Conservation Genetics Resources
https://doi.org/10.1007/s12686-021-01220-9

TECHNICAL NOTE

Development of highly validated SNP markers for genetic analyses of chestnut species

Clément Larue[1,2] · Erwan Guichoux[1] · Benoît Laurent[1] · Teresa Barreneche[3] · Cécile Robin[1] · Marie Massot[1] · Adline Delcamp[1] · Rémy J. Petit[1]

Received: 10 March 2021 / Accepted: 14 June 2021

Abstract

To better study and manage chestnut trees and species, we identified nuclear single nucleotide polymorphism (SNP) markers using restriction-associated DNA sequencing. Out of 343 loci tested, 68 SNP markers were selected that withhold stringent quality criteria such as quasi-systematic amplification across species and Mendelian segregation in both purebred and hybrid individuals. They provide sufficient power for species, hybrids and backcross characterization as well as for clonal identification, as shown by a comparison with single sequenced repeat (SSR) loci.

Keywords *Fagaceae* · *Castanea* sp. · Single nucleotide polymorphism · MassARRAY

Chestnuts are self-incompatible and insect-pollinated Fagaceae trees from the Northern hemisphere (Stout 1926; Xiong et al. 2019; Barreneche et al. 2019; Larue et al. 2021a, b). Three species are widely cultivated for their nutritious nuts, the Japanese (*Castanea crenata*), Chinese (*C. mollissima*) and European (*C. sativa*) chestnuts (Barreneche et al. 2019). *C. sativa* is very vulnerable to ink disease and chestnut blight caused by pathogenic agents originating from Asia (Gonthier and Robin 2019). Hybrids between Asiatic species and *C. sativa* proved resistant to ink disease and were selected for cultivation in Europe. Genetic markers could help differentiate chestnut species, hybrids and other introgressed material as well as varieties, thereby facilitating the management of orchards. Such markers could also clarify the status of natural chestnut stands threatened by the propagation of diseases and by genetic pollution. A number of molecular markers have been previously developed in chestnuts, especially SSRs (e.g. Buck et al. 2003; Marinoni et al. 2003; Durand et al. 2010; Laurent et al. 2020) and were then widely used (e.g. Barreneche et al. 2004; Casasoli et al. 2006; Bodénès et al. 2012; Fernández-Cruz and Fernández-López 2012; Mattioni et al. 2013; Pereira-Lorenzo et al. 2017, 2020; Bouffartigue et al. 2020; Nishio et al. 2021). However, SNPs have some important advantages over SSRs (Guichoux et al. 2011). First, genotyping errors are much rarer with SNPs than with SSRs, facilitating standardisation across laboratories. Second, SNP genotyping platforms make it possible to quickly characterize and score a large number of samples at reduced costs. Although SNPs have already been developed in chestnuts (Santos et al. 2017; Garcia et al. 2018; Nunziata et al. 2020; Sun et al. 2020), no SNP assay has been designed and optimized for the mentioned applications.

We identified SNPs in the nuclear genome of chestnuts using restriction-site associated DNA sequencing (Miller et al. 2007). We used three samples from the INRAE chestnut germplasm collection, one *C. mollissima* (Ca 737), one *C. crenata* (Ca 04) and one *C. sativa* (Ca 577). We also

✉ Clément Larue
clement.larue@inrae.fr

[1] BIOGECO, INRAE, Univ. Bordeaux, 69 route d'Arcachon, 33610 Cestas, France

[2] INVENIO, Maison Jeannette, 24140 Douville, France

[3] BFP, INRAE, Univ. Bordeaux, 33140 Villenave d'Ornon, France

Published online: 22 June 2021

included SNPs originating from *C. mollissima* CCall_Unigene_V2 assembly data and from *C. dentata* AC454_Unigene_V3 contig data (Santos et al. 2017; http://www.hardwoodgenomics.org/).

We first tested the markers on a set of 95 samples including the three sequenced parents, the offspring of two interspecific crosses (Ca 577 × Ca 737 and Ca 577 × Ca 04) and nine French cultivars. Their DNA was isolated from leaves dried in silica gel with Qiagen DNeasy 96 Plant kit. We further checked the markers on another set of 95 unique genotypes from the INRAE chestnut germplasm collection, which includes the three chestnut species and several F1, F2 and advanced hybrids. Their DNA was isolated from frozen leaves with a modified CTAB DNA isolation protocol (Supplementary 1) adapted from Doyle and Doyle (1987).

We selected 343 candidate SNPs, including 37 loci from Santos et al. (2017) and 306 loci originating from a restriction-associated DNA sequencing experiment (García et al. 2018). These loci were successfully sequenced in all three parents, were heterozygous in at least one of them, and lacked variation within at least 50 bp around the SNP position. We designed nine MassARRAY multiplexes (Assay Design Suite v2.0, Agena Bioscience, San Diego, USA) of up to 40 loci. Data analysis relied on MassARRAY Typer Analyzer 4.0.26.75 (Agena Biosciences). We excluded all monomorphic SNPs, loci with weak or ambiguous signal (i.e., displaying more than three genotypes clusters or with unclear cluster delimitation) and loci with > 10% missing data. Out of the 343 loci tested, 237 were validated (Larue 2021, File 1).

We then combined the best markers in two new multiplexes of 40 and 39 loci (Larue 2021, File 2) and used them to genotype the second set of 95 individuals. All 79 loci were polymorphic. Extensive testing with a specially developed software MismatchFinder (https://github.com/jschoete/mismatchfinder, Supplementary 2) of over 3000 progeny identified 68 markers segregating according to Mendelian expectations that worked well on all species. The mean amplification rate was 99.9% (97.8–100%) per SNP locus and 99.9% (98.7–100%) per sample. By contrast, the mean amplification rate was 92.2% (76.9–100%) per SSR locus and 91.7% (41.5–100%) per sample (Laurent et al. 2020).

These markers are listed in Table 1, and 66 of the 68 SNPs are located on the chromosome assembly of *C. mollissima* genome (Sun et al. 2020). Table 1 also includes allelic frequencies for the three chestnut species and for *C. sativa* × *C. crenata* hybrids, computed using Genalex 6.51 (Peakall and Smouse 2012).

To evaluate the utility of these markers for species and hybrid identification, we used the Bayesian clustering analysis software Structure (Pritchard et al. 2000) and compared the results with those obtained with SSRs (Laurent et al. 2020). A total of 91 unique genotypes were characterized with both types of markers and used for the comparison (Larue 2021, Files 3 and 4; Supplementary 3). Three clear-cut genetic clusters were identified with both markers, matching well with the known identity of the trees and confirming the taxonomic utility of these SNPs (Fig. 1). We also computed the probability of identity for the 68 SNPs and the 94 SSRs (Supplementary 4). For the SNPs, they were all close to zero, showing that all chestnut genotypes can be easily differentiated with these markers. To conclude, the developed SNPs are suitable for identification of chestnut cultivars, species and hybrids. They should help manage production orchards and monitor the few remaining wild European chestnut stands.

Table 1 Allele frequencies of the 79 polymorphic SNPs

Name	Chromosome[1]	Position	Alleles[2]	*crenata*	*mollissima*	*sativa*	*s* × *c*
10090_56	03	66.9	A/G	0.714	0.235	0.863	0.500
10271_144	02	66.6	A/G	1.000	1.000	0.325	0.735
1156_97	08	19.8	G/A	1.000	1.000	0.438	0.676
11811_126	03	51.0	A/T	1.000	0.559	1.000	1.000
12533_73	07	21.4	G/A	1.000	1.000	0.154	0.618
13102_76	11	0.8	G/A	1.000	1.000	0.400	0.588
14353_126	12	12.6	G/A	0.786	1.000	0.150	0.500
14391_73	09	9.4	A/T	0.214	0.971	1.000	0.706
14608_73	01	48.7	T/G	1.000	1.000	0.175	0.588
1473_122	09	6.1	G/A	0.643	1.000	1.000	0.824
14743_107	05	54.0	T/C	0.429	0.000	1.000	0.824
15233_98	11	26.1	C/G	1.000	1.000	0.263	0.588
16018_58	07	10.9	T/C	1.000	0.441	0.150	0.618
16460_134	04	18.7	A/T	1.000	1.000	0.564	0.824
17159_26	02	37.7	A/G	1.000	1.000	0.550	0.735
177_149	08	44.3	C/A	1.000	1.000	0.825	0.912
18967_46	07	49.6	T/C	1.000	0.912	0.775	0.941
2219_98	02	62.4	G/A	1.000	1.000	0.788	0.882
22561_111	02	76.3	G/A	0.571	1.000	1.000	0.794
23240_125	12	3.5	A/T	1.000	0.382	1.000	1.000
24122_43	01	13.1	A/G	0.214	0.912	1.000	0.559
24959_44	06	14.9	C/T	1.000	1.000	0.488	0.588
262_54	05	13.4	C/T	0.714	0.559	0.925	0.735
26669_93	07	26.5	A/C	0.786	1.000	1.000	0.912
26674_123	01	69.0	C/T	1.000	1.000	0.638	0.912
27408_32	08	32.6	C/T	1.000	1.000	0.313	0.618
28714_26	02	17.6	G/A	1.000	0.735	1.000	1.000
30876_169	07	29.8	T/C	0.286	1.000	1.000	0.588
3239_136	04	26.7	A/G	0.643	1.000	1.000	0.794
3252_33	08	37.1	C/T	0.643	1.000	1.000	0.706
33254_153	04	31.4	C/T	1.000	1.000	0.738	0.912
36048_114	01	69.5	C/T	1.000	1.000	0.250	0.618
37241_49	10	13.9	A/T	0.643	1.000	1.000	0.824
3876_115	–	–	T/C	1.000	0.559	1.000	1.000
38812_93	10	19.5	A/C	1.000	1.000	0.200	0.588
39014_71	04	45.7	A/G	0.786	1.000	1.000	0.676
4137_29571	07	19.4	C/G	1.000	0.000	0.650	0.853
4285_31	05	24.3	C/G	1.000	1.000	0.538	0.882
435_64	12	37.7	C/G	1.000	1.000	0.775	0.941
4856_83	12	15.6	C/T	1.000	0.529	1.000	1.000
6083_144	02	66.6	A/G	1.000	1.000	0.325	0.735
6207_157	01	76.4	G/C	0.786	1.000	1.000	0.706
6505_48	01	73.9	T/A	0.500	1.000	1.000	0.676
6519_41_95_2	04	40.6	G/A	0.714	0.471	1.000	0.853
6803_119	07	10.9	C/T	0.857	0.441	1.000	0.647
774_155	04	20.2	A/T	1.000	1.000	0.838	0.941
8143_154	08	10.9	C/G	0.571	1.000	1.000	0.824
8363_141	07	9.7	G/A	1.000	1.000	0.138	0.676
AC_32934_470	02	71.7	C/T	0.429	0.824	0.950	0.529
AC_39247_1551	05	49.6	T/C	1.000	1.000	0.550	0.824
b1_SNP_higher_path_12837	02	52.8	G/A	0.429	1.000	0.363	0.412

Table 1 (continued)

Name	Chromosome[1]	Position	Alleles[2]	*crenata*	*mollissima*	*sativa*	*s* × *c*
b1_SNP_higher_path_18882	02	21.2	A/C	1.000	1.000	0.438	0.706
b1_SNP_higher_path_5736	04	7.1	G/A	0.929	0.853	0.625	0.706
CC_22194_867	12	41.6	G/A	1.000	0.618	0.838	0.971
CC_23658_1308	11	42.8	G/A	1.000	0.794	0.750	0.912
CC_34458_544	04	36.2	G/A	1.000	0.529	0.738	0.941
CC_4168_418	01	39.3	C/T	0.714	0.941	0.000	0.471
CC_45599_593	02	78.3	T/C	0.857	0.294	1.000	0.971
CC_46354_1005	–	–	C/T	1.000	0.529	0.713	0.824
CC_47186_942	04	7.7	G/A	1.000	0.147	0.625	0.941
SNP_10570_98_A_T	07	31.8	T/A	0.500	0.000	0.863	0.735
SNP_11428_101_A_G	05	27.6	G/A	1.000	0.000	0.525	0.882
SNP_12453_74_G_A	11	0.8	G/A	1.000	1.000	0.400	0.588
SNP_15209_73_A_G	12	38.5	A/G	1.000	1.000	0.500	0.706
SNP_15252_70_A_G	01	66.2	A/G	1.000	1.000	0.325	0.588
SNP_17902_74_C_T	06	40.6	C/T	1.000	1.000	0.588	0.735
SNP_7453_81_A_T	04	14.1	A/T	0.429	0.735	1.000	0.853
SNP_higher_path_9439722	06	5.3	G/A	1.000	1.000	0.650	0.794

[1]Location on the 12 chromosomes of the reference genome of *C. mollissima*

[2]The two alleles of each SNP are indicated and the frequency of the first allele (the most abundant in the overall sample) is provided for the three chestnuts species and for the *C. sativa* x *C. crenata* (s x c) hybrids

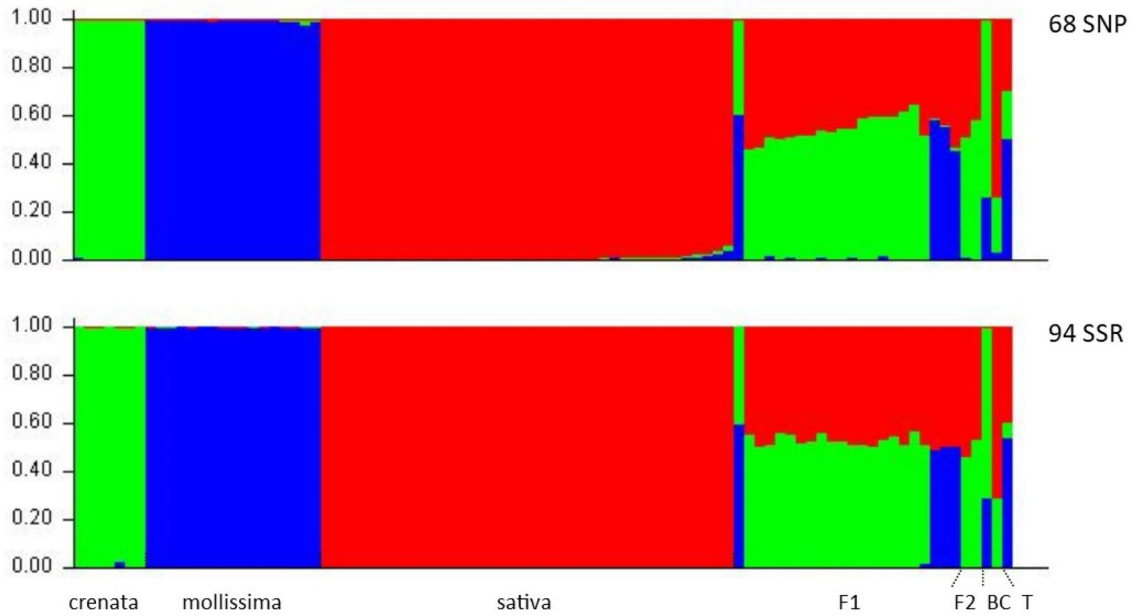

Fig. 1 Comparison of species assignment for 91 chestnut trees characterized with 68 SNPs (top) and 94 SSRs (bottom). Results are very similar even though the admixture values of hybrids and other crosses are closer to the expected proportions with SSRs than with SNPs, except for the three-way hybrid (T)

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12686-021-01220-9.

Acknowledgements We thank J. Schoeters for developing MISMATCHFINDER software, C. Lalanne for DNA isolation protocol development and L. Dubois, C. Boury and C. Mouden for SNP identification and genotyping. We thank INRAE experimental unit Vigne Bordeaux (UEVB) in Villenave d'Ornon, and INRAE experimental unit Forêt Pierroton (UEFP) for their continuous support for chestnut orchards management.

Author contributions CL: Conceptualization, investigation, resources, formal analysis, writing—original draft. EG: Conceptualization, methodology, formal analysis, writing—review & editing. BL: Formal analysis, writing—review & editing. TB: Conceptualization, funding acquisition, writing—review & editing. CR: Conceptualization, funding acquisition, writing—review & editing. MM: Investigation, resources, writing—review & editing. AD: Investigation, resources, writing—review & editing. RJP: Conceptualization, funding acquisition, supervision, writing—review & editing.

Funding This paper is part of the PhD of CL. The work was supported by ANRt funding under CIFRE PhD program to CL. Financial support to Invenio for this project was provided by the Conseil Régional d'Aquitaine for CL PhD work, as part of the REGINA project, and for BL work (Grant Number 2018-1R20204 project DIGIE "chestnut dieback: vulnerability and genetic determinism of ink disease resistance"). SNP development and genotyping were performed at the Genome Transcriptome Facility of Bordeaux (Grants from the Conseil Régional d'Aquitaine No. 20030304002FA and 20040305003FA, from the European Union FEDER No. 2003227 and from Investissements d'Avenir ANR-10-EQPX-16-01).

Data availability All data are available on Data INRAE: Larue C (2021) Development of SNP markers for the identification of chestnut species, hybrids and varieties. Portail Data INRAE, V2, UNF:6:eJQwJJ12X/74swiSKl9V2A== [fileUNF]. https://doi.org/10.15454/XEMDLD

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Informed consent This work does not contain any studies with human participants performed by any of the authors. This work does not contain any studies with human participants performed by any of the authors.

Research on human and animal participants This technical note does not contain any studies with animals performed by any of the authors.

References

Barreneche T, Botta R, Robin C (2019) Advances in breeding of chestnuts. In: Serdar U, Fulbright D (eds) Achieving sustainable cultivation of tree nuts. Burleigh Dodds Science Publishing Limited, Cambridge, pp 317–348

Barreneche T, Casasoli M, Russell K et al (2004) Comparative mapping between *Quercus* and *Castanea* using simple-sequence repeats (SSRs). Theor Appl Genet 108:558–566. https://doi.org/10.1007/s00122-003-1462-2

Bodénès C, Chancerel E, Gailing O et al (2012) Comparative mapping in the *Fagaceae* and beyond with EST-SSRs. BMC Plant Biol 12:153. https://doi.org/10.1186/1471-2229-12-153

Bouffartigue C, Debille S, Fabreguettes O et al (2020) Two main genetic clusters with high admixture between forest and cultivated chestnut (*Castanea sativa* Mill.) in France. Ann for Sci 77:74. https://doi.org/10.1007/s13595-020-00982-w

Buck EJ, Hadonou M, James CJ et al (2003) Isolation and characterization 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

Casasoli M, Derory J, Morera-Dutrey C et al (2006) Comparison of quantitative trait loci for adaptive traits between oak and chestnut based on an expressed sequence tag consensus map. Genetics 172:533–546. https://doi.org/10.1534/genetics.105.048439

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Durand J, Bodénès C, Chancerel E et al (2010) A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genom 11:570. https://doi.org/10.1186/1471-2164-11-570

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

García C, Guichoux E, Hampe A (2018) A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. turbinata). Tree Genet Genomes 14:87. https://doi.org/10.1007/s11295-018-1301-x

Gonthier P, Robin C (2019) Diseases. In: Beccaro G, Alma A, Bounous G, Gomes-Laranjo J (eds) The chestnut handbook: crop and forest management. CRC Press, Boca Raton, pp 297–315

Guichoux E, Lagache L, Wagner S et al (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

Larue C (2021) Development of SNP markers for the identification of chestnut species, hybrids and varieties. Portail Data INRAE, V2, UNF:6:eJQwJJ12X/74swiSKl9V2A== [fileUNF]. https://doi.org/10.15454/XEMDLD

Larue C, Austruy E, Basset G, Petit RJ (2021a) Revisiting pollination mode in chestnut (*Castanea* spp.): an integrated approach. Bot Lett. https://doi.org/10.1080/23818107.2021.1872041

Larue C, Barreneche T, Petit RJ (2021b) Efficient monitoring of phenology in chestnuts. Sci Hortic 281:109958. https://doi.org/10.1016/j.scienta.2021.109958

Laurent B, Larue C, Chancerel E et al (2020) Microhaplotype genotyping-by-sequencing of 98 highly polymorphic markers in three chestnut tree species. Conserv Genet Resour. https://doi.org/10.1007/s12686-020-01157-5

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

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

Miller MR, Dunham JP, Amores A et al (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248. https://doi.org/10.1101/gr.5681207

Nishio S, Takada N, Terakami S et al (2019) Estimation of effective pollen dispersal distance for cross-pollination in chestnut orchards by microsatellite-based paternity analyses. Sci Hortic 250:89–93. https://doi.org/10.1016/j.scienta.2019.02.037

Nishio S, Takada N, Terakami S et al (2021) Genetic structure analysis of cultivated and wild chestnut populations reveals gene flow from cultivars to natural stands. Sci Rep 11:240. https://doi.org/10.1038/s41598-020-80696-1

Nunziata A, Ruggieri V, Petriccione M, De Masi L (2020) Single nucleotide polymorphisms as practical molecular tools to support European chestnut agrobiodiversity management. Int J Mol Sci 21:4805. https://doi.org/10.3390/ijms21134805

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

Pereira-Lorenzo S, Bischofberger Y, Conedera M et al (2020) Reservoir of the European chestnut diversity in Switzerland. Biodivers Conserv 29:2217–2234. https://doi.org/10.1007/s10531-020-01970-2

Pereira-Lorenzo S, Ramos-Cabrer AM, Barreneche T et al (2017) Database of European chestnut cultivars and definition of a core collection using simple sequence repeats. Tree Genet Genomes 13:114. https://doi.org/10.1007/s11295-017-1197-x

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Santos C, Nelson CD, Zhebentyayeva T et al (2017) First interspecific genetic linkage map for *Castanea sativa* x *Castanea crenata* revealed QTLs for resistance to *Phytophthora cinnamomi*. PLoS ONE 12:e0184381. https://doi.org/10.1371/journal.pone.0184381

Stout AB (1926) Why are chestnuts self-fruitless? J N Y Bot Gard 27:154–158

Sun Y, Lu Z, Zhu X, Ma H (2020) Genomic basis of homoploid hybrid speciation within chestnut trees. Nat Commun 11:3375. https://doi.org/10.1038/s41467-020-17111-w

Xiong H, Zou F, Guo S et al (2019) Self-sterility may be due to prezygotic late-acting self-incompatibility and early-acting inbreeding depression in Chinese chestnut. J Am Soc Hortic Sci 144:172–181. https://doi.org/10.21273/JASHS04634-18