Patterns of genetic variation and the potential origin of sweet chestnut (*Castanea sativa* Mill.) stands far from its natural northern distribution edge

Oliver Caré[1,3,*], Oleksandra Kuchma[1], Bernhard Hosius[1], Wolfgang Voth[2], Eric A. Thurm[2], Ludger Leinemann[1,3]

[1] ISOGEN GmbH & Co KG; Büsgenweg 2; 37077 Göttingen
[2] Department of Forest Planning/ Forest Research/ Information Systems; Landesforstanstalt Mecklenburg-Vorpommern AöR; Zeppelinstraße 3; 19061 Schwerin
[3] Buesgen Institute Forest Genetics and Forest Tree Breeding; Faculty of Forest Sciences and Forest Ecology; University of Göttingen; Büsgenweg 2; 37077 Göttingen.

*Corresponding author: Oliver Caré, E-mail: oliver.care@isogen.de

Abstract

Castanea sativa Mill. is an important European tree species that has been used for timber and nut production since Greek and Roman times. In Germany, the species accounts for only a small proportion of forest stands but it is attracting increasing interest due to its possible potential for climate tolerance and resilience. However, the status of German sweet chestnut stands in terms of genetic diversity and adaptive potential is poorly investigated. We took a representative sample of 520 individuals from 13 stands in northern Germany (Mecklenburg Western Pomerania; MWP) and genotyped these at 11 genomic SSRs and EST-SSRs. In addition, we compared our data with the available literature, focusing on the possible origin of the MWP populations. The analysis of genetic structure identified three distinct genetic groups within MWP, suggesting different origins or seed mixtures of various provenances. The comparison with reference data at the EST-SSR FIR059 suggests that the MWP stands originate from the western gene pool. The genetic variation observed in most MWP stands is at approximately the same level as stands in the natural range. Therefore, these stands with high diversity seem to be particularly suitable for further production of seed and plant material in MWP. This study contributes to the understanding of the genetic diversity and origin of sweet chestnut stands outside their natural distribution in MWP. It provides valuable knowledge for the evaluation and conservation of genetic resources of this tree species, which is expected to receive more attention under warmer and drier conditions.

Keywords: *chestnut, genetic diversity, population structure, Germany, provenance, microsatellite (SSR)*

Introduction

The natural distribution area of *Castanea sativa* Mill. (Fagaceae) is found in the area of Mediterranean to sub-Mediterranean-temperate climatic conditions on the western European continent. Nevertheless, it has also successfully spread north of the Alps. *C. sativa* is a species of warm-temperate regions, where it generally occurs in mild, late and winter frost-protected locations, growing best on well drained neutral to acidic soils and is considered to be slightly drought sensitive. *C. sativa* is a medium-sized deciduous tree. It reaches heights of up to 30/35 m, trunk diameters at breast height of up to 2 m and can live for 500 to 600 years. In addition to valuable timber and nuts, the sweet chestnut provides many ecosystem services and serves as an important resource for bees (Conedera et al. 2016, Hein et al. 2013, Kleber et al. 2020).

Sweet chestnut survived the last ice age in refugial areas across the Iberian Peninsula, southern France, Italy, the Balkans, Greece, Turkey, and the Caucasus (Huntley and Birks, 1983; Krebs et al., 2019). Genetic analyses of natural *C. sativa* populations have described three distinct gene pools: "eastern gene pool" (eastern Turkey to the Caucasus), "central gene pool" (western Turkey, Greece, Bulgaria), and "western gene pool" (Italy, southern Switzerland, southern France, Spain, Portugal) (Mattioni et al., 2013, 2017; Castellana et al., 2021;

DOI:10.2478/sg-2023-0020
edited by the Thünen Institute of Forest Genetics

Alessandri et al., 2022). Recent studies (e.g., Fernández-Cruz & Fernández-López, 2012; Lusini et al., 2014; Martin et al., 2010; Mattioni et al., 2017; Pereira-Lorenzo et al., 2017; Roces-Díaz et al., 2018) describing the redistribution of genotypes through natural colonization and anthropogenic translocation support the native status of *C. sativa* populations in parts of continental Europe and their origin from southern European regions. Hotspots of genetic diversity within the range vary between studies. Mostly Turkish, Italian and Bulgarian populations showed high diversity, while Spanish and Greek populations showed contrasting results (Castellana et al. 2021, Chiocchini et al. 2016, Mattioni et al. 2017). *C. sativa* has already been used for timber and fruit production before ancient Greek and Roman times. Since that time, sweet chestnut cultivation has become more widespread across Europe with involvement of monasteries, royal and noble estates, farmers, and foresters throughout Europe (e.g., Conedera et al., 2004, 2016). Genetic studies show that this anthropogenic transfer and spread has led to cultivated-to-wild introgression in the sweet chestnut (Poljak et al. 2017). However, genetic variation and differentiation among sweet chestnut cultivars is poorly studied in Germany (Faust and Fussi, 2009). Consequently, the influence of the above factors on the genetic variation of sweet chestnut in Germany remains largely unknown.

Although the sweet chestnut is mainly found in the warmer temperate regions, its distribution extends to the cooler areas of Germany (Thurm et al., 2018). Sweet chestnut was repeatedly introduced into German forests beginning in Roman times. The distribution area of sweet chestnut and the location of the investigated stands in Mecklenburg Western Pomerania (MWP) is shown in Figure 1.

Thurm et al. (2018) estimated the suitability for future cultivation in different tree species using a modelling approach combining the European-wide occurrence of species with climate data. Following this approach, Thurm and Heitz (2018) showed that suitable climatic constellations for the cultivation of sweet chestnut are already available in many areas of Germany today. The distribution and growth of existing stands in Germany confirmed the high heat tolerance and the adaptive potential of the sweet chestnut in MWP.

Within the investigated stands in this study, the older and vital stands in MWP can also be regarded as evidence for the tree species' suitability for cultivation. This suggests that it may be a promising forest tree species in light of climate change in MWP, where temperature increases are projected, and precipitation is assumed to remain the same.

The present study aims to describe patterns of genetic variation of selected stands of sweet chestnut in MWP. In comparison with data from the literature, especially Castellana et al. (2021), it will be analysed whether these patterns can be assigned to provenances from Southern Europe. The results should provide a basis for gene conservation and silvicultural management with special emphasis on seed production in sweet chestnut in MWP.

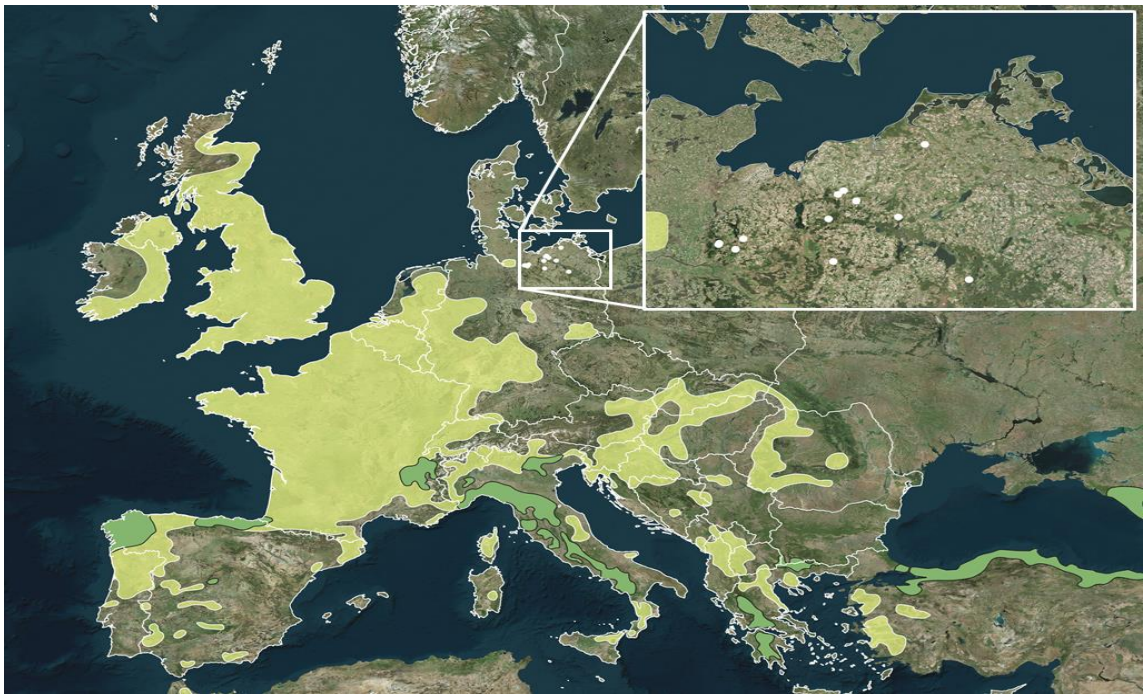


Figure 1
The distribution of sweet chestnut in Europe (green: native; yellow: historically introduced) (Caudullo et al. 2017) and the locations of investigated stands in MWP (white dots)

Material and Methods

Plant material

In summer 2022, 13 stands of sweet chestnut in MWP were sampled (Table 1). For each stand, 40 individual tree samples were obtained, distributed representatively over the area. Forest stands SK-1/SK-2 are contiguous and have the same age and stand structure. Thus, identical origin and planting was assumed, and samples were pooled for analysis. The geographical location, silvicultural and yield data of the stands can be seen in Table 1.

The number of valuable trees per hectare (zf/ha) corresponds to the specifications for Z-tree selection according to quality and vitality in silvicultural concepts without consideration of spatial distribution (cf. Thurm et al. 2022).

Laboratory analyses

Genomic DNA was extracted from leaves using the DNeasy 96 Plant Kit (QIAGEN GmbH, Hilden, Germany). Five genomic (g)- and six EST-SSRs (nuclear and expressed sequence tag simple sequence repeats) CsCAT1, CsCAT3, CsCAT6, CsCAT16, GOT021, VIT057, FIR080, GOT045, FIR059, FIR094 and EMCs38 (Buck et al., 2003; Marinoni et al., 2003; Durand et al., 2010) were selected for genotyping of all individuals. Microsatellite fragments were separated electrophoretically on the ABI Prism Genetic Analyzer 3130xl along with the Applied Biosystems GeneScan 500 ROX internal size standard. Microsatellite size was determined using the GeneMapper v4.0 software package (Applied Biosystems Inc., Foster City, USA).

Data analyses

First we analysed the MWP dataset regarding diversity and populations structuring. In these analysis five genomic (g)- and six EST-SSRs were used, with the markers EMCs38 and GOT021 excluded for most analysis due to polymorphic patterns. The relationship between identified genetics groups and the number of quality trees was then analysed using a regression model. Secondly, we aimed to standardize datasets from the literature (Pereira-Lorenzo et al., 2017; Bouffartigue et al., 2020; Mattioni et al., 2020; Castellana et al., 2021) at five SSR genetic markers investigated in all studies mentioned (CsCAT1, CsCAT16, CsCAT3, CsCAT6, EMCs38) for comparative analyses. As the results did not contribut meaningful results, the results and discussion for this part can be found in the supplementary material. Third, in our attempt to identify refugial areas of origin, we sought to use the original dataset of Castellana et al. (2021) (https://zenodo.org/record/4384029). It shares 10 SSR gene markes with our dataset from MWP. We used data of 8 SSR markers CsCAT1, CsCAT16, CsCAT3, CsCAT6, FIR080, FIR059, FIR094, and GOT045 and excluded EMCs38 and GOT021 due to polymorphic patterns in the MWP dataset. On this basis, genetic parameters were calculated for the corresponding stands and regions in Southern Europe in comparison to the stands from Mecklenburg Western Pomerania based on the initial allele scoring. In addition, we used the genetic marker FIR059, which exhibits differential and possibly adaptive variation among different refugial regions in southern Europe where Alcaide et al. (2019) and Castellana et al. (2021) found a adaptive correlation to drouht stress with allele distribution. For comparison, we used STRUCTURE (Pritchard et al. 2000) to compare genetic clusters between stands in Mecklenburg Western Pomerania and potential refugial regions using this loci.

GenAlEx 6.51b2 (Peakall and Smouse, 2012) was used to calculate the allelic diversity (N_e), heterozygosity (H_o) and fixation index parameters. Rarefied allelic richness (A_r) was caluclated using "hierfstat" 0.5-11 (Goudet and Jombart, 2015). Polymorphic information content (PIC) was calculated in "PopGenUtils" 0.1.8 (Tourvas 2023). Allelic diversity parameters, rarefied allelic richness (A_r) on the eight shared SSRs markes to the Castellana et al. (2021) data were also calculated the package "hierfstat" 0.5-11 (Goudet and Jombart, 2015). Rarefaction sample size in the MWP dataset was N=39 and in the combined analysis with the southern European dataset N=16. For tests on significance between MWP stands and reference data from provenances from Southern Europe the Student's t-test was used. Analysis of molecular variance (AMOVA) and permutation testing to asses significance was performed in "poppr" 2.9.4 (Kamvar et al. 2015 a,b) on the separate and combined set of gSSR and EST-SSR. Cluster analysis of Nei's genetic distance on the population level with bootstrap support based on 5000 replicates was also calculated in "poppr" 2.9.4 for the seperat and combined SSR sets. Regarding the above described analysis the markers EMCs38 and GOT021 were excluded due to polymorphic patterns in the MWP dataset. The number of valuable trees per hectare (zf/ha) was tested against the explanatory variables of the identified three genetic groups and the mean diameter (dg) in a linear additive model (zf/ha~group+dg) with "mgcv" 1.8-42 (Wood 2017). Model testing was performed in "DHARMa" 0.4.6 (Hartig 2022). These analyses were performed in the free programming language for statistical computing and graphics "R" version 4.1.0 (R Core Team, 2021).

Population structure analysis was performed using the STRUCTURE 2.3.4 program (Pritchard et al., 2000) using all 11 SSRs. Due to polyploid patterns the markers EMCs38 and GOT021 were each split into two regions by a threshold allele size. Thus, each of the two marker was then representet by two artifical markers comprised of either smaller or bigger alleles. Further, separate analysis for both only genomic and only EST-SSR were performed. The program calculates membership probabilities for genetic clusters based on variation differences in genetic markers. These are based on interspecific or intraspecific changes in frequencies at the genetic markers examined. For the analysis in STRUCTURE, the assignments for the cluster numbers (K = genetic clusters) were assumed to range from K = 1 to K =15, and each was repeated 10 replicates, based on a Markov chain Monte Carlo procedure, with 5,000 burn-in and 10,000 steps. The results of the 10 replicates per assumed number of clusters were summarized using CLUMPAK (Kopelman et al., 2015). To assess the population structure at hand, clustering with the most statistically probable value for K was chosen using the method of Puechmaille (2016). The calculation for this was performed in StructureSelector (Li and Liu, 2018). The

Table 1
Information on stands sampled in MWP. Abbreviations, forest district, coordinates, age, number of samples (N_{sample}), mean diameter at breast height (dg), mean height (hg), trees per hectare (N/ha), basal area per hectare (G), volume per hectare (V), Number of valuable trees in stand per hectare (zf/ha).

Abbrev.	Forest district	Coordinates (WGS84)	Age	Nsample	dg (cm)	hg (m)	N/ha	G (m2/ha)	V (m3/ha)	zf/ha
BiKu	Billenhagen	N54.13421 E12.51771	98	40	57.4	28.4	147	37.9	547.4	40
GaJu	Gädebehn	N53.66433 E11.70672	23	40	19	19.5	883	25.1	235	58
GuWi	Güstrow	N53.67599 E12.29468	31	40	28.3	19.7	420	26.4	262.2	120
KaSp	Karbow	N53.39236 E11.74750	56	40	41	28.2	230	30.2	425.3	28
MiPe	Mirow	N53.27747 E12.88980	60	40	27.4	22.9	844	49.9	565.7	89
RaLa	Radelübbe	N53.53763 E10.99015	85	40	43.1	26.3	316	46	605	66
SG343	Schildfeld	N53.50141 E10.78216	85	40	37.4	23.2	349	38.4	443.4	70
SG361	Schildfeld	N53.50904 E10.78719	85	40	35.8	21.6	229	23.1	249.3	43
SK-1/SK-2	Schildfeld	N53.47082 E10.92367	79	80	43.65	26.8	153	22.95	309.75	49.5
SQ40	Schlemmin	N53.84331 E11.83440	25	40	21.3	19.9	571	20.3	198.5	95
SQ49	Schlemmin	N53.81915 E11.78740	53	40	35.5	24.4	240	23.8	287.5	63
SeTa	Schlemmin	N53.77843 E11.94032	27	40	21.3	18.4	707	25.3	227.5	83

comparative analysis on the FIR059 marker was calculated in the analogue way.

Results

Genetic variation within populations in MWP

Our results show that the methods used provide clear and easy assessment of alleles, resulting in reliable and easily scored genotypes. EMCs38 and GOT021 showed polymorphic patterns and were therefore excluded from genetic diversity estimates. PIC ranged between 0.18 (GOT045) and 0.84 (CsCAT3) overall stands with a total average of 0.58 (Table S2). An overview of common genetic parameters is shown in Table S1. Overall, the minimum and maximum values vary between 4.222 and 7.111 (N_a) and 2.506 and 3.863 (N_e). Calculation of rarefied allele richness (A_r) for SK1+2 resulted in a lower value of 5.757 compared to N*a* due to sample size differences to the other populations. Overall, the lowest diversity values were observed in the stands KaSp, SeTa, SQ49 and SQ40 and MiPe. KaSp, SeTa, SQ49 and SQ40 were later identified as one (groups 2) out of three genetic groups in MWP (c.f.Fig.2 & S2). Highest diversity was found in the stands BiKu and SG343 (genetic group 3) and GuWi, SG361 and RaLa (genetic group 1). All populations show F-values close to or not far from zero.

Genetic variation between stands in MWP

The population genetic structures based on all genetic markers (EST-SSRs and gSSRs) distinguished four genetic clusters as the statistically most likely classification. When considering the population structure separately for EST-SSRs and gSSRs, 3 and 4 genetic clusters, respectively, were identified in MWP (Fig. 2, Fig. S2). The proportions of these genetic groups are clearly differentiated between stands in MWP (Fig. 2). AMOVA confirmed high significance (p = 0.01) between stand variances, with the covariance component at the population level being 7.4 %, 5.16 % and 9.09 % for all SSRs, EST-SSRs and gSSRs respectively (Table S3).

Basically, stands can be classified into three genetic groups based on the proportions of genetic cluster assignment for all SSRs. (Fig. 2a): Group 1 consists of stands with dominant "orange" cluster assignment. These are the stands GaJa, GuWi, RaLa, SG361 and SK1+2. Group 2 comprises stands with a significant share of the "light blue" cluster. These are the stands KaSp, MiPe, SeTa, SQ49 and SQ40. In group 3, two stands occur in which the "dark blue" cluster dominates. These are the stands BiKu and SG343. This grouping persists both considering only EST-SSRs (Fig. 2b) and gSSRs (Fig. 2c), with the EST-SSRs showing weaker clustering than the gSSRs. UPGMA cluster analysis divided the stands into the same three major groups. Where group 3 (BiKu and SG343) is genetically closer to group 1 (GaJa, GuWi, RaLa, SG361 and SK1+2) considering all SSRs and only gSSRs, but closer to group 2 (KaSp, MiPe, SeTa, SQ49 and gSSR SQ40) considering only EST-SSRs (Fig. S2).

No clear relationship between genetic structures and spatial distribution of chestnut stands in MWP can be observed. For stands where the orange cluster is dominant and those where the light blue cluster is most abundant, local geographic clumping can be observed. For the orange cluster, stands SG361, RaLa and SK-1/SK-2 are adjacent. The same is true for the three stands of the light blue dominated cluster (SQ40, SQ49, SeTa). Other stands with high proportions of these two clusters are distant from each other, with no discernible geographical pattern (Figure 3).

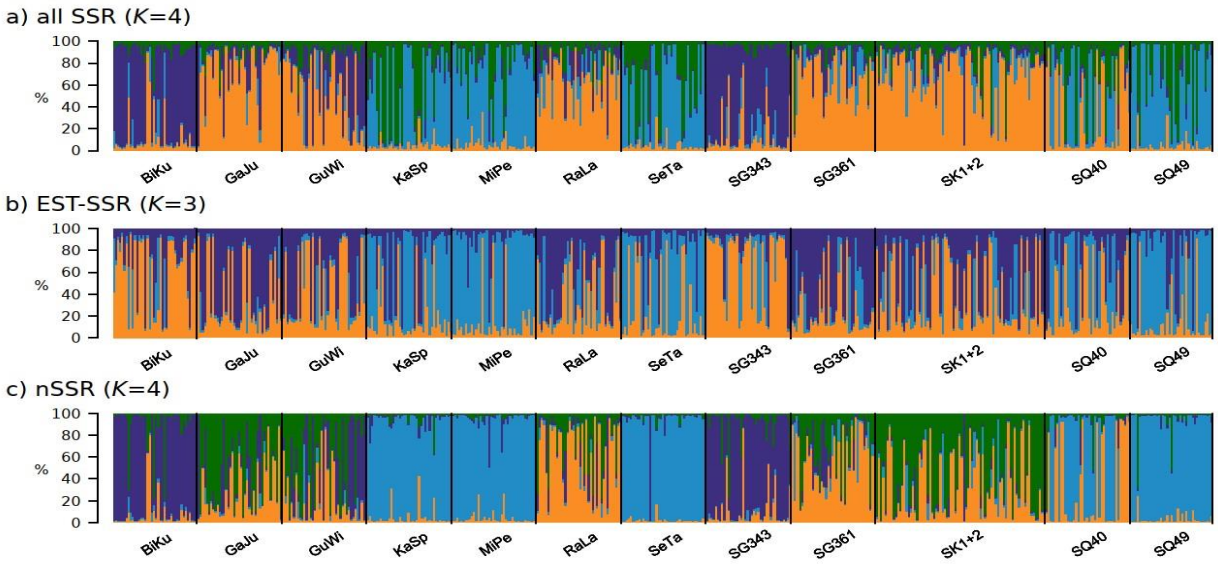


Figure 2
Mean cluster assignment of individuals from the 10 replicates per cluster number (K) for each most likely cluster number (K) for a) all 11 genetic markers examined, b) the EST SSRs, and the c) genomic SSRs.

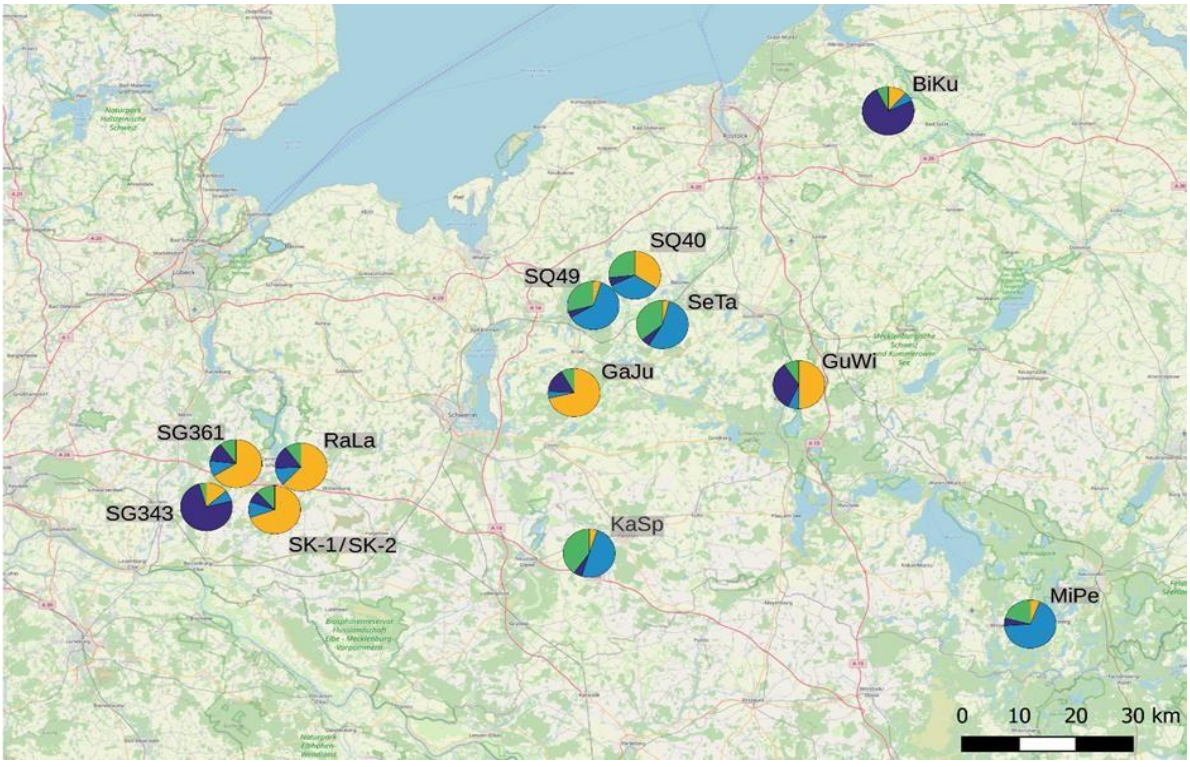


Figure 3
Mean population structure of the study populations for K = 4 genetic clusters. The proportions in the pie charts represent the mean membership of individuals in a population to the four genetic clusters (orange, light blue, dark blue, green). Geographical position of the pie charts corresponds approximately to the location of the study populations to avoid overlapps.

No significant relationship between the identified genetic groups and the stand quality, measured by valuable trees per hectare (zf/ha), was observed. Nevertheless, genetic group 2 shows the lowest number of valuable trees per hectare compared to the other two genetic groups (Tab. S4). Test statistics on model assumptions are displayed in table S5. Initial model showed non uniform distributed residuals. This was improved by adding a random effect, e.g. N/ha or H*e*, with no change in the results displayed in table S4.

Genetic variation within MWP and provenance from Southern Europe

In the following sections the dataset of Castellana et al. (2021) is re-analysed with the collected data in MWP on the basis of 8 overlapping SSRs in our and the study of Castellana et al. (2021).

Mean values of genetic parameters, based on the 8 overlapping SSRs are shown in Table 2. In comparison to average values of refugial regions in Spain, Italy, Greece and Turkey (N_a = 6.79, N_e = 3.83, A_r = 6.35) the stands in MWP show slightly less genetic variation (N_a = 6.21, N_e = 3.21, A_r = 5.32). The Greece provenances showed to be the less variable (N_a = 5.56, N_e = 3.78). The highest genetic variation was observed in Turkey for N_a and A_r of 7.56 and 7.05, respectively and in Spain on basis of N_e of 4.21. Deviation from Hardy-Weinberg expectations were lowest in MWP and Greece with an average *F* of 0.05, the highest value was detected in Turkey at *F* = 0.24.

Boxplots depicting N*a* (number of different alleles, Fig. 4a) indicate widely overlapping distributions between MWP and the refugial regions with higher values observed among southern European provenances. However, no significant difference was observed. In case of A*r* (rarefied allelic richness, Fig. 4b) and N*e* (number of effective alleles, Fig. 4c) the distributions showed more prevelant and statistically significant differences between the MWP stands and the southern European provenances.

Allele frequency distribution in MWP and Southern Europe with special emphasis on SSR-marker FIR059

Concerning the EST-SSR FIR059 gene locus, we standardized genotyping in our survey and the dataset of Castellana et al. (2021) for this marker.

Based on the assumption that up to now adaptive properties of the MWP provanances have not yet been significantly impacted by additional gene transfer and local adaptation, we used FIR059 as an indicator for the origin of MWP seed stocks and their corresponding adaptive traits. Figure 5 shows the allele frequencies for this marker in the MWP stands and dataset from Castellana et al. (2021) with standardized alleles.

Table 2

Comparison of maximum, mean and minimum values based on 8 overlapping SSR genetic markers analyzed in this study and in Castellana et al. (2021). Genetic parameters are: number of analyzed trees per site (N), the number of different alleles (N_a),rarefied allelic richness (A_r), number of effective alleles (N_e), observed Heterozygosity (H_o), expected Heterozygosity (H_e) and Fixation Index (*F*) between stands of MWP and Southern Europe.

Region		N	N_a	A_r	N_e	H_o	H_e	*F*
	Max	80	7.75	6.65	4.19	0.63	0.65	0.10
MWP	Mean	43.3	6.21	5.32	3.21	0.58	0.61	0.05
	Min	39	4.50	4.2	2.65	0.53	0.58	-0.03
SPAIN	Mean*	30	7.09	6.55	4.21	0.58	0.65	0.10
ITALY	Mean*	23	6.63	6.32	3.78	0.52	0.61	0.12
GREECE	Mean*	24.8	5.56	5.26	2.93	0.52	0.55	0.05
TURKEY	Mean*	25.9	7.56	7.05	4.04	0.49	0.66	0.24
	Max	31	8.13	7.4	4.83	0.63	0.70	0.30
S. Europe	Mean	26.7	6.79	6.35	3.83	0.54	0.62	0.12
	Min	20	5.25	5.25	2.73	0.41	0.55	0.01

* Because of few sampled provenances per country mean-values only.

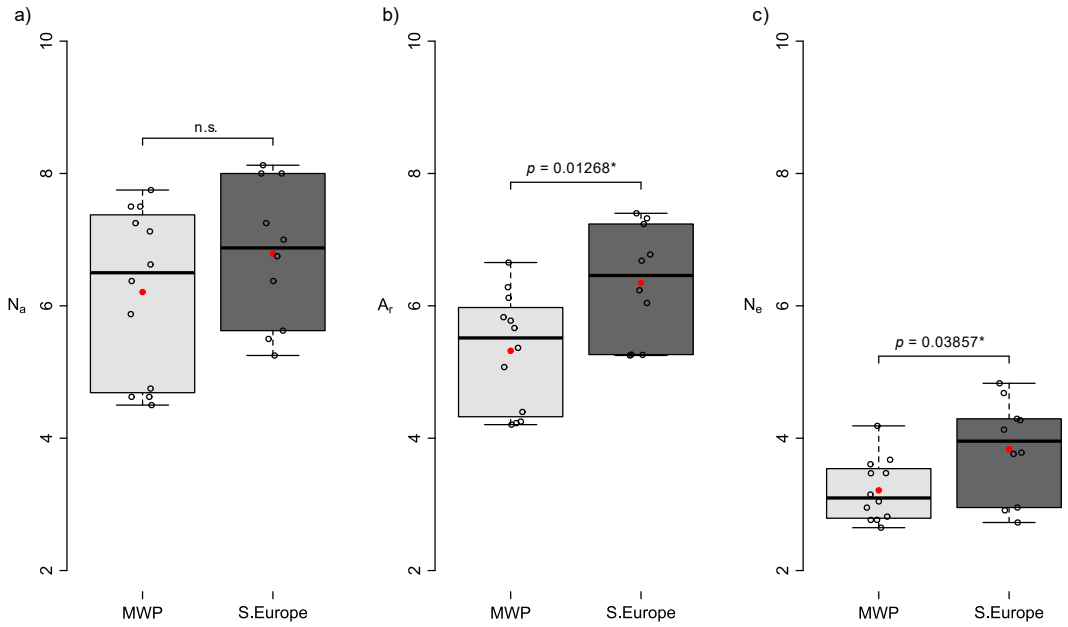


Figure 4a-c
Boxplot of a) the number of different alleles (N_a), b) rarefied allelic richness (A_r) and c) number of effective alleles (N_e) in investigated MWP-stands (grey) and provenances from Southern Europe (dark grey) acording to data from Catellana et al. (2021). Mean values are indicated as red dots, single stand datapoints are indicated as circles. Significance according to a two-sided t-test.

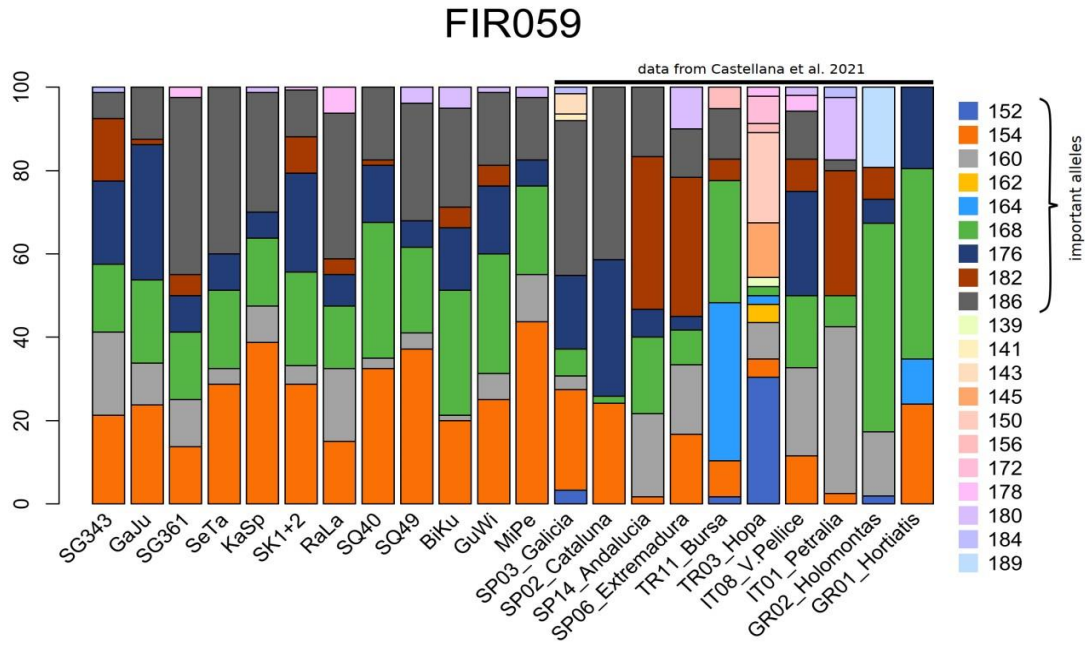


Figure 5
Frequency distribution of alleles at gene locus FIR059 of the studied populations and the data from Castellana et al. (2021). The data were adjusted for allele shifts between the data sets. The proportions of alleles per population identified as enviromentally associated in Castellana et al (2021) and Alcaide et al. (2019) and important for geographic assignment are shown in bold colors, others in pale colors. Abbreviations SP: Spain, TR: Turkey, GR: Greece and IT: Italy.

The frequency distributions of alleles between MWP stands are comparable to provenances from Southern Europe based on the reference data from Castellana et al. 2021. All stands in MWP contain alleles 154, 160, 168, 176 and 186 in varying proportions. The main difference in MWP, despite the different allele frequencies, is that allele 182 is absent in the stands SeTa, KaSP, SQ49 and GuWi. Similarities can be observed in the provenances SP03 Galicia and SP02 Cataluna where allele 182 is also absent and 168 has much lower proportions compared to MWP. Additonally, allele 160 is either absent or has lower proportion. Further similarities can be seen to the provenance IT08 V. Pellice, where largely the proportion of alleles 160,176 and 186 differ from MWP stands.

To identify genetic clusters based on FIR059 we used STRUCTURE to compare the standardized data from Castellana et al. 2021 and from MWP stands. In Figure S3 results for *K*=2 – *K*=4 are shown, with the most likely assignment being *K*=4. This assignment (*K*=4) is in agreement with the allocation according to the frequency distributions (Fig. 5). Consequently, the genetic differentiation between northern (SP02, SP03 IT08) and southern stands (SP06, SP014, IT01) in Spain and Italy is evident. The two Greek stands are minimally differentiated from the northern stands in Spain and Italy, but feature a higher proportion of the dark blue cluster. Notably distinct is the genetic structure in provenances from Turkey at FIR059. In comparison, all but one stands in MWP show less differentiation and similar genetic pattern to northern provenances in Spain (SP02, SP03). The stand SG343 resembles more the structure of the northern Italian provenance (IT08) at FIR059. Due to the cluster assignment, all MWP-stands can be assigned to the northern part of the western gene pool within the southern European provenances.

Discussion

Sampling, genotyping, and parameters of genetic variation

To our knowledge, we have conducted the first study of sweet chestnut stands growing far from the north-eastern limit of their natural distribution. The study encompassed stands within the federal state of Mecklenburg Western Pomerania, Germany. The sample size utilized was equivalent to or greater than those of other studies (Mattioni et al., 2013; Janfaza et al., 2017; Bouffartigue et al., 2020; Castellana et al., 2021).

Our estimate of average genetic diversity (H_e) in MWP stands is within the range of other analyses. The minimum and maximum values of H_e in our analysis are 0.53 and 0.621, respectively, whereas other studies have reported values between 0.43 and 0.56 (Janfazu et al. 2017) or 0.64 and 0.81 (Castellana et al. 2021). Usually, the comparison of genetic parameters based on SSR markers is difficult because these estimates are strongly influenced by sample size, marker set composition, etc.. These and other drawbacks have already been mentioned by Barrandeguy and García (2021). Therefore, a comparison of genetic variation parameters based on different microsatellite analyses from the literature should be considered as a rough estimate.

To minimize differences due to the use of different marker sets, we compare our data with the accessible genotype data of Castellana et al. (2021) based on 8 SSR markers analysed in both studies. The calculated H_e values gave very similar results in our study (H_e= 0.58 — 0.65) and Castellana et al. (2021) (H_e =0.56 — 0.70). Overall, the mean number of different alleles (N_a) in MWP is lower (N_a = 6.21) than in most provenances. Only the Greek provenances and the Italian provenance from Petralia have fewer alleles (N_a = 5.25— 5.63). However, the comparison between all MWP stands and southern European provenances showed a significant decrease in N_e and A_r in MWP. It is possible that this reduction in the number of different alleles is due to the transfer of forest reproductive material from one provenance to another geographic region. Loss of genetic diversity is one amongst several risks in transfer of reproductive material (Chen et al., 2022). Also, during harvest and production of seeds and plants, genetic structures can be strongly influenced by genetic drift (Hosius et al., 2006). Thus, substantial differences in genetic composition can occur between seed lots of different years.

Assuming genetic drift by seed transfer, N_a is usually more sensitive than N_e. However, our data show that N_e seems to be more affected. Barrandeguy and Garcia (2021) indicate that the sensitivity of N_e to genetic drift depends on the frequency distribution of the gene loci analysed. Moreover, our results support the observation of high allelic richness reported for refugial areas (Widmer and Lexer 2001). Thus, our results simply indicate that only a portion of the total natural range of sweet chestnut has been sampled and used for plantings in MWP. Nevertheless, there is a wide overlap of genetic variation between stands of MWP and provenances from Southern Europe. Therefore, the gene pool of chestnut stands in MWP shows comparable genetic variation to at least some stands from refugial regions which should be sufficient for further production of seed and plant material. The highest diversity was measured in stands of genetic group 3 (BiKu, SG343), followed by genetic group 1 which still realized medium to high values of genetic diversity. Concerning the overall stem quality, the stand GuWi shows exceptionally high quality (cf. Table 1) and belongs in tendency to these genetic groups with higher diversity (cf. Table S2). Therefore, the stocks of genetic groups 1 and 3 are best suited as local seed sources. But since group 2 is highly differentiated from the other two genetic groups, its exclusion from seed production could reduce the overall genetic diversity of the *C. sativa* gene pool in MWP. Also, the stands are still comparable to French stands on the basis of effective number of alleles (Bouffartigue et al. 2020). Further, imports of seed material with tested higher diversity from southern provenances should be considered to enrich the genetic diversity of *C. sativa* in MWP. In addition, the establishment and evaluation of provenance trials for quality, yield and climate resistance/ resilience are needed to assess the suitability of specific origins.

Analysis of geographic genetic pattern and possible origin of stands in MWP

Our structural analysis of 13 stands in MWP identified substantial genetic differentiation among the stands that could indicate different places of origin, growing stocks, or seed mixtures.

Three main genetic groups were identified. Considering that the stands originated from introduced seed and plant material, a visible clustering of similar genetic structures (stands SG361, RaLa, SK-1/SK-2 and SQ40, SQ49 and SeTa) is noticeable. This could possibly be attributed to the use of different but locally similar seed sources. Studies of red oak stands in Germany based on cpDNA haplotypes (Pettenkofer et al., 2020) have shown that the introduction history of a plant species can lead to such genetic/geographic patterns. There also seems to be an influence of different phases of introduction for sweet chestnut in MWP. For example, the stands of three clumped stands within the "orange" cluster (genetic group 1) are significantly older (79-85 years) than those of the "dark blue" cluster (25-53 years). This could indicate different origins of the used forest reproductive material.

In contrast to the examination of the FIR059 marker alone, the complete SSR analysis by Castellana et al. (2021) provided a much more detailed picture of the genetic structure between regions, distinguishing Spanish, Italian and Greek populations, as well as the northern and southern provenances of these regions. The genetic differences between Spanish and Italian provenances, and their northern and southern parts, were also found by Alessandri et al. (2022). Therefore, more emphasis should be placed on the use of reference material in future studies to increase the accuracy of assignment.

Regarding the potentially adaptive genetic marker FIR059, we were able to standardize the genotyping between our study and Castellana et al. (2021). This EST-SSR is located near the DEAD-box ATP-dependent RNA helicase 7 / RH7 gene, which has been reported to have a function in the stress adaptation processes, including salt or drought stress (Vashisht and Tuteja, 2006; Kim et al., 2008; Macovei et al., 2012). For the provenances from Spain, Italy, Greece, and Turkey, our results are partially similar to those based on the full marker set of Castellana et al. (2021). We found a clear genetic differentiation between northern and southern provenances in Spain and Italy and a separation of the eastern Turkish provenances. This is in agreement with the results of Mattioni et al. (2017). They identified three genetic clusters with a western cluster including populations from Spain, UK, Italy and Hungary, a central cluster including Romania, Bulgaria and Greece as well as western parts of Turkey, and an eastern cluster including populations from eastern Turkey, Georgia and more eastern parts of the distribution.

Comparison at FIR059 between southern European provenances (Castellana et al. 2021) and the introduced stands in MWP indicates a possible northern origin within the western gene pool. Hereby the assumption that the adaptive properties of FIR059 have not changed significantly was made; i.e., that the genetic structure has not been significantly affected to date. Potentially the region of origin of the MWP stands is in the northern parts of Spain and/or Italy. Nevertheless, a French origin of the MWP stands can also not be ruled out. According to Mattioni et al. (2017) Spanish and/or Italian provenances share possible genetic similarities to French provenance. The found genetic similarities between Portuguese, Italian, French and English provenances were concluded to be attributed to the human seed translocation and consecutive mating (Mattioni et al., 2017). Unfortunately, data from French provenances were not investigated in Castellana et al. (2021).

Considering the adaptive characteristics of FIR059 investigated in Alcaide et al. (2019) and Castellana at al. (2021), both studies showed some overlapping results. However, the particular alleles associated with drought response or climate variables were not completely identical between these studies. Additionally, standardization of marker data without reference samples may pose a challenge. This could be the case when comparing the above-mentioned studies and applies to the comparison of MWP stands with those of Castellana et al. (2021) at FIR059. Thus, only a general assignment to a possible northern origin within the western gene pool is possible. For further clarification of the origin and a conclusive assessment on climate adaptation in MWP stands, further studies on the association of genetic markers to drought tolerance in chestnut based on provenance trials or comparative test cultivation are needed.

Conclusions

Our analysis of *C. sativa* stands in Mecklenburg-Western Pomerania was the first genetic evaluation of populations far from the north-eastern distribution limit. The stands harbour a comparable genetic variation to stands from refugial regions. Thus, it can be concluded that the stands are sufficient for further production of seed and plant material in MWP. Although, from a genetic point of view, many of the studied stands in MWP are suitable as seed stocks, seed harvests in areas of southern provenance with a high degree of documented genetic variation can also be considered to increase the genetic variation of the gene pool in MWP. However, other regions with *C. sativa* in Germany have not yet been genetically mapped. Further collaborative research efforts are needed for seed production and evaluation of seed zones for this alternative tree species in Germany. It is possible that a highly differentiated pattern could be observed, as suggested by the remarkable regional differentiation in MWP.

Tracing the origin of introduced stands to a particular region using literature data is challenging. Our results highlight the importance and need of standardized genotyping through the use of reference samples to improve the accuracy of dataset comparisons. Overall, the findings suggest that the sweet chestnut stands in Mecklenburg Western Pomerania likely originated from introduced seed and plant material from the northern part of the western gene pool, potentially from various southern European provenances. These finding must be carefully considered as a first estimate of origin. Further

analyses based on additional samples from provenance trials can substantiate these results and then serve as a basis for targeted gene conservation management and seed production in MWP. Further research is needed to explore the adaptive properties of specific genetic markers and their association with drought tolerance in sweet chestnut.

Funding

This research was funded by the Ministry of Climate Protection, Agriculture, Rural Areas and the Environment Mecklenburg-Western Pomerania (Funding/ Research plan number WG.1995.1.3 and WW.2016.1.4).

Acknowledgements

The authors are grateful for the support in sample collection by Fritz von Baumbach and the forest rangers of the Landesforstanstalt Mecklenburg Western Pomerania.

Supporting Information

The SSR datasets of the individuals from MWP and supplementary tables and figures are available with the online version of the current paper.

References

Alcaide F, Solla A, Mattioni C, Castellana S, Martín MÀ (2019) Adaptive diversity and drought tolerance in Castanea sativa assessed through EST-SSR genic markers. Forestry 92(3): 287–296. https://doi.org/10.1093/forestry/cpz007

Alessandri S, Cabrer AMR, Martín MÁ, Mattioni C, Pereira-Lorenzo S, Dondini L (2022) Genetic characterization of Italian and Spanish wild and domesticated chestnut trees. Scientia Horticulturae 295: 1–28. https://doi.org/10.1016/j.scienta.2022.110882

Barrandeguy ME, García MV (2021) The sensitiveness of expected heterozygosity and allelic richness estimates for analyzing population genetic diversity. In: Trindade MR and CM de Araújo (eds) Genetic Variation. IntechOpen, ISBN 978-1-83881-097-9. https://doi.org/10.5772/intechopen.95585

Bouffartigue C, Debille S, Fabreguettes O, Cabrer AR, Pereira-Lorenzo S, Flutre T,L Harvengt L (2020) Two main genetic clusters with high admixture between forest and cultivated chestnut (Castanea sativa Mill.) in France. Annals of Forest Science. Annals of Forest Science 77(3). https://doi.org/10.1007/s13595-020-00982-w

Buck EJ, Hadonou M, James CJ, Blakesley D, Russell K (2003) Isolation and characterization of polymorphic microsatellites in European chestnut (Castanea sativa Mill.). Molecular Ecology Notes 3(2): 239–241. https://doi.org/10.1046/j.1471-8286.2003.00410.x

Castellana S, Martín MÀ, Solla A, Alcaide F, Villani F, Cherubini M, Neale D, 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). https://doi.org/10.1007/s13595-021-01027-6

Caudullo G, Welk E, San-Miguel-Ayanz J (2017) Chorological maps for the main European woody species. Data in Brief 12: 662-666. https://doi.org/10.1016/j.dib.2017.05.007

Chen Z, Grossfurthner L, Loxterman JL, Masingale J, Richardson BA, Seaborn T, Smith B, Waits LP, Narum SR (2022) Applying genomics in assisted migration under climate change: Framework, empirical applications, and case studies. Evolutionary Applications 15(1): 3–21. https://doi.org/10.1111/eva.13335

Chiocchini F, Mattioni C, Pollegioni P, Lusini I, Martín MÀ, Cherubini M, Lauteri M, Villani F (2016) Mapping the genetic diversity of Castanea sativa: exploiting spatial analysis for biogeography and conservation studies. Journal of Geographic Information System 08:248–259.

Conedera M, Manetti MC, Giudici F, Amorini E (2004) Distribution and economic potential of the sweet chestnut (Castanea sativa Mill.) in Europe. Ecologia mediterranea 30(2): 179–193. https://doi.org/10.3406/ecmed.2004.1458

Conedera M, Tinner W, Krebs P, de Rigo D, Caudullo G (2016) Castanea sativa in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, D de Rigo, G Caudullo, T Houston Durrant and A Mauri (eds.), European atlas of forest tree species. Publications Office of the European Union, Luxembourg, pp. 78-79 ISBN 978-92-76-17290-1

Durand J, Bodénès C, Chancerel E, Frigerio JM, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn HP, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrán A, Ikaran Z, Cabané C, Ueno S, Alberto F, Dumoulin PY, Guichoux E, de Daruvar A, Kremer A, Plomion C (2010) A fast and cost-effective approach to develop and map EST-SSR markers: Oak as a case study. BMC Genomics 11(1). https://doi.org/10.1186/1471-2164-11-570

Faust K, Fussi B (2009) Genetik und Vermehrungsgut der Esskastanie. LWF Wissen 81: 14–19

Fernández-Cruz J, Fernández-López J (2012) Morphological, molecular and statistical tools to identify Castanea species and their hybrids. Conservation Genetics 13(6): 1589–1600. https://doi.org/10.1007/s10592-012-0408-0

Goudet J, Jombart T (2015) hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.04-22. R package version 0.04-22

Hartig F (2022) DHARMa: Residual diagnostics for hierarchical (multi-Level/ mixed) regression models. R package version 0.4.6. https://CRAN.R-project.org/package=DHARMa

Hosius B, Leinemann L, Konnert M, Bergmann F (2006) Genetic aspects of forestry in the central Europe. European Journal of Forest Research 125(4): 407–417. https://doi.org/10.1007/s10342-006-0136-4

Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13000 Years Ago Cambridge University Press

Janfaza S, Yousefzadeh H, Hosseini Nasr SM, Botta R, Asadi Abkenar A, Marinoni DT (2017) Genetic diversity of castanea sativa an endangered species in the hyrcanian forest. Silva Fennica 51(1): 1–15. https://doi.org/10.14214/sf.1705

Kamvar ZN, Brooks JC, Grünwald NJ (2015a) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6: 1–10. https://doi.org/10.3389/fgene.2015.00208.

Kamvar ZN, Tabima JF, Brooks JC, Grunwald NJ (2015b) Genetic analysis of populations with mixed reproduction Package ' poppr '.

Kim JS, Kim KA, Oh TR, Park CM, Kang H (2008) Functional characterization of DEAD-Box RNA helicases in Arabidopsis thaliana under abiotic stress conditions. Plant and Cell Physiology 49(10): 1563–1571. https://doi.org/10.1093/pcp/pcn125

Kleber A, Reiter P, Ehrhart H-P, Matthes U (2020) Steckbriefe Ergänzende Baumarten. FAWF/RLP Kompetenzzentrum für Klimawandelfolgen. https://www.klimawandel-rlp.de/fileadmin/website/klimakompetenzzentrum/Klimawandelinformationssystem/Handlungsfelder/Wald/Ergaenzende_Baumarten/CASA_Steckbrief.pdf.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15(5): 1179–1191. https://doi.org/10.1111/1755-0998.12387

Krebs P, Pezzatti GB, Beffa G, Tinner W, Conedera M (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. https://doi.org/10.1016/j.quascirev.2019.01.002

Li YL, Liu JX (2018) StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular Ecology Resources 18(1): 176–177. https://doi.org/10.1111/1755-0998.12719

Lusini I, Velichkov I, Pollegioni P, Chiocchini F, Hinkov G, Zlatanov T, Cherubini M, Mattioni C (2014) Estimating the genetic diversity and spatial structure of Bulgarian Castanea sativa populations by SSRs: Implications for conservation. Conservation Genetics 15(2): 283–293. https://doi.org/10.1007/s10592-013-0537-0

Macovei A, Vaid N, Tula S, Tuteja N (2012) A new DEAD-box helicase ATP-binding protein (OsABP) from rice is responsive to abiotic stress. Plant Signaling and Behavior 7(9): 1138–1143. https://doi.org/10.4161/psb.21343

Marinoni D, Akkak A, Bounous G, Edwards KJ, Botta R (2003) Development and characterization of microsatellite markers in Castanea sativa (Mill.). Molecular Breeding 11(2): 127–136. https://doi.org/10.1023/A:1022456013692

Martin MA, Mattioni C, Cherubini M, Taurchini D, Villani F (2010) Genetic diversity in European chestnut populations by means of genomic and genic microsatellite markers. Tree Genetics and Genomes 6(5): 735–744. https://doi.org/10.1007/s11295-010-0287-9

Mattioni C, Martin MÀ, Pollegioni R, Cherubini M, Villani F (2013) Microsatellite markers reveal a strong geographical structure in European populations of Castanea sativa (Fagaceae): Evidence for multiple glacial refugia. American Journal of Botany 100(5): 951–961. https://doi.org/10.3732/ajb.1200194

Mattioni C, Martin MÀ, Chiocchini F, Cherubini M, Gaudet M, Pollegioni P, Velichkov I, Jarman R, Chambers FM, Paule L, Damian VL, Crainic GC, Villani F (2017) Landscape genetics structure of European sweet chestnut (Castanea sativa Mill): indications for conservation priorities. Tree Genetics & Genomes 13(2): 39. https://doi.org/10.1007/s11295-017-1123-2

Mattioni C, Ranzino L, Cherubini M, Leonardi L, La Mantia T, Castellana S, Villani F, Simeone MC (2020) Monuments unveiled: Genetic characterization of large old chestnut (Castanea sativa Mill.) trees using comparative nuclear and chloroplast DNA analysis. Forests 11(10): 1–20. https://doi.org/10.3390/f11101118

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28(19): 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pereira-Lorenzo S, Ramos-Cabrer AM, Barreneche T, Mattioni C, Villani F, Díaz-Hernández MB, Martín LM, Martín À (2017) Database of European chestnut cultivars and definition of a core collection using simple sequence repeats. Tree Genetics & Genomes 13(5). https://doi.org/10.1007/s11295-017-1197-x

Pettenkofer T, Finkeldey R, Müller M, Krutovsky KV, Vornam B, Leinemann L, Gailing O (2020) Genetic variation of introduced red oak (Quercus rubra) stands in Germany compared to North American populations. European Journal of Forest Research 139(2): 321–331. https://doi.org/10.1007/s10342-019-01256-5

Poljak I, Idžojtić M, Šatović Z et al. (2017) Genetic diversity of the sweet chestnut (Castanea sativa Mill.) in Central Europe and the western part of the Balkan Peninsula and evidence of marron genotype introgression into wild populations. Tree Genetics & Genomes 13, 18 https://doi.org/10.1007/s11295-017-1107-2

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Puechmaille SJ (2016) The program STRUCTUR does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources 16(3): 608–627. https://doi.org/10.1111/1755-0998.12512

R Core Team (2021) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

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

Tourvas N (2023) PopGenUtils: A collection of useful functions to deal with genetic data in R. R package version 0.1.8.

Thurm EA, Werning M, Nagel R (2022) Tsuga und Thuja als Nadelholz-Alternativen im Klimawandel? - Standörtliches und waldwachstumskundliches Potential der westlichen Hemlocktanne und des Riesenlebensbaums in Deutschland. In: Nagel R, Schmidt M (eds) Tagungsband 2022: Deutscher Verband Forstlicher Forschungsanstalten. Sektion Ertragskunde. online Jahrestagung 2022.pp 138–152.

Thurm EA, Heitz R (2018) Anbaueignung der Edelkastanie in Deutschland. LWF Wissen 81: 31–40

Thurm EA, Hernandez L, Baltensweiler A, Ayan S, Rasztovits E, Bielak K, Zlatanov TM, Hladnik D, Balic B, Freudenschuss A, Büchsenmeister R, Falk W (2018) Alternative tree species under climate warming in managed European forests. Forest Ecology and Management 430: 485–497. https://doi.org/10.1016/j.foreco.2018.08.028

Vashisht AA, Tuteja N (2006) Stress responsive DEAD-box helicases: A new pathway to engineer plant stress tolerance. Journal of Photochemistry and Photobiology B: Biology 84(2): 150–160. https://doi.org/10.1016/j.jphotobiol.2006.02.010

Wood SN (2017) Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.Zhang L, Wu S, Chang X, Wang X, Zhao Y, Xia Y, Trigiano RN, Jiao Y, Chen F (2020) The ancient wave of polyploidization events in flowering plants and their adaptation to environmental stress. Plant Cell Environ 2020:1-10. https://doi.org/10.1111/pce.13898