Reshaping the *Hexagone*: the genetic landscape of modern France

Simone Andrea Biagini[1], Ángel Carracedo[2,3,4], David Comas[1], Francesc Calafell[1,*]

1. Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSIC-UPF), Universitat Pompeu Fabra, Barcelona, Catalonia, Spain.

2. Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Spain

3. Grupo de Medicina Xenómica (GMX), Faculty of Medicine, University of Santiago de Compostela, Spain

4. Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia

* Address correspondence to Francesc Calafell, francesc.calafell@upf.edu, Departament de Ciències Experimentals i de la Salut, Institute of Evolutionary Biology (CSIC-UPF), Universitat Pompeu Fabra, Carrer Doctor Aiguader 88, 08005 Barcelona, Catalonia, Spain.

Abstract

Unlike other European countries, Metropolitan France is surprisingly understudied. In this work, we combined newly genotyped samples from various zones in France with publicly available data and applied both allele frequency and haplotype-based methods in order to describe the internal structure of this country, taking advantage of the Human Origins SNP array, specifically designed for human population genetics studies. We found out that French Basques are genetically distinct from all other populations in the *Hexagone* and that the populations from southwest France (namely the Franco-Cantabrian region) are intermediate between Basques and other populations. Moreover, Bretons slightly separated from the rest of the groups and a link with the historical gene flow from the British Isles has been found. Results from the allele frequency analyses point to a general background that appears to be a mixture of two components, one closer to Southern Italy and the other to Ireland. This combination may be the result of a contact happened in two different moments: in the Early Neolithic, and then Ireland would be a proxy for the continental pathway for the Neolithic wave of advance and South Italy for the coastal penetration, or the Iron Age, when the Celtic and the Mediterranean worlds met in France. On the other hand, results from the haplotype-based methods describe a

more structured landscape, highlighting the presence of areas characterized by differential links with the neighboring populations, possibly reflecting a more recent history.

Introduction

Located in the center of Western Europe, Metropolitan France acts as a bridge connecting Northern Europe to the Mediterranean and the Iberian spaces. Nowadays, France is a cosmopolitan country whose society is shaped by a plurality of lifestyles and truly different ethno-cultural diversity. Without any doubt, the impact immigrations from colonised countries to mainland France, such as the migration of Arabs and Berbers from Algeria which was the most extensive of all colonial migrations to Western Europe before the 1960s (MacMaster, 1997), enriched the modern genetic landscape of the French territory. However, it is beyond our intention to explore this genetic contribution here, which can be quantified much more precisely with demographic analyses. Instead, we are more keen to dig into a deeper and ancient genetic background. The geographical position of France strongly affected the history of the settlement of the different parts of the territory, whose continuous fragmentation through time is attested by the large number of populations and cultures that settled this area. Greeks, Romans and Celtic tribes from central Europe shaped a first internal structure between the 6th and the 1st centuries BCE, while waves of barbarian invasions (Alamanni, Burgundians, Visigoths, Franks, and Celts) harshly impacted the population landscape of France during the 5th century CE (Haine, 2000). During the 9th and 10th centuries CE, foreign invasions from all sides also influenced the territory: Muslims and Saracens from North Africa coming through Iberia, Hungarian Magyar from the east, and Vikings (*Northmen*) from the north (Haine, 2000). At the light of this complex past, the genetic landscape of France has been poorly analyzed, especially in recent times. First studies with classical markers defined a general heterogeneous pattern considering different geographical arrangements such as military districts, historical provinces, and regions (Kherumian *et al.*, 1967; Cambon-Thomsen *et al.*, 1988). With his synthetic maps, Cavalli-Sforza proposed that this heterogeneity was a consequence of differential Neolithic influences between northern and southern France, and also pointed out a differentiation for Brittany and the Franco-Cantabrian region (Cavalli-Sforza *et al.*, 1994). More recently, studies on mitochondrial DNA highlighted a general homogeneity when the samples were distributed among the 22 regions

established in 1982 and historic provinces (Dubut *et al.*, 2004; Richard *et al.*, 2007). Generally, the mtDNA haplogroup composition of French people did not differentiate neither internally, nor from the surrounding European genetic landscape (Dubut *et al.*, 2004; Richard *et al.*, 2007). On a microgeographical scale, Brittany showed affinity with Scandinavia and Britain, while French Basques stood out for a high frequency of haplogroup H, suggesting a link with the Neolithic diffusion in Europe (Dubut *et al.*, 2004; Richard *et al.*, 2007). In agreement with the homogeneity described by mtDNA studies, Y-chromosome strongly pointed out a lack of differentiation between the distinct groups when samples were organized on a regional scale. Even in this case, Brittany represented an exception, showing a lower Y-chromosome diversity that was interpreted as consequence of a possible founder effect, plus an isolation process (Ramos-Luis *et al.*, 2014). Based on autosomal variants, a genome-wide study on Western France did not find any differentiation among the distinct groups organized on a regional geographical distribution. Even in this case, the only outlier was Brittany, whose higher linkage disequilibrium suggested a lower effective population size, thus supporting the hypothesis of isolation inferred by the outcomes of the Y-chromosome analyses (Karakachoff *et al.*, 2015). Furthermore, in agreement with mitochondrial studies, Bretons were found to be admixed with individuals from the British Isles (Karakachoff *et al.*, 2015). In this work, we present a comprehensive genome-wide study on France with the aim to define patterns of internal differentiation using both allele frequency and haplotype-based methods, presenting a more heterogeneous geogenetical landscape.

Material and Methods

Dataset arrangement and genotypes

In this study, informed consent was obtained from 331 individuals from different French departments. Internal Review Board approval for this work was granted by CEIC-PSMAR ref. 2016/6723/I. These samples were first reported in an analysis of Y-chromosome markers in Ramos-Luis *et al.* 2014. DNA was extracted from blood samples as described in Ramos-Luis *et al.* (Ramos-Luis *et al.*, 2014). A total of four Axiom ®

Genome-Wide Human Origins Arrays (~629 K SNPs) (Patterson *et al.*, 2012) were genotyped at the Centro Nacional de Genotipado - Universidade de Santiago de Compostela facility. Genotype calling was performed running four different batches according to the Affymetrix Best Practices Workflow implemented in the software Axiom™ Analysis Suite 2.0. Out of 331 samples, 52 failed the genotyping process and a total of 279 samples were retained. Three additional samples were removed following an Identity-by-descent analysis (IBD) since they displayed a Proportion IBD value ≥ 0.125 (minimum threshold for removing relatedness equal or higher than a third degree). Eventually, 276 samples were retained. To complete the French dataset, 79 additional samples from a public source (Lazaridis *et al.*, 2016a) and 60 from unpublished data (from an ongoing study on the Basque Country and the Franco-Cantabrian region; samples are subset from those in Martínez-Cruz *et al.*, 2012) were added to the original 276, leading to a total of 415 samples. In a preliminary part of this work, 20 out of the 276 samples were identified as outliers and removed from the study (see Supplementary Figure1 and caption). Thus, the complete dataset included 256 newly genotyped samples, plus 139 additional ones, for a final group of 395 samples distributed among 20 different French departments (see Figure 1 for the geographical distribution). For the allele frequency analyses, as comparison with external populations, 218 samples among Germany, Norway, Spain, Italy, England, Ireland, and Scotland were used from public data (Lazaridis *et al.*, 2016a), together with 107 samples from the Spanish autonomous communities of Catalonia, Valencian Community, and Balearic islands from public data (Biagini *et al.*, 2019), and 8 additional samples from South Italy (Naples) newly genotyped with Axiom ® Genome-Wide Human Origins Arrays (~629 K SNPs) and presented in this study for the first time. A further 797 samples from external populations (Lazaridis, Nadel, Rollefson, Merrett, *et al.*, 2016) were added to the previous ones when applying haplotype-based methods (see Supplementary Figure 2 for the external samples distribution, and Supplementary table 1 for a summary of the dataset composition).

Data Quality Control

Data were prepared using PLINK1.9 (Purcell *et al.*, 2007). Uniparental markers and X-chromosome variants were excluded. For the French dataset, a preliminary set of filters were applied to each group separately before the merging process. We filtered out all variants with missing call rates greater than 5%, those that failed Hardy-Weinberg test at

$p < 10^{-5}$, and samples with more than 10% missing genotype data. After merging, only variants common to the three datasets were retained and SNPs with a minor allele frequency (MAF) below 5% were excluded, resulting in a final 343,884 variants used for haplotype-based methods. For the analyses that needed a set of independent markers, SNPs were pruned setting a pairwise linkage disequilibrium maximum threshold of 0.5, a window of size 200 and a shift step of 25. Eventually, the pruned data retained 142,803 variants. In the analyses that included the external populations, only the pruned dataset, consisting in 154,889 SNPs, was used for the allele frequency analyses, while a set of 380,697 variants was retained in the haplotype-based methods.

Statistical analyses

Eigenvectors were computed using the SmartPCA program in Eigenstrat software package (v. 13050) (Patterson *et al.*, 2006). Results were plotted in R (v 3.0.1).

The F_{ST} Fixation index was computed using the SmartPCA tool (v. 13050) from the Eigenstrat software package. Results were produced in Rstudio (RStudio Team, 2015) using R version 3.4.4 (R Core Team, 2018). The F_{ST} matrix was used together with a geographic distance matrix produced with The Geographic Distance Matrix Generator (v. 1.2.3, available from http://biodiversityinformatics.amnh.org/open_source/gdmg) in order to perform a Mantel test correlation using the ade4 (Dray *et al.*, 2007) library in R. Results were displayed using ggplot2 (Wickham, 2009) and reshape (Wickham, 2007) libraries.

Based on different hierarchical levels (within Departments, Between Departments within Areas/Regions, Between Areas/Regions; see Supplementary Figure 3 for a visual representation of the used Areas and Regions), AMOVA was performed using the *poppr.amova* function in R package poppr (v. 2.8.1) (Kamvar *et al.*, 2014, 2015) and significance was tested with the *randtest* function implemented in R package ade4. For every percentage of variance, a p-value was calculated based on 1000 permutations.
Patterns of population structure were explored using ADMIXTURE (Alexander *et al.*, 2009) testing from K=2 to K=10 ancestral clusters and using 10 independent random seeds. Results were represented using the software pong (Behr *et al.*, 2016). Admixture

was formally tested with f3 statistics computed using the qp3Pop function implemented in Admixtools (Patterson *et al.*, 2012).

EEMS (Estimated Effective Migration Surface)

EEMS (Petkova *et al.*, 2016) analysis was run on the 395 samples French dataset using 142,803 variants from the pruned file. With a matrix of average pairwise genetic dissimilarities calculated using the internal program bed2diffs, a sample coordinates file, and a habitat coordinates file generated using Google Earth Pro (v. 7.3.2.5495), we performed 10 pilot runs of 6 million MCMC iterations each, with 3 million burn-in, and a thinning interval of 30,000. A second set of 5 runs was then performed restarting the chain with the highest likelihood with 4 million MCMC iterations, 1 million burn-in, and thinning interval of 10,000. The density of the population grid was set to 300 demes, and random seeds were used for each one of the runs. We used the default hyperparameter values but tuned some of the proposed variances to improve convergence in the second set of runs. Results for the chain with the highest likelihood were displayed using eems.plots function in the R package rEEMSplots.

Haplotype-based analysis

Two different analyses were performed: one on the internal French population only (including 395 samples and 343,884 variants), and one also including external populations (consisting in a total of 1527 individuals and 380,697 variants). In both cases, phasing was performed using the software Shapeit (v. v2.r837) (Delaneau *et al.*, 2014; O'Connell *et al.*, 2014). When running ChromoPainter (Lawson *et al.*, 2012), all samples were used as both recipients and donors, (Lawson *et al.*, 2012)without any population specification (-a option) and not allowing self-copying. First, the parameters for the switch rate and global mutation probability were estimated using the EM algorithm implemented in ChromoPainter using the parameters -i 15 -in -iM for chromosomes 1, 7, 14, and 20 for all the samples. This step allows to estimate the two parameters that will be then averaged for all chromosomes. The outcome for the average weighted values for the global mutation probability and the switch rate parameters were respectively 0.000745 and 266.67196 for the dataset with 395 samples, and 0.000586 and 237.50784 for the dataset with 1527 samples. In a second step, ChromoPainter was run for all chromosomes

using the two fixed parameters. Later, the final coancestry matrices for each chromosome were combined using the tool Chromocombine. The latter also estimates the C parameter which is needed for the normalization of the coancestry matrix data when we run fineSTRUCTURE in order to identify the population structure. The MCMC of fineSTRUCTURE was run using 1000000 burn-in iterations (flag -x), 2000000 iterations sampled (flag -y), and thinning interval of 10000 (flag -z). Eventually, the fineSTRUCTURE tree was estimated running three different seeds and using the flags -X -Y -m T that allow to build the sample relationship tree. In the analysis with the external populations, the work was then divided in two paths. In the first one, ChromoPainter and fineSTRUCTURE were rerun, this time silencing France in order to define the external groups only. In the second case, fineSTRUCTURE was rerun using the "force file" option (-F), using "continents" as donor groups, a function that allows to exclude the donor representation in the building tree phase and focus on the distribution of the recipient groups, represented by the French samples only. We then applied the non-negative-least-squares (nnls) function from GLOBETROTTER in order to describe the ancestry profiles for the French groups we detected with the "force file" option.

Results

Internal genetic structure in France

In order to define the best geographical partitioning of genetic differentiation, a hierarchical analysis of molecular variance (AMOVA) was performed with areas or regions as major grouping factors. In the first AMOVA we determined the proportion of genetic variation partitioned among geographic areas, among departments within geographic areas, and within departments. In the second one we tested the proportion of genetic variation partitioned among regions (considering the 13 regions established in 2016), among departments within regions, and within departments. A further AMOVA was performed only testing the proportion of genetic variation partitioned among and within departments. As shown in Table 1, in all cases the main contribution to the genetic variance was found at the lowest hierarchical level (variation within departments), while differences among regions resulted in a negative value that could be interpreted as zero,

meaning absence of any structure at this level. Conversely, differences among areas displayed positive values, supporting the role of areas as more reliable grouping factors of genetic variations when considering wider sample distributions. Finally, the results for the variation between departments, also supported by significant p-values in all the AMOVA analyses, pointed to the fact that this level of stratification might be a better representation for the minimal unit of genetic differentiation. Based on these results, samples were distributed on the map according to the departmental locations (Figure 1) and all the subsequent analyses considered this grouping factor. A first Principal Component Analysis (PCA) showed two distinct groups separated along the first PC (Figure 2A): the Basque samples on the right part of the plot, against most of the rest of the samples on the left one, within which a structure cannot be defined. These two major groups are connected by a "bridge" of samples represented by non-Basque-speaking individuals from the Franco-Cantabrian region in the southwestern corner of France. When we averaged the eigenvalues for the first two PCs and represented the same PCA, together with standard deviation (SD) values for each group, no evident pattern could still be discerned (Supplementary Figure 4A). When we removed both Basque- and non-Basque-speaking Franco-Cantabrian samples from the analysis (Figure 2B), the resulting PCA showed some internal pattern of differentiation, more clearly defined by the average PCA (Supplementary Figure 4B), in which samples from the departments belonging to the northwestern region of Brittany seem to form a cluster on the left part of the plot.

Patterns of gene flow within France

In the genetic variation computed with the F_{ST} analysis, a general homogeneous pattern was found, with fine scale values of differentiation between some departments. The Franco-Cantabrian samples showed the highest values of differentiation with the northwestern departments reaching scores between 0.008 and 0.009 for the Basque-speaking samples, and between 0.004 and 0.006 for the non-Basque-speaking ones (Supplementary Figure 5A, left), followed by lower values of differentiation with the northern and northeastern departments. Without the Franco-Cantabrian samples, the main differentiation was recorded between the northwestern departments and the southeastern corner of the country, with a highest value of differentiation around 0.002 between the southeastern department of Bouches-du-Rhône (BdR) and the northwestern Breton department of Côtes-d'Armor (CdA) (Supplementary Figure 5B, left). Lower levels of

differentiation were locally found among the departments in the northwest, among those in the north together with the northeastern ones, and among the Basque-speakers Franco-Cantabrian groups. A Multidimensional Scaling analysis (MDS) based on the F_{ST} matrices clearly showed how the Franco-Cantabrian samples separate from the rest of the groups (Supplementary Figure 5A, right), and how the Breton departments do the same once the Franco-Cantabrian samples are removed (Supplementary Figure 5B, right). A Mantel test of isolation by distance (IBD) between the F_{ST} values and the geographical distances showed a positive and statistically supported correlation (R^2=0.332, P=0.001) (Supplementary Figure 6A), moving to even more positive values when the Franco-Cantabrian samples were removed (R^2=0.432, P=0.001) (Supplementary Figure 6B). Next, we used the EEMS analysis, a method for visualizing genetic diversity patterns, and found that the resulting effective migration surface mirrors the outcomes of genetic differentiation detected by the F_{ST} analyses (Figure 3); a higher effective migration was locally found in northern, northeastern and northwestern France among departments belonging to the same geographical areas, while a major barrier was discovered along the western side of France.

Sources of gene flow into France

When we added external sources from the surrounding populations (yellow dots in Supplementary Figure 2 representing the 333 samples for the allele frequency analyses) to test for signatures of admixture events, the configuration observed pointed to a general homogeneous picture. The only exception was represented by the samples belonging to the Breton departments whose configuration was more like the one observed for the Irish, Scottish, and English groups. Moving through the different K ancestral components, this behavior clearly characterizes the northwestern departments, separating them from the rest of the French groups since from the very first K ancestral components (Supplementary Figure 7). Thus, we formally tested for admixture events using the f3-statistics with the test groups being the different departments, and the external surrounding populations as sources. We only retained the negative f3 values for those departments represented at least by two individuals. Results are shown in Table 2 were only significant Z-scores < -3 are reported, while results for those departments passing all the requested filters but with higher Z-score values are shown in Supplementary Table

2. Notably, in 9 departments, a combination of sources that was highly significant was Ireland-Southern Italy (see discussion below).

Haplotype sharing patterns within France

Using haplotype-based methods, we looked for patterns of haplotype sharing, illustrating relations between departments. In this first step, cutting the fineSTRUCTURE tree at the very base, allowed us to describe a fine scale haplotype sharing distribution on a departmental scale; the outcome is a picture of the haplotype configuration within France (Figure 4). The resulting map is clearly consistent with the results seen so far; we can define at least four distinct groups, plus a more widespread component. In the southwestern corner, the Franco-Cantabrian samples clearly split in two groups, represented by the Basque-speaking and non-Basque-speaking subjects, respectively. In the northwestern vertex, the Breton departments exhibit their very own haplotypic signature, in agreement with the lower level of differentiation detected with the F_{ST} analysis and the higher internal effective migration rate detected with EEMS. The same was found for the northern and northeastern departments that display a clearly shared haplotypic configuration. Furthermore, a more generally spread "French haplotype pattern" is found on the north-south axis. Lastly, also the southwestern department of Haute-Garonne (HG) and the southeastern one of Bouches-du-Rhône (BdR) present higher frequencies for some local haplotypes that in other departments reached only lower frequencies.

Haplotype sharing patterns with external sources

Based on the haplotype sharing with external sources it was possible to redefine the French haplotype configuration. After merging the 395 French samples with the 1132 external ones, we first defined the external groups by silencing France when rerunning ChromoPainter and fineSTRUCTURE. The result was represented by 35 different external groups (Supplementary Figure 8a). Secondly, focusing on our target, we redefined the French internal clusters (Supplementary Figure 8b). The 11 different clusters we found where then represented as separate maps (Supplementary Figure 9);

each map is a heatmap showing the number of samples falling in the different departments. We named the clusters based on the fact that a specific area was more represented than others in terms of sample size. Three of the 11 clusters were excluded as targets in the following analysis, since they were both poorly represented in terms of sample size and also not representative of a specific main geographic area within France (the exclusion of these three clusters did not represent a significant issue, since only the 10% of the French samples were then not included as target in the ancestry profile analysis). As in the analysis described in the previous paragraph, even in this case the main areas of interest were mostly representing the Northwest, the Southwest (divided in Basque-speaking and non-Basque-speaking groups), the Northern and Northeastern areas, the Southeast, and a central/southwestern part of France (Supplementary Figure 10). These eight main areas represented the targets for the GLOBETROTTER nnls analysis that we used to describe the ancestry profiles for these same clusters.

Ancestry profiles

The results from the application of the nnls algorithm are displayed in Figure 5 and Supplementary Figure 11. Figure 5 is showing the ancestry profiles for each one of the eight French clusters as a doughnut chart; the different colors represent proportions of haplotype sharing with specific groups (for each ancestry profile, only contributions above 2.5% are shown). All the pink shades represent proportions of the different French groups, while the remaining colors describe the ones from external sources. Focusing on the latter (lower panel in Figure 5), it is possible to define differential patterns of contributions from those neighboring populations that are more geographically close to specific areas within the French territory. The southwestern areas are receiving more from the Spanish side, the Northwestern part is sharing more with the external cluster named “Irish_Scottish” (with a proportion of 30.37%), the northeastern departments are more connected to the external clusters representing central and eastern European countries (a cluster we named “Central/Eastern_Europe”), while the southeastern area is mostly connected to the Italian clusters and other Mediterranean countries, and the central/southwestern part of France is clearly receiving more from both Spain and Italy.

Discussion

In order to describe the internal structure of Metropolitan France, in this work we used both allele frequency and haplotype-based methods. While the first described a more homogeneous landscape, the latter unveiled patterns of local differentiation with some connections with the surrounding European populations. In previous works about France, samples were differently arranged into the geographical space and no consensus had been reached on what system would work better; apart from less canonical systems of organization like the military districts (Kherumian *et al.*, 1967), historical provinces (Cambon-Thomsen *et al.*, 1988; Richard *et al.*, 2007) and old regions (Ramos-Luis *et al.*, 2014) are the most used so far. Thus, our first goal was to search for the best geographical level of genetic stratification before arranging our samples on a map. After the French Revolution in 1790, in order to weaken the old loyalties, the ancient provinces of France were subdivided into departments, whose overall configuration has been mostly conserved so far (Forstenzer, 1981). Furthermore, in 1982, a system of 22 regions was established by grouping different departments into wider areas (Sowerwine, 2009). However, in 2016, the number of the regions was reduced to 13, with the consequent rearrangement of the departments (OECD, 2017). Given this background, our AMOVA results support the idea that regions, as a new internal reorganization, are not a suitable model for the genetic compartmentalization and point to the absence of any contribution to the total genetic variation, possibly implying that regions are separating genetically similar departments into different groups. On the other hand, departments, as result of a more conserved internal geographical structure, represent the best minimal unit of genetic stratification. Using this system of internal organization, we found a truly remarkable connection between the different results we achieved, but it was with fineSTRUCTURE that we could really define the first ever detected internal subdivision of France (Figure 4). A general widespread "French haplotype" moving through the north-south axis was detected; possibly the overall homogeneity found with the principal component analysis can be linked to the fact that on an allele frequency scale, such widespread pattern may represent a confounding factor. Indeed, only the two Franco-Cantabrian groups were not reached by this common "French haplotype". These two populations clearly differentiated from the rest of the French groups for both allele frequency and haplotype-based methods. It is interesting to notice that the presence of two distinct groups in the Franco-Cantabrian region stressed the outcome of the isolation the Basque-speaking groups experienced, splitting from their non-Basque-speaking neighbors from the very

same departments. This finding is in agreement with their recognized distinct cultural entity (Calafell *et al.*, 1994) and their genetic outlier position in the European landscape (Rodríguez-Ezpeleta *et al.*, 2010), as also with the lower internal levels of differentiation we detected with the F_{ST} analysis, and the low effective migration rates evidenced by EEMS, resulting in a barrier to migration in the southwestern corner of France. The latter was found to span along the entire western French coast, defining a second barrier in the northwestern corner, justifying the presence of another distinct group represented by the Breton departments. This group was firstly detected, on a coarser scale, with the removal of the Franco-Cantabrian samples from the first PCA, and its outstanding position is in agreement with different studies on both uniparental and autosomal markers (Dubut *et al.*, 2004; Richard *et al.*, 2007; Ramos-Luis *et al.*, 2014; Karakachoff *et al.*, 2015). However, based on the fineSTRUCTURE results, in our work we detected a stronger evidence of differentiation based on haplotypic data. Furthermore, the barrier detected with EEMS on the western side of France is supported by the highest level of genetic differentiation between the Breton and the Franco-Cantabrian groups. Such levels of F_{ST} separated geographically more distant groups, in a corner-to-corner scheme inside the *hexagone*, thus justifying the positive results obtained with the Mantel test. Also, the genetic distances the Franco-Cantabrian and Breton samples show span through the entire geographical space, skewing the more homogeneous samples in the lower part of the distribution (Supplementary Figure 6A), defining two distinct groups, separated by the trend bar. Furthermore, the barrier in the northwestern corner is consistent with the lower levels of allele sharing we found between the Breton departments and the southeastern corner of France. Apart from some possible artifacts due to the presence of unsampled regions (House *et al.*, 2018), we found the EEMS results consistent with the rest of the data collected in our study. Indeed, another group with a higher internal migration rate according to the EEMS results is represented by the north and the northeastern departments. Samples from these areas showed a shared haplotypic configuration as detected by fineSTRUCTURE and lower levels of internal differentiation according to the F_{ST} results, pointing to a preferential internal gene flow in this area. In the southern part of France, two departments, Haute-Garonne (HG) in the southwest and Bouches-du-Rhône (BdR) in the southeast also showed higher frequencies for some local haplotypes.

In order to understand whether these internal patterns of differentiation are due to recent events or whether they reflect a more ancient history, we relied on two different analyses

obtaining distinct information. On the one hand, we performed an admixture analysis, and, on the other hand, we described the ancestry profiles for eight different French clusters defined by the haplotype sharing with external sources. The results we obtained highlight the different power of resolution allele frequency and haplotype-based methods allow to reach in human population genetics studies.

Echoes from the past

With the admixture analysis, a more ancient and homogeneous background was found. The homogeneity we detected affected all the French departments, with one important exception: Brittany, whose connection to the Irish samples completely agrees with previous findings (Dubut *et al.*, 2004; Ramos-Luis *et al.*, 2014; Karakachoff *et al.*, 2015). Historical migrations from Ireland to Brittany are well recorded since the 4th century CE (Monnier, 1997), as well as the emigration of Irish people during the War of Ireland (1641-1651) into the present day departments of Finistère (FI) and Côte d'Armor (CdA), within which a higher integration of the Irish immigrants is proved by records of marriage, birth and death certificates (Dubut *et al.*, 2004). Additionally, the Irish admixture signal may be a proxy for the Celtic British (mostly Welsh and Cornish) migration into Brittany also in the 4th century CE (see below). Furthermore, a Celtic root for the Breton language links the Breton departments to the Insular Celtic languages from the British Isles (Forster *et al.*, 2003). One of the most surprisig outcomes of our work is undoubdetly represented by the f3-statistics; 9 out of 22 distinct targets we tested against different external sources gave significant results with the lowest Z-scores detected for the same couple represented by the Italian South and Irish sources. Z-scores lower than -3 indicate that our test populations are admixed from sources not necessarily identical but related to the sources we used in the analysis (Lazaridis *et al.*, 2016b). Thus, these two sources may be proxies, respectively, for Mediterranean and continental components. Further analyses are required to restrict the time frame when this admixture occurred, but we may hypothesize an Iron Age infusion of Mediterranean Greeks and continental Celts, or two different paths followed by the Neolithic expansion, either along the Mediterranean or through the Balkans and Central Europe. According to this model, the general outcome is a widespread homogenous background that link the entire French territory to pre-historic events; by 600 BCE Greeks established a colony on the Mediterranean coastline of France in the city of Massalia (present-day Marseille) (Fine, 1985); around the same time (ca.

700 BCE) Celtic tribes from central Europe (subsequently known as Gauls) expanded across the entire territory bringing their culture, advanced farming tecniques, and their language (Hubert, 1989).

A more recent history

As we previously pointed out, the application of haplotype-based methods produced a more complex outcome, adding different leyers of information on the more recent history of France. The ancestry profiles described in Figure 5 are informative of differential migratory patterns (Leslie *et al.*, 2015) into each of the eight French genetic clusters. The ancestry profiles are a way to describe the DNA of each one of the eight French clusters as a mixture of DNA from other groups, without inferring any admixture (Hellenthal, 2015). With this analysis, each cluster is described as a composition of different proportions of haplotype sharing with other groups (both French and external ones), excluding the contribution of the group that we want to explain (no self-copying allowed). With this analysis we detected the presence of a strong French identity (all the pink shades in Figure 5) that somehow dominates the entire country. However, higher proportions of external contributions differently affected mostly peripheral areas of France, strongly connecting each one of them to the more geographically close neighboring countries, and with the history they share. Further analyses will help to define the time frame for the different events but so far, the knowledge of history can definetly help in suggesting an interpretation. Apart from the northwestern departments of Brittany, the external contributions to the remaining clusters could be a reflection of the long history of immigration that France experienced in recent times. With the Industrial Revolution in the 19th century a deep labor-related immigration brought people from different countries, mostly from Poland, Italy, Belgium, Portugal, and Spain (Hassell, 1992). Between the second half of the 19th century and the end of WWI, the percentage of immigrants permanently living in the French territory tripled (Schain, 2008).The role France had during those times as asylum for political refugees fleeing repressive regimes in the Western world is undeniable, and our results might reflect this rich heritage. With a proportion of 12.27% from the external cluster named “Central/Eastern_Europe” (also including samples from Germany), the northeastern departments of Bas-Rhin and Moselle in our dataset recall the long history of the Alsace-Lorraine territory: a fuzzy border between France and Germany for a long time, and only recently retroceded to

France in 1945 (Blumenthal, 2012). As for the northwestern departments of Brittany, the higher contribution (30.37%) from the external source we named "Irish_Scottish" suggests a more ancient history, possibly associated to the British Celtic migrations (from Cornwall and south-west Britain) into the Gaulish Armorica (then renamed Brittany) from the 3rd to 9th centuries CE, with a higher flow between the 5th and the 6th centuries CE (Koch, 2005).

Conclusions

In conclusion, our work aims to povide a first genetic profile of France through the application of methods with different power of resolution, gaining a higher support from the Human Origins Arrays, reliable and recognized tool for anthropological genetics studies. According to the allele frequency analyses, the French genetic landscape seems to be homogeneously dominated by an ancient background, with cultural isolates in the southwest and a more recent gene flow into the northwest, but also with subsequent isolation. On the other hand, the application of haplotype-based methods allowed us to look beyond the more homogeneous "French identity", discovering connections with the neighbouring populations, possibly reflecting the more recent history of this understudied European country.

ACKNOWLEDGMENTS

We thank Inés Quintela (CEGEN - USC) for her assistance in genotyping the samples. Of course, this work would not have been possible without the kind collaboration of all

the sample donors. Funding was provided by the Agencia Estatal de Investigación and Fondo Europeo de Desarollo Regional (FEDER) (grant CGL2016-75389-P), Agència de Gestió d'Ajuts Universitaris i de la Recerca (Generalitat de Catalunya) grant 2014 SGR 866, and "Unidad de Excelencia María de Maeztu", funded by the MINECO (ref: MDM-2014-0370). SAB was supported by the Agencia Estatal de Investigación FPI grant BES-2014-069224.

References

Alexander, D. H. and Novembre, J. (2009) 'Fast Model-Based Estimation of Ancestry in Unrelated Individuals', *Genome Research*, 19(9), pp. 1655–1664. doi: 10.1101/gr.094052.109.vidual.

Behr, A. A. *et al.* (2016) 'Pong: Fast analysis and visualization of latent clusters in population genetic data', *Bioinformatics*, 32(18), pp. 2817–2823. doi: 10.1093/bioinformatics/btw327.

Biagini, S. A. *et al.* (2019) 'People from Ibiza: an unexpected isolate in the Western Mediterranean', *European Journal of Human Genetics*. Springer US. doi: 10.1038/s41431-019-0361-1.

Blumenthal, D. (2012) *Alsace-Lorraine: A Study of the Relations of the Two Provinces to France and to Germany, and a Presentation of the Just Claims of Their People*. Forgotten Books.

Calafell, F. and Bertranpetit, J. (1994) 'Principal component analysis of gene frequencies and the origin of Basques', *Am J Phys Anthropol*, 93(2), pp. 201–215. doi: 10.1002/ajpa.1330930205.

Cambon-Thomsen, A. and Ohayon, E. (1988) 'Practical Application of Population Genetics: The Genetic Survey "Provinces Françaises"', in Mayr, W. (ed.) *Advances in Forensic Haemogenetics. Advances in Forensic Haemogenetics, vol2*. Berlin, Heidelberg: Springer.

Cavalli-Sforza, L. L., Menozzi, P. and Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton: Princeton University Press.

Delaneau, O. *et al.* (2014) 'Integrating sequence and array data to create an improved

1000 Genomes Project haplotype reference panel', *Nature Communications*. Nature Publishing Group, 5, pp. 1–9. doi: 10.1038/ncomms4934.

Dray, S. and Dufour, A.-B. (2007) 'The **ade4** Package: Implementing the Duality Diagram for Ecologists', *Journal of Statistical Software*, 22(4). doi: 10.18637/jss.v022.i04.

Dubut, V. *et al.* (2004) 'mtDNA polymorphisms in five French groups: Importance of regional sampling', *European Journal of Human Genetics*, 12(4), pp. 293–300. doi: 10.1038/sj.ejhg.5201145.

Fine, J. V. A. (1985) *The Ancient Greeks: A Critical History*. Belknap Press: An Imprint of Harvard University Press.

Forstenzer, T. R. (1981) *French Provincial Police and the Fall of the Second Republic: Social Fear and Counterrevolution*. Princeton University Press.

Forster, P. and Toth, A. (2003) 'Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European', *Proceedings of the National Academy of Sciences*, 100(15), pp. 9079–9084. doi: 10.1073/pnas.1331158100.

Haine, W. S. (2000) *The History of France*. Greenwood Press.

Hassell, J. E. (1992) *Russian Refugees in France and the United States Between the World Wars (Transactions of the American Philosophical Society)*. Amer Philosophical Society.

Hellenthal, G. (2015) 'Instruction Manual for " GLOBETROTTER : a program for identifying , dating and describing admixture events in population data " for GLOBETROTTER', pp. 1–24.

House, G. L. and Hahn, M. W. (2018) 'Evaluating methods to visualize patterns of genetic differentiation on a landscape', *Molecular Ecology Resources*. doi: 10.1111/1755-0998.

Hubert, H. (1989) *The Greatness and Decline of the Celts*. Marboro Books.

Kamvar, Z. N., Brooks, J. C. and Grünwald, N. J. (2015) 'Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality', *Frontiers in Genetics*, 6(JUN), pp. 1–10. doi: 10.3389/fgene.2015.00208.

Kamvar, Z. N., Tabima, J. F. and Grünwald, N. J. (2014) 'Poppr : an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction ', *PeerJ*, 2, p. e281. doi: 10.7717/peerj.281.

Karakachoff, M. *et al.* (2015) 'Fine-scale human genetic structure in Western France', *European Journal of Human Genetics*, 23(6), pp. 831–836. doi: 10.1038/ejhg.2014.175.

Kherumian, R., Moullec, J. and Van Cong, N. (1967) ‘Groupes sanguins érythrocytaires A_1, A_2, BO, MN, Rh (CcDE) et sériques, Hp, Tf, Gm dans quatre régions militaires françaises.’, *Bulletins et Mémoires de la Société d’Anthropologie de Paris*, 1(XII), pp. 377–384. doi: DOI : https://doi.org/10.3406/bmsap.1967.1396.

Koch, J. T. (2005) ‘Breton migrations’, *Celtic Culture : A Historical Encyclopedia*. ABC-CLIO.

Lawson, D. J. *et al.* (2012) ‘Inference of Population Structure using Dense Haplotype Data’, *PLoS Genetics*, 8(1), pp. 11–17. doi: 10.1371/journal.pgen.1002453.

Lazaridis, I. *et al.* (2016a) ‘Genomic insights into the origin of farming in the ancient Near East’, *Nature*. Nature Publishing Group, 536(7617), pp. 1–22. doi: 10.1038/nature19310.

Lazaridis, I. *et al.* (2016b) ‘Genomic insights into the origin of farming in the ancient Near East’, *Nature*. doi: 10.1038/nature19310.

Leslie, S. *et al.* (2015) ‘The fine-scale genetic structure of the British population’, *Nature*, 519(7543), pp. 309–314. doi: 10.1038/nature14230.

MacMaster, N. (1997) *Colonial Migrants and Racism: Algerians in France, 1900–62*. Palgrave Macmillan.

Martínez-Cruz, B. *et al.* (2012) ‘Evidence of pre-roman tribal genetic structure in basques from uniparentally inherited markers’, *Molecular Biology and Evolution*, 29(9), pp. 2211–2222. doi: 10.1093/molbev/mss091.

Monnier, J. (1997) ‘Chapitre 6 : L’immigration bretonne en Armorique’, in Monnier, J. and Cassard, J. (eds) *Toute l’histoire de Bretagne*. Skol Vreizh, pp. 97–106.

O’Connell, J. *et al.* (2014) ‘A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness’, *PLoS Genetics*, 10(4), p. e1004234. doi: 10.1371/journal.pgen.1004234.

OECD (2017) *OECD Multi-level Governance Studies Multi-level Governance Reforms Overview of OECD Country Experiences*.

Patterson, N. *et al.* (2012) ‘Ancient Admixture in Human History’, *Genetics*, 192(November), pp. 1065–1093. doi: 10.1534/genetics.112.145037.

Patterson, N., Price, A. L. and Reich, D. (2006) ‘Population structure and eigenanalysis’, *PLoS Genetics*, 2(12), pp. 2074–2093. doi: 10.1371/journal.pgen.0020190.

Petkova, D., Novembre, J. and Stephens, M. (2016) ‘Visualizing spatial population structure with estimated effective migration surfaces’, *Nature Genetics*, 48(1), pp. 94–

100. doi: 10.1038/ng.3464.Visualizing.
Purcell, S. *et al.* (2007) 'PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses', *The American Journal of Human Genetics*, 81(3), pp. 559–575. doi: 10.1086/519795.
R Core Team (2018) 'R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.'
Ramos-Luis, E. *et al.* (2014) 'Y-chromosomal DNA analysis in French male lineages', *Forensic Science International: Genetics*. Elsevier Ireland Ltd, 9(1), pp. 162–168. doi: 10.1016/j.fsigen.2013.12.008.
Richard, C. *et al.* (2007) 'An mtDNA perspective of French genetic variation', *Annals of Human Biology*, 34(1), pp. 68–79. doi: 10.1080/03014460601076098.
Rodríguez-Ezpeleta, N. *et al.* (2010) 'High-density SNP genotyping detects homogeneity of Spanish and French Basques, and confirms their genomic distinctiveness from other European populations', *Human Genetics*, 128(1), pp. 113–117. doi: 10.1007/s00439-010-0833-4.
RStudio Team (2015) 'RStudio Team (2015). RStudio: Integrated Development for R.' Available at: http://www.rstudio.com/.
Schain, M. (2008) *The Politics of Immigration in France, Britain, and the United States: A Comparative Study (Perspectives in Comparative Politics)*. Edited by AIAA.
Sowerwine, C. (2009) *France since 1870: Culture, Society and the Making of the Republic*. 2nd edn. Palgrave Macmillan.
Wickham, H. (2007) 'Reshaping Data with the \pkgreshape Package', *Journal of Statistical Software*, 21(12), pp. 1–20. doi: 10.1016/S0142-1123(99)00007-9.
Wickham, H. (2009) 'ggplot2: Elegant Graphics for Data Analysis', *Media*, 35(July), p. 211. doi: 10.1007/978-0-387-98141-3.

Main Figures and Tables

	Groupings		% Total variance	Φ-statistics	p
A	Variations	Between Areas	0.07	Φ_{ST}= 0.0007	0.2417
	Variations	Between Departments Within Areas	0.17	Φ_{ST}= 0.0017	0.0009
	Variations	Within Departments	99.75	Φ_{ST}= 0.0024	0.0009
B	Variations	Between Regions	-0.073	Φ_{ST}= -0.0007	0.7462
	Variations	Between Departments Within Regions	0.3	Φ_{ST}= 0.003	0.0009
	Variations	Within Departments	99.76	Φ_{ST}= 0.0023	0.0009
C	Variations	Between Departments	0.23	Φ_{ST}= 0.0023	0.0009
	Variations	Within Departments	99.76		

Table 1. Hierarchical analysis of molecular variance (AMOVA). Results for percentage of total variance, Φ-statistics, and p-values are reported for the three distinct analyses. **A)** proportion of genetic variation partitioned among geographic areas, among departments within geographic areas, and within departments; **B)** proportion of genetic variation partitioned among regions, among departments within regions, and within departments; **C)** proportion of genetic variation partitioned among departments and within departments.

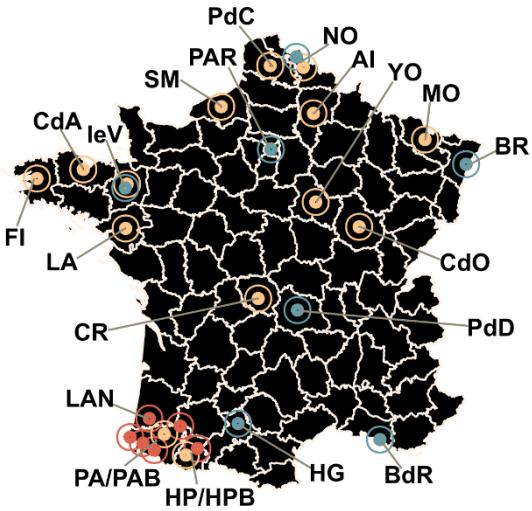

Figure 1. Map showing sample distribution among the different departments. Geographical coordinates are averages among samples. Different colors define the three datasets used in this work (blue dots correspond to the 256 samples genotyped for this work; yellow dots correspond to the 79 samples from Lazaridis et al., 2016; red dots correspond to the 60 samples from unpublished data). Sample size and acronyms for the departments are: PdD, Puy-de-Dôme (33); CR, Creuse (25); CdO, Côte-d'Or (1); AI, Aisne (1); NO, Nord (47); PdC, Pas-de-Calais (4); PAR, Paris (22); YO, Yonne (1); MO, Moselle (8); BR, Bas-Rhin (48); IeV, Ille-et-Vilaine (45); CdA, Côtes-d'Armor (3); FI, Finistère (5); SM, Seine-Maritime (2); LA, Loire-Atlantique (1); BdR, Bouches-du-Rhône (21); LAN, Landes (10); HG, Haute-Garonne (43); PA, Pyrénées-Atlantiques (15); PAB, Pyrénées-Atlantiques Basque (31); HP, Hautes-Pyrénées (9); HPB, Hautes-Pyrénées Basque (20).

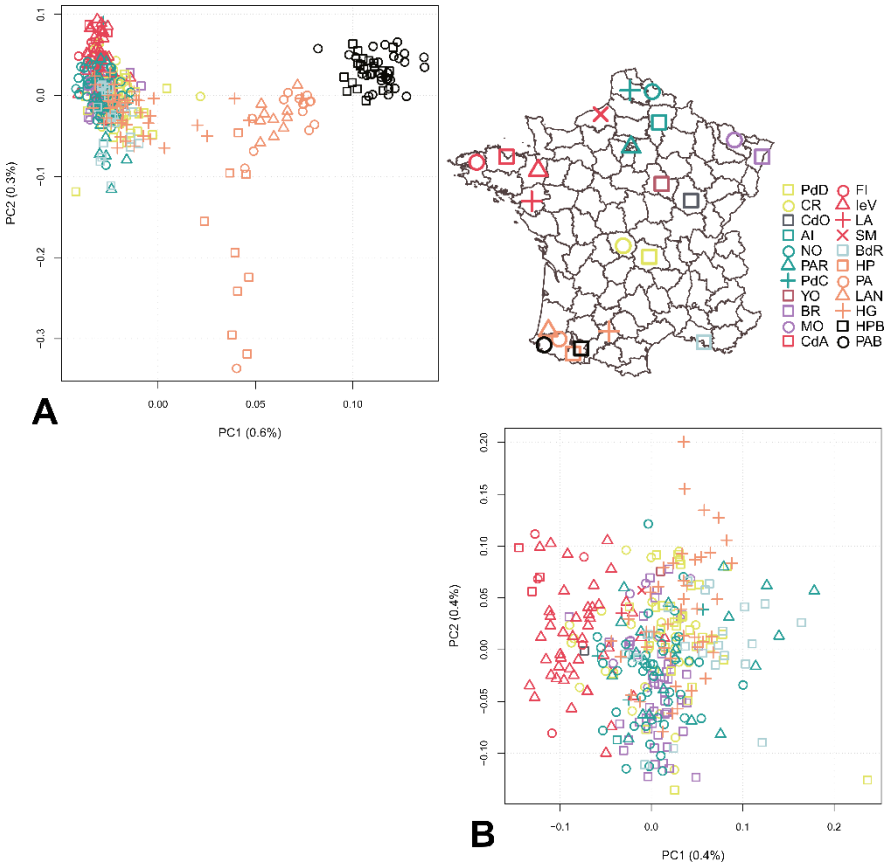

Figure 2. Principal Component Analysis with **A)** Franco-Cantabrian samples, and **B)** without them. Colors correspond to distinct geographic areas, while different symbols with the same color represent distinct departments in each area (See map distribution). However, Basques are colored differently than the non-Basque-speaking samples from that same area, but symbols recall the departments they share with the non-Basque-speaking groups.

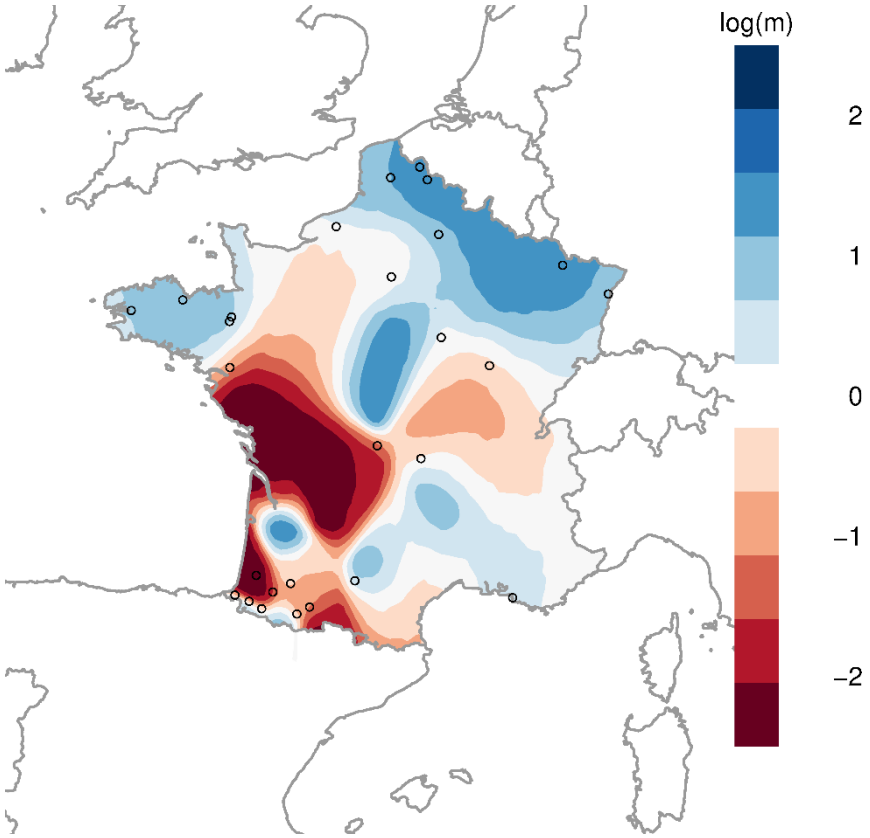

Figure 3. EEMS plot based on 395 French samples. Different shades of the same color represent differential levels of high (blue) or low (red) effective migration rates. The zero value indicates the average effective migration rate. Geographical locations for the different departments are averages of the coordinates among samples.

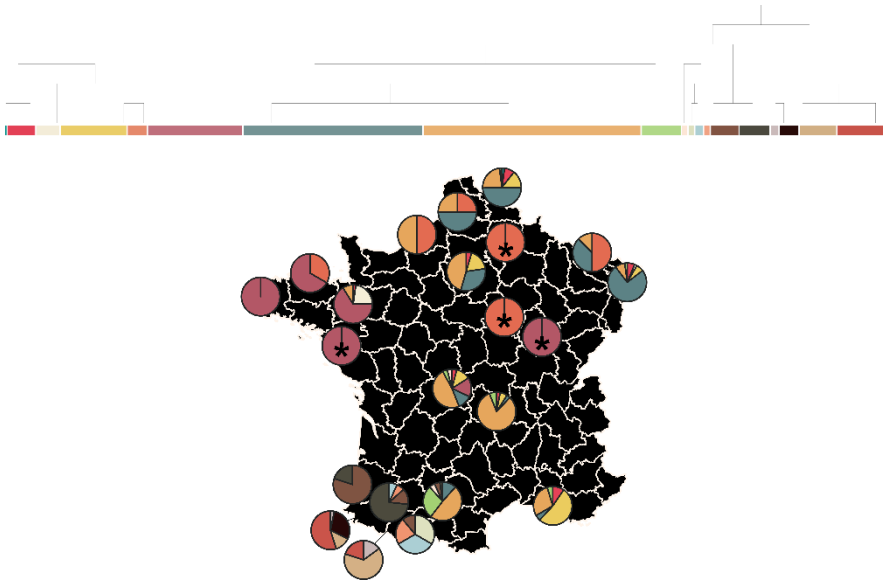

Figure 4. Pie charts showing the spatial distribution of haplotypes inferred by the fineSTRUCTURE tree. Each pie chart is a department, while colors correspond to the clusters described in the tree above the map. See Figure 1 for department names. Asterisks indicate departments with only one sample.

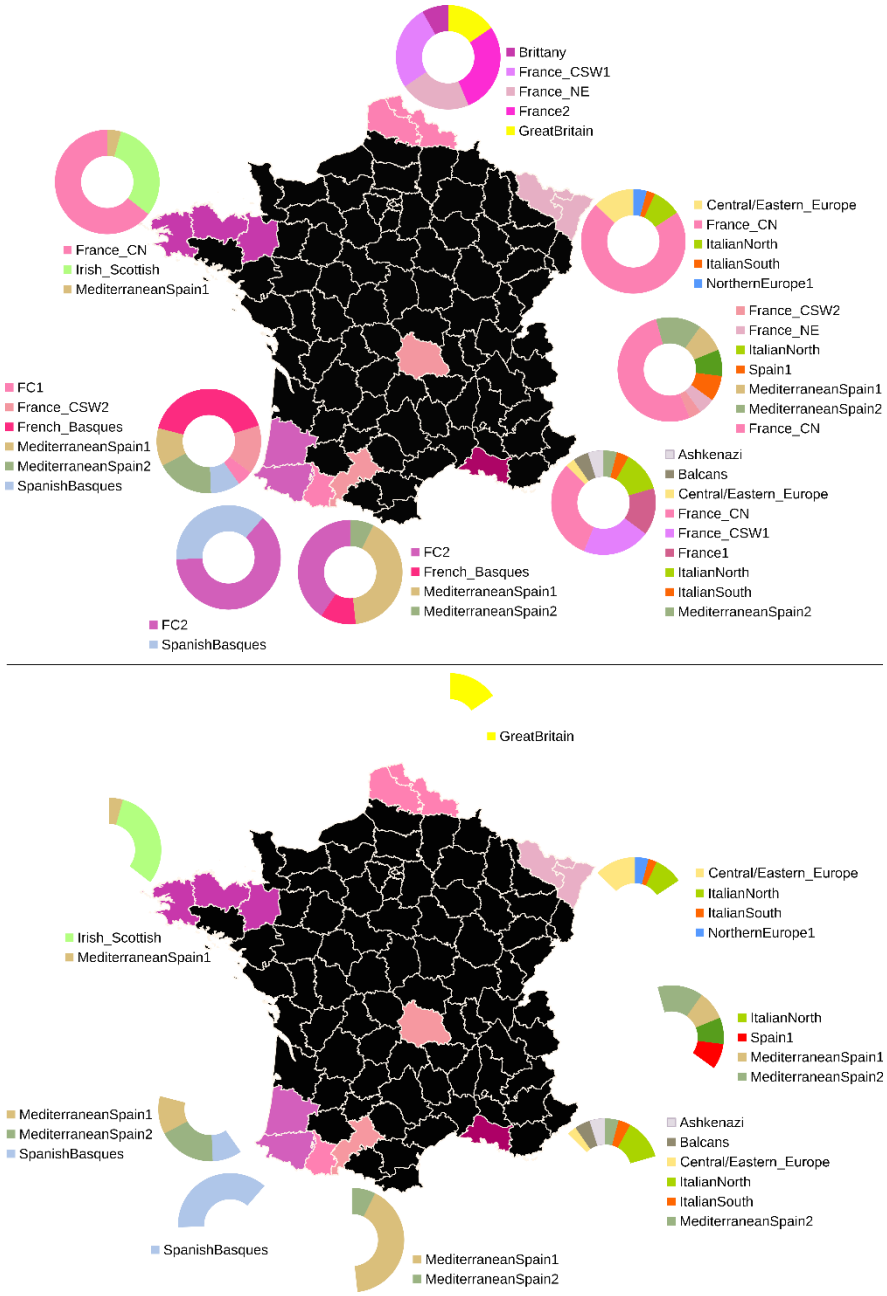

Figure 5. Ancestry profiles for the 8 French clusters. Each donut chart is representing the ancestry profile for one of the 8 French clusters shown in Supplementary Figure 10. The different colors represent proportions of haplotype sharing with a specific group and correspond to the values in Supplementary Figure 11 (above the 2.5%). All the pink shades refer to contributions coming from the different French groups, while other colors represent external contributions (See the lower panel for a representation without the French contributions).

Target	Source1	Source2	f_3(Target; Source1, Source2)	Z-score
BdR	Italian_South	Irish	-0.001401	-12.843
BdR	Italian_South	Scottish	-0.001232	-12.129
BdR	Italian_South	Basque_Country	-0.000967	-7.427
BdR	Norwegian	Italian_South	-0.000973	-7.001
BdR	German	Italian_South	-0.000942	-6.728
BdR	English	Italian_South	-0.000879	-6.188
BdR	Italian_North	Irish	-0.000626	-5.34
BdR	Italian_North	Scottish	-0.000508	-4.708
BdR	Norwegian	Italian_North	-0.00063	-4.502
BdR	South_Spain	Norwegian	-0.000536	-3.469
BdR	Norwegian	Basque_Country	-0.000475	-3.165
BdR	South_Spain	Scottish	-0.000341	-3.065
BdR	South_Spain	Irish	-0.000371	-3.016
BR	Italian_South	Irish	-0.001114	-14.112
BR	Italian_South	Scottish	-0.000886	-13.011
BR	Italian_North	Irish	-0.000783	-9.907
BR	Italian_North	Scottish	-0.000608	-9.575
BR	Norwegian	Italian_North	-0.000797	-7.74
BR	Norwegian	Italian_South	-0.000694	-6.563
BR	German	Italian_South	-0.000693	-6.457
BR	Norwegian	Mediterranean_Spain	-0.000388	-5.612
BR	English	Italian_South	-0.000499	-4.921
BR	German	Italian_North	-0.000495	-4.832
BR	German	Mediterranean_Spain	-0.0003	-4.252
BR	South_Spain	Norwegian	-0.000506	-4.18
BR	South_Spain	Irish	-0.00033	-3.868
BR	Central_Spain	German	-0.000412	-3.781
BR	Norwegian	Basque_Country	-0.000404	-3.385
BR	South_Spain	Scottish	-0.000243	-3.341
BR	Northwestern_Spain	Scottish	-0.000259	-3.27
BR	German	Northwestern_Spain	-0.000425	-3.266
BR	English	Italian_North	-0.000327	-3.175
BR	Mediterranean_Spain	Irish	-0.000164	-3.043
BR	Mediterranean_Spain	Scottish	-0.000145	-3.005
CR	Italian_South	Irish	-0.001268	-12.482
CR	Italian_South	Scottish	-0.001006	-10.597
CR	Italian_North	Irish	-0.001018	-10.162
CR	Italian_North	Scottish	-0.000808	-8.695
CR	Norwegian	Mediterranean_Spain	-0.000747	-8.007
CR	Norwegian	Italian_North	-0.000993	-7.904
CR	South_Spain	Irish	-0.000814	-7.629
CR	South_Spain	Scottish	-0.000693	-7.451
CR	Norwegian	Central_Spain	-0.00088	-7.091
CR	Mediterranean_Spain	Irish	-0.000562	-7.027
CR	Mediterranean_Spain	Scottish	-0.000509	-6.924
CR	Central_Spain	Irish	-0.000715	-6.91
CR	South_Spain	Norwegian	-0.000951	-6.735
CR	Norwegian	Basque_Country	-0.000987	-6.656
CR	Norwegian	Italian_South	-0.000809	-6.305
CR	Central_Spain	Scottish	-0.000609	-6.245
CR	Central_Spain	English	-0.00073	-5.409
CR	English	Italian_South	-0.000692	-5.33

Target	Source1	Source2	f_3(Target; Source1, Source2)	Z-score
CR	Basque_Country	Irish	-0.000563	-4.905
CR	English	Basque_Country	-0.000718	-4.895
CR	Central_Spain	German	-0.000667	-4.779
CR	South_Spain	English	-0.000652	-4.73
CR	English	Mediterranean_Spain	-0.000471	-4.727
CR	Basque_Country	Scottish	-0.000499	-4.687
CR	English	Italian_North	-0.000601	-4.538
CR	Northwestern_Spain	Scottish	-0.000497	-4.509
CR	Northwestern_Spain	Irish	-0.000533	-4.455
CR	German	Italian_South	-0.000507	-3.794
CR	German	Mediterranean_Spain	-0.000358	-3.559
CR	English	Northwestern_Spain	-0.000522	-3.43
CR	South_Spain	German	-0.000453	-3.238
CR	German	Italian_North	-0.00039	-3.118
CR	German	Basque_Country	-0.000471	-3.104
HG	Italian_South	Irish	-0.000918	-11.28
HG	Italian_South	Scottish	-0.000796	-11.014
HG	Italian_North	Scottish	-0.00057	-8.014
HG	Italian_North	Irish	-0.00064	-7.831
HG	Mediterranean_Spain	Scottish	-0.000364	-7.517
HG	South_Spain	Scottish	-0.00057	-7.472
HG	Basque_Country	Scottish	-0.000586	-7.09
HG	Norwegian	Basque_Country	-0.000847	-6.77
HG	Italian_South	Basque_Country	-0.000667	-6.276
HG	South_Spain	Irish	-0.00055	-6.256
HG	English	Basque_Country	-0.000749	-5.718
HG	Basque_Country	Irish	-0.00051	-5.447
HG	Northwestern_Spain	Scottish	-0.000447	-5.224
HG	Norwegian	Mediterranean_Spain	-0.000375	-5.146
HG	Mediterranean_Spain	Irish	-0.000277	-5.1
HG	South_Spain	Norwegian	-0.0006	-4.935
HG	German	Basque_Country	-0.000601	-4.834
HG	Norwegian	Italian_North	-0.000527	-4.832
HG	Central_Spain	Scottish	-0.00035	-4.65
HG	Italian_North	Basque_Country	-0.000469	-4.505
HG	South_Spain	English	-0.000472	-4.029
HG	English	Italian_South	-0.000424	-3.876
HG	Central_Spain	German	-0.00045	-3.78
HG	Central_Spain	Irish	-0.000315	-3.628
HG	Central_Spain	English	-0.000413	-3.613
HG	Northwestern_Spain	Irish	-0.000342	-3.607
HG	English	Mediterranean_Spain	-0.00027	-3.514
HG	German	Mediterranean_Spain	-0.000256	-3.451
HG	Norwegian	Central_Spain	-0.000393	-3.322
HG	Norwegian	Italian_South	-0.000371	-3.292
HG	English	Northwestern_Spain	-0.000414	-3.097
HG	South_Spain	German	-0.000372	-3.074
IeV	Italian_North	Irish	-0.000493	-6.175
IeV	Italian_South	Irish	-0.00051	-6.05
IeV	South_Spain	Irish	-0.000423	-4.624
IeV	Mediterranean_Spain	Irish	-0.000257	-4.498
IeV	Central_Spain	Irish	-0.000345	-4.365
MO	Italian_South	Irish	-0.001174	-5.138

Target	Source1	Source2	f_3(Target; Source1, Source2)	Z-score
MO	Italian_South	Scottish	-0.001093	-4.962
MO	South_Spain	Norwegian	-0.001053	-4.098
MO	Norwegian	Italian_South	-0.00098	-4.003
MO	German	Italian_South	-0.000935	-3.941
MO	Norwegian	Italian_North	-0.000855	-3.476
MO	Norwegian	Basque_Country	-0.000885	-3.332
MO	South_Spain	German	-0.000813	-3.316
MO	Norwegian	Mediterranean_Spain	-0.00072	-3.276
MO	South_Spain	Scottish	-0.000713	-3.159
NO	Italian_South	Irish	-0.001196	-14.818
NO	Italian_South	Scottish	-0.001029	-14.549
NO	Italian_North	Scottish	-0.000773	-11.819
NO	Italian_North	Irish	-0.000888	-11.416
NO	Mediterranean_Spain	Scottish	-0.000445	-9.474
NO	Norwegian	Mediterranean_Spain	-0.000657	-9.21
NO	Norwegian	Italian_North	-0.000931	-8.896
NO	Mediterranean_Spain	Irish	-0.000404	-7.64
NO	South_Spain	Scottish	-0.000571	-7.552
NO	Norwegian	Italian_South	-0.000807	-7.383
NO	South_Spain	Irish	-0.000598	-6.817
NO	German	Mediterranean_Spain	-0.000486	-6.797
NO	South_Spain	Norwegian	-0.000803	-6.682
NO	German	Italian_South	-0.000721	-6.65
NO	Central_Spain	Scottish	-0.000449	-6.191
NO	Norwegian	Central_Spain	-0.000693	-6.129
NO	Central_Spain	German	-0.000698	-6.102
NO	English	Italian_South	-0.000581	-5.536
NO	Central_Spain	Irish	-0.00046	-5.515
NO	Norwegian	Basque_Country	-0.000646	-5.249
NO	German	Italian_North	-0.000546	-5.174
NO	Northwestern_Spain	Scottish	-0.000403	-4.954
NO	South_Spain	German	-0.000523	-4.418
NO	Central_Spain	English	-0.000435	-4.092
NO	English	Italian_North	-0.000431	-3.938
NO	English	Mediterranean_Spain	-0.000273	-3.789
NO	Northwestern_Spain	Irish	-0.000344	-3.661
NO	South_Spain	English	-0.000396	-3.408
NO	German	Northwestern_Spain	-0.000453	-3.381
PAR	Italian_South	Irish	-0.001536	-14.131
PAR	Italian_South	Scottish	-0.001384	-13.931
PAR	Norwegian	Italian_South	-0.001225	-9.242
PAR	Italian_North	Irish	-0.000911	-8.574
PAR	Italian_North	Scottish	-0.000811	-8.326
PAR	Norwegian	Italian_North	-0.001032	-8.075
PAR	German	Italian_South	-0.001055	-8.024
PAR	English	Italian_South	-0.001026	-7.991
PAR	Norwegian	Mediterranean_Spain	-0.000686	-6.825
PAR	South_Spain	Norwegian	-0.000877	-6.035
PAR	South_Spain	Scottish	-0.000583	-5.506
PAR	South_Spain	Irish	-0.000595	-5.134
PAR	Mediterranean_Spain	Scottish	-0.000411	-5.098
PAR	Norwegian	Central_Spain	-0.000634	-4.494
PAR	Northwestern_Spain	Scottish	-0.000501	-4.465

Target	Source1	Source2	f_3(Target; Source1, Source2)	Z-score
PAR	German	Italian_North	-0.000563	-4.205
PAR	English	Italian_North	-0.000559	-4.15
PAR	Mediterranean_Spain	Irish	-0.000355	-4.036
PAR	German	Mediterranean_Spain	-0.00043	-4.009
PAR	Central_Spain	German	-0.000554	-3.898
PAR	Norwegian	Basque_Country	-0.000601	-3.828
PAR	Italian_South	Basque_Country	-0.000512	-3.755
PAR	South_Spain	German	-0.000513	-3.543
PAR	Northwestern_Spain	Irish	-0.000428	-3.477
PAR	South_Spain	English	-0.000498	-3.439
PAR	German	Northwestern_Spain	-0.000531	-3.341
PAR	English	Mediterranean_Spain	-0.000329	-3.092
PAR	Central_Spain	Scottish	-0.000327	-3.03
PAR	English	Northwestern_Spain	-0.00048	-3.02
PdD	Italian_South	Irish	-0.001146	-12.844
PdD	Italian_South	Scottish	-0.000978	-12.316
PdD	Italian_North	Irish	-0.000688	-7.723
PdD	Italian_North	Scottish	-0.000572	-7.324
PdD	German	Italian_South	-0.000684	-6.048
PdD	Norwegian	Italian_South	-0.000693	-6.016
PdD	Norwegian	Mediterranean_Spain	-0.000466	-5.992
PdD	Mediterranean_Spain	Scottish	-0.000316	-5.786
PdD	Norwegian	Italian_North	-0.000669	-5.677
PdD	South_Spain	Scottish	-0.00045	-5.496
PdD	Italian_South	Basque_Country	-0.000624	-5.473
PdD	English	Italian_South	-0.00062	-5.3
PdD	Norwegian	Basque_Country	-0.000714	-5.239
PdD	South_Spain	Irish	-0.000477	-4.847
PdD	South_Spain	Norwegian	-0.000619	-4.81
PdD	German	Mediterranean_Spain	-0.00037	-4.492
PdD	Central_Spain	German	-0.000553	-4.317
PdD	Mediterranean_Spain	Irish	-0.000275	-4.269
PdD	Norwegian	Central_Spain	-0.000473	-3.945
PdD	Northwestern_Spain	Scottish	-0.000368	-3.932
PdD	German	Basque_Country	-0.000491	-3.584
PdD	English	Basque_Country	-0.000489	-3.455
PdD	Basque_Country	Scottish	-0.000314	-3.455
PdD	South_Spain	German	-0.000415	-3.332
PdD	Central_Spain	Scottish	-0.00029	-3.309
PdD	German	Italian_North	-0.000359	-3.172
PdD	Central_Spain	Irish	-0.000302	-3.124
PdD	German	Northwestern_Spain	-0.000433	-3.044

Table 2. Results for the f_3(Test; Source1, Source2) statistics. Test populations are the departments with at least 2 individuals, Sources are groups of external populations. Only tests with Z scores < -3 are listed.