimé showing no particular excess of WHG ancestry, as described above. Weighted LD-decay profiles are provided in Figure 4E, together with average and standard deviation estimates for the respective times of admixture inferred.

Microbial profiling

Microbial taxonomic profiles of each individual DNA library were determined using the metaBIT automated computational package[19], and restricting analyses to the fraction of collapsed reads (i.e., collapsed truncated and uncollapsed pairs were disregarded), and ignoring virus and eukaryotes. This package leverages the MetaPhLAn2 diversity database that comprises a total of 1,036,027 biomarkers including 896,781 markers for 12,926 bacterial taxa. By-library profiles were merged per individual, and bacterial taxa supported by abundances lower than 1% were disregarded. Principal Coordinate Analyses were carried out based on Bray-Curtis distances to compare these taxonomic profiles to those obtained using a similar methodology: 15 profiles from soil samples[121] and 689 profiles published by the Human Microbiome Project Consortium[122] for the mouth (n = 382), skin (n = 26), nose (n = 87) and vagina (n = 56) environments.

DNA methylation

The characterization of two individual genomes at high-coverage allowed us to assess DNA methylation patterns in Late Neolithic individuals from France. DNA methylation was inferred using DamMet[48] within genomic windows of 1,000, 1,500 and 2,000 bases centered around 27,578 genomic positions corresponding to the Illumina 27k methylation Bead Chip system. The age-at-death of the two ancient individuals was then estimated using these profiles and leveraging the DNA Methylation Age Calculator statistical model available at http://dnamage.genetics.ucla.edu/submit, which integrates age-dependent DNA methylation profiles at 353 sites to return age estimates. This analysis, and all following analyses, were repeated for an extended comparative panel including two Neolithic and two Mesolithic individual genomes also sequenced to high coverage-of-depth (Stuttgart and Loschbour[4], as well as Jerpoint14 and Sramore62[12]; note that sequence data generated in the absence of USER treatment were disregarded). These analyses were restricted to those individuals characterized from teeth in order to avoid possible DNA methylation differences across tissues. Despite returning point estimates for the age-at-death, calculations also indicated poor correlation with a Gold standard panel of DNA methylation values used internal for data normalization (Data S3B). They were, thus, considered only indicative.

Independent estimates for the age-at-death of this panel of individuals were obtained considering the methylation data reported by Bekaert and colleagues[53] for 11 individual loci (ASPA: 17:3379066-3380067; PDE4C: 19:18343388-18344389, 19:18343400-18344401, 19:18343414-18344415; EDARADD:1:236557181-236558182; ELOVL2: 6:11044393-11045394, 6:11044387-11045388, 6:11044379-11045380, 6:11044376-11045377, 6:11044374-11045375, 6:11044372-11045373). Eight additional loci analyzed in the original publication were disregarded either as they could not be positioned on the GRCh37 reference unambiguously or as their DNA methylation profiles were not associated with age. Following the original publication, we developed multiple regression models between the age of 207 modern individuals and either linear or quadratic values of their DNA methylation profiles at the 11 loci. Cross-validation using the R crossval library indicated excellent prediction performance (0.9378 and 0.9366 for the linear and quadratic multiple regression models, respectively). We applied similar models to the panel of ancient individuals, considering sites

Please cite this article in press as: Seguin-Orlando et al., Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France, Current Biology (2020), https://doi.org/10.1016/j.cub.2020.12.015

for which the DNA methylation value estimate returned by DamMet showed a confidence range where at least one boundary was included within the 0-1 frequency range, so as to include only those sites where sufficient data may be collected. The resulting individual age-at-death estimates are provided in Data S1A and S3C.

Finally, we estimated DNA methylation profiles within promoter regions and gene bodies. Promoters were defined as the one kilobase upstream of the Transcription Start Site (TSS) while gene bodies were restricted to the two kilobases downstream. TSS and gene models were downloaded from Ensembl (GRCh37). Only the longest transcript was considered at a given locus. A total of 10,659 gene bodies provided DNA methylation values with confidence ranges where at least one boundary was included within the 0-1 interval for all six individuals analyzed (i.e., d2H10, d2H11, Stuttgart, Loschbour, Jerpoint14 and Sramore62). Pairwise correlations varied between 0.40 and 0.62, but were lower for those two individuals sequenced to lower coverage (Sramore62, 13.68X and d2H10, 13.90X respectively; Figure 2B). This indicated that inferred DNA methylation values may be of lower accuracy in those two individuals, due to limited sequence data. Additionally, the Stuttgart and Loschbour individuals, which represent a Neolithic woman and a Mesolithic man genome-sequenced in the same laboratory[4], showed correlation values of 0.55 which is superior to those observed in Neolithic pairs from the same site (d2H10 and d2H11: 0.53) or from different sites (d2H10 and Stuttgart: 0.52). This indicated possible technical batch effect impacting on DNA methylation inference. To limit such technical batch artifacts, we restricted the following analyses to only three individual genomes generated in different laboratories, and including one Mesolithic individual. These include only males, and display the maximal pairwise correlation values observed within gene bodies (0.62, Figure 2B). Although the exact sources of those technical batch artifacts are unknown, we noticed that the procedures underlying data production encompass a range of experimental conditions, both at the levels of DNA extraction, USER treatment, library construction, sequencing and sequence post-processing. For example, reduced efficacy of USER-treatment at terminal ends or *in silico* trimming of read termini is expected to affect those sequence positions that are the most subject to post-mortem Cysotine deamination[123]. This, in turn, is expected to directly impact DNA methylation measurements[48]. For example, fully efficient USER treatment leaves CpG→TpG substitutions only at CpG sites that were methylated, while treatments of lower efficacy leave variable proportions of CpG→TpG substitutions at CpG sites that were not methylated. The latter procedures will, thus, limit the specificity and sensitivity of any statistical method aimed at DNA methylation inference based on CpG→TpG counts. *In silico* trimming of read ends is expected to reduce sensitivity only, as it removes the total number of occurrences of CpG→TpG substitutions. Further work is required to quantify the exact importance of those possible sources of bias in the inference of DNA methylation patterns in ancient individuals.

DNA methylation profiles for a total of 6,355 promoters showing credible confidence range (see above) were compared between the two pairs of Mesolithic and Neolithic individuals (i.e., Loschbour versus d2H11 on the one hand, and Loschbour versus Jerpoint14 on the other hand). Those promoters showing (1) average DNA methylation shifts of at least 50% across pairs, (2) average DNA methylation shifts at least twice as large as the standard deviation across pairs, and (3) that showed consistent directions were shortlisted as a candidate list of promoters that experience significant DNA methylation changes during the Mesolithic to Neolithic transition. Functional enrichment in Gene Organizer ([71], http://geneorganizer.huji.ac.il/), Panther ([70], http://www.pantherdb.org/) and David 6.8 ([72], https://david.ncifcrf.gov/), however, indicated that no particular functional category was significantly over-represented in the 611 promoter shortlisted relative to the full background of 6,355 promoters (Data S3A). We noted that the list of candidates included significantly more gene promoters that were hyper-methylated in Neolithic individuals, than in Mesolithic individuals (404 versus 207,chi-square test *p-value*s = $1.59.10^{-15}$).

ADDITIONAL RESOURCES

Our study has not generated or contributed to a new website/forum and it is not part of a clinical trial.