

Supplemental information

Origin and mobility of Iron Age

Gaulish groups in present-day France

revealed through archaeogenomics

Claire-Elise Fischer, Marie-Hélène Pemonge, Isaure Ducoussau, Ana Arzelier, Maïté Rivollat, Frederic Santos, Hélène Barrand Emam, Alexandre Bertaud, Alexandre Beylier, Elsa Ciesielski, Bernard Dedet, Sophie Desenne, Henri Duday, Fanny Chenal, Eric Gailledrat, Sébastien Goepfert, Olivier Gorgé, Alexis Gorgues, Gertrud Kuhnle, François Lambach, Anthony Lefort, Amandine Mauduit, Florent Maziere, Sophie Oudry, Cécile Paresys, Estelle Pinard, Suzanne Plouin, Isabelle Richard, Muriel Roth-Zehner, Réjane Roure, Corinne Thevenet, Yohann Thomas, Stéphane Rottier, Marie-France Deguilloux, and Mélanie Pruvost

Supplemental Information

Legends' STAR Methods Figure

Figure S1 –Principal Component Analysis performed on present-day western Eurasians, related to Figure 1.

Figure - S2 – Evolution of mitochondrial lineages frequencies measured for ancient groups originating from the French territory and dated from the Palaeolithic to the Iron Age periods, related to STAR Methods

Figure S3 – Evolution of Y-chromosome lineages for ancient groups originating from the French territory and dated from the Palaeolithic to the Iron Age periods, related to STAR methods

Figure S4 – READ estimates for the French IA samples. Sorted non-normalised average $P0$ values for all pairwise comparisons between French IA individuals, related to STAR methods

Figure S5 – Pairwise $qpWave$ testing to detect outliers. Grey-coloured models have a p value of less than 0.05 and were rejected, pink-coloured models have a p value of more than 0.05. Orange circles represent outliers from the chronological-cultural group from which they originate (p value < 0.05), related to Figure 2

Figure S6 – Heatmap of the f3-outgroup in the form $f3(Mbuti, Ind1, Ind2)$. We only plot data when we have at least 10,000 SNPs per pairwise. White square means that we do not have enough SNPs between individuals, related to STAR methods

Figure S7 – Biplot of the f3-outgroup statistics in the form $f3(Mbuti, Anatolia N, Ind/Region)$ in the x axe and in the form $f3(Mbuti, Yamnaya, Ind/Region)$ in the y axis. Each square represents a region, and each star represents an outlier, as defined in the previous figure, related to Figure 3

Figure S8 : Evolution of $qpAdm$ values between the Bronze and Iron Age periods for the Early neolithic farmer ancestry component, related to Figure 3

Figure S9: Distribution and average of ancestral component in IA samples from western Europe. Each colour corresponds to a region (France) or a country. Biplot of $qpAdm$ value within regional groups for the Anatolia Neolithic ancestry component (A) and for the Yamnaya ancestry component, related to Figure 3

Fig S10 – $qpAdm$ value for each individual with a 3 sources model (WHG, Anatolia N and Yamnaya), related to Figure 3

Fig S11- Sum of ROH (sROH) measured for all published and newly sequenced individuals from France dated between 9000BP and 2000 BP. Each circle in the main panels represents sROH for one ancient Individual. We also show mean estimates calculated from a Gaussian Process model (solid black line), and 95% empirical confidence intervals for individuals (light gray) and for the estimated mean (dark gray). Horizontal lines depict the theoretical expectations for sum of ROH blocks 4-8 cM for panmictic population sizes calculated using analytical formulas (see Ringbauer *et al.*, 2021), related to STAR methods

Fig S12 - ROH in COL336. We depict all ROH of COL336 (right panel), with map length annotated in Morgan. We also show a histogram of the ROH lengths, together with expected densities of ROH for certain degrees of parental relationship), related to STAR methods

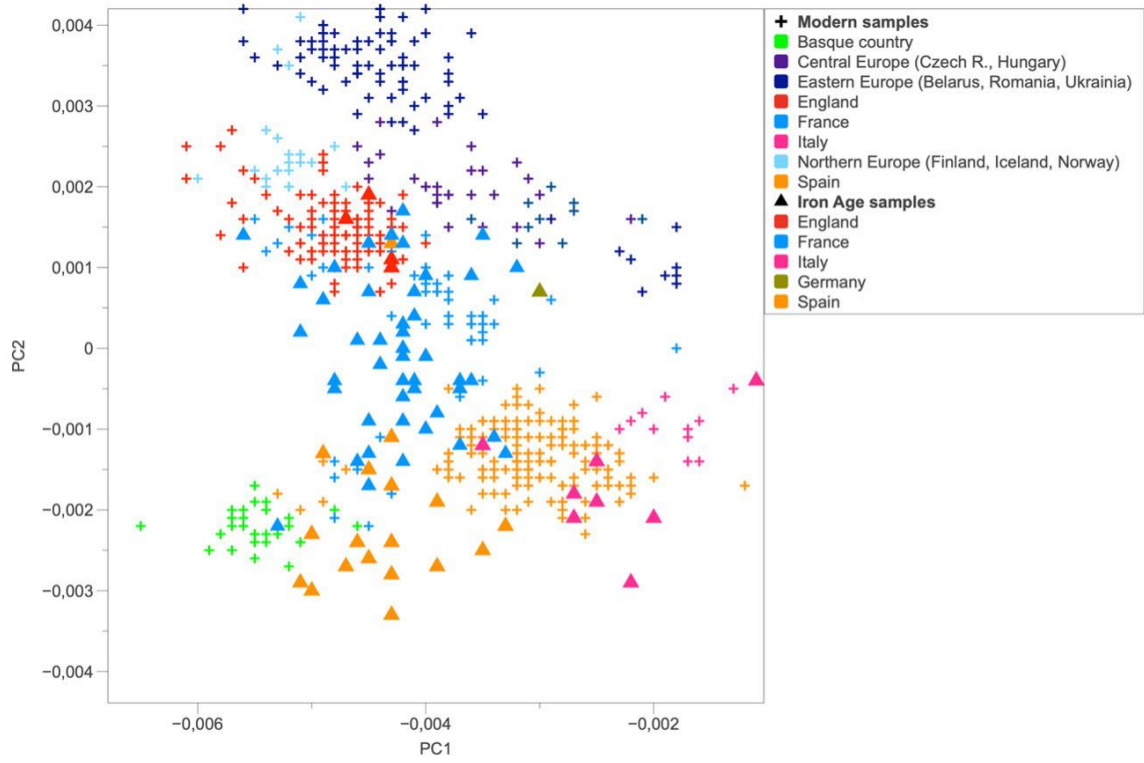


Figure S1 –Principal Component Analysis performed on present-day western Eurasians, related to Figure 1.

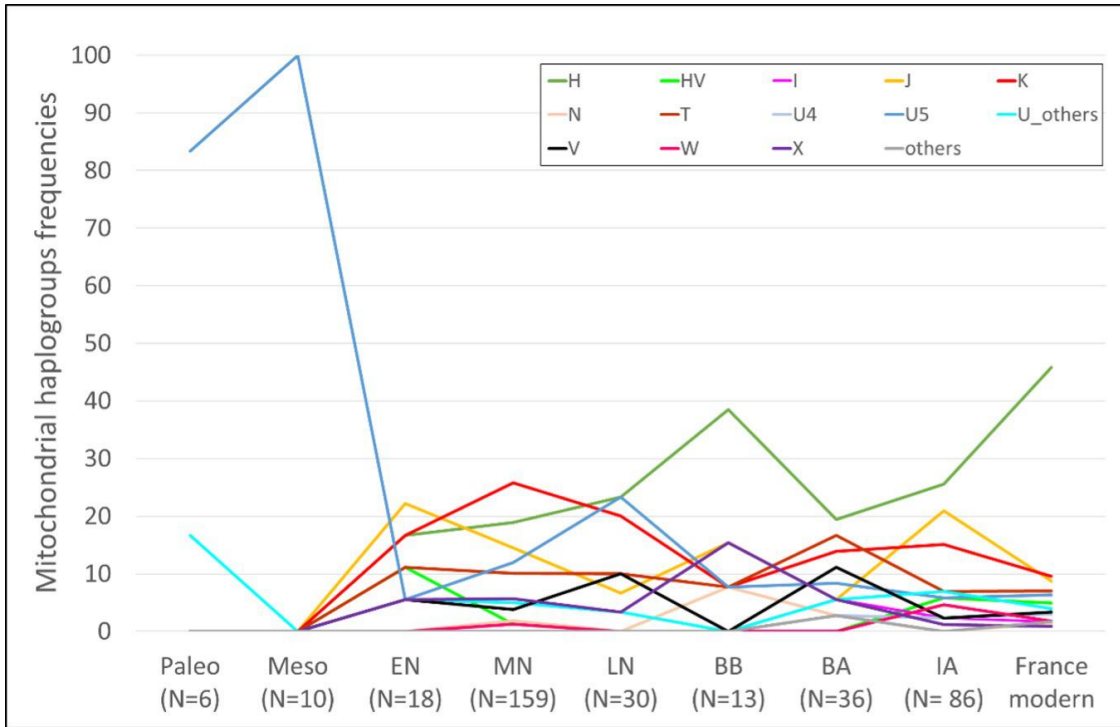


Figure - S2 – Evolution of mitochondrial lineages frequencies measured for ancient groups originating from the French territory and dated from the Palaeolithic to the Iron Age periods, related to STAR Methods

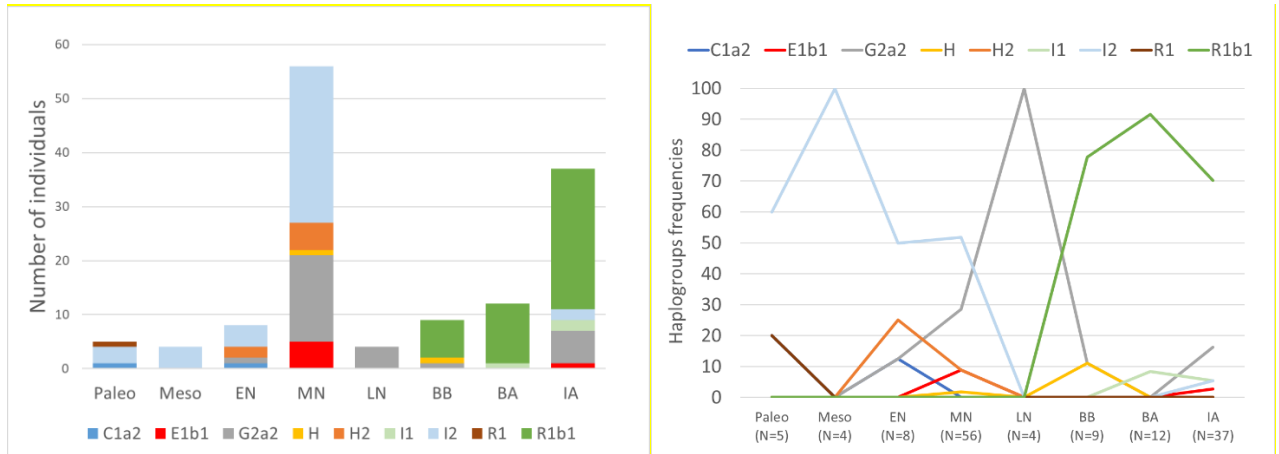


Figure S3 – Evolution of Y-chromosome lineages for ancient groups originating from the French territory and dated from the Palaeolithic to the Iron Age periods, related to STAR methods

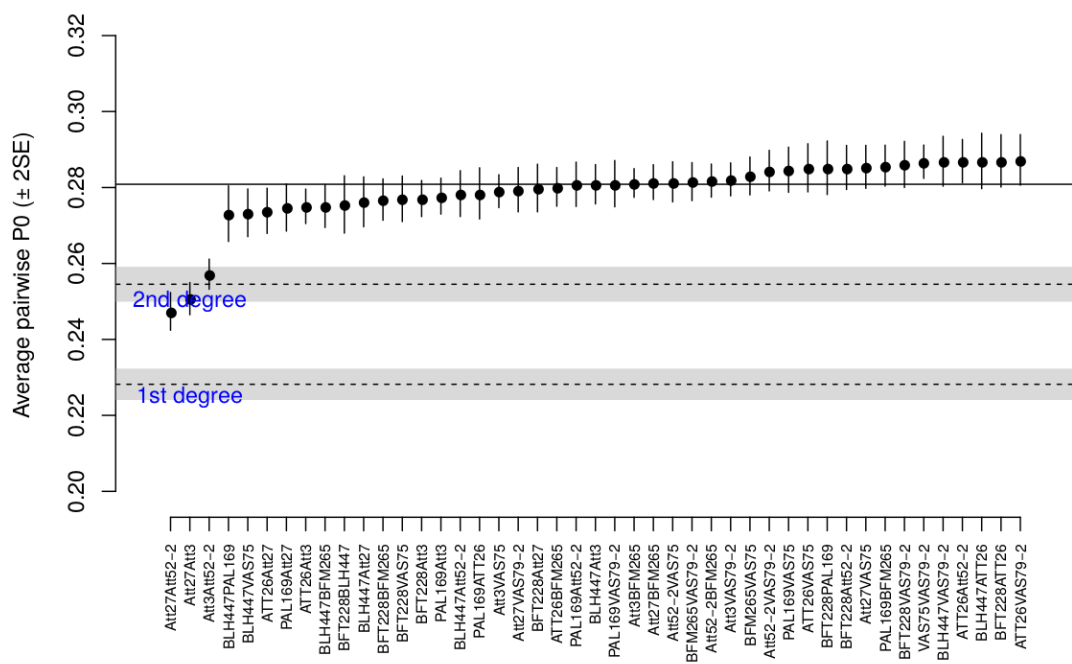


Figure S4 – READ estimates for the French IA samples. Sorted non-normalised average P0 values for all pairwise comparisons between French IA individuals, related to STAR methods

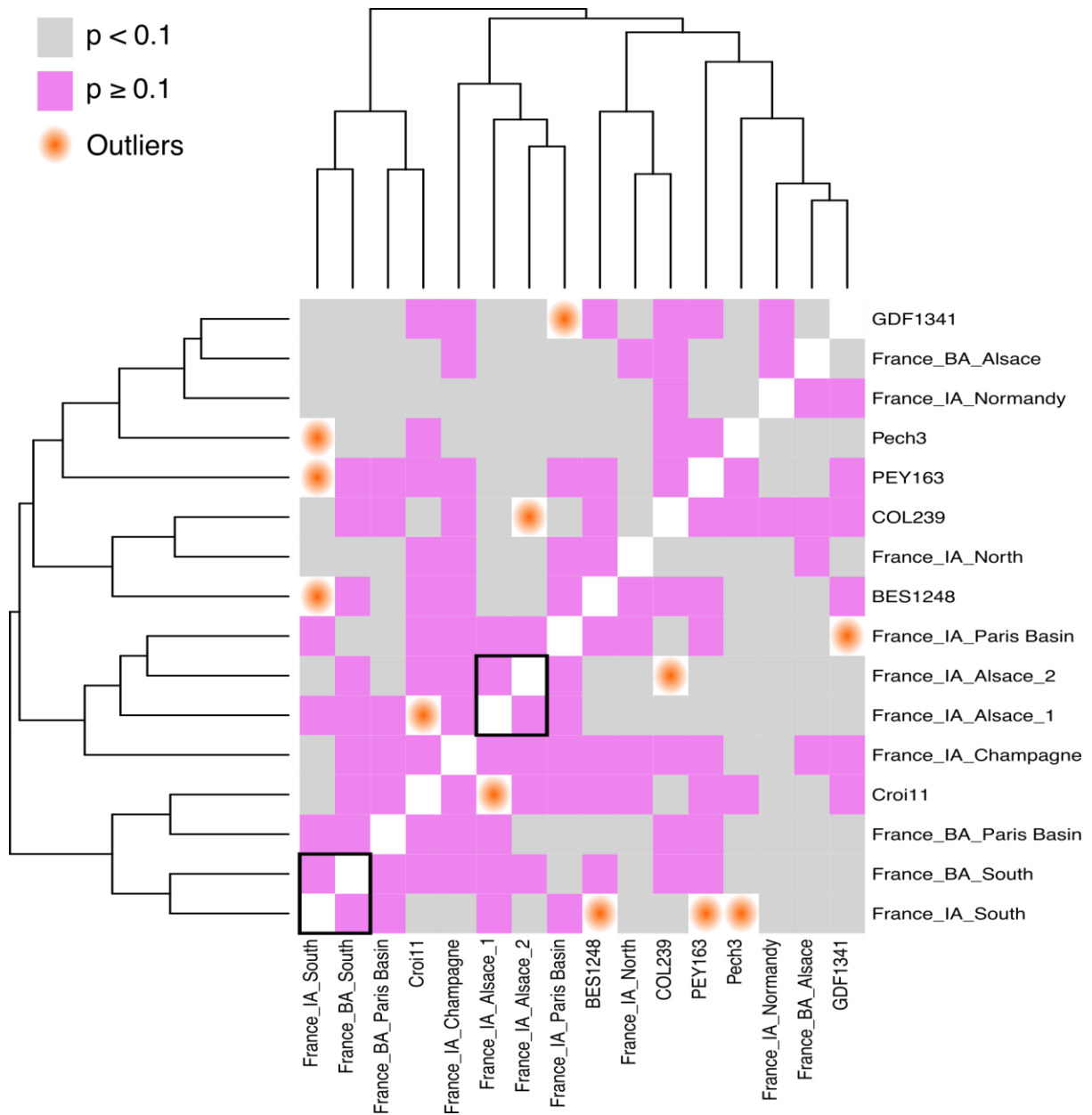


Figure S5 – Pairwise *qpWave* testing to detect outliers. Grey-coloured models have a p value of less than 0.05 and were rejected, pink-coloured models have a p value of more than 0.05. Orange circles represent outliers from the chronological-cultural group from which they originate ($p < 0.05$), related to Figure 2

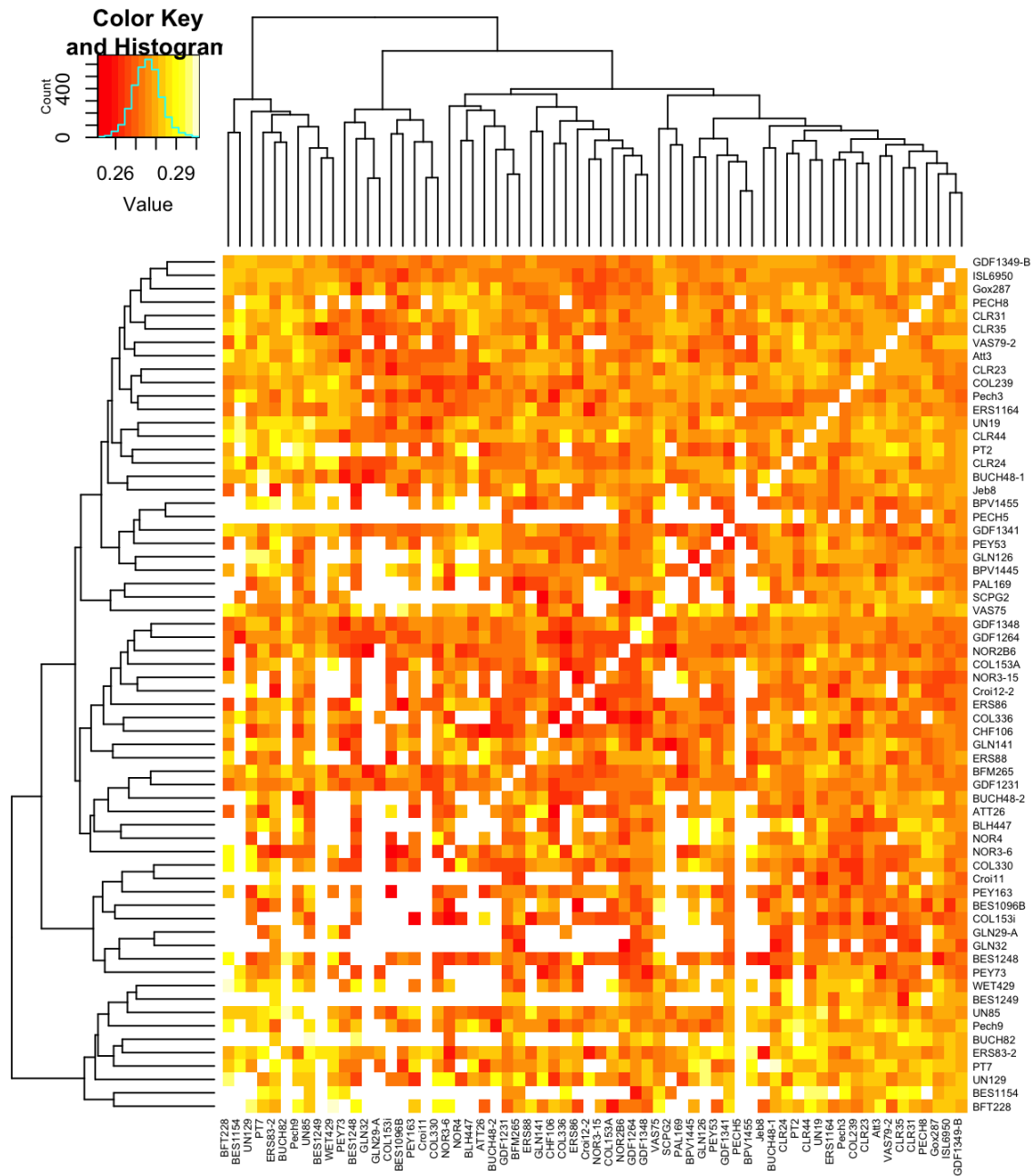


Figure S6 – Heatmap of the f_3 -outgroup in the form $f_3(\text{Mbuti}, \text{Ind1}, \text{Ind2})$. We only plot data when we have at least 10,000 SNPs per pairwise. White square means that we do not have enough SNPs between individuals, related to STAR methods

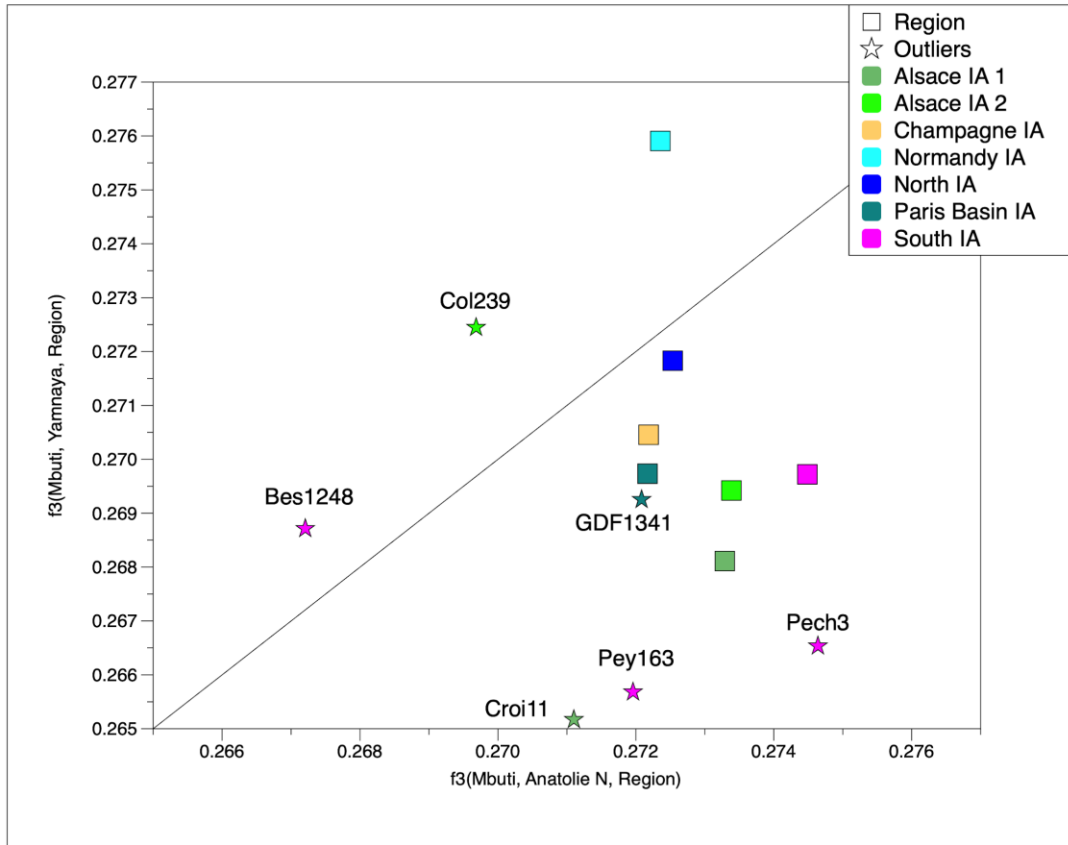


Figure S7 – Biplot of the f_3 -outgroup statistics in the form $f_3(\text{Mbuti, Anatolia N, Ind/Region})$ in the x axis and in the form $f_3(\text{Mbuti, Yamnaya, Ind/Region})$ in the y axis. Each square represents a region, and each star represents an outlier, as defined in the previous figure, related to Figure 3

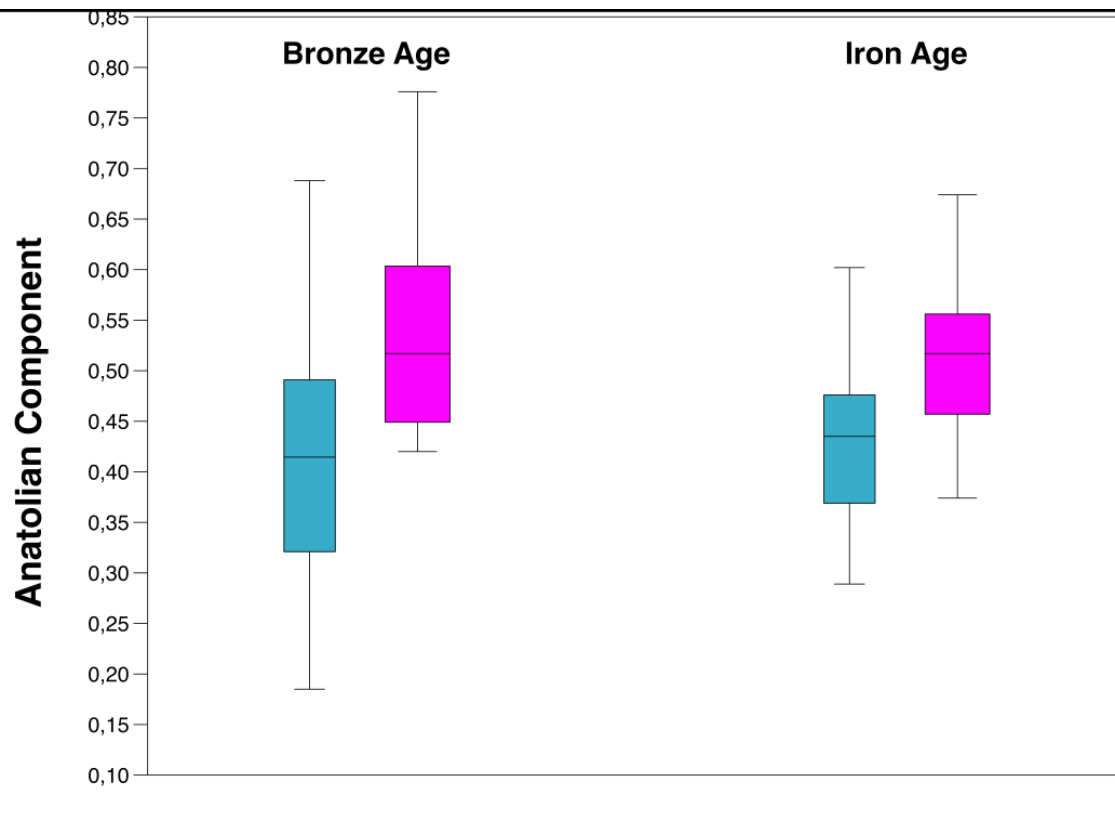


Figure S8 : Evolution of $qpAdm$ values between the Bronze and Iron Age periods for the Early neolithic farmer ancestry component, related to Figure 3

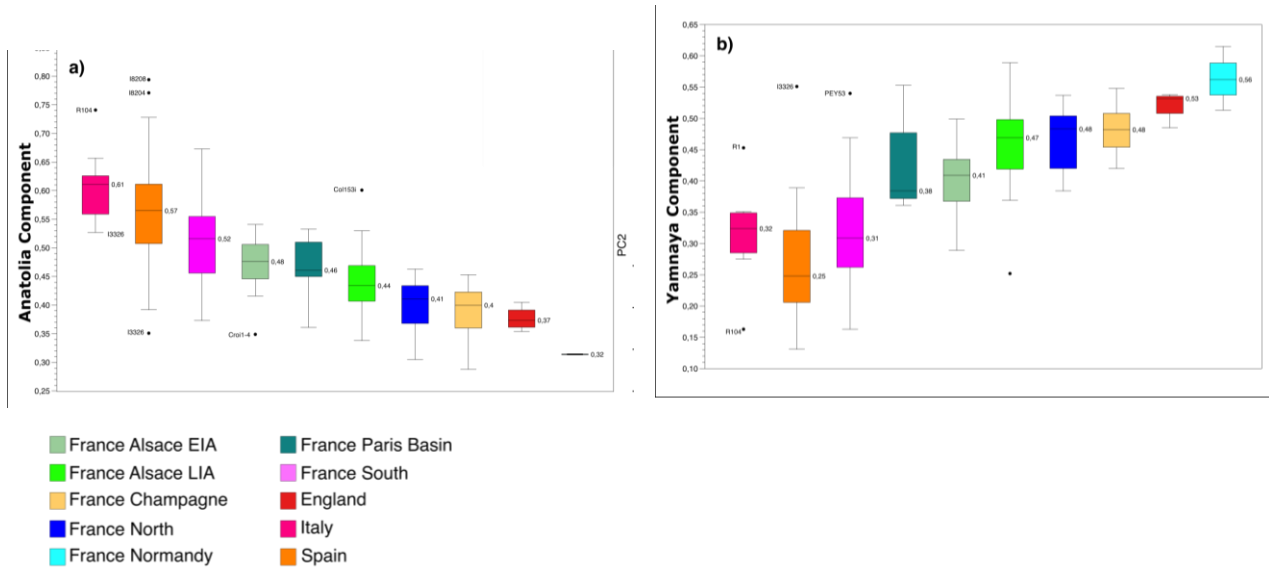


Figure S9: Distribution and average of ancestral component in IA samples from western Europe. Each colour corresponds to a region (France) or a country. Biplot of $qpAdm$ value within regional groups for the Anatolia Neolithic ancestry component (A) and for the Yamnaya ancestry component, related to Figure 3

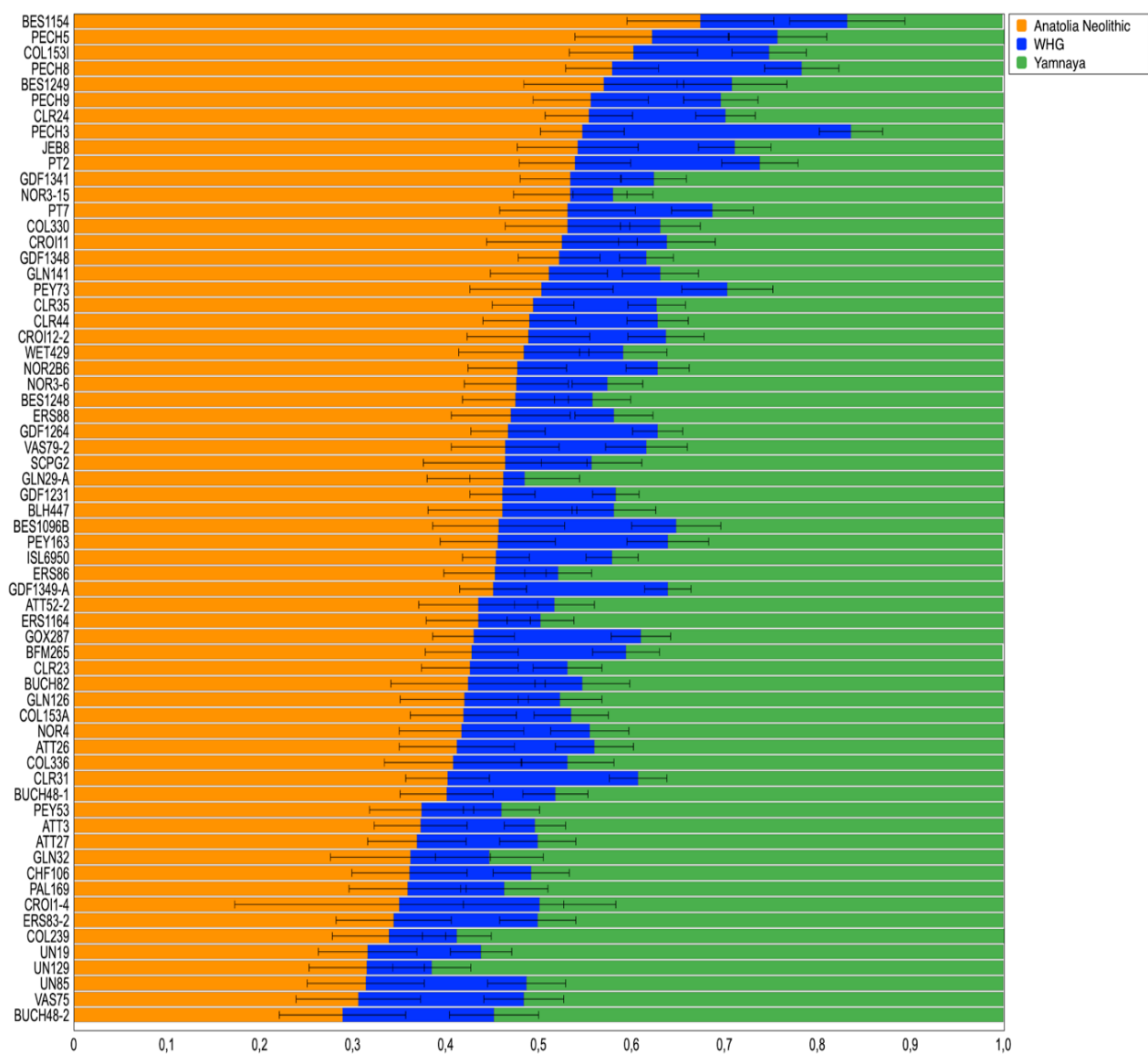


Fig S10 – qpAdm value for each individual with a 3 sources model (WHG, Anatolia N and Yamnaya), related to Figure 3

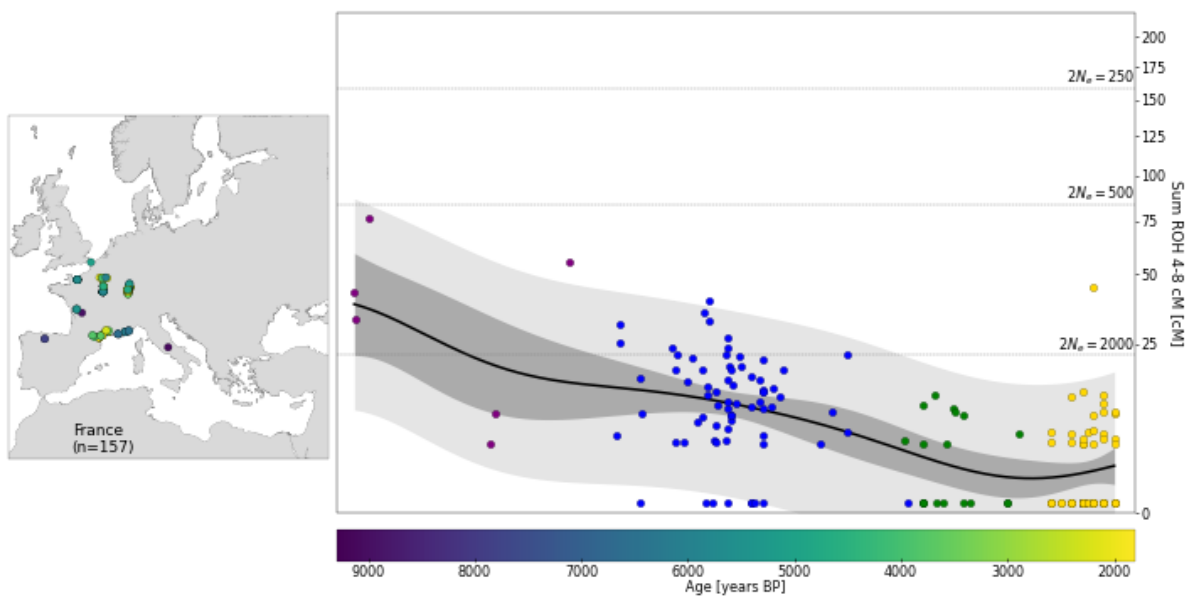


Fig S11- Sum of ROH (sROH) measured for all published and newly sequenced individuals from France dated between 9000BP and 2000 BP. Each circle in the main panels represents sROH for one ancient Individual. We also show mean estimates calculated from a Gaussian Process model (solid black line), and 95% empirical confidence intervals for individuals (light gray) and for the estimated mean (dark gray). Horizontal lines depict the theoretical expectations for sum of ROH blocks 4-8 cM for panmictic population sizes calculated using analytical formulas (see Ringbauer *et al.*, 2021), related to STAR methods

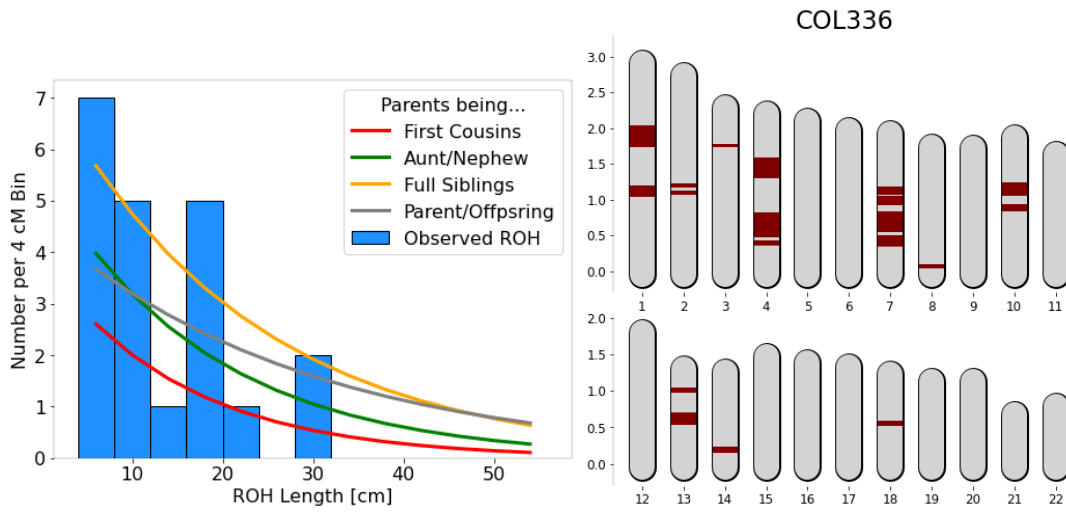


Fig S12 - ROH in COL336. We depict all ROH of COL336 (right panel), with map length annotated in Morgan. We also show a histogram of the ROH lengths, together with expected densities of ROH for certain degrees of parental relationship, related to STAR methods