Genetic diversity and structure of sweet chestnut (*Castanea sativa* Mill.) in France: At the intersection between Spain and Italy

Cathy Bouffartigue[a]*, Sandrine Debille[b], Olivier Fabreguette[c], Ana Ramos Cabrer[d], Santiago Pereira-Lorenzo[d], Timothée Flutre[e]†, Luc Harvengt[b]†

a AGIR, Université de Toulouse, INRA, INPT, INP-EI PURPAN Castanet Tolosan, France.

b Genetics and Biotechnology team, Biotech and Advanced Forestry Department, FCBA tech institute, Campus Recherche Forêt-Bois de Pierroton, 71 route d'Arcachon, 33610 Cestas, France.

c INRA - Université de Bordeaux, UMR 1202 BIOGECO Biodiversité, Gènes et Communautés. Centre de recherche Nouvelle-Aquitaine-Bordeaux, Cestas, France. INRA - Bordeaux Sciences Agro, UMR 1065 SAVE Santé et Agroécologie du Vignoble. Centre de recherche de Bordeaux Aquitaine, Villenave D'Ornon, France.

d Departamento de Producción Vegetal y Proyectos de Ingeniería, Escola PolitécnicaSuperior, Campus de Lugo, Universidade de Santiago de Compostela, Lugo, Spain.

e INRA, UMR Amélioration Génétique et Adaptation des Plantes méditérranéennes et tropicales, F-34060 Montpellier, France and GQE – Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France.

† Contributed equally

* Corresponding author

Email address: cathy.bouffartigue@inra.fr; sandrine.debille@fcba.fr; olivier.fabreguettes@inra.fr ; ana.ramos@usc.es ; santiago.pereira.lorenzo@usc.es ; timothee.flutre@inra.fr; luc.harvengt@fcba.fr

Key message

This paper presents the results of the first assessment of genetic diversity and structure of wild and cultivated sweet chestnut in France. It reveals high diversity, a low but significant structure, and strongly suggests that the French gene pool is at the intersection between the Italian and Spanish gene pools.

Abstract

Context

Renewed interest in European chestnut in France is focussed on finding locally adapted populations partially resistant to ink disease and identifying local landraces.

Aims

We genotyped trees to assess (i) the genetic diversity of wild and cultivated chestnut across most of its range in France, (ii) their genetic structure, notably in relation with the sampled regions, and (iii) relations with its neighbors in Spain and Italy.

Methods

A total of 1,401 trees in 17 sampling regions in France were genotyped at 13 SSRs, and a subset of 693 trees at 24 SSRs.

Results

Genetic diversity was high in most sampling regions, with redundancy between them. No significant differentiation was found between wild and cultivated chestnut. A genetic structure analysis with no *a priori* information found a low, yet significant structure, and identified three clusters. Two clusters of sampling regions, south east France and Corsica, were

less admixed than the others. A substructure was detected in the admixed cluster suggesting differentiation in wild chestnut trees in Finistère and Aveyron sampling regions.

Conclusion

The genetic structure within and between our sampling regions is likely the result of natural events (recolonization after the last glaciation) and human activities (migration and exchanges). Notably, we provide evidence for a common origin of most French and Iberian chestnut trees, except those from, south east France that were associated with the Italian gene pool. This advance in our knowledge of chestnut genetic diversity and structure will benefit conservation and help our local partners' valorization efforts.

Keywords

Castanea sativa Mill.; chestnut; genetic diversity; genetic structure; microsatellite markers; landraces

1. Introduction

Sweet chestnut (*Castanea sativa* Mill.) is an endemic, multi-purpose tree species cultivated for its wood and nuts. It is the third broad-leaved tree species in France in forest area (750,000 ha) and in 2016, accounted for 5% of land used for fruit production (FranceAgriMer 2017). With an annual production of 7,000-9,000 tons in the last 10 years, France is the fifth European producer (FAO 2018). Sweet chestnut has been intensively cultivated in coppices and orchards for centuries in France. However, since the beginning of the 18th century, it has suffered from abandonment, leading to a sharp decrease in production (Pitte 1986; Sauvezon et al. 2000). Many landraces and associated knowledge were lost. In the 1960s, the French National Institute for Agricultural Research (INRA) started a breeding program to develop interspecific hybrids resistant to ink disease caused by a Phytophtora fungus, by crossing two Asian tolerant species, *Castanea crenata* and *Castanea mollissima*, with local landraces from regions with an oceanic climate. These hybrids are now mainly used for fruit production and as rootstock (particularly Marigoule and Bouche de Bétizac varieties, and more recently BelleFer). However, they are not adapted to continental and Mediterranean conditions (Martin et al. 2017; Míguez-Soto et al. 2019). Their fruit quality has also been criticized by some growers and by chestnut lovers, particularly in comparison with landraces. Action was thus taken by these actors, involving survey of old chestnut trees, phenotypic observations and the establishment of conservatory orchards.

Strong geographical structure was reported in wild populations in Italy, Spain, Greece and Turkey (Mattioni et al. 2013). A study of wild chestnut in Spain, Italy and Greece (Fernández-Cruz and Fernández-López 2016) found two main gene pools in Europe, and another study of wild, natural or naturalized populations (Mattioni et al. 2017), found three. These findings agree with evidence of spontaneous establishment originating from the Last Glacial Maximum refugia in the north of the Iberian, Italian and Balkan peninsulas, and in northern Anatolia (Krebs et al. 2004, 2019; Roces-Díaz et al. 2018). In southern France, there is possible evidence for chestnut refugia in palaeo-botanical data (Krebs et al. 2019). The preferred hypothesis is therefore that most pre-cultivation Castanea in France are the result of the spontaneous spread of the species from neighboring southern European refugia, i.e. in Spain and Italy. However, the most recent genetic analyses conducted exclusively on French populations were published in the 1990s on wild chestnut and at a regional scale (Frascaria et al. 1991, 1992; Frascaria and Lefranc 1992) the and the results obtained in the CASCADE project (Eriksson et al. 2005) have not yet been published (T. Barreneche pers. com.). Mattioni et al. (2008) compared naturalized, coppice and orchard populations in Italy, Greece, Spain, the UK and France, and showed differences in within-population genetic parameters between fruit orchards and other types of chestnut management. This result implies that long-term management techniques can influence the genetic makeup of the populations. Differences between and within countries have also been reported (Pereira-Lorenzo et al. 2016). For these reasons, specifically French, finer-scale sampling of both wild (forest) and cultivated chestnut trees (orchards and alignments) was needed to help distinguish between natural and anthropogenic evolutionary factors.

In terms of sampling, many authors have genotyped tree collections *ex situ*, i.e., in conservatories (Martín et al. 2010a), and *in situ* (Pereira-Lorenzo and Fernandez-Lopez 1997; Gobbin et al. 2007; Martin et al. 2010b; Pereira-Lorenzo et al. 2010, 2019; Beccaro et al. 2012; Mellano et al. 2012, 2018; Beghè et al. 2013; Quintana et al. 2014; Fernández-López and Fernández-Cruz 2015). In this study, we used both in- and ex-situ sources to assess the known and currently used genetic diversity of sweet chestnut. As a result, we often sampled several individuals belonging to the same landrace. Hereafter, we use the term “landrace” as defined by Villa et al. (2005) rather than “variety” or “cultivar”, as it better covers the variety of sampling situations we encountered in the field. However, we do use the term “cultivar” when known cultivars were encountered.

The main aims of this work were to assess (i) the genetic diversity of wild and cultivated chestnut in most of its range in France, (ii) their genetic structure, notably in relation with the sampling regions, and (iii) relations between French chestnut and its neighbors in Spain and Italy. For this purpose, we sampled natural chestnut populations, ancient grafted chestnut identified *in situ* by local partners and *ex situ* local landraces in conservatories in the main nut-producing regions and in most of the distribution of natural chestnut forests in France. We used microsatellite markers from the EU chestnut database to genotype all sampled trees at 13 SSRs and a subset at 24 SSRs (Pereira-Lorenzo et al. 2017). By also including Iberian samples cited in the Pereira-Lorenzo et al. publication, we also provide some evidence for the origin of the trees we sampled.

Fixation of genotypes by grafting from spontaneous chestnut, or “instant domestication” as defined by (Harris et al. 2002), is reported in the literature (Aumeeruddy-Thomas et al. 2012), and was recently documented in Italy and Spain (Pereira-Lorenzo et al. 2019). As a working hypothesis, this suggested a possible lack of genetic structure between wild and cultivated chestnut. It is common knowledge that grafts and nuts travel by means of markets, historically via occupational travelers such as glass blowers (Pitte 1986) and now via local and internet-mediated exchange fairs.

However, the extent and impact of this phenomenon on the genetic structure of cultivated chestnut was previously unknown in France. We hypothesized that it is sufficiently frequent to have a significant impact, leading to a low genetic structure of cultivated chestnut in France. As reported in (Pereira-Lorenzo et al. 2019), we also expected to find a high overall genetic diversity, but without marked differences between the wild and cultivated sets. In addition, we expected *in situ* local landraces to be multiclonal due to repeated grafting over the centuries and the accumulation of mutations or the use of seedlings from the landrace.

2. Materials and Methods

Terminology: we avoid the use of "population" and instead use "sampling region" to describe a geographically or socially meaningful region where a non-profit association has prospected and conserved chestnut, or a group of sampling sites located close by. We use "genetic cluster" to denote a cluster of genotyped trees resulting from the analysis of genetic structure. "Chestnut type" is used as a category with two levels, "forest" and "cultivated".

2.1. Geographical sampling

In forest stands, trees were chosen randomly, located several dozen meters apart in the middle of forest patches. Their exact locations were recorded by GPS. In Brittany, Auvergne-Rhône-Alpes, Occitanie, Provence-Alpes-Côtes d'Azur (PACA) and Corsica, mature leaves were sampled and immediately enclosed in plastic bags with silicagel. In Gironde, dormant buds were sampled from trees close to the laboratory to facilitate frequent re-sampling when assessing the accuracy of genotyping protocols. In Corsica, nuts and dried leaves were also sampled in the field, whereas cultivated chestnut was provided as DNA extract. Whenever we sampled offspring as groups of half sib fruits, we also sampled leaves from their mothers. Nuts harvested in the Finistère, Corsica, Basque Country and Aveyron forest sampling regions were germinated and sown in the greenhouse.

2.2. Expert-based sampling

Field surveys of cultivated chestnut were conducted in 2016-2017 in collaboration with producer and amateur organizations. In 2016, we focused our sampling effort on the landraces they knew and were interested in. In 2017, we expanded sampling to most known landraces and grafted trees, supplemented by random sampling in a few chestnut orchards. Associative conservatories were also sampled. We sampled several chestnut trees that had the same name to test the genetic diversity of landraces. When attributing sampled trees to a given landrace, when known, we followed the field expert's determination.

2.3. SSR genotyping

A total of 1401 trees were genotyped at 13 SSRs, and a subset of 693 were genotyped at 24 SSRs. Total genomic DNA was extracted from fresh leaves, silica-dried leaves or dormant buds using the DNeasy 96 Plant kit (Qiagen, Hilden Allemagne). Twenty-four SSR markers previously selected to study chestnut genetic diversity were used for this study (Buck et al. 2003; Gobbin et al. 2007; Kampfer et al. 1998; Marinoni et al. 2003; Steinkellner et al. 1997) based on the protocol of Pereira-Lorenzo et al. (2017). We amplified these 24 SSRs into 5 multiplex and 2 singleplex PCRs using one of the FAM, NED, PET, VIC fluorophore-labeled primers (PE Applied Biosystems, Warrington, UK) modified following (Pereira-Lorenzo et al. 2017, 2019). The PCR final reaction volume was 15 µl (7.5 µl of QIAGEN Multiplex Master Mix, 0.075 to 0.3 µM of each primer, 4 to 4.9 µl RNase Free Water and 2 µl of ADN at 5-10 ng/µl). The amplification conditions were 95°C for 5 min, followed by 30 cycles at 95°C for 30 s, annealing at a specific temperature depending on the multiplex set, for 1.5 min, and 1 min at 72°C, and final extension at 60 °C for 30 min. Negative controls were included in all PCR reactions to enable detection of cross contamination of the samples.

Amplifications at 13 SSRs corresponded to sets 1, 2 and 3. Amplifications at 24 SSRs corresponded to all sets. Set 1 (57°C): EmCs14-VIC, EmCs15-FAM, EmCs38-FAM, EmCs2-NED, CsCAT14-PET, CsCAT2-VIC. Set 2: (50°C): CsCAT-16-PET, CsCAT41-FAM, QpZAG110-PET, QpZAG36-VIC, CsCAT3-NED. Set 3: post-PCR multiplexing: QrZAG4-NED (48°C) and QrZAG96-NED (52°C). Set 4: (50°C): CsCAT6-NED, CsCAT1-PET, CsCAT15-FAM, CsCAT8-VIC. Set 5: (58°C): RIC-FAM, CsCAT17-PET, EmCs22-VIC. Set 6: (60°C): EmCs25-FAM, CIO-NED, OCI-PET and OAL-VIC. Amplification products were diluted with water, 2 µl of the diluted amplification product was added to 0.12 µl of 600LIZ size standard (Applied Biosystems, Foster City, USA) and 9.88 µl of formamide.

Genotyping was performed partly on an ABI 310 capillary sequencer (Applied Biosystems, Foster City, CA, USA) at the Xylobiotech FCBA facility of Cestas-Pierroton with further work on an ABI 3500 XL capillary sequencer (Applied Biosystems, Foster City, CA, USA) at the CIRAD GenSeqUM platform in Montpellier, France. Allele sizes were read independently by two investigators using GENEMAPPER 4.1 and 5.0 respectively (Applied Biosystem, Foster City, USA). The output files in the fsa format were made compatible for GENEMAPPER 4.1 using a Python script from the Montpellier platform.

2.4. Data analysis

2.4.1. Detection of clonal groups and null alleles

All individuals with more than 20% of missing alleles were removed along with with individuals showing Asian alleles (Pereira-Lorenzo et al. 2010). CsCAT41 is known to amplify two sites: the CsCAT41A (Pereira-Lorenzo et al. 2010) locus was thus removed before analysis. The presence of uninformative loci was tested with the informloci function in

the R/poppr package version 2.8.3 (Kamvar et al. 2015; Kamvar et al. 2014) in both data sets. The percentages of missing data were obtained using the info_table function in R/poppr. The frequency of null alleles per locus was calculated with the R/PopGenReport package version 3.0.4 (Adamack and Gruber 2014) based on Brookfield formula (Brookfield 1996). Following (Lassois et al. 2016), we discarded loci with more than 10% of null alleles. After removing loci, the genotype curve function implemented in the R/poppr. was applied to both data sets to determine the minimum number of loci necessary to discriminate between individuals. Redundant genotypes were searched within each sampling region to identify multi-locus genotypes (MLGs) for each data set, using the clonecorrect function in R/poppr.

2.4.2. Genetic diversity

The observed number of alleles (Na) and observed heterozygosity (Ho) were calculated at each locus using the summary function in the R/adegenet package 2.1.1 (Jombart 2008). The effective number of alleles (Ne) was calculated using the expected heterozygosity (He) from the summary function for genind object in R/adegenet, with Ne=1/(1-He). The Fst and corrected Fst (Fstp), Fis and Dest per locus (Jost 2008; Nei 1987) were calculated using the basic.stats function in the R/hierfstat package version 0.04-22 (Goudet 2005). The poppr function in R/poppr was used to report other basic statistics per sampling region including the Shannon-Weiner diversity index (H), the index of association (Ia), and the standardized index of association (rbarD) (Agapow and Burt 2001). The significance of Ia and rbarD were tested with 1000 permutations, shuffling the genotypes at each locus while maintaining the heterozygosity and allelic structures. Deviation from the Hardy-Weinberg equilibrium (HWE) was tested on both loci and populations with 1000 permutations using the hw.test function in the R/pegas package version 0.11 (Paradis 2010). The Chi2 statistic was calculated over the entire data set and two p values were computed, one analytical and one derived from 1000 Monte-Carlo permutations.

2.4.3. Population structure

In each data set, using the find.clusters function in R/adegenet, SSR genotypes were transformed by a principal component analysis (PCA), followed by the k-means algorithm applied to the principal components (PCs) to identify groups of individuals we call “genetic clusters” (Jombart et al. 2010). The number of clusters was determined using the BIC. Discriminant analysis of principal components (DAPC, Jombart et al. 2010) was then performed based on this grouping. The number of principal components (PCs) to keep was chosen by cross-validation using the xvalDapc function in R/adegenet with 30 repetitions and a maximum of 80 PCs. Hierarchical analysis of molecular variance (AMOVA, Excoffier and Smouse 1992) as implemented in the poppr.amova function in R/poppr was performed using all loci with less than 5% missing data on the preset hierarchy of chestnut types and sampling regions, and on genetic clusters. Fis, pairwise Fst and hierarchical F-statistics were calculated, and 95% confidence intervals were obtained by bootstrapping with 1000 samples over loci using the boot.ppfis, boot.ppfst and boot.vc functions. Differences between hierarchy levels were tested by randomization with the function `randtest` in the R/ade4 package version 1.7-13 (Excoffier and Smouse 1992; Chessel et al. 2004). Some components of covariance could have slightly negative estimates due to the absence of significant genetic structure at the corresponding hierarchical level (FAQ List for Arlequin 2.000).

2.4.5. Reproducibility

To facilitate method reproducibility (Goodman et al. 2016), all our analyses were performed in R (R Core Team 2019); the scripts are available at https://data.inra.fr/privateurl.xhtml?token=8c03a83c-be4d-4984-972f-7808558b4539.

3. Results

3.1 Sampling scheme

To characterize and understand the genetic diversity and population structure of the European chestnut (*Castanea sativa* Mill.) in France, we genotyped 1,401 trees in 17 sampling regions in both forest and cultivated areas. Table 1 lists sampling details and Figure 1 shows the location of the sampling regions (GPS of sampled trees are available upon request).

<table 1>

<figure 1>

3.2. Detection of null alleles and redundant multi-locus genotypes

After filtering genotyped trees for missing alleles, 1,214 trees genotyped at 13 SSRs (respectively 642 at 24 SSRs) remained for further analysis (Table 1). Moreover, some SSRs were known to often have a high null allele frequency, such as EmCs25 (Lusini et al. 2014) and CsCAT14, CsCAT2, CsCAT41, QrZAG4 and CIO (Pereira-Lorenzo et al. 2017). In our data, EmCs38 null allele frequency was higher than 10% in the 13 SSR data set (respectively EmCs38 CIO and EmCs25 in the 24 SSR data set) and was discarded (Online Resource 1). After filtering uninformative loci and those with more than 5% of missing values, the resulting data sets had 19 SSRs, hereafter called *19All*, and 10 SSRs, hereafter called *10All*. Redundant multi-locus genotypes (MLGs) were then discarded in each sampling region, as they could be the result of both practices (grafting) and sampling choices, and had to be removed to avoid the artefactual detection of genetic structure resulting from the sampling strategy. The resulting data sets (Table 1) are called *10Unik* (1050 trees) and *19Unik* (521 trees). In both data sets, the discriminating power of the polymorphic markers to differentiate between genotypes was sufficient to discriminate all individuals irrespective of the number of loci and individuals (Online Resource 2).

3.3. Description of SSR diversity per sampling region

The 19 SSRs analyzed in this study varied greatly in allele diversity (Online Resource 3). The *10Unik* data set (respectively *19Unik*) had a total of 113 alleles (respectively 186), with an average of 11.3 alleles per locus (respectively 9.8). This ranged from 3 for EMCs2 to 33 for CsCAT3 (respectively 2 for QrZAG4 to 31 for CsCAT3). In terms of expected heterozygosity (He), EMCs2 showed the lowest diversity with 0.66 in *10Unik* (respectively QrZAG4 with 0.17 in *19Unik*) and CsCAT3 the highest diversity with 0.85 in *10Unik* (respectively 0.83 in *19Unik*). The within-population inbreeding coefficient (Fis) ranged from -0.437 to 0.134 in *10Unik* (respectively -0.439 to 0.152 in *19Unik*), with a mean of -0.069 in *10Unik* (respectively. -0.116 in *19Unik*). Across all sampling regions, in *10Unik*, it was not possible to reject the HWE for CsCAT3 and QpZAG110. In *19Unik*, only QpZAG110 and QrZAG4 were in the HWE (Online Resource 4). When tested per sampling region, only ForGard, ForHerault and ForBasque were in the HWE in both data sets. ForAveyron and ForCantal were in the HWE only in *10Unik*. Moreover, in both data sets, HWE was rejected for all SSR loci in at least one sampling region except OCI in *19Unik*

3.4. Redundant diversity among sampling regions and no differentiation between chestnut types

Genetic diversity indices calculated for each sampling region genotyped at 10 SSRs without MLGs are listed in Table 2 (results at 19 SSRs are presented in Online Resource 5). The aim of sampling ForGironde was not to be representative of the region, but to facilitate resampling. In the *19Unik* data set ForBasque had a single individual. Therefore, diversity and differentiation are discussed excluding ForGironde in the *10Unik* data set, and excluding ForGironde and ForBasque in the *19Unik* data set. The highest effective number of alleles per sampling region was found in the Finistère forest sampling regions in *10Unik* (ForFinistere, north west of France) and the lowest was found in the cultivated sampling region in Var (CultVar, south east of France). The mean observed heterozygosity was 0.681 and the mean expected heterozygosity was 0.658. The sampling regions with the lowest (respectively highest) observed heterozygosity were ForVar in the south east of France (respectively the forest sampling region in Hérault, CultHerault). The sampling regions with the lowest (respectively highest) expected heterozygosity were CultVar (respectively the forest sampling regions in Finistère, ForFinistere). Excluding ForGironde, no positive and significant inbreeding (Fis) was found in any region. The highest Ia and rbarD were found in CultVar and the lowest (but not significant) were found in ForFinistere. The results of AMOVA (Table 3 and Online Resource 6), revealed no substantial difference in structure in chestnut type between forest stands and cultivated orchards (the variance component did not significantly differ from zero and Fct with confidence intervals excluding zero, although very close). Instead, more than 80% of the variance was found within each sampling region. At a threshold of 0.001, we rejected the null hypothesis of panmixia,

both among sampling regions within chestnut types and within sampling regions. Among sampling regions within chestnut types, the Phi test statistic of the AMOVA indicated greater variance than expected under the null hypothesis. This suggested an underlying structure at this hierarchical level that was confirmed by a positive bootstrap-derived confidence interval for Fst (7%-9.3%). Within sampling regions, the Phi test statistic indicated lower variance than expected under the null hypothesis. This suggested some inbreeding at this hierarchical level, but this hypothesis was invalidated by a bootstrap-derived confidence interval for Fis including zero.

3.5. Highly admixed genetic structure

In addition to analyzing genetic diversity per sampling region, we also evaluated the overall genetic structure to detect genetic clusters, if any, and to assess their congruence with respect to each sampling region. The number of genetic clusters was determined using the BIC after running the k-means algorithm. For each data set, this criterion started by decreasing sharply (Online Resource 7), demonstrating the presence of genetic structure. However, the signal was not clear for all the data sets, making the choice of the number of genetic clusters rather difficult. But based on the results and motivated by the parsimony principle, we chose K=3 for the remaining analyses and for each data set (except for the cultivated data set genotyped at 10 SSRs where K=6).

On the *10Unik* data set, in the first step of the DAPC, 70 principal components were selected by cross-validation, collectively representing 99.2% of the total variance (Figure 2). In the second step, two linear discriminant functions were used to discriminate the three genetic clusters. The first discriminant function separated clusters 1 and 3 most strongly, and 78.6 % of the individuals from Corsica were grouped in cluster 3. The second discriminant function separated clusters 1 and 2. Cluster 2 grouped most individuals in Var (ForVar and CultVar) and some in Ardèche (CultArdech). However, overall, most individuals (79.8%) of the cultivated and forest types were grouped in cluster 1, pointing to an overall admixed genetic structure in our sample. This was confirmed by the relatively low pairwise Fst calculated between clusters and, as can be seen in the assignment plot (Online Resource 8). Nineteen samples out of 1,050 had a posterior assignment probability for a given genetic cluster of less than 80%. Similar results were obtained with the DAPC at 19 SSRs without MLGs (Online Resource 9, plot 1). Cluster 2 represented 66.4% of all genotyped individuals in most sampling regions. Clusters of forest and cultivated sampling regions in Var and some in Ardèche (n° 3) and Corsica (n° 1) were identified, showing that there was no clear genetic differentiation between the forest and cultivated stands in either of these two sampling regions. A hierarchical AMOVA of the *10Unik* data set (respectively *19Unik*) corroborated this finding(Table 4 and Online Resource 6) and showed that 84.3% of the variance (respectively 84%) was found among samples within clusters. No substantial difference in structure was found between clusters: the variance component at this level was not significantly different from zero, although the Fst confidence interval excluded zero.

When the inbreeding coefficient was calculated per cluster (Table 5 and Online Resource 10), the 95% confidence interval of all the clusters included zero. The mean observed heterozygosity was 0.693 for *10Unik* (respectively 0.703 for *19Unik*) and the mean expected heterozygosity was 0.688 (respectively. 0.666 for *19Unik*).

As cluster 1 in figure 2 contained 79.8% of all the samples, we investigated its sub-structure by performing a DAPC on its samples (Online Resource 9, plot 3). BIC showed an optimal structure at K=3 (Online Resource 7). The resulting sub-clusters were all admixed with low Fst (0.045 – 0.055) even though the confidence intervals excluded zero. Moreover, this sub-structure separated samples from the two most frequently represented sampling regions: 91% of ForAveyron samples belonged to sub-cluster 1 and 82% of ForFinistere samples belonged to sub-cluster 3.

4. Discussion

4.1. Sampling

This work is the first comprehensive survey of genetic diversity and structure of *Castanea sativa* Mill. in France. As such, it fills the sampling gap in France for the benefit of future studies of chestnut structure in Europe. Our study benefited from two projects (The first author's PhD and the FCBA project) which had different goals but whose sampling regions partially overlapped ours, and which used the same genotyping and allele scoring procedures. Combining these projects resulted in a large sampling effort to better assess the overall diversity of cultivated and forest chestnut in France, a crucial component of landscape genetics (Schwartz and McKelvey 2009), although not respective abundance in each sampling regions, which was not our aim in this particular study.

4.2. Diversity indices

The levels of diversity in our sampling regions are comparable with those reported in other studies (Lusini et al. 2014; Mattioni et al. 2017; Mattioni et al. 2013; Skender et al. 2017), similarly, the mean number of alleles per locus are comparable with those obtained in other European regions (Lusini et al. 2014; Pereira-Lorenzo et al. 2017). The high observed heterozygosity in two of our sampling regions, CultArdech and CultLimousin, could be explained by the fact that they were sampled in several local conservatories.

4.3. Redundant diversity among sampling regions and no differentiation between chestnut types

The absence of significant genetic structure between forest and cultivated stands, and the high variance found within sampling regions, implies that each sampling region hosts substantial diversity, mostly shared with the other sampling regions. Such redundancy between sampling regions can be interpreted as the result of human exchanges (Bruneton-Governatori 1999; Conedera et al. 2016; Krebs et al. 2019; Pitte 1986). Concerning Var and Corsica, some information made us think that the sampled forest in these regions may previously have been used as chestnut orchards. One MLG in ForVar region was equivalent to one in CultVar, and forest and cultivated trees from sampling regions of Var and Corsica were grouped in the same genetic cluster. This could be explained by the multipurpose past uses of the forests, as attested by the current owner of the Corsica stands. After performing the AMOVA on the *10Unik* data set, this time after removing the forest sampling regions of Var and Corsica, the Fct among chestnut types had a confidence interval including zero (Online Resource 6).

Redundant genetic diversity in our sampling regions should ensure backup diversity, as long as information about landraces is shared among stakeholders in the different sampling regions. *In situ* sampling revealed that many landraces are multi-clonal. This source of diversity and hence of potential adaptation argues in favor of not reducing a landrace to one arbitrary clone. Even clones should be carefully evaluated, as morphological differences between clones were reported during our field trips, as has been the case in other species (Cipriani et al. 2010). All this is particularly interesting at a time when chestnut valuation tends to be based on heritage, with significance and quality marks based on local landraces (e.g., AOC Châtaigne d'Ardèche, AOC Farine de châtaigne Corse – Farina castagnina corsa, Label rouge Marron du Périgord). Genetics could provide authorities with arguments to justify certifying landraces are "local". On the other hand, even if a landrace has been cultivated for centuries in a particular place, this may also be the case elsewhere. Therefore, one might rightfully ask whether the quality of local chestnut comes from its locality. For crops like chestnut, usage and practices may be at least as important as genetics to give value to chestnut for growers and consumers (Dupré 2002, 2005; Martin et al. 2017).

4.4. A highly admixed genetic structure

Paralleling the high redundancy between sampling regions, the genetic structure from the DAPC remained low or moderate among subgroups. The main finding here was the high admixture between the regions we sampled, both forest and cultivated. There was thus no clear-cut distinction between sampling regions considered as forest or as cultivated, as confirmed by the AMOVA. This result was not completely unexpected given that chestnut is an outcrossing species and that gene flow between forest and cultivated stands is known to occur, together with changes in usage over time and in certain practices such as forests being used as a source of seedlings for rootstock, good quality fruits as a source of seedlings to plant forests, peasant woods in Limousin (personal communication), and "instant domestication" (Pereira-Lorenzo et al. 2019).

Characterizing genetic diversity (respectively structure) as high (strong) or low (weak) can be particularly risky as it has to be in relative terms. Like (Pereira-Lorenzo et al. 2019), we found a Fct close to zero between wild and cultivated chestnut. When characterizing the genetic diversity and structure of wild chestnut from Italy, Spain, Greece and Turkey,

Mattioni et al. (2013) obtained a molecular variance among three clusters of 11.58%, i.e. lower than our 15.7%. They also found a Fst of 12.6% between genetic clusters representing Italian and Spanish samples, higher than our 9%.

4.5. A hypothetical common glacial refugia for French and Iberian chestnut

The genetic structure inferred from our samples did not necessarily match the sampling regions. This result was also expected for a continuously dispersed species affected by human management like European chestnut. Moreover, an admixed genetic structure was consistent with the known patterns of divergence and distribution of chestnut (Mattioni et al. 2017), combined with evidence from fossil pollen of several tree species suggesting that chestnut populations originating from Italy or the Balkans spread into the Iberian Peninsula from the north (Grivet and Petit 2003; Petit 2003).

In the EU database (2017) and in (Pereira-Lorenzo et al. 2019), « Luguesa » was classified with the Italian group of cultivars. In our analyses, it was found in the south-eastern cluster (cluster n°3 in Online Resource 9, plot 2) grouped together with « Puga » and « Raigona », which were both originally classified in the Iberian group, whereas the other Spanish cultivars were found in cluster n°2. Therefore, the majority of the Iberian group seems to match the main French group, suggesting that both originated in the same glacial refugia.

Before removal of hybrid individuals, the Basque sampling region was represented in the 10 (respectively 19) dataset by 119 (respectively 10) successfully genotyped non-redundant individuals. This high number of admixed individuals is an important feature of the actual chestnut forest there, resulting from the long history of interspecific hybridization in this region which extends on both sides of the border between Spain and France (Pereira-Lorenzo et al. 2017). It is further substantiated by the high prevalence of trees tolerant to ink disease, as found in artificial inoculation experiments (Robin et al., in preparation).

A European analysis of the genetic structure of European chestnut including a significant French sampling remains to be done.

4.6. Future outlook of SNP genotyping

The markers we used were selected after an extensive review of the literature (by us for the 13 SSR, and independently by Pereira-Lorenzo et al. 2018), and allele scoring was the subject of a recent optimization by Pereira-Lorenzo et al. (2018). Nevertheless, we faced the usual difficulties and drawbacks of microsatellites, i.e., errors and uncertainties in allele calling, difficulty in data comparison and transferability across labs and collaborators over time, and the huge amount of time needed to perform the analysis, as emphasized in previous studies (reviewed by (Guichoux et al. 2011)).

We consequently set up a small project to define nuclear SNPs, at least to check clear duplicates (in the case of good quality genotyping results) and putative duplicate (in the case of low quality results) among samples from variety repositories. In a few months, we re-genotyped about 500 samples with up to 160 SNPs and confirmed all suspected duplicates. A detailed description of this work will the subject of a separate article.

5. Conclusion

In conclusion, this study revealed the genetic diversity and structure of French forest and cultivated chestnut across most of its range. We showed high diversity redundancy between sampling regions and a weak genetic structure. Based on external knowledge, the influence of human activity is the most probable explanation for this finding. Three main clusters were found, one in Corsica, one in the south east of France, probably partially matching a previously-described Italian group of cultivars, and one main admixed cluster matching the Iberian cultivars. This confirms existing historical knowledge on land use changes, the movement of landraces, and « instant domestication » landraces. Furthermore, we provide evidence for a common origin of most of the French and Iberian chestnut, except those from the south east of France, which were associated with the Italian gene pool. We believe our work provides useful information for conservation planning purposes and for cooperation between chestnut non-profit associations and groups of growers interested in landrace conservation and diffusion.

References

Adamack AT, Gruber B (2014) POPGENREPORT: simplifying basic population genetic analyses in R. Methods in Ecology and Evolution 5:384–387. doi: 10.1111/2041-210X.12158

Agapow P-M, Burt A (2001) Indices of multilocus linkage disequilibrium. Molecular Ecology Notes 1:101–102. doi: 10.1046/j.1471-8278.2000.00014.x

Aumeeruddy-Thomas Y, Therville C, Lemarchand C, et al (2012) Resilience of Sweet Chestnut and Truffle Holm-Oak Rural Forests in Languedoc-Roussillon, France: Roles of Social-Ecological Legacies, Domestication, and Innovations. Ecology and Society 17:. doi: 10.5751/ES-04750-170212

Beccaro GL, Torello-Marinoni D, Binelli G, et al (2012) Insights in the chestnut genetic diversity in Canton Ticino (Southern Switzerland). Silvae Genetica 61:292–300

Beghè D, Ganino T, Dall'Asta C, et al (2013) Identification and characterization of ancient Italian chestnut using nuclear microsatellite markers. Scientia Horticulturae 164:50–57. doi: 10.1016/j.scienta.2013.09.009

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 453–455

Bruneton-Governatori A (1999) Le pain de bois□: Ethnohistoire de la châtaigne et du châtaignier, Lacour. C. Lacour, Nîmes

Buck EJ, Hadonou M, James CJ, et al (2003) Isolation and characterization of polymorphic microsatellites in European chestnut (Castanea sativa Mill.). Molecular Ecology Notes 3:239–241. doi: 10.1046/j.1471-8286.2003.00410.x

Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I□: One-table methods. R news 4:5–10

Cipriani G, Spadotto A, Jurman I, et al (2010) The SSR-based molecular profile of 1005 grapevine (Vitis vinifera L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theoretical and Applied Genetics 121:1569–1585. doi: 10.1007/s00122-010-1411-9

CIRAD Plateau Génotypage CIRAD / Contacts - GPTR Génotypage. https://www.gptr-lr-genotypage.com/contacts/plateau-genotypage-cirad

Conedera M, Tinner W, Krebs P, et al (2016) Castanea sativa in Europe: distribution, habitat, usage and threats

395 Dupré L (2002) Du marron à la châtaigne d'Ardèche. La relance d'un produit régional. Éditions du CTHS.

Dupré L (2005) Classer et nommer les fruits du châtaignier ou la construction d'un lien à la nature. Natures Sciences Sociétés 13:395–402. doi: 10.1051/nss:2005060

Eriksson G, Pliura A, Fernández-López J, et al (2005) Management of genetic resources of the multi-purpose tree species Castanea sativa Mill. In: Acta Horticulturae. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp 373–386

Excoffier L, Smouse PE (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. 13

FAO (2018) Faostat data for chestnut. http://www.fao.org/faostat/. Accessed 19 Apr 2019

FAQ List for Arlequin 2.000 FAQ List for Arlequin 2.000. http://cmpg.unibe.ch/software/arlequin/software/2.000/doc/faq/faqlist.htm#negative%20variance%20components. Accessed 31 May 2019

Fernández-Cruz J, Fernández-López J (2016) Genetic structure of wild sweet chestnut (Castanea sativa Mill.) populations in northwest of Spain and their differences with other European stands. Conserv Genet 17:949–967. doi: 10.1007/s10592-016-0835-4

Fernández-López J, Fernández-Cruz J (2015) Identification of traditional Galician sweet chestnut varieties using ethnographic and nuclear microsatellite data. Tree Genetics & Genomes 11:. doi: 10.1007/s11295-015-0934-2

FranceAgriMer (2017) chiffres clés Filière Fruits et Légumes 2016

Frascaria N, Blaise S, Guittet J, Lefranc M (1991) Analysis of the spatial genotype distribution in a small chestnut tree population (Castanea sativa MILL.). Spatial autocorrelation and F-statistics. In: Fineshi, S and Malvoti, ME and Cannata, F and HATTEMER, HH (ed) Biochemical markers in the population genentics of forest trees. S P B Academic Publ, The Hague, p 219

Frascaria N, Chanson B, Thibaut B, Lefranc M (1992) Gene diversity and wood quality charasteristics in chestnut (Castanea sativa Mill.). Annales des sciences forestières 49:49–62. doi: 10.1051/forest:19920105

Frascaria N, Lefranc M (1992) Chestnut trade - A new aspect of the differenciation fo chestnut tree (Castanea sativa MILL.) populations in France. Annales des sciences forestières 49:75–79. doi: 10.1051/forest:19920107

Gobbin D, Hohl L, Conza L, et al (2007) Microsatellite-based characterization of the Castanea sativa cultivar heritage of southern Switzerland. Genome 50:1089–1103. doi: 10.1139/G07-086

Goodman SN, Fanelli D, Ioannidis JPA (2016) What does research reproducibility mean? Science Translational Medicine 8:341ps12-341ps12. doi: 10.1126/scitranslmed.aaf5027

Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes 5:184–186. doi: 10.1111/j.1471-8286.2004.00828.x

Grivet D, Petit RJ (2003) Chloroplast DNA phylogeography of the hornbeam in Europe: Evidence for a bottleneck at the outset of postglacial colonization. 10

Guichoux E, Lagache L, Wagner S, et al (2011) Current trends in microsatellite genotyping: TRENDS IN MICROSATELLITE GENOTYPING. Molecular Ecology Resources 11:591–611. doi: 10.1111/j.1755-0998.2011.03014.x

Harris SA, Robinson JP, Juniper BE (2002) Genetic clues to the origin of the apple. Trends in Genetics 18:426–430. doi: 10.1016/S0168-9525(02)02689-6

IGN (2007) BD Forêt® V2. http://professionnels.ign.fr/bdforet. Accessed 19 Apr 2019

IGN (2016) RPG. http://professionnels.ign.fr/rpg#tab-1

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. doi: 10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94. doi: 10.1186/1471-2156-11-94

Jost L (2008) GST and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026. doi: 10.1111/j.1365-294X.2008.03887.x

Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of (GA)n Microsatellite Loci from Quercus Robur. Hereditas 129:183–186. doi: 10.1111/j.1601-5223.1998.00183.x

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics 6:. doi: 10.3389/fgene.2015.00208

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. doi: 10.7717/peerj.281

Krebs P, Conedera M, Pradella M, et al (2004) Quaternary refugia of the sweet chestnut (Castanea sativa Mill.): an extended palynological approach. Vegetation History and Archaeobotany 13:. doi: 10.1007/s00334-004-0041-z

Krebs P, Pezzatti GB, Beffa G, et al (2019) Revising the sweet chestnut (Castanea sativa Mill.) refugia history of the last glacial period with extended pollen and macrofossil evidence. Quaternary Science Reviews 206:111–128. doi: 10.1016/j.quascirev.2019.01.002

Lassois L, Denancé C, Ravon E, et al (2016) Genetic Diversity, Population Structure, Parentage Analysis, and Construction of Core Collections in the French Apple Germplasm Based on SSR Markers. Plant Molecular Biology Reporter 34:827–844. doi: 10.1007/s11105-015-0966-7

Lusini I, Velichkov I, Pollegioni P, et al (2014) Estimating the genetic diversity and spatial structure of Bulgarian Castanea sativa populations by SSRs: implications for conservation. Conservation Genetics 15:283–293. doi: 10.1007/s10592-013-0537-0

Marinoni D, Akkak A, Bounous G, et al (2003) Development and characterization of microsatellite markers in Castanea sativa (Mill.). Molecular Breeding 11:127–136. doi: 10.1023/A:1022456013692

Martín MA, Mattioni C, Cherubini M, et al (2010a) Genetic characterisation of traditional chestnut varieties in Italy using microsatellites (simple sequence repeats) markers. Annals of Applied Biology 157:37–44. doi: 10.1111/j.1744-7348.2010.00407.x

Martin MA, Mattioni C, Cherubini M, et al (2010b) Genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genetics & Genomes 6:735–744. doi: 10.1007/s11295-010-0287-9

Martin MA, Monedero E, Martín LM (2017) Genetic monitoring of traditional chestnut orchards reveals a complex genetic structure. Annals of Forest Science 74:. doi: 10.1007/s13595-016-0610-1

Mattioni C, Cherubini M, Micheli E, et al (2008) Role of domestication in shaping Castanea sativa. Tree Genetics & Genomes 4:563–574. doi: 10.1007/s11295-008-0132-6

Mattioni C, Martin MA, Chiocchini F, et al (2017) Landscape genetics structure of European sweet chestnut (Castanea sativa Mill): indications for conservation priorities. Tree Genetics & Genomes 13:39. doi: 10.1007/s11295-017-1123-2

Mattioni C, Martin MA, Pollegioni P, et al (2013) Microsatellite markers reveal a strong geographical structure in European populations of Castanea sativa (Fagaceae): Evidence for multiple glacial refugia. Am J Bot 100:951–961. doi: 10.3732/ajb.1200194

Mellano MG, Beccaro GL, Donno D, et al (2012) Castanea spp. biodiversity conservation: collection and characterization of the genetic diversity of an endangered species. Genetic Resources and Crop Evolution 59:1727–1741. doi: 10.1007/s10722-012-9794-x

Mellano MG, Torello-Marinoni D, Boccacci P, et al (2018) Ex situ conservation and characterization of the genetic diversity of Castanea spp. Acta Horticulturae 1–6. doi: 10.17660/ActaHortic.2018.1220.1

Míguez-Soto B, Fernández-Cruz J, Fernández-López J (2019) Mediterranean and Northern Iberian gene pools of wild Castanea sativa Mill. are two differentiated ecotypes originated under natural divergent selection. PLOS ONE 14:e0211315. doi: 10.1371/journal.pone.0211315

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, NY, USA

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420. doi: 10.1093/bioinformatics/btp696

Pereira-Lorenzo S, Costa RML, Ramos-Cabrer AM, et al (2010) Variation in grafted European chestnut and hybrids by microsatellites reveals two main origins in the Iberian Peninsula. Tree Genetics & Genomes 6:701–715. doi: 10.1007/s11295-010-0285-y

Pereira-Lorenzo S, Fernandez-Lopez J (1997) Description of 80 cultivars and 36 clonal selections of chestnut (Castanea sativa Mill) from Northwestern Spain. Fruit varieties journal 51:13–27

Pereira-Lorenzo S, Costa R, Anagnostakis S, et al (2016) Interspecific Hybridization of Chestnut. In: Annaliese S. Mason (ed) Polyploidy and Hybridization for Crop Improvement, CRC Press. Boca Raton, USA, p (p. 377-407)

Pereira-Lorenzo S, Ramos-Cabrer A, Barreneche T, et al (2017) Database of European chestnut cultivars and definition of a core collection using simple sequence repeats. Tree Genetics & Genomes 13:. doi: 10.1007/s11295-017-1197-x

Pereira-Lorenzo S, Ramos-Cabrer AM, Barreneche T, et al (2019) Instant domestication process of European chestnut cultivars. Annals of Applied Biology. doi: 10.1111/aab.12474

Petit RJ (2003) Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. Science 300:1563–1565. doi: 10.1126/science.1083264

Pitte J-R (1986) Terres de Castanide. Hommes et paysages du Châtaignier de l'Antiquité à nos jours, Fayard. Paris

Quintana J, Contreras A, Merino I, et al (2014) Genetic characterization of chestnut (Castanea sativa Mill.) orchards and traditional nut varieties in El Bierzo, a glacial refuge and major cultivation site in northwestern Spain. Tree Genetics & Genomes 11:0. doi: 10.1007/s11295-014-0826-x

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Roces-Díaz JV, Jiménez-Alfaro B, Chytrý M, et al (2018) Glacial refugia and mid-Holocene expansion delineate the current distribution of Castanea sativa in Europe. Palaeogeography, Palaeoclimatology, Palaeoecology 491:152–160. doi: 10.1016/j.palaeo.2017.12.004

Sauvezon R, Sauvezon A, Sunt C (2000) Châtaignes et Châtaigniers, Edisud

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conservation Genetics 10:441–452. doi: 10.1007/s10592-008-9622-1

Skender A, Kurtovic M, Pojskic N, et al (2017) Genetic structure and diversity of European chestnut (Castanea sativa Mill.) populations in western Balkans: On a crossroad between east and west. Genetika 49:613–626. doi: 10.2298/GENSR1702613S

Steinkellner H, Fluch S, Turetschek E, et al (1997) Identification and characterization of (GA/CT)n- microsatellite loci from Quercus petraea. Plant Mol Biol 33:1093–1096. doi: 10.1023/A:1005736722794

UPOV (1989) Guidelines for the conduct of tests for distinctnes□: chestnut

Villa TCC, Maxted N, Scholten M, Ford-Lloyd B (2005) Defining and identifying crop landraces. Plant Genetic Resources: Characterization and Utilization 3:373–384. doi: 10.1079/PGR200591

Tables

Table 1: Description of the sampling regions and sampled trees

French department	Sampling region	Contributing partners to the sampling	*In/Ex-situ* Off.	n10 (n19)	Sum N_10 (Sum N_19)	u10 (u19)	Sum U_10 (Sum U_19)
Hérault, Gard, Drôme and Ardèche	CultArdech	CDA Ardèche + CRA Occitanie + SDCA	E	74 (75)	492 (336)	47 (49)	333 (220)
Ariège	CultAriege	Renova	I	89 (60)		64 (35)	
Aveyron	CultAveyron	ACRC + P.Rance	E+I	97 (37)		70 (24)	
Corsica	CultCorsica	GRPTCMC	I	48 (48)		38 (39)	
Hautes-Pyrénées	CultHtpyr	Châtaigne des Pyrénées	I	51 (29)		42 (25)	
Corrèze + Haute-Vienne	CultLimousin	C.Pommes + PNR	E	113 (83)		59 (44)	
Var	CultVar	SPCV	I	20 (4)		13 (4)	
Isère	ForArdech	FCBA	I	86 (0)	722 (306)	86 (0)	717 (301)
Aveyron	ForAveyron	FCBA	Off.	140 (29)		140 (29)	
Pyrénées-Atlantiques	ForBasque	FCBA	Off.	24 (1)		24 (1)	
Cantal	ForCantal	FCBA	I	24 (24)		22 (22)	
Corsica	ForCorsica	FCBA	Off.	116 (71)		116 (71)	
Finistère	ForFinistere	FCBA	I + Off.	248 (97)		248 (97)	
Gard	For Gard	FCBA	I	30 (30)		30 (30)	
Gironde	ForGironde	FCBA	I	8 (8)		5 (5)	
Hérault	ForHerault	FCBA	I	16 (16)		16 (16)	
Var	ForVar	FCBA	I	30 (30)		30 (30)	
Total					1214 (642)		1050 (521)

Each sampling region contains one or several sampling sites in geographically close stands (For = high forest, Cult = cultivated). *In/Ex situ*/Off.: I = *in situ*, E = *ex situ*, Off. = offspring originating from nuts harvested in forests. Even though the total number of genotyped trees was 1,401 at 10 SSR (respectively 693 at 19 SSRs), the numbers of trees listed in Table 1 are limited to those with no more than 20% of missing alleles and after detected interspecific hybrids were removed. The number of SSRs are those remaining after the removal of loci with null alleles and more than 5% of missing data. n10 (respectively n19): refers to the number of samples genotyped at 10 SSRs (respectively 19 SSRs). Sum N_10 (respectively Sum N_19): refers to the number of samples per chestnut type genotyped at 10 SSRs corresponding to the *10All* data set (respectively 19 SSRs corresponding to the *19All* data set). u10 (respectively u19): refers to the number of unique samples genotyped after the removal of loci with null alleles, at 10 SSRs (respectively 19 SSRs). Sum U_10 (respectively U_19): refers to the number of unique samples per chestnut type at 10 SSRs corresponding to the *10Unik* data set (respectively at 19 SSRs corresponding to the *19Unik* data set).

425 **Table 2: Genetic diversity indices for 17 French sampling regions at 10 loci without MLGs (*10Unik* data set)**

Sampling Regions	N	Na	Ne	Ho	He	H	Ia	rbarD	Fis
CultArdech	47	64	3.125	0.717	0.673	3.85	0.489*	0.055*	-0.104 [-0.191;-0.009]
CultAriege	64	69	3.448	0.561	0.705	4.16	0.259*	0.029*	0.031 [-0.080;0.042]
CultAveyron	70	68	3.215	0.7	0.684	4.25	0.467*	0.052*	-0.054 [-0.163;0.042]
CultCorsica	38	63	3.448	0.715	0.701	3.64	0.281	0.031	-0.048 [-0.153;0.030]
CultHtPyr	42	63	3.195	0.665	0.678	3.74	0.444*	0.05*	-0.016 [-0.132;0.084]
CultLimousin	59	66	3.367	0.731	0.697	4.08	0.556*	0.062*	-0.064 [-0.139;0.0009]
CultVar	13	40	2.262	0.637	0.537	2.56	2.880*	0.336*	-0.158 [-0.369;0.141]
ForArdech	86	66	3.205	0.645	0.684	4.45	0.218	0.024	0.012 [-0..094;0.122]
ForAveyron	140	59	2.786	0.643	0.638	4.94	0.237*	0.026*	-0.044 [-0.146;0.010]
ForBasque	24	45	2.513	0.629	0.590	3.18	0.391	0.044	-0.073 [-0.143;0.010]
ForCantal	22	48	2.618	0.664	0.604	3.09	0.293	0.033	-0.121 [-0.205 ;-0.050]
ForCorsica	116	70	3.472	0.712	0.709	4.75	0.210*	0.023*	-0.041 [-0.130 ;-0.046]
ForFinistere	248	88	3.704	0.722	0.728	5.51	0.086	0.01	-0.036 [-0.123;0.032]
ForGard	30	57	2.915	0.743	0.646	3.40	0.247	0.028	-0.164 [-0.260 ;-0.095]
ForGironde	5	42	3.788	0.620	0.662	1.61	0.824	0.098	0.149 [0.005;0.325]
ForHerault	16	50	3.236	0.769	0.669	2.77	0.279	0.031	-0.139 [0.227 ;-0.050]
ForVar	30	48	2.457	0.624	0.583	3.40	0.203	0.023	-0.109 [-0.285;0.251]
Total	1050	113	3.846	0.681	0.658	6.94	0.173	0.019	-0.058

N: number of unique individuals genotyped per sampling region; Na: number of alleles; Ne: mean number of effective alleles; Ho: observed heterozygosity; He: expected heterozygosity; H: Shannon-Weiner diversity index; Ia: index of association; rbarD: standardized index of association; Fis: inbreeding coefficient, with 95% confidence interval (CI). Asterisks indicate significant *p* values at the 0.001 threshold. The "Total" row contains the sum of N, total Na and total
430 H, and the mean for the other indices.

Table 3: Hierarchical AMOVA and F-statistics for 17 French sampling regions at 10 loci without MLGs (*10Unik* data set). df: degree of freedom, Alter: alternative hypothesis, 95% confidence intervals, ***: p value ≤ 0.001

Source of variation	df	Variance component	% variation	*p* value	Alter.	F statistic
Among chestnut type	1	-0.055	-1.36	0.794	greater	Fct -0.007 [-0.011; -0.002]
Among sampling regions within chestnut types	15	0.613	15.10	0.001***	greater	Fst 0.082 [0.070; 0.093]
Within sampling regions	1033	3.501	86.26	0.001***	less	Fis -0.002 [-0.023; 0.024]
Total	1049	4.059	100.00			Fit 0.074 [0.049; 0104]

Table 4: Hierarchical AMOVA and F-statistics for three genetic clusters at 10 loci without MLGs.
df: degrees of freedom, Alter: alternative hypothesis, 95% confidence intervals, ***: p value ≤ 0.001

Source of Variation	df	Variance component	% Variation	*p* value	Alter.	F statistic
Among clusters	2	0.702	15.7	0.702	greater	Fst 0.090 [0.069; 0.111]
Among samples within clusters	1047	3.782	84.3			Fis 0.037 [0.015; 0.065]
Total	1050		100.00			Fit 0.123 [0.091; 0.163]

Table 5: **Within-cluster genetic variability at 10 loci without MLGs**
N: number of unique individuals genotyped per cluster; Na: number of alleles; Ho: observed heterozygosity; He: expected heterozygosity; Fis: inbreeding coefficient with 95% confidence interval. The "total" row contains the sum of N, the total number of alleles (Na), and the mean for the other indices.

Clusters	N	Na	Ho	He	Fis
Cluster 1	838	103	0.686	0.724	0.008 [-0.081; 0.066]
Cluster 2	81	70	0.666	0.621	-0.125 [-0.254; 0.047]
Cluster 3	131	82	0.728	0.720	-0.054 [-0.136; 0.019]
Total	1049	113	0.693	0.688	-0.054

Figure captions

Fig. 1: Map of sampling regions. The distribution of chestnut forest areas where chestnut accounts for at least 75% of the leaf cover, which represents about 50% of the total chestnut-comprising forest area in France (IGN 2007) (in green, and the distribution of cultivated chestnut and orchards (IGN 2016) (in orange). Each dot represents a sampling region where chestnut forest or orchard is present (this qualitative information does not reflect the relative areas or number of trees).

Fig. 2: Genetic clustering of French chestnuts at 10 loci without MLGs. A: Plot of linear discriminant analysis with genetic clusters (in color) and forest/cultivated status of chestnut individuals (symbols). B: Plot of principal components retained in the analysis. C: Plot of discriminant components retained in the analysis. D: Number of samples per sampling region assigned to each genetic cluster. E: Pairwise Fst between clusters.

Contributions of the co-authors

Conceptualization: Cathy Bouffartigue, Timothée Flutre and Luc Harvengt; **Methodology**: Cathy Bouffartigue, Sandrine Debille, Ana Ramos Cabrer, Timothée Flutre, Luc Harvengt; **Software**: Cathy Bouffartigue, Timothée Flutre; **Validation**: Cathy Bouffartigue, Timothée Flutre; **Formal Analysis**: Cathy Bouffartigue, Ana Ramos Cabrer, Santiago Pereira-Lorenzo, Timothée Flutre; **Investigation**: Cathy Bouffartigue, Sandrine Debille, Olivier Fabreguette, Ana Ramos Cabrer**; Resources**: Cathy Bouffartigue, Sandrine Debille, Olivier Fabreguette, Ana Ramos Cabrer, Santiago Pereira-Lorenzo, Luc Harvengt; **Data curation**: Cathy Bouffartigue, Sandrine Debille; **Writing-original draft**: Cathy Bouffartigue, Timothée Flutre; **Writing-review&editing**: Cathy Bouffartigue, Sandrine Debille, Ana Ramos Cabrer, Santiago Pereira Lorenzo, Timothée Flutre, Luc Harvengt ; **Visualization**: Cathy Bouffartigue; **Supervision**: Timothée Flutre, Luc Harvengt; **Project administration**: Cathy Bouffartigue, Timothée Flutre, Luc Harvengt; **Funding acquisition**: Cathy Bouffartigue, Luc Harvengt.

Acknowledgments

This paper is part of the PhD of the first author who is grateful of her supervisors Laurent Hazard and Nathalie Couix of the INRA, AGIR and Timothée Flutre, INRA, AGAP.

We would thank the local partners who introduced the first author to the chestnut, helped in sampling and shared their knowledge. In particular Loïc Vincent and Laëticia Faliez from ACRC, Stéphane Artigues from Châtaigne des Pyrénées, Francis Michaux, Brigitte Boitel and Théo Churoux from Rénova, Michel Gauthier and Michel Chauprade from Croqueurs de Pommes du Limousin, Romain Barret from SPCV, Guy Ginisti and Yvon Viala from Paysans du Rance.

We would like to thank the following people who provided access to forest samples or sampling regions:

In Corsica: Carine Franchi from GRPTCMC and François Luro for INRA Corsica,

In Limousin: Laure Dangla who coordinated contribution from all partners involved in regional chestnut conservation (with Croqueurs de pommes du Limousin and the Parc Naturel Regional Perigord-Limousin)

In Ardèche and Eastern Occitanie: Eric Bertoncello, Helina Deplaude and Anne Boutitie (chambre départementale d'agriculture d'Ardèche (CDA Ardèche), Syndicat des producteurs de châtaigne d'Ardèche (SDCA) et chambre régionale d'agriculture d'Occitanie (CRA Occitanie)

Other public and private landowners who granted us access to their forest stands in Brittany, Gard, Aveyron, Basque country and Cantal

We are very grateful to advisors and partners taking part to other aspects of our work on chestnut, particularly Sébastien Cavaignac from Invenio, Bernard Hennion and Fabrice Lheureux from CTIFL, Patrick Léger, Xavier Capvielle and Cécile Robin from INRA, Biogeco,Teresa Barreneche from INRA, Biologie du Fruit et Pathologie, Josephina Fernandez-Lopez from CIF-Lourizan (Spain), and the members of the national workgroup on chestnut forestry set up by the National Center for Private Forest owners (CNPF) and particularly its chairman René Lempire as well as, Jean Lemaire and Sabine Girard.

We would also like to thank FCBA colleagues involved in sampling: Jean-Mathieu de Boisesson, Francis Melun, Marjorie Vidal, Francis Canlet, Thierry Fauconnier and Stéphane Grulois.

The authors thank the anonymous reviewers for critically reading the manuscript and suggesting substantial improvements.

Funding

A grant of The Fondation de France and a regional research program for and on rural development (PSDR4-Occitanie, France) fund the PhD of the first author.

In 2016/2017, local partners (Rénova, ACRC and Châtaigne des Pyrénées) collected subsidies from the Conservatory of the Regional Biological Patrimony (CPBR) for the genotyping.

The FCBA project on forest samples was funded by the Conseil Régional de Nouvelle Aquitaine through the "sélection châtaignier à bois" project (2015-2016) contrat number 15007198-046 and grant "FEDER-FSA 2014-2020 Dossier 47010", and "sélection châtaignier" project, grant n°16008302-043 and "FEDER-FSE 2014-2020 – AXE 1 n° 3296610". The two FEDER grant parts are funds obtained from European Union by the Conseil Régional de Nouvelle Aquitaine. We thank the members of the FCBA external advisory board on Forestry and members of the National Center for private forest owners for its help to setup and manage the above-mentioned projects

The Xylobiotech platform is funded through the Equipe Xyloforest project (grant ANR-10-EQPX-16).

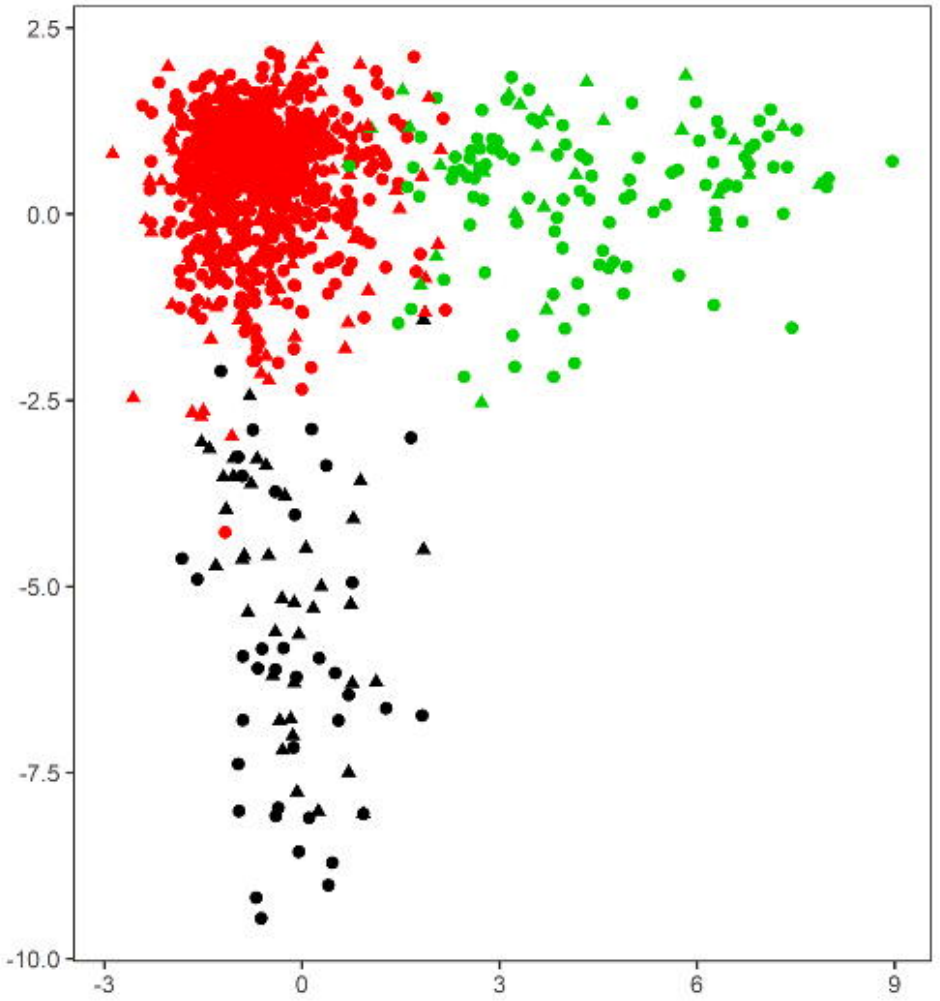
Data availability:

The datasets analysed during the current study are available in the data.inra repository at a private link https://data.inra.fr/privateurl.xhtml?token=8c03a83c-be4d-4984-972f-7808558b4539. The link will be published after the review process will be completed.

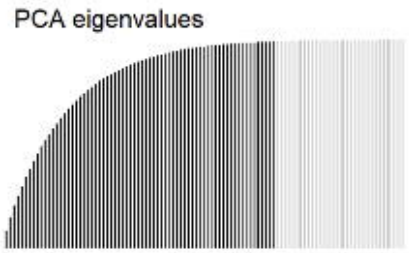
Declaration on conflicts of interest:

The authors declare that they have no conflict of interest.

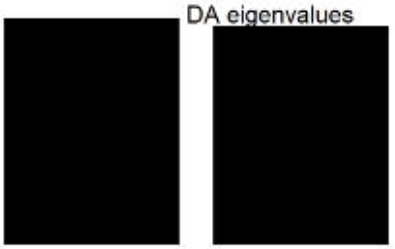
A



B



C

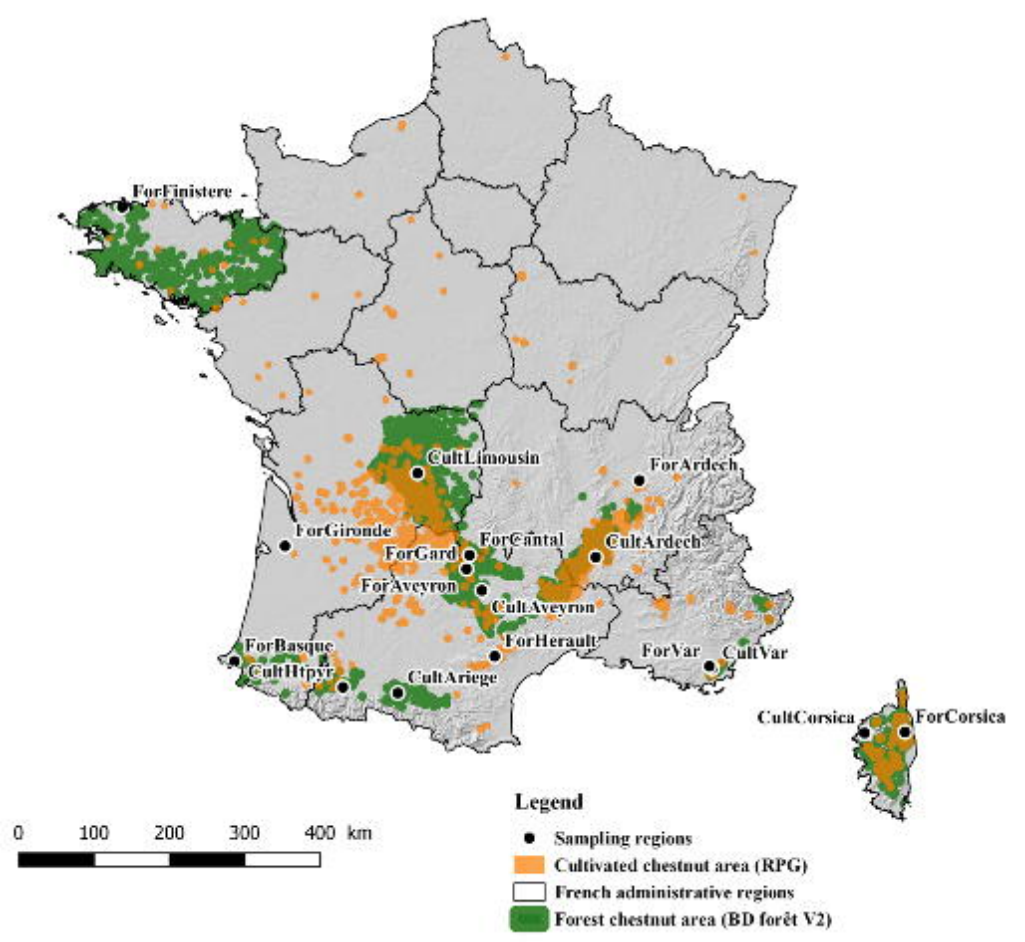


D

Sampling Regions	1	2	3	Sum
CultArdech	24	23		47
CultAriege	61	1	2	64
CultAveyron	63	7		70
CultCorsica	13		25	38
CultHtPyr	42			42
CultLimousin	59			59
CultVar		13		13
ForArdech	78	7	1	86
ForAveyron	140			140
ForBasque	23		1	24
ForCantal	22			22
ForCorsica	19	1	96	116
ForFinistere	246		2	248
ForGard	30			30
ForGironde	3		2	5
ForHerault	14		2	16
ForVar	1	29		30
Sum	838	81	131	1050

E

Clusters	1	2
2	0.101 [0.067;0.139]	
3	0.065 [0.044;0.084]	0.130 [0.067;0.213]


ForFinistere
CultLimousin
ForArdech
ForGironde
ForCantal
ForGard
CultArdech
ForAveyron
CultAveyron
ForBasque
ForHerault
ForVar
CultVar
CultHtpyr
CultAriege
CultCorsica
ForCorsica
0
100
200
300
400 km
Legend
Sampling regions
Cultivated chestnut area (RPG)
French administrative regions
Forest chestnut area (BD forêt V2)