Genetic characterization of Italian and Spanish wild and domesticated chestnut trees

(Article begins on next page)

10 July 2024

1 **Genetic characterization of Italian and Spanish wild and domesticated chestnut**

2 **trees**

3 **Sara Alessandri[1], Ana Maria Ramos Cabrer[2], M. Angela Martìn[3], Claudia Mattioni[4], Santia-**

4 **go Pereira-Lorenzo[2], Luca Dondini[1]**

5 [1]DISTAL - Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy

6 [2]USC - Department of Crop Production and Projects of Engineering, University of Santiago de

7 Compostela, Lugo, Spain

8 [3]UCO – Department of Genetics, University of Cordoba, Cordoba, Spain

9 [4] IRET - Research Institute on Terrestrial Ecosystems, National Research Council, Porano, Italy

10 **Corresponding author:** Sara Alessandri - sara.alessandri4@unibo.it

11

12 ## Abstract

13 The only native species of the *Castanea* genus in Europe is C*astanea sativa* Mill., a widespread and

14 important multipurpose tree species in the Mediterranean area that provides fruit, wood and shelter

15 for hives. With the aim of expanding the knowledge of the genetic variability of the chestnut spe-

16 cies (wild trees and varieties) and promoting the traceability of local products, an analysis based on

17 16 SSRs was carried out on 630 single trees from Italy and Spain. 319 unique genotypes were iden-

18 tified. A Bayesian approach combined with the Markov Chain Monte Carlo (MCMC) simulation

19 method revealed the existence of two genetically distinct groups of chestnuts: Cluster 1 (Spain) and

20 Cluster 2 (Italy), with a clear separation between the cultivars from (northern and southern) Spain

21 and from Italy. The results also confirmed a common genetic structure between chestnut popula-

22 tions from southern Spain and southern Italy, which is the result of historical events and long-term

1

23 human impact. The results showed no genetic differentiation between chestnut cultivars (grafted

24 trees) and wild chestnut trees, probably as a consequence of the proximity of orchards and natural

25 populations, which resulted in a gene flow between them.

26

29

30 1 Introduction

31 The *Castanea sativa* Mill. species belongs to the *Fagaceae* family and is the only native species of

32 the genus *Castanea* in Europe. Nowadays, its widespread distribution is the result of both natural

33 and anthropogenic factors that affected the species over time (Conedera et al., 2004). Climatic con-

34 ditions have influenced chestnut distribution since the Pleistocene; during this glacial period, the

35 species was confined to limited and climatically stable areas called refugia (Krebs et al., 2004;

36 Mattioni et al., 2013). Subsequently, during the postglacial period, the natural recolonization of

37 chestnuts started from these areas.

38 However, chestnut distribution was also influenced by human colonization and migration. In south-

39 ern Italy, chestnut trees were introduced by the ancient Greeks about 5,000 years ago. Later, the

40 Romans spread this species in Europe (Huntley and Birks, 1983; Bernetti, 1995; Krebs et al., 2004

41 and 2019; Roces-Diaz et al., 2018).

42 In the last 20 years, several studies have described in detail the fundamental role of the historical

43 and natural processes that resulted through time in the establishment of the varietal genetic diversity

44 of chestnuts that exists to this day (Martín et al., 2009; Pereira-Lorenzo et al., 2010, 2011; Marinoni

45 et al., 2013; Lusini et al., 2014, Villani et al., 1999; Martín et al., 2012; Fernández-Cruz and Fer-

46 nández-López, 2016; Mattioni et al., 2013 and 2017). While the genetic pool of the domesticated

2

trees was influenced by human selection, propagation and hybridization (to improve fruit or wood quality traits, as well as resistance to abiotic and biotic stress), wild trees were subjected to natural selection to adapt to different environmental and geographical conditions (Barrett and Schluter, 2008; Nishio et al., 2021). In particular, the selective pressure to which wild populations were subjected favored different allele pools involved in the trees' adaptation to different regional conditions, such as adaptation to drought (Soto et al., 2019; Alcaide et al., 2019; Castellana et al., 2021). Furthermore, wild chestnut genetic variability has also been affected by the proximity of orchards, mainly due to pollen diffusion and male-sterility, which have contributed to the inclusion of local and nonlocal new alleles (Lopez et al., 2021). It is also worth mentioning the importance of non-grafted giant chestnuts as a reservoir of genetic diversity, which represent the basis from which the selection and cultivation process commenced (Pereira-Lorenzo et al., 2019).

Despite the different selection process, domesticated chestnuts have preserved most of the diversity found in the oldest wild tree populations (non-grafted giant chestnuts; Pereira-Lorenzo et al., 2019). However, the hybridization conducted by humans improved chestnut genetic variability. For instance, the crosses between *C. sativa* and the Asiatic species allowed the creation of new hybrid cultivars that are resistant to several pathogens (e.g., to the pathogenic fungus *Gnomoniopsis pascoe* and the Chinese wasp *Dryocosmus kuriphilus*) and tolerant to biotic stress (Sartor et al., 2009 and 2015; Dini et al., 2012; Botta et al., 2012; Alcaide et al., 2020 and 2021).

Low differences in genetic variability have been observed between chestnut cultivars and wild trees (Pereira-Lorenzo et al., 2019; Bouffartigue et al., 2019 and 2020).

Therefore, in this study, we assessed a genetic characterization among chestnut cultivars and wild chestnut trees from the Iberian Peninsula and Italy. The main aims of this research were: a) to investigate the gene pools of chestnut trees from Italy and the Iberian Peninsula based on the reference SSR; b) to estimate the genetic diversity between wild chestnut trees and chestnut varieties.

71

2 Materials and methods

2.1 Plant material

A total of 630 wild chestnut trees and chestnut varieties were analyzed: in particular, 244 were new-
ly collected samples and 386 were derived from previous studies (Table S1). More in detail, 520
samples were varieties and the remaining 110 derived from wild chestnut trees, as single and isolat-
ed trees.

13 representative samples of *Castanea pumila, C. crenata, C. mollissima* and the 'Volos' cultivar
were added to the analysis in order to collect information on interspecific hybrids versus Italian and
Spanish chestnut samples.

The 630 genotypes were standardized with the SSR profiles of the unique accessions available in
the European Chestnut Database (Pereira-Lorenzo et al., 2017).

2.2 DNA extraction and PCR amplification of microsatellites (SSRs)

Young leaves used for DNA extraction were frozen in liquid nitrogen and stored at -80 ° C or ly-
ophilized. The extraction was performed on samples of 0.1-0.5 grams of fresh leaves previously
ground in liquid nitrogen, or on 5 mg of ground lyophilized leaves. The DNA was extracted follow-
ing the CTAB protocol developed by Maguire et al. (1994). dsDNA was quantified using a
Nanodrop™ ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and diluted
to 10 ng/μl.

The microsatellites chosen were selected from the CsCAT and EMCs series, and OAL (Marinoni et
al., 2003; Buck et al., 2003 and Gobbin et al., 2007) and QrZAG (Kampfer et al., 1998) series. The
primers were used in multiplex sets according to Pereira-Lorenzo et al. (2017).

The PCR reaction was performed in 10 μl final volume containing 6.45 μl of sterile H_2O, 1 μl of
GeneAmp® 10X reaction buffer, 0.8 μl of $MgCl_2$, 0.25 of dNTPs and 0.1 units AmpliTaq Gold™

4

96 DNA polymerase. Amplification products were sequenced on a 3130 Genetic Analyzer capillary

97 sequencer (Applied Biosystems, USA). The internal GeneScan TM size standard 500 LIZ (-250)

98 was included in each sample. The allele sizes were detected using Peak Scanner TM software (Ap-

99 plied Biosystems). The samples collected were suitably standardized with the alleles found in the

100 European dataset for the 16 SSR being assessed.

101

102 2.3 Cluster analysis

103 The genetic distance between the 630 samples was calculated by using the R software. The con-

104 struction of the genetic distance dendrogram was elaborated using the Unweighted Pair-Group

105 Method (UPGMA) by R software with the function 'hclust, method = 'average'', packages 'adege-

106 nett''.

107

108 2.4 Genetic diversity and gene pools structure

109 The genetic diversity indices were assessed by the Cervus software, version 3.0.3 (Kalinowski et

110 al., 2007): the number of alleles per locus (k), the expected and the observed heterozygosity (H_e and

111 H_o), polymorphism information content (PIC) and the probability of the allele null (F-null) were es-

112 timated. A PIC threshold of 0.7 was considered for defining loci as highly polymorphic and in-

113 formative. The frequency of the null alleles (F-null) for each locus was calculated using the maxi-

114 mum likelihood (ML) estimator of Kalinowski (2007) implemented in Cervus. The Marker index

115 (MI) was calculated with iMEC: Online Marker Efficiency Calculator (Amirvousefi et al., 2018).

116 The MI allowed the measurement of polymorphism information for the individual markers used.

117 The STRUCTURE 2.3.3 (Pritchard et al., 2000) software was used to evaluate the genetic pool sub-

118 division of 319 genotypes (varieties and wild trees) and to calculate the estimated membership coef-

119 ficient (Q-value) that indicates the membership of each individual in each cluster. This analysis was

120 conducted with a Bayesian approach combined with the Markov Chain Monte Carlo (MCMC) sim-

121 ulation method and was performed using an "admixture model" and correlated allele frequencies.

5

122   Following the protocol of Pereira-Lorenzo et al. (2019), 30 replicate runs of STRUCTURE were

123   performed by setting the number of clusters (K) from 1 to 15. Each run consisted of a burning peri-

124   od of 200,000 steps followed by 200,000 MCMC replicates, with the usage options: locprior=0,

125   popinfo = 0, popflag = 0 (Pereira-Lorenzo et al., 2019; Porras-Hurtado et al., 2013).

126   In order to check if the inclusion of hybrid samples affected the STRUCTURE analysis, the hybrid

127   samples were removed, and a second analysis was performed on a total of 306 samples with the

128   same conditions described above (Figure S3). A Q threshold of 0.8 was used to infer an accession to

129   a specific cluster. The ΔK value (defined as the most probable number of clusters in the population)

130   was calculated through Structure Harvester v.09.93 (Earl, 2012) by testing the change of the log-

131   likelihood between K values (ΔK) as described by Evanno (2005). If a sample has a Q-value < 0.8,

132   it is considered an admixed sample.

133

134   2.5 Genetic differentiation

135   To validate the genetic structure revealed by the Bayesian model-based clustering, a multivariate

136   Principal Coordinate Analysis (PCoA) was elaborated with GenAlEx version 6.502 (Peakall and

137   Smouse, 2006). The PCoA representation was determined on the genetic distance measured by Jac-

138   card coefficient, based on the estimates of ΔK from STRUCTURE for 306 samples (hybrid samples

139   were not considered). A set of analysis to estimate the population differentiation was conducted un-

140   der four scenarios: a) the two main groups (Cluster 1 vs Cluster 2) resulting from the Structure

141   analysis; b) the sub-groups (K=3, K=4), c) chestnut varieties versus wild trees (200 vs 106, respec-

142   tively) and d) chestnut varieties and the wild trees separated in the two main clusters (Cluster 1 and

143   Cluster 2). Pairwise F_{ST} values and private alleles (Np) were estimated for the different partitioning

144   levels considered using GeneAlEx version 6.502; missing data were coded as 0. The F_{ST} value

145   ranges between -1 (absent inbreeding, excess of heterozygous) and 1 (non-random reproduction,

146   excess homozygous).

6

147 The gene flow (Nm) was estimated for the different partitioning levels considered (K=2, 3 and 4)

148 using GenAlEx version 6.502 (Peakall and Smouse, 2012). Hierarchical analysis of molecular vari-

149 ance (AMOVA) was implemented in the GeneAlEx version 6.502 (Peakall and Smouse, 2006) in

150 order to evaluate the genetic variation among and within Clusters. Tests of significance were per-

151 formed using 9999 permutations within the total dataset of 306 samples. The 13 hybrids were re-

152 moved from this analysis.

153 **3 Results**

154 3.1 Genetic variability of the 16 microsatellites

155 The 16 SSRs showed high levels of polymorphism and discriminating power and revealed a total of

156 212 alleles, with an average number of 13.25 alleles per locus (Table 1). The average PIC was

157 0.735, ranging between 0.879 for *CsCAT3* and 0.593 for *EMCs2*. Furthermore, expected

158 heterozygosity varied between 0.889 for *CsCAT3* and 0.619 for *CsCAT1,* with a mean value of

159 0.763 (Table 1). *CsCAT41* was known to amplify two different genomic sites (A and B); for this

160 reason, the *CsCAT41A* locus was removed from the dataset before the analyses (Pereira-Lorenzo et

161 al., 2010). *CsCAT2* and *EMCs38* showed a high frequency value of null alleles (0.209 and 0.118,

162 respectively); consequently, these two loci were removed from the subsequent analysis.

163

164

165

166 **Table 1:** Number of alleles (k), observed (Ho) and expected (He) heterozygosity, polymorphic information content

167 (PIC), null alleles frequencies (F[null]) and marker index (MI) for 319 unique genotypes of *C. sativa* accessions evalu-

168 ated with 16 SSRs.

Locus	k	Ho	He	PIC	F(Null)	MI
CsCAT41B	12	0.690	0.812	0.793	0.071	0.270

CsCAT16	12	0.777	0.812	0.785	0.018	0.262
CsCAT6	19	0.846	0.872	0.858	0.013	0.186
CsCAT1	16	0.596	0.619	0.601	0.022	0.197
CsCAT3	26	0.803	0.889	0.879	0.050	0.165
QrZAG96	12	0.596	0.718	0.694	0.082	0.311
EMCs15	7	0.599	0.661	0.605	0.053	0.385
EMCs38	19	0.693	0.883	0.871	0.118	0.204
EMCs2	6	0.624	0.661	0.593	0.030	0.427
EMCs22	10	0.652	0.682	0.654	0.020	0.301
CsCAT2	16	0.555	0.855	0.841	0.209	0.240
CsCAT17	13	0.771	0.844	0.827	0.042	0.284
CsCAT14	10	0.721	0.753	0.711	0.015	0.313
CsCAT15	11	0.665	0.666	0.605	-0.003	0.266
CsCAT8	11	0.727	0.843	0.821	0.068	0.341
OAL	12	0.586	0.651	0.629	0.048	0.268
Mean	13.25	0.681	0.763	0.735	0.005	0.257

169

170 3.2 Cluster analysis

171 A dendrogram was constructed using the UPGMA method with R software to evaluate the genetic

172 diversity and relatedness between the 630 wild and grafted chestnut trees.

173 Clustering according to the UPGMA method allowed the authentication of accessions (true-to-type

174 cultivars), as well as the indication of possible homonyms, synonyms and incorrect denominations.

175 The 630 samples corresponded to 319 unique chestnut genotypes (Figure S1).

176

177 3.3 Gene pools

178 The genetic structure of the 319 unique genotypes (including varieties and wild chestnut trees) was

179 evaluated using the 14 loci that did not show the presence of null alleles. In the first set of STRUC-

180 TURE analysis (Figure S2), the ΔK statistics gave a maximum value of K=2 ($\Delta K = 78.61$), alt-

181 hough a small peak of ΔK was also observed for K=3 ($\Delta K=18.42$), K=6 ($\Delta K=6.21$) and for K=8

182 ($\Delta K=6.05$).

183 For K=2, genotypes were grouped into two main clusters with a clear distinction between Spanish

184 genotypes, represented by Cluster 1 (with 102 accessions), and Italian genotypes, represented by

185 Cluster 2 (with 163 accessions). The threshold for membership determination was Q > 80%. Addi-

186 tionally, 54 admixed samples were found (Q < 80%) (Figure 1A).

187 For K=3, a further separation of hybrid samples was observed. Cluster 3 included the Euro-Chinese

188 hybrids (Figure 1B) and some significant Italian cultivars ('Marrone Fiorentino'; 'Madonna', 'Lu-

189 cente' and 'Inserta') as well as two varieties from the Andalucía region ('Tomasa' and 'Capilla').

190
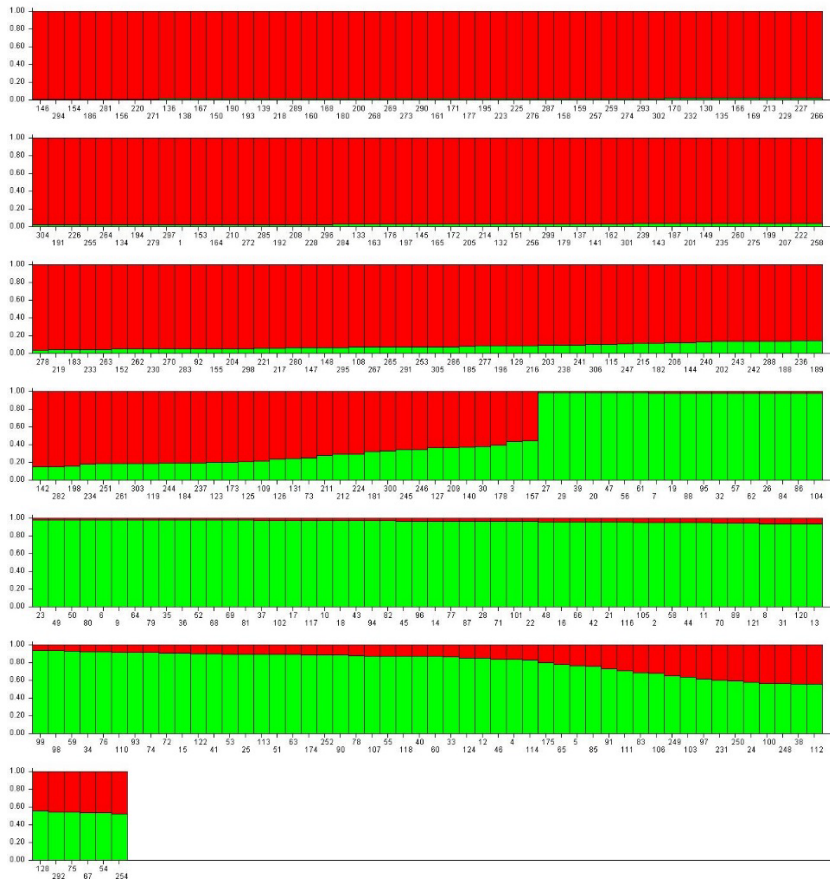
Figure 1A: Representation of 319 samples for K=2 by STRUCTURE Software. Each individual is represented by a

vertical line and each cluster by different colors: Cluster 1 in red, Cluster 2 in green.

194
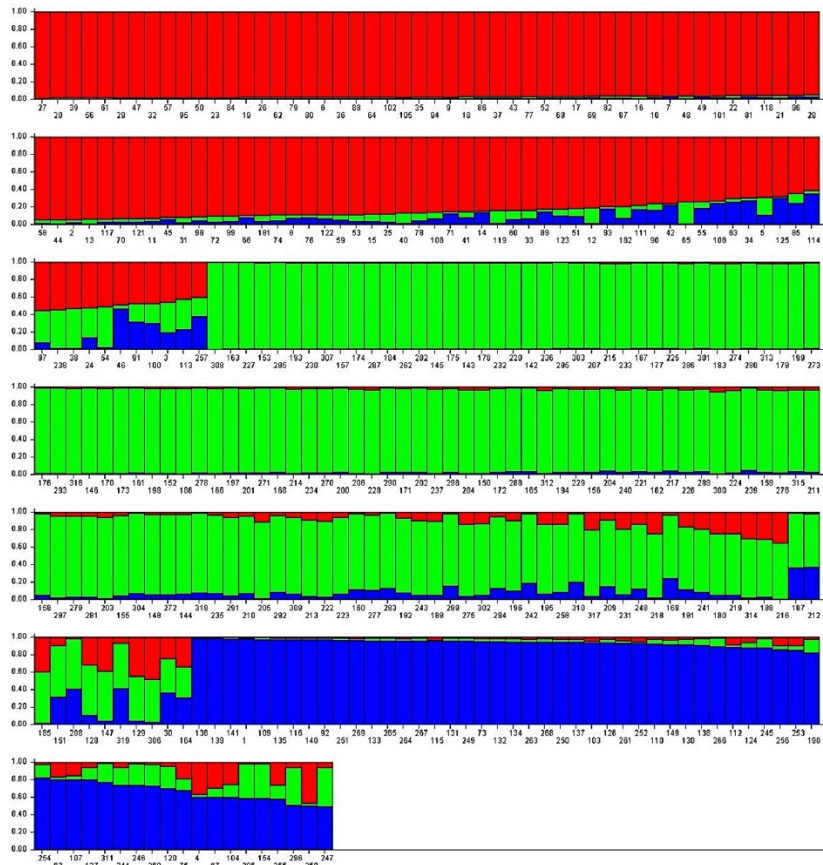
Figure 1B: Representation of 319 samples for K=3 by STRUCTURE Software. Each individual is represented by a vertical line and each cluster by different colors: Cluster 1 in red, Cluster 2 in green and Cluster 3 in blue.

3.4 Genetic structure of chestnut accessions of the Iberian and Italian Peninsulas (varieties and wild trees)

The results confirmed the clear separation between the two main clusters (Cluster 1 represented by 167 Spanish samples; Cluster 2 represented by 104 Italian and 35 admixed samples). The comparison of chestnut varieties and wild chestnut trees within each cluster did not highlight genetic structural differences (Table S3). The separation for K=3 and K=4, with ΔK=27.34 and 25.72 respectively, was also considered (Figure S3, Table S3). Accordingly, the separation between the Italian and Spanish varieties was also maintained in further subdivisions. The Italian cluster was represented as follows: Cluster 1 for subdivision K=3 included 89 samples; Cluster 4 for subdivision K=4 had 86 samples. The Italian cluster included both wild chestnut trees (such as samples named 'Matildico',

11

209 ‘Legno’) and varieties from the Tuscan-Emilian Apennines (such as samples named ‘Pastanese’,

210 ‘Ceppa’, ‘Pistolese’, ‘Piusela’ and ‘Lisanese’).

211 An important group of chestnut varieties was Cluster 2 for subdivision K=3 with different relevant

212 varieties: (e.g. ‘Marrone Fiorentino’) from northern Italy (Emilia-Romagna, Trentino Alto Adige

213 and Piedmont); (e.g., ‘Riggiola’) from southern Italy (Calabria and Campania); (e.g., “Temprana”)

214 from southern Spain (Canary Islands, Extremadura, Andalucía). Varieties from Galicia Asturias and

215 Andalucía, such as “Luguesa”, “Longal”, “Temprana” and “Miguelina”, as well as “Martahiña”,

216 were also represented in Cluster 2 (Table S3).

217 Most of the varieties contained in Cluster 2 for subdivision K= 3 were also included in Cluster 1 for

218 subdivision K=4. In Cluster 2, only a few wild chestnut trees were present (Table S3).

219 Finally, Cluster 3 for subdivision K=3 included the main varieties from northern Spain: Galicia

220 (‘Famosa’, ‘Inxerta’), Castilla-León (‘Negral’), Asturias (‘Parede’, ‘Rapuca’ and ‘Chamberga’),

221 Cantabria, some accessions from Extremadura (‘Verata’) and from Canary Islands (‘Mollar’, ‘Mu-

222 lata’ and ‘Armentina’; Table S3). In addition, most part of the wild chestnut trees from Galicia,

223 Castilla-León and Cantabria were grouped in Cluster 3 (such as ‘Pesaguero’, ‘Alcobilla’,

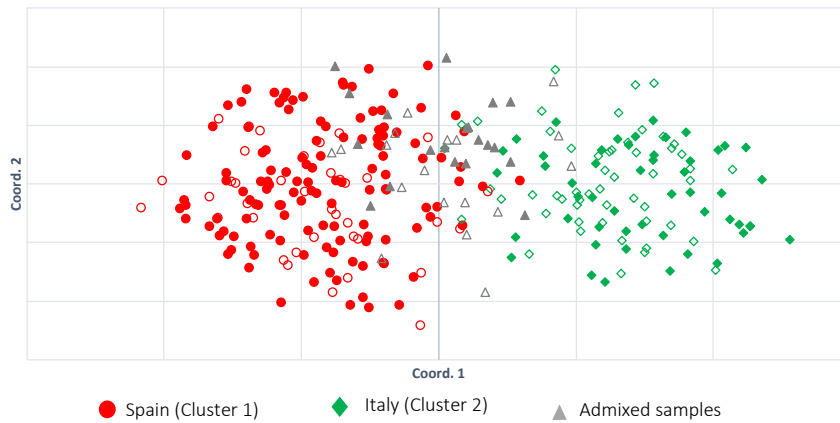224 ‘Peixeroos’).

225

226 3.5 Genetic differentiation

227 The PCoA analysis performed on subdivision K=2 corroborated the results of the STRUCTURE

228 analysis by showing a clear separation between Italian (green) and Spanish (red) accessions (Figure

229 2A). Some admixed samples were observed as reported by the STRUCTURE analysis.

230 Results obtained for subdivision K=3 separated northern (blue) from southern (red) Spanish varie-

231 ties. The northern Italian cluster appeared particularly well separated (green) from the two Spanish
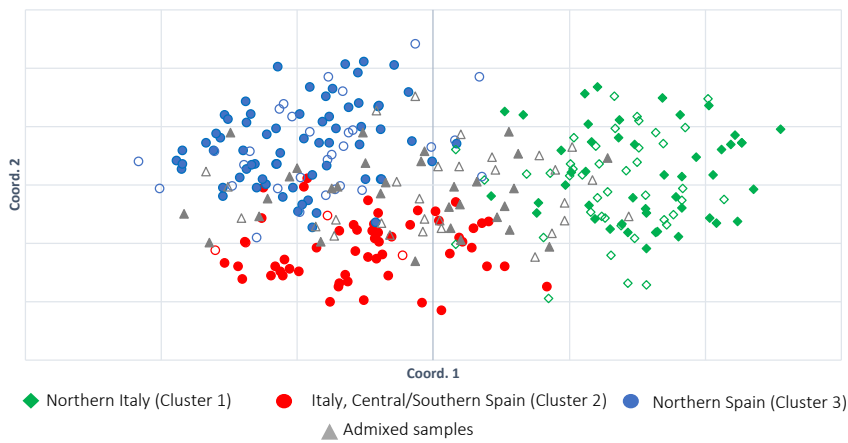
232 clusters (Figure 2B).

233 Furthermore, a larger number of wild chestnut trees was found in the northern Spanish and Italian

234 clusters than in the one from southern Spain, in which most samples were varieties (Figure 2B).

12

235   Subdivision K=4 showed the presence of two further (sub) clusters (mainly chestnut varieties) in

236   southern Spain: the first is a cluster deriving from southern Spain and Italy and the second a cluster

237   deriving from central-southern Spain (represented by Cluster 1 and Cluster 2) (Figure 2C).

238   The genetic differentiation between the two main clusters was F_{ST} = 0.077, P < 0.001 (Table S4.A),

239   suggesting a genetic structure for the chestnut at Italian and Spanish level, also confirmed by the

240   AMOVA results (8%) (Table 2A). A higher gene flow (Nm) (9.235) was detected for subdivision

241   K=2. In addition, the number of private alleles detected for subdivision K=2 and the observed val-

242   ues are quite similar (0.8571 and 0.7142 respectively; Table S5).

243   Similar AMOVA results were found for K=3 and K=4, with a 6% and 7% of variance component

244   among the populations, respectively (Table 2, B and C). In contrast, high gene flows (Nm) (6.118

245   and 3.302 respectively) were detected for subdivisions K=3 and K=4.

246   The largest differentiation between pairs of groups was found between the northern Italian cluster

247   (Cluster 4), with samples mainly from the Tuscan-Emilian Apennines, and the northern-central

248   Spanish cluster (Cluster 3) for K=4 (F_{ST}=0.133, P<0.001), as shown in Table S4.C.

249   A high F_{ST} value was observed also between Cluster 1, containing cultivars from Italy and southern

250   Spain, and Cluster 4, represented by the Italian cluster with F_{ST}=0.113, P<0.001; similarly, between

251   Cluster 1 and Cluster 2, which included central Spain's varieties (F_{ST}=0.112, P<0.001).

252   In addition, AMOVA analysis, conducted between wild trees and chestnut varieties, showed no dif-

253   ferences between them, in agreement with the STRUCTURE division into two main clusters (Table

254   2. D and E). The variance components among populations were 1% and 5% respectively, confirmed

255   also by the F_{ST} index (0.0012 with P<0.001, Table S4. D and E). The Nm values between wild

256   chestnut trees and varieties and between the two main clusters subject of the STRUCTURE analysis

257   were 8.112 and 6.364 respectively. Notably, the number of private alleles (Np) in chestnut varieties

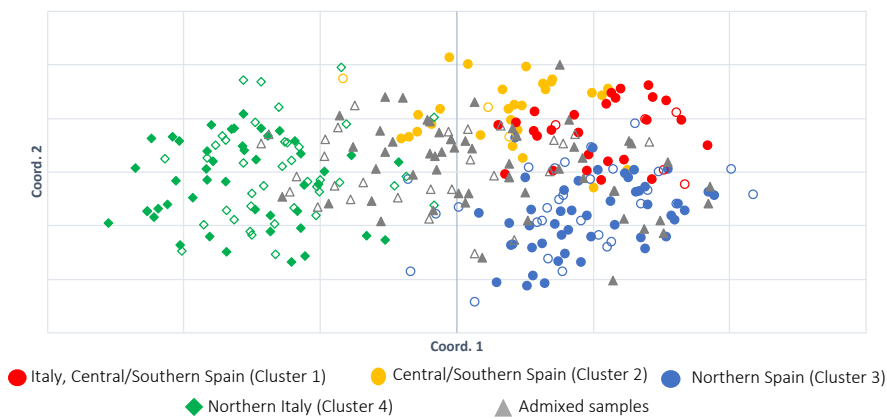258   (0.9285) was higher than the number of private alleles in wild chestnut trees (0.5714, Table S5).

13

A) Principal Coordinates (PCoA) K=2

Coord. 2

Coord. 1

● Spain (Cluster 1)   ◆ Italy (Cluster 2)   ▲ Admixed samples

B) Principal Coordinates (PCoA) K=3

Coord. 2

Coord. 1

◆ Northern Italy (Cluster 1)   ● Italy, Central/Southern Spain (Cluster 2)   ● Northern Spain (Cluster 3)
▲ Admixed samples

C) Principal Coordinates (PCoA) K=4

Coord. 2

Coord. 1

● Italy, Central/Southern Spain (Cluster 1)   ● Central/Southern Spain (Cluster 2)   ● Northern Spain (Cluster 3)
◆ Northern Italy (Cluster 4)   ▲ Admixed samples

Figure 2: Principal Coordinate Analysis (PCoA) based on polymorphism at 14 SSR loci for 306 unique genotypes. A),
K=2; B), K=3; C) K=4. Accession color reflects the consistent assignment using Bayesian analysis to the sub-groups
defined in Fig. 3. Wild chestnut trees are represented by empty symbols inside the sub-groups.

14

Populations	Df		Variance components (%)		p-value
	Among Populations	Within Populations	Among Populations	Within Populations	
A) Structure Cluster K=2, No hybrids included (306 samples)	2	609	8	92	0.077
B) Structure Cluster K=3, No hybrids included (306 samples)	3	608	6	94	0.060
C) Structure Cluster K=4, No hybrids included (306 samples)	4	607	7	93	0.070
D) 306 chestnut trees divided in chestnut varieties and wild	1	610	1	99	0.012
E) 306 chestnut trees divided in chestnut varieties and wild in Cluster 1/Cluster 2	4	607	5	95	0.048

266

4 Discussion

268 Our results confirmed the high degree of variability of *C. sativa* and the selected 16 SSRs markers

269 as powerful tools to evaluate the genetic diversity of EU chestnut germplasm (Pereira-Lorenzo et

270 al., 2011, 2017; Martín et al., 2017a). In particular, the *CsCAT3* (PIC=0.879) and *EMCs15*

271 (PIC=0.604) loci appeared to be the most and least informative loci respectively, as previously re-

272 ported by Pereira-Lorenzo et al. (2010, 2011) and Martín et al. (2012). The average number of al-

273 leles per locus in this study was 13.25 for 16 SSRs, as in Pereira-Lorenzo et al. (2017), with 8.92

274 average using 24 SSRs.

275 Our results highlighted the complex structure and genetic diversity of chestnut trees. Genetic diver-

276 sity is linked to climatic conditions, mainly temperature and precipitation gradient (Pereira-Lorenzo

277 et al., 2010) and to the domestication process carried out by humans through the centuries (Pereira-

278 Lorenzo et al., 2011, 2019).

279 Hybridization could also have played an important role in the diversification process, as previously

280 suggested by Pereira-Lorenzo et al. (2011), and, furthermore, it explains the great diversity found in

281 small geographical areas such as the Tuscan-Emilian Apennines (central-northern Italy) and Galicia

282 (northern Spain), as shown by the STRUCTURE and PCoA analysis. In these regions, wild and

283 domesticated chestnuts were found to be genetically similar and cannot be separated. This is in

15

284 agreement with previous studies, in which no substantial differences between chestnut varieties and

285 wild chestnut trees were found (Fst=0.007 in Pereira-Lorenzo et al., 2019; Fst=0.008 in Bouffar-

286 tigue et al., 2020). This evidence is supported by the AMOVA analysis that showed a low F_{ST} value

287 among wild trees and cultivars (F_{ST}= 0.012 with P<0.001). The Fst and the Nm values observed in

288 samples from the Tuscan-Emilian Apennines and Galicia can be explained considering these areas

289 as the probable sites where the domestication process started, favored by the high number of au-

290 tochthonous chestnut populations. In addition, private alleles were also detected in higher values in

291 chestnut varieties, due probably to selection practices, and they have evolutionary significance (Petit

292 et al., 1998).

293 Our results therefore highlight a low gene flow between chestnut varieties and wild trees. The gene

294 flow between wild trees and chestnut varieties is maintained by changes in forest use over time and

295 the practices related to how the forests were used (Pereira-Lorenzo et al., 2019).

296 At least three major results were obtained by the STRUCTURE analysis.

297 Firstly, the STRUCTURE analysis evidenced a clear separation between the Spanish and the Italian

298 chestnut trees with admixed samples. Similar results, differentiating Spanish and Italian varieties,

299 were previously reported for adaptive markers (Martín et al., 2017b).

300 The genetic differentiation between the two clusters and the admixed samples was low (F_{ST}=0.019

301 and 0.032, P<0.001 respectively). In addition, the AMOVA analysis which compared varieties and

302 wild trees within the two clusters did not show genetic structural differences (F_{ST} =0.012, P<0.001).

303 The large number of admixed samples can indicate a hybridization between the two clusters as

304 suggested also by Pereira-Lorenzo et al. (2012). Furthermore, the aforesaid author evidenced that

305 the hybridization process occurred before the 15th century by considering the oldest giant trees from

306 Andalucía with Italian genetic background (Pereira-Lorenzo et al., 2019).

307 Secondly, the STRUCTURE analysis evidenced a separation among samples belonging to the

308 northern and central/southern Spanish Clusters. In particular, the Northern Cluster contained most

309 of the wild chestnut trees while the Central/Southern Cluster included far more chestnut varieties,

310 probably due to human selection (and vegetative propagation of the selected genotypes) related to

311 the use of new seedlings from local cultivars selected for the superior traits of their nuts (Auge and

312 Brandl, 1997; Forneck, 2005; Pereira-Lorenzo et al., 2010). In this respect, Pereira-Lorenzo et al.

313 (2010), reported that the distribution of seedlings from the main cultivar groups from the northern

314 and central Iberian Peninsula was used to create new orchards in southern Spain, in Andalucía and

315 the Canary Islands (e.g., the "Longal", "Reborda" and "Dieguina" cultivars). Our results confirmed

316 the relationship between "Longal" and the main different varieties from southern Spain, such as

317 "Laga", "Temprana" and "Pelona", and from Extremadura, with "Injerta". This was also in line

318 with previous studies (Pereira-Lorenzo et al., 2006, and Costa et al., 2008), which identified "Lon-

319 gal" as a cultivar used for genetic contribution to create new cultivars in different regions of Spain.

320 This may also explain the huge number of admixed samples between the sub-clusters being as-

321 sessed.

322 Finally, the STRUCTURE analysis evidenced that part of the varieties from southern Spain (Anda-

323 lucía) shared a higher number of alleles with both varieties from southern Italy (Calabria and Cam-

324 pania regions) and with "Marrone Fiorentino", the most important northern Italian variety.

325 As shown in Figure 3, northern Italy produced a high introgression (30% of the genotypes) in cen-

326 tral-southern Spain and an even higher introgression in central-southern Italy (50%). Moreover, in-

327 trogressions from both the north and the south of Italy were also noticed in northern Spain (16%),

328 with 5% from northern and 11% from southern Italy.

329 The above results are in line with Pereira-Lorenzo et al. (2019), who supports an early introduction

330 of chestnut cultivations from Italy into Spain. In particular, the introduction of chestnut cultivations

331 started in the Andalucía and Extremadura regions, with contacts also in Castilla-León and Galicia

332 (see the results related to the "Luguesa" cultivar, which was included in Cluster 2 for K=3 and

333 Cluster 1 for K=4 with the main southern Italian varieties).

334
335
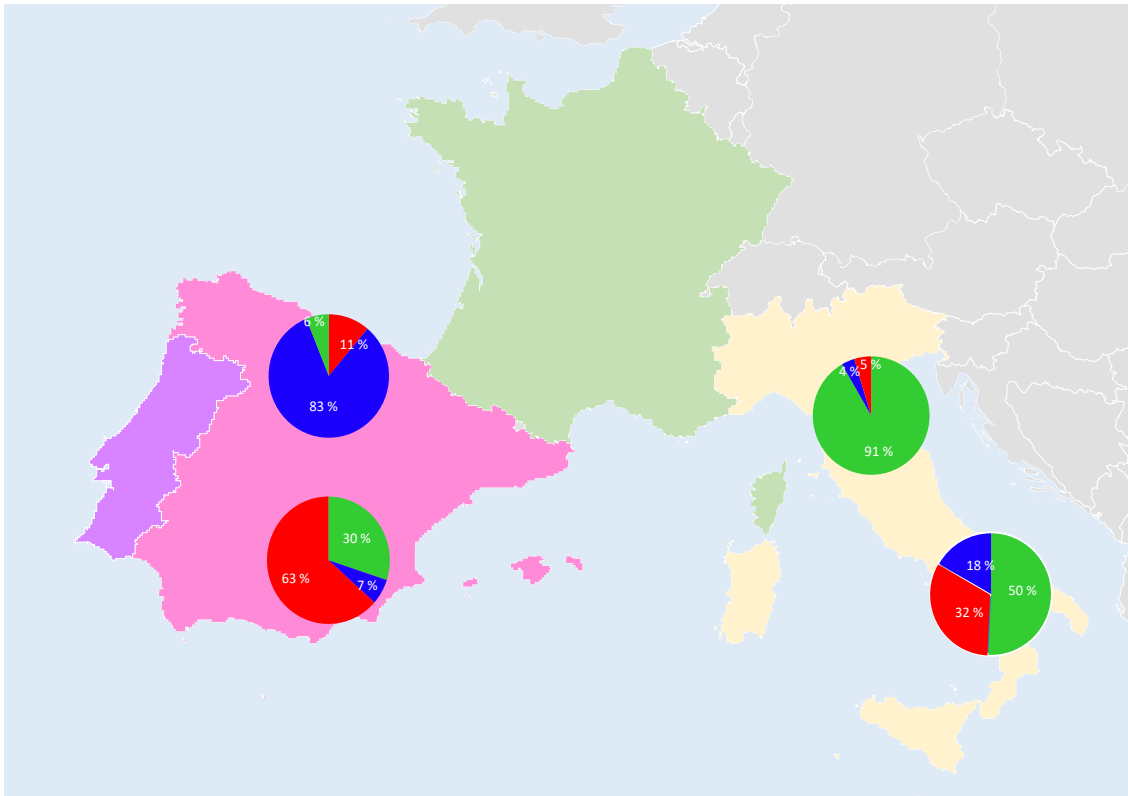
17

Figure 3: Gene pool distribution of chestnut cultivars for K=3 between the north and the south, both of Italy and Spain. Green – Mainly northern Italian and France (Cluster 1); Red – Italy, central and southern Spain (Cluster 2); Blue – northern Spain (Cluster 3).

5 Conclusions

This study contributes to improving knowledge on the genetic relationships between chestnut varieties and wild chestnut trees in the Iberian and Italian Peninsula and to expanding the Chestnut European Genetic Dataset. The genetic variations between and within Italian and Spanish clusters, including both chestnut varieties and wild trees, reflect a combination of historical migrations and selection processes. This is highlighted by the high number of admixed in relation to the chestnut populations of central-southern Spain and central-southern Italy. Adaptation to different environments and hybridization led to a wide genetic variability in limited areas such as the Tuscan-Emilian Apennines and the Galicia region. In these areas, no genetic structure differences between wild chestnut trees and chestnut varieties was found.

18

Bibliography

Alcaide, F.; Solla, A.; Cuenca, B.; Martín, M. Á. (2021). Molecular evidence of introgression of Asian germplasm into a natural *Castanea sativa* forest in Spain. Forestry: An International Journal of Forest Research.

Alcaide, F.; Solla, A.; Cherubini, M.; Mattioni, C.; Cuenca, B.; Camisón, Á.; Martín, M. Á. (2020). Adaptive evolution of chestnut forests to the impact of ink disease in Spain. Journal of Systematics and Evolution, 58(4), 504-516.

Alcaide, F.; Solla, A.; Mattioni, C.; Castellana, S.; Martín, MA (2019) Adaptive diversity and drought tolerance in *Castanea sativa* assessed through EST-SSR genic markers. Forestry 92:287–296. https:// doi.org/10.1093/forestry/cpz007

Amiryousefi, A.; Hyvönen, J.; Poczai, P. (2018). iMEC: online marker efficiency calcula-tor. Applications in plant sciences, 6(6), e01159.

19

376

Auge, H.; Brandl, R. (1997). Seedling recruitment in the invasive clonal shrub, *Mahonia aquifolium*

Pursh (Nutt.). Oecologia (Berl.), 110(2): 205–211. doi: 10.1007/s004420050151.

379

Bernetti, G. (1995). Selvicoltura speciale. Unione tipografico-editrice Torinese.

381

Bouffartigue, C.; Debille, S.; Fabreguette, O.; Ramos Cabrer, A.; Pereira-Lorenzo, S.; Flutre, T.;

Harvengt, L. (2019). Genetic diversity and structure of sweet chestnut (*Castanea sativa* Mill.) in

France: At the intersection between Spain and Italy. bioRxiv 792259; doi:

https://doi.org/10.1101/792259

386

Bouffartigue, C.; Debille, S.; Fabreguettes, O.; Cabrer, A. R.; Pereira-Lorenzo, S.; Flutre, T.;

Harvengt, L. (2020). High admixture between forest and cultivated chestnut (*Castanea sativa* Mill.)

in France. bioRxiv, 792259.

390

Buck, E.J.; Hadonou, M.; James, C.J.; Blakesley, D.; Russell, K. (2003) Isolation and characteriza-

tion of polymorphic microsatellites in European chestnut (*Castanea sativa Mill.*) Mol Ecol Notes

3:239–241. https://doi.org/10.1046/j.1471-8286.2003.00410.x

394

Castellana, S.; Martin, M. Á.; Solla, A; Alcaide, F.; Villani, F.; Cherubini, M.; Mattioni, C. (2021).

Signatures of local adaptation to climate in natural populations of sweet chestnut (Castanea sativa

mill.) from southern Europe. Annals of Forest Science, 78(2), 1-21.

398

Conedera, M. (1996). Die Kastanie: der Brotbaum. Vergangenheit, Gegenwart und Zukunft der

‘‘Waldfrucht par excellence’’‘. Bu¨ndnerwald, 49: 28–46.

401

20

402 Conedera, M.; Krebs, P.; Tinner, W.; Pradella, M.; Torriani, D. (2004). The cultivation of *Castanea*

403 *sativa* (Mill.) in Europe, from its origin to its diffusion on a continental scale. Vegetation History

404 and Archaeobotany, 13(3), 161–179. https://doi.org/10.1007/s00334-004-0038-7

405

406 Costa, R.; Ribeiro, C.; Valdiviesso, T.; Afonso, S.; Borges, O.; Soeiro, J.; Costa, H.; Fonseca, L.;

407 Augusta, C.; Cruz, M.H.; Salazar, M.; Matos Soares, F.; Sequeira, J.; Assunção, A.; Correia, P.;

408 Lima, M.J. (2008). Variedades de Castanha das Regiões Centro e Norte de Portugal. INRB.I.P.

409

410 Earl, D.A.; Vonholdt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visual-

411 izing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–

412 361. https://doi.org/10.1007/s12686-011-9548-7

413

414 Evanno, G.; Regnaut, S.; Goudet, J. (2005). Detecting the number of clusters of individuals using

415 the software STRUCTURE: A simulation study. Molecular Ecology, 14(8), 2611–2620.

416 https://doi.org/10.1111/j.1365-294X.2005.02553.x

417

418 Fernández Cruz, J.; Fernández López, J. (2016) Genetic structure of wild sweet chestnut (*Castanea*

419 *sativa* Mill.) populations in northwest of Spain and their differences with other European stands.

420 Conserv Genet 17:949–967. https://doi.org/10.1007/s10592-016-0835-4

421

422 Fernández-López, J.; Fernández-Cruz, J.; Míguez-Soto, B. (2021). The demographic history of *Cas-*

423 *tanea sativa* Mill. in southwest Europe: a natural population structure modified by translocations.

424 Molecular Ecology.

425

426 Fideghelli, C. (2017). Atlante dei fruttiferi autoctoni italiani. Crea - Centro di frutticoltura, Roma.

427

21

428   Forneck, A. (2005). Plant breeding: clonality — a concept for stability and variability during vege-

429   tative propagation. Prog. Bot. 66: 164–183. doi: 10.1007/3-540-27043-4_8

430

431   Gobbin, D.; Hohl, L.; Conza, L.; Jermini, M.; Gessler, C.; Conedera, M. (2007) Microsatellite-

432   based characterization of the *Castanea sativa* cultivar heritage of southern Switzerland. Genome

433   50:1089–1103. doi: 10.1139/G07-086

434

435   Hummer, K.E.; Dempewolf, H.; Bramel, P.; Markham, R.; Stover, E. (2015) Status of global strate-

436   gies for horticultural fruit crops. Acta horticulturae

437

438   Huntley, B. and Birks, H. J. B. (1983). Atlas of past and present pollen maps for Europe, 0-13,000

439   years ago. Cambridge University Press.

440

441   iMEC: Online Marker Efficiency Calculator, Ali Amiryousefi, Jaakko Hyvönen, Péter Poczai Appl

442   Plant Sci. (2018); 6(6): e01159.

443

444   Jarman, R.; Mattioni, C.; Russell, K.; Chambers, F. M.; Bartlett, D.; Martin, M. A.; Webb, J.

445   (2019). DNA analysis of *Castanea sativa* (sweet chestnut) in Britain and Ireland: Elucidating Euro-

446   pean origins and genepool diversity. PloS one, 14(9), e0222936.

447

448   Kalinowski, S. T.; Taper, M. L.; Marshall, T. C. (2007). Revising how the computer program

449   CERVUS accommodates genotyping error increases success in paternity assignment. Molecular

450   Ecology, 16(5), 1099–1106. https://doi.org/10.1111/j.1365294X.2007.03089.x

451

22

452 Kampfer, S.; Lexer, C.; Glossl, J.; Steinkellner, H. (1998). Characterization of (GA)(n) microsatel-

453 lite loci from *Quercus robur*. Hereditas, 129(2), 183–186. https://doi.org/10.1111/j.1601-

454 5223.1998.00183.x

455

456 Krebs, P.; Conedera, M.; Pradella, M.; Torriani, D.; Felber, M.; Tinner W. (2004) Quaternary refu-

457 gia of the sweet chestnut (*Castanea sativa* Miller): an extended palynological approach. Vegetation

458 History and Archaeobotany, 13, 145–160.

459

460 Krebs, P.; Pezzatti, G. B.; Beffa, G.; Tinner, W.; Conedera, M. (2019). Revising the sweet chestnut

461 (*Castanea sativa* Mill.) refugia history of the last glacial period with extended pollen and macrofos-

462 sil evidence. Quaternary Science Reviews, 206, 111-128.

463

464 Lusini, I.; Velichkov, I.; Pollegioni, P.; Chiocchini, F.; Hinkov, G.; Zlatanov, T. et al. (2014). Esti-

465 mating the genetic diversity and spatial structure of Bulgarian *Castanea sativa* populations by

466 SSRs: implications for conservation. Conservation genetics, 15(2), 283-293.

467

468 Maguire, T. L.; Collins, G. G.; Sedgley, M. (1994). A modified CTAB DNA extraction procedure

469 for plants belonging to the family proteaceae. Plant Molecular Biology Reporter, 12(2), 106–109.

470 https://doi.org/10.1007/BF02668371

471

472 Marinoni, D. T.; Akkak, A.; Beltramo, C.; Guaraldo, P.; Boccacci, P.; Bounous, G.; Botta, R.

473 (2013). Genetic and morphological characterization of chestnut (*Castanea sativa* Mill.) germplasm

474 in Piedmont (north-western Italy). Tree genetics & genomes, 9(4), 1017-1030. DOI

475 10.1007/s11295-013-0613-0

476

23

477   Marinoni, D.; Akkak, A.; Bounous, G.; Edwards, K.J.; Botta, R. (2003). Development and charac-

478   terization of microsatellite markers in *Castanea sativa* (Mill.). Mol. Breed. 11 (2), 127–136

479   https://doi.org/10.1023/ A:1022456013692.

480

481   Martín, M.A.; Alvarez, J.B.; Mattioni, C.; Cherubini, M.; Villani, F.; Martín, L.M. (2009) Identifi-

482   cation and characterization of traditional chestnut varieties of southern Spain using morphological

483   and simple sequence repeat (SSRs) markers. Ann Appl Biol 154:389–398.

484   https://doi.org/10.1111/j.1744-7348.2008.00309.x

485

486   Martín, M.A.; Mattioni, C.; Molina, J.R.; Alvarez, J.A.; Cherubini, M.; Herrera, M.A. et al. (2012)

487   Landscape genetics structure of chestnut (*Castanea sativa* Mill.) in Spain. Tree Genet Genomes

488   8:127–136. https://doi.org/10.1007/s11295-011-0427-x

489

490   Martín, M. A.; Monedero, E.; Martín, L. M. (2017a). Genetic monitoring of traditional chestnut or-

491   chards reveals a complex genetic structure. Annals of Forest Science, 74(1).

492   https://doi.org/10.1007/s13595-016-0610-1

493

494   Martín, M. A.; Mattioni, C.; Cherubini, M.; Villani, F.; Martín, L. M. (2017b). A comparative study

495   of European chestnut varieties in relation to adaptive markers. Agroforestry systems, 91(1), 97-109.

496

497   Mattioni, C.; Martin, A.M.; Pollegioni, P.; Cherubini, M.; Villani, F. (2013). Microsatellite markers

498   reveal a strong geographical structure in European populations of *Castanea sativa* (Fagaceae): Evi-

499   dence for multiple glacial refugia. American Journal of Botany, 100(5), 951–961.

500   https://doi.org/10.3732/ajb.1200194

501

502 Mattioni, C.; Martin, M. A.; Chiocchini, F.; Cherubini, M.; Gaudet, M.; Pollegioni, P. et al. (2017).

503 Landscape genetics structure of European sweet chestnut (*Castanea sativa* Mill): indications for

504 conservation priorities. Tree Genetics and Genomes, 13(2). https://doi.org/10.1007/s11295-017-

505 1123-2

506

507 Mellano, M.G.; Torello-Marinoni, D.; Boccacci, P. (2018) Ex situ conservation and characterization

508 of the geneticdiversity of Castanea spp. Acta Horticulturae 1–6. doi:

509 10.17660/ActaHortic.2018.1220.1

510

511 Míguez-Soto, B.; Fernández-Cruz, J.; Fernández-López, J. (2019). Mediterranean and Northern Ibe-

512 rian gene pools of wild *Castanea sativa* Mill. are two differentiated ecotypes originated under natu-

513 ral divergent selection. PloS one, 14(2), e0211315.

514

515 Neri, L.; Dimitri, G.; & Sacchetti, G. (2010). Chemical composition and antioxidant activity of

516 cured chestnuts from three sweet chestnut (*Castanea sativa* Mill.) ecotypes from Italy. Journal of

517 Food Composition and Analysis, 23(1), 23–29. https://doi.org/10.1016/j.jfca.2009.03.002

518

519 Nishio, S.; Takada, N.; Terakami, S.; Takeuchi, Y.; Kimura, M. K.; Isoda, K.; Saito, T.; Iketani, H.

520 (2021). Genetic structure analysis of cultivated and wild chestnut populations reveals gene flow

521 from cultivars to natural stands. Scientific Reports, 11(1), 1-13.

522

523 Peakall, R.; Smouse, P.E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software

524 for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-

525 8286.2005.01155.x

526

25

527   Peakall, R., Smouse, P.E., (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic soft-

528   ware for teaching and research-an update. Bioinformatics 28, 2537e2539.

529   https://doi.org/10.1093/bioinformatics/bts460.

530

531   Petit, R.J., El Mousadik, A., Pons, O., (1998). Identifying populations for conservation on the basis

532   of genetic markers. Conserv. Biol. 12, 844e855. https://doi.org/10.1111/j.1523-1739.1998.96489.x.

533

534   Pereira-Lorenzo, S.; Costa, R.; Ramos-Cabrer, A.M.; Ribeiro, C.; Serra da Silva, C.; Manzano, G.;

535   Barreneche, T. (2010) Variation in grafted European chestnut and hybrids by microsatellites reveals

536   two main origins in the Iberian Peninsula. Tree Genet Genomes 6:701–715.

537   https://doi.org/10.1007/s11295-010-0285-y

538

539   Pereira-Lorenzo, S.; Lourenço Costa, R.M.; Ramos-Cabrer, A.M.; Ciordia-Ara, M.; Marques Ribei-

540   ro, C.A.; Borges, O.; Barreneche, T. (2011) Chestnut cultivar diversification process in the Iberian

541   Peninsula, Canary Islands, and Azores. Genome 54:301–315. https://doi.org/10.1139/g10-122

542

543   Pereira-Lorenzo, S.; Ramos-Cabrer, A. M.; Barreneche, T.; Mattioni, C.; Villani, F.; Díaz-

544   Hernández, M.B. et al. (2017). Database of European chestnut cultivars and definition of a core col-

545   lection using simple sequence repeats. Tree Genetics and Genomes, 13(5).

546   https://doi.org/10.1007/s11295-017-1197-x

547

548   Pereira-Lorenzo, S.; Ramos-Cabrer, A. M.; Barreneche, T.; Mattioni, C.; Villani, F.; Díaz-

549   Hernández, B.; Martín, A. (2019). Instant domestication process of European chestnut cultivars.

550   Annals of Applied Biology, 174(1), 74–85. https://doi.org/10.1111/aab.12474

551

26

552 Pereira-Lorenzo, S.; Bischofberger, Y.; Conedera, M.; Piattini, P.; Crovadore, J.; Chablais, R. et al.

553 (2020) Reservoir of the European chestnut diversity in Switzerland. Biodivers Conserv (2020).

554 https://doi.org/10.1007/s10531-020-01970-2

555

556 Pitte, J.R. (1986). Terres de castanide. Hommes et paysages du chataignier de l''Antiquite´a nos

557 jours. Librairie Artheme Fayard.

558

559 Porras-Hurtado, L.; Ruiz, Y.; Santos, C.; Phillips, C.; Carracedo, A.; Lareu, M. (2013) An overview

560 of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:98.

561 https://doi.org/10. 3389/fgene.2013.00098

562

563 Pritchard, J.K.; Stephens, M.; Donnelly, P. (2000) Inference of population structure using multi-

564 locus genotype data. Genetics 155:945–959

565

566 Roces-Díaz, J. V.; Jiménez-Alfaro, B.; Chytrý, M.; Díaz-Varela, E. R.; Álvarez-Álvarez, P. (2018).

567 Glacial refugia and mid-Holocene expansion delineate the current distribution of *Castanea sativa* in

568 Europe. Palaeogeography, Palaeoclimatology, Palaeoecology, 491, 152-160.

569

570 Silvanini, A.; Marinoni, D. T.; Beccaro, G. L.; Ganino, T. (2011). La caratterizzazione varietale del

571 germoplasma di *Castanea sativa* Mill. Italus Hortus, 18, 47-61.

572

573 Villani, F.; Sansotta, A.; Cherubini, M.; Cesaroni, D.; Sbordoni, V. (1999). Genetic structure of

574 natural populations of *Castanea sativa* in Turkey: evidence of a hybrid zone. JOURNAL OF EVO-

575 LUTIONARY BIOLOGY, 12(2), 233-244.

576

577

Supporting information

Table S1: unique genotypes analyzed in this study (319 samples).

Table S2: list of the different varieties collected with their countries of origin, the STRUCTURE subdivision for K=2 and K=3 for 319 samples. A Q threshold of 0.8 was used to infer an accession to a specific cluster.

Table S3: list of the different varieties collected with their countries of origin, the STRUCTURE subdivision for K=2, K=3 and K=4 for 306 samples (no hybrids included) and varieties/wild subdivision. A Q threshold of 0.8 was used to infer an accession to a specific cluster.

Table S4: pairwise estimate of Fst value based on the 14 SSR loci of 306 chestnut accessions defined by STRUCTURE analysis: A), K=2; B), K=3; C) K=4; D) the total population (n= 306) divided into wild trees and chestnut varieties; E) wild trees between chestnut varieties in K=2. All estimates were highly significant - $P < 0.001$.

Figure S1: UPGMA tree indicating the relationships among 630 chestnut samples.

Figure S2: Estimates of Δ k calculated based on Evanno et al., (2005), based on k-subdivision for 319 samples.

Figure S3: Estimates of Δ k calculated as described by Evanno et al., (2005), based on k-subdivision for 306 samples.