ANNALS of SILVICULTURAL RESEARCH

Log traceability and supply-chain verification by DNA markers in Fagus sylvatica L. in Italy

Journal:	*Annals of Silvicultural Research*
Manuscript ID	ASR-2022-0009.R1
Manuscript Type:	Research Paper
Date Submitted by the Author:	05-Jul-2022
Complete List of Authors:	Teani, Angela; CREA, Research Centre for Forestry and Wood; CREA Research centre for forestry and wood de Dato, Giovanbattista; CREA, Research Centre for Forestry and Wood
Keywords:	traceability, Fagus sylvatica, molecular markers

Log traceability and supply-chain verification by DNA markers in *Fagus sylvatica* L. in Italy

Abstract

Verifying the match of a timber log with its population of origin represents a meaningful improvement for the preservation and valorisation of the wood chain.

In this study, we tested the applicability of the molecular approach for tracing the supply chain of timber of European beech (*Fagus sylvatica* L.) to check if the declared source of timber truly complied with the declared origin.

Samples were collected in four sawmills and in the declared forest standing populations. Different assignment tests for the "unknown" timber were used. Ordination pattern and STRUCTURE analyses identified three clusters and one case of mismatch between the analysed wood and the declared source. The Bayesian method (GeneClass2) and the machine learning approach (AssignPop) assigned 42-72% of the cases to the proper source, with higher uncertainty in one site.

This study permitted to detect the lack of congruence of the declared wood source with the actual logging site by DNA markers, screening sampling reference populations (standing trees) and timber from sawmills. Thus, the application of DNA markers confirmed to be able to find failure in the wood supply chain. This application could act as a benchmark to further develop an independent supply-chain verification system that should be implemented as part of the inventory process.

Keywords: Traceability, *Fagus sylvatica*, molecular markers, value-chain.

Introduction

Timber certification consists of a statement (certificate) attesting to the origin of raw wood material and its status and/or qualifications. Certification can serve different purposes. It can be used to demonstrate compliance of timber with standards related to forest management or environmental issues. Furthermore, it can provide information about the forest of origin, which is not generally disclosed by the producer or manufacturer (Lowe and Cross 2011). This second use case requires several control points

covering the supply chain of domestic and export markets and has implications for the process's applicability. Controlling the origin of wood is frequently one of the primary requirements for avoiding trading timber that was illegally logged. A related problem is the certification of timber logged outside a registered concession or within a protected area (Lawson and MacFaul 2010).

Illegal logging and trade in illegal timber and wood products cause economic and ecological problems, as they create market disadvantages for products sourced from legal and sustainable forestry. Verifying the origin of a timber log with its declared population of origin would have positive consequences in terms of preserving and valorising the wood chain.

Isotopic methods can be applied to differentiate wood samples originating from different geographical origins, since wood biomass incorporates elements (mainly carbon, hydrogen, oxygen and nitrogen) with specific signatures based on the environment where it was formed. In practice, this approach can be impacted by the variability of local environmental conditions (Dormontt et al. 2015), such as altitude, distance from the shoreline, rainfall, soil processes, geology, etc. An individual, absolute isotopic "fingerprint" is difficult to detect, and it is possible that samples from different geographical origins could show similar isotope discrimination. There could also be cases that result in a discrepancy between individuals from the same population or even between different tissues from the same individual (Lowe and Cross 2011, Dormontt et al. 2015, Vlam et al. 2018).

Genetic methods based on the analysis of shared ancestry or frequency of alleles can provide information on the geographic region of origin within species through population genetics or phylogeographic analyses. Fingerprinting can even identify individuals or associate seized wood material to its stump (Dormontt et al. 2015). The ability of population genetic approaches and phylogeographic analyses to determine the geographic provenance of trees is due to the existence of a spatial genetic structure within species, which can usually be found at different spatial scales. In this regard, molecular markers, such as chloroplast DNA (cpDNA) and nuclear microsatellite (nSSR) markers and RAPD (random amplified polymorphic DNA) can identify the geographical structure of populations (Deguilloux et al. 2002, Degen et al. 2013). Additionally, species-specific nSSRs can be used to generate DNA profiling databases for individual/species

identification (Jolivet and Degen 2012, Nithaniyal et al. 2014). Most of the studies have been focussed on the timber of tropical South America, Africa and Malaysia (Degen et al. 2001, Tnah et al. 2009, Tnah et al. 2010, Degen et al. 2017, Chaves et al. 2018, Finch et al. 2020, Honorio Coronado et al. 2020). Furthermore, tracing single individual and species identification is a frequent issue that has been solved by DNA fingerprinting and DNA barcoding (Degen et al. 2013, Nithaniyal et al. 2014). At the European level, significant efforts have been made for broadleaf species, such as the oaks *Quercus robur* L.and *Q. petraea* (Matt.) Liebl (Blanc-Jolivet and Liesebach 2015) - but also other oak species in temperate areas of the globe (Schroeder et al. 2016) - and ash (*Fraxinus excelsior* L.); (Tereba et al. 2017).

In this study, we aimed to test the applicability of the molecular approach for tracing the supply chain of timber of European beech (*Fagus sylvatica* L.) to check if the claimed source of timber complied with the real origin. Our main goal was to develop a three-step method that could be fast, reliable, and easy to be applied:

(i) DNA extraction from samples (logs or boards) with a declared origin, and from their reference samples (cores or leaves) from standing tree populations, tested with different extraction protocols;

(ii) nSSR amplification and fragment analysis;

(iii) genotype assignment.

To our knowledge, this is the first case study in Italy aiming at identifying and tracing logs of commercial interest by a molecular marker technique on a species of wide economic importance.

Material and methods

Collection of raw material

Samples were collected in four sawmills and in the corresponding forest standing populations declared as their source, used as reference.

The standing reference populations were in four geographical regions in Italy (Fig. 1; Tab. S1).

[Here the Fig.1]

Two populations were in the Alps in northern Italy (Val di Susa, VDS, number of samples n=23, and Sella Nevea, SNV, n=21). One population was in the central Apennines (Simbruini, SIM, n=31), and another was in the southern Apennines (Serra San Bruno, SSB, n=34). Sampled material consisted of young leaves or wood microcores collected from adult trees at least 100 m apart. The samples were immediately placed in individual bags with silica gel pearls and stored until DNA extraction.

Samples collected in sawmills consisted in wood cores extracted from boards (or logs). According to the declared source, we named the samples collected in the sawmills as VDSt (n = 28), SNVt (n = 25), SIMt (n = 14), SSBt (n = 26). The differences in sample size were due to the availability of wood material and to the effective samples that were amplified.

DNA extraction

Extraction from leaves was performed on dry material using a commercial kit and the protocol recommended by the provider (DNeasy Plant mini kit ®, Qiagen).

Extraction from wood was performed testing two commercial kits:

(i) DNeasy Plant mini kit ® (Qiagen) with the addiction of Polyvinylpyrrolidone (PVP) 2.5% w/v in the AP1 buffer and an extended incubation time of 1 hour;

(ii) DNeasy PowerPlant ® Pro kit (Qiagen).

Extraction yield was assessed spectrophotometrically (Eppendorf Biophotometer®) and quality was assessed by the ratio of absorbances at 260 nm and 280 nm.

Marker selection and laboratory analyses

Microsatellite markers were selected among the most informative ones available in the literature (Weising and Gardner 1999, Pastorelli et al. 2003, Sebastiani et al. 2004), taking into consideration the length of the amplicon (Tab. 1).

[Here the Tab.1]

Genomic DNA was amplified using a master mix designed for multiplexed microsatellites (Type-it ® Microsatellite PCR Kit, Qiagen) with the following protocol: 2X Type-it Multiplex PCR Master mix, 10x primer mix (2 mM of each primer), DNA 15 ng, with water added up to a total volume of 25 ml. Amplification was performed in a thermal cycler (Eppendorf Mastercycler pro) under the following conditions: 95 °C for 5 minutes; 3-step cycling (28 for leaves, 32 for wood) at 95 °C for 30 seconds, 60 °C for 1.5 minutes, and 72 °C for 30 seconds; final elongation 60 °C for 30 minutes. Two fluorescently labelled primer (Life Technologies) mixes were prepared for nSSR: Mix 1 (FS1-15 (VIC®), FS3-04 (PET), FCM5 (FAM)) and Mix 2 (FS1-03 (NED), FS1-11 (VIC®), FS4-46 (FAM)).

PCR products were diluted to a concentration of 1:60, and 1 ml of each sample was added to a mixture of 15 ml Hi-Di ® Formamide (Applied Biosystems) and 0.2 ml size standard 500 LIZ ® (Applied Biosystems). After denaturation at 95 °C for 5 minutes, samples were run in a capillary sequencer (ABI PRISM 310, Life Technologies). Results were analysed by GeneMapper® Software (Applied Biosystems), and allelic profiles were scored by automatic binning and visual checking.

Statistical analyses

DNA extraction protocols comparison

Differences in the yields and quality of DNA extracted from wood between the two commercial kits with the modified protocols, were tested with a paired z-test.

Genetic diversity and markers' evaluation

The numbers of observed (N_A) and effective (N_E) alleles as well as expected heterozygosity (H_E) (Nei 1973) were calculated using GeneAlEx v6.5 (Peakall and Smouse 2012, Peakall and Smouse 2006). The fixation index (F) at each locus was obtained by computing a hierarchical AMOVA using Arlequin v3.5.2.2 (Excoffier and Lischer 2010), and its statistical significance was determined by a non-parametric approach using 1,000 permutations Tab. 1). Null allele frequency (*f*na, Tab. 1) for each locus was estimated via FreeNA (Chapuis and Estoup 2007).

We assessed the statistical power of our markers to detect genetic differentiation among populations given the observed genetic diversity and sample sizes using theprogram

POWSIM v4.1 (Ryman and Palm 2006). POWSIM simulates genetic drift in populations of effective size Ne over a specified number of generations t to produce a level of genetic differentiation described by:

$Fst = 1 - (1 - 1/2N_e)^t$ (eq. 1)

The given FST (resulting from the AMOVA, Tab. S2) may be obtained through many combinations of Ne and t, with care to maintain overall allele frequencies reasonably similar to those of the base population, and to prevent excessive loss of low frequency alleles. We set the program with Ne = 800, t = 50. This combination of Ne and t was also selected to simulate an infinitely large base population drifting over a prolonged period. Each set of populations is then sampled, a test for genetic homogeneity is applied, and the proportion of significant outcomes (χ^2 and Fisher test, $p<0.05$) represents the estimate of power (ranging from low power 0 to the maximum of 1) the probability of detecting the minimum amount of genetic differentiation) under the actual conditions. Simulation scenarios were completed with 999 replications.

Ordination pattern and genotype assignment

To delineate the major ordination pattern of the reference material collected from the standing populations and the unknown material using a multivariate approach, a Discriminant Analysis of Principal Components (DAPC) was performed using the package adegenet (v2.1.3) in R (v4.0, R Development Core Team 2020). The method transforms the genetic data into principal components and then uses k-means clustering to define groups of individuals to minimise within-group variation while maximising among-group variation. We kept axes for principal component analysis that explained about 90% of the variation, and k was selected based on the Bayesian Information Criterion (BIC).

Genetic assignment methods aim to assign or exclude reference populations as possible origins of individuals. Thus, they can infer whether wood samples from sawmills did or did not originate from the reputed forest concession.

Individual assignment tests were performed using the Bayesian approach in STRUCTURE (Pritchard et al. 2000) and GeneClass2 (Piry et al. 2004). A machine-learning function implemented in the R package assignPOP (v1.2; Chen et al. 2016) was also applied.

STRUCTURE was used to infer the populations' genetic structure and assign each individual to a population using the origin of the reference samples to assist clustering. Twenty independent runs were performed by setting the number of groups between 1 and 10, each run consisting of a burn-in period of 10,000 iterations followed by 100,000 Monte Carlo Markov Chain (MCMC) iterations, assuming an admixture model and correlated allele frequencies. The most likely number of groups (K) was chosen based on the ΔK statistic, analysed by the Evanno method (Evanno et al. 2005), using the Structure Harvester (Earl and vonHoldt 2012). The runs were averaged with CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) and graphically represented using DISTRUCT (Rosenberg 2003)

GENECLASS2 uses jack-knifing (leave-one-out) as its default method to assign individuals to the source population. The simulation algorithm generates population samples of the same size as the reference population sample. The assignment criterion is then computed for each individual from the newly simulated population minus itself (leave-one-out procedure). The program iterates until the total number of simulated assignment criterion values is reached. We set an assignment threshold score of 0.05. Computation was executed according to the Bayesian method (Rannala and Mountain 1997, Cornuet et al. 1999), enabling probability computation through a resampling algorithm (1,000 simulated individuals).

In AssignPOP, a set of individuals from source populations were used as the training dataset, with the remaining individuals allocated to the test dataset. Genetic training data are reduced in dimensionality using PCA, which is used to build a user-chosen machine-learning classification function. The classification function is used to assign "test" individuals to source populations. This entire process was automatically repeated 1,000 times by a resampling cross-validation procedure (Monte Carlo) to validate the baseline reference data. When the predictive model built from the baseline was satisfactory, it was applied to perform the assignment test on unknown individuals.

Results

DNA extraction and extraction protocols comparison

Comparison of the two commercial kits used for DNA extraction from wood cores showed an average concentration of 22.2 ng/μl for samples extracted with the Dneasy

PowerPlant Pro kit and 3.9 ng/μl for samples extracted with the Dneasy Plant Mini kit (z value 7.85 versus a critical z value of ±1.95). No significant difference was detected in the quality of the DNA extracted – as determined by the ratio of absorbances 260/280 – with an average value around 1.2 for Dneasy extracted samples, and around 1 for Power plant extracted samples. A z value of -2.01 was obtained versus a critical z value of ±1.95.

Genetic diversity

Based on the analysis of the standing tree populations (reference material), a total of 106 alleles were observed across the 6 loci (Tab. 1). The number of alleles per locus varied from 4 to 24. The frequency of the null allele was on average 10% over all loci. High values of expected heterozygosity were observed for each locus, while the fixation index was very low.

The POWSIM power analysis (Tab. 2) showed that simulated datasets with the same number of markers (six) and comparable genetic diversity (106 loci) would be able to detect significant FST values as low as the true FST = 0.03 (Tab. S2), as resulted from power = 1, $p < 0.05$, in both χ^2 and Fisher tests (Tab. 2). When the drift steps were omitted (t = 0), the power when assuming divergence among populations was 0.036 (χ^2 test) and 0.051 (Fisher test) (Tab. 2).

[Here Tab. 2]

Discriminant analysis of principal components (DAPC)

The cluster analysis of principal components (DAPC; Fig. 2) of the samples derived from both the source sites (standing trees) and the sawmills revealed four main clusters.

[Here Fig.2]

One cluster constituted the VDS site and sawn material VDSt, together with the reference samples from SNV; another two clusters were formed by the material from SIM along with SIMt (site and sawmill), and SSB together with SSBt (site and sawmill), respectively; the sawn material from SNVt formed a single group.

Genotype assignment

Individual assignment by Bayesian methods

The STRUCTURE analysis showed k = 3 as the most likely clustering (Fig. 3). One cluster consisted of the two northern sites VDS and SNV, with their respective sawn material, VDSt and SNVt. The second cluster accounted for the central Italy site SIM and its relative sawn material SIMt. The third cluster constituted the southern SSB site and its sawn material SSBt. In SNVt in particular, a high degree of mixture of all gene pools could be observed.

[Here Fig.3]

GENECLASS2 assignment (Tab. 3(A)) allocated 72% of the samples from SIMt to the SIM reference material. The correct assignment rate was medium for VDSt and SSBt timber: 42% and 49%, respectively assigned to VDS and SSB. Only 8% of the samples of SNVt were assigned to SNV.

Individual assignment by machine learning

The assignment in assignPOP (Tab. 3(B)) resulted in a high to medium percentage of association of the "unknown" individuals (sawn material) from SIMt, VDSt and SSBt with the proper provenance SIM (54%), VDS (48%) and SSB (46%), respectively. For SNVt sawn material there was a low rate of correct assignment (28%), with an almost equal probability of assignment to the other origins.

[Here Tab. 3]

Discussion

This work aimed to develop a reliable method to genetically verify the congruence of the declared wood source with the actual logging site (reference sample). We used DNA markers to trace timber along the wood supply chain, sampling reference populations (standing trees) and timber from sawmills.

Molecular markers are useful in timber tracking, as the high levels of polymorphism can be used to sufficiently differentiate individuals and provide enough information for

identifying population origin (Lowe et al. 2010, Ogden and Linacre 2015). Generally, microsatellites with amplified fragments usually not larger than 500 bps, as those used in the current study, can be successfully used in this analysis (Rachmayanti et al. 2006, Tani et al. 2003). Despite the capability of molecular markers in tracing timber, DNA extraction from wood material, both processed and unprocessed, is challenging for different reasons: a small amount of genetic material is obtained; the DNA extracted from wood can be excessively fragmented; and the DNA is highly contaminated by polymerase inhibitors (Finkeldey et al. 2010). Protocols for the extraction from vegetable samples have been tested to extract good quality and adequate amount of DNA from wood. The use of a commercial kit with a standard extraction protocol can decrease the risk of human error and can help in certifying the methodology. Nevertheless, the usual addition of PVP to the standard procedure generally brings satisfactory improvements in terms of DNA yields per gram of dried wood sample (Rachmayanti et al. 2009, 2006, Shabrina et al. 2019), as PVP can effectively reduce the amount of PCR inhibitory substances (e.g., cellulose and lignin). Thus far, however, no single method can be applied to all species, likely for reasons related to the variable biochemical composition of plant tissues of different species. The fact that the PVP addition did not enhance the DNA yield and quality in the species under investigation in this study can be ascribed to this case. Conversely, when using a kit explicitly aimed at processing difficult samples, even without additional PVP, an elevated quantity of fragments in the range expected for the selected markers was produced, resulting in an adequate substrate for further amplification.

Factors influencing the isolation success have been found to include the length of the DNA fragments, how much the wood has been processed, which part of the wood has been sampled, and the wood's age. Molecular markers used to identify timber origin must be short and simple sequences, as a negative correlation has been observed between amplicon size and amplification success (Lowe et al. 2010, Rachmayanti et al. 2009, 2006). Additionally, unprocessed wood results in higher success rates since the thermal treatments and chemicals applied in wood processing are deleterious for DNA integrity (Rachmayanti et al. 2009). Amplification success is higher if DNA is isolated from outer sapwood than older wood (transition zone between sapwood and heartwood and the inner heartwood). However, when sampling from outer parts, it is crucial to avoid surface

tissues likely contaminated by other organisms' nucleic acids (Jiao et al. 2015, Rachmayanti et al. 2006). In general, the more the material is aged, the more fragmented DNA can be extracted (Deguilloux et al. 2002, Dumolin-Lapegue et al. 1999, Jiao et al. 2015, Speirs et al. 2009, Tani et al. 2003). In this study, the unknown material consisted of recently sized logs (SIMt) or recently sized tables sampled in sawmills. Only material from SSBt was aged and sampled from splints. However, the extraction, even from this more-aged material, did not show evident issues, as usually reported in studies (Deguilloux et al. 2002, Jiao et al. 2015, 2014) dealing with ancient wood.

Typically, in natural forests, a genetic structure with a clear differentiation can be observed at local and regional spatial scales. In this study, the analysis of the sampled standing tree populations matched the expected Italian beech populations' spatial variability. In fact, we could detect an Alpine cluster and two different Apennine groups, in agreement with the picture widely reported in previous papers (Emiliani et al. 2004, Magri et al. 2006, Vettori et al. 2004), mainly caused by past glacial periods and the repeated retreat-colonisation events of tree populations (Hewitt 2004, Petit et al. 2003). The presence of this well-defined genetic structure of beech populations in Italy made it possible to properly detect the provenance of the unknown material with good to high probability in three out of four cases, with the timber from Sella Nevea (SNVt) proving the exception. We observed correct assignment probabilities ranging from 42% to 72%. These values are comparable to assignment percentages evaluated by nSSR for a wide set of species (Degen et al. 2013) ranging from 50% to 70% when working at the population level. Higher rates have been reported in studies attempting to determine the origin at the country level. It is unlikely that the observed low assignment of SNVt to the proper source is caused by the low number of markers used. In fact, for the given level of samples, differentiation and alleles, our nuclear markers had a significantly low probability of error in rejecting the null hypothesis of the lack of differentiation. Nor did the use of chloroplast markers (results not shown), which confirmed the regional differentiation, result in a more detailed assignment.

The assignment can also depend on the applied method (Degen et al. 2017). In fact, in this study, we could observe that the PCoA and STRUCTURE analyses revealed a possible overlapping of VDSt with both VDS and SNV. Thus, it was impossible to assign

SNVt to a specific geographic location in the Alps because there is no differentiation between individuals from different reference locations in this genetic cluster, indicating gene flow. Differently, the Bayesian method (GeneClass2) and the machine learning approach (AssignPop) showed to be complementary and could allocate better unknown samples to the site of origin. Both the Bayesian method and the machine learning approach detected a higher percentage for the putative origin site for VDSt, SIMt and SSBt. In the case of SNVt, GeneClass2 allocated almost the same percentage to VDS and SSB, while AssignPop did not detect differences in the assignment distribution. According to some studies, the low success of assignment can be related to high levels of intrapopulation diversity, as the more alleles the tested population has, the higher the chances of a population having alleles in common with other populations (Chaves et al. 2018). However, this was not the case for SNV, which recorded the lowest allelic richness and genetic diversity. The low probability observed in the case of Sella Nevea might be due to an error in the indication of the provenance site, but we cannot totally exclude the lack of information caused by missing data that led to the omission of some samples in the reference population. Since for this kind of investigation, based on the probability that a genotype could belong to a specific provenance, the availability of an exhaustive and geographically widespread reference database is crucial (Cavers et al. 2005, Ogden and Linacre 2015), this uncertainty could have been overcome by extending the reference data sampled on site and to bordering areas and possibly in neighbouring countries to exclude the provenance of material from abroad.

Conclusions

This study permitted to detect the lack of congruence of the declared wood source with the actual logging site by the application of DNA markers sampling reference populations (standing trees) and timber from sawmills. Thus, the application of DNA markers confirms to be able to find failure in the wood supply chain.

Regrettably, we could not extend the analysis to other sites. For this reason, we reckon that genetic inventories with extensive and systematic samples across the whole species distribution area are the basis for identifying the country of timber origin. Higher spatial resolution is needed to control the source of timber at the level of logging concessions.

A solid traceability system should rely on different types of information. Implementing regular genetic monitoring in forest inventories would be desirable since it would integrate with collected data (e.g., dendrometric data, stand structure, etc.) and extend the reference database. Furthermore, sharing existing data on an accessible database of reference materials to enable effective tests would be a primary instrument for stakeholders involved in the certification process and help ensure a reliable and applicable certification system.

Acknowledgements

This research was funded by "PRIN bando 2015 – La filiera corta nel settore biomasse-legno: approvvigionamento, tracciabilità, certificazione e sequestro di Carbonio. Innovazioni per la bioedilizia e l'efficienza energetica", financed by the Ministry of University and Research (MIUR) – D.M. 1826 del 20/9/2016.

References

Blanc-Jolivet C., Liesebach M. 2015 - *Tracing the origin and species identity of Quercus robur and Quercus petraea in Europe: a review.* Silvae Genetica 64 (1–6): 182–93. doi:10.1515/sg-2015-0017

Cavers S., Degen B., Caron H., Lemes M. R., Margis R., Salgueiro F., Lowe A. J. 2005 - *Optimal sampling strategy for estimation of spatial genetic structure in tree populations.* Heredity 95 (4): 281–89. doi:10.1038/sj.hdy.6800709

Chapuis M.-P., Estoup A. 2007 - *Microsatellite Null Alleles and Estimation of Population Differentiation.* Molecular Biology and Evolution 24 (3): 621–31. doi:10.1093/molbev/msl191

Chaves C. L., Degen B., Pakull B., Mader M., Honorio E., Ruas P., Tysklind N., Sebbenn A. M. 2018 - *Assessing the Ability of Chloroplast and Nuclear DNA Gene Markers to Verify the Geographic Origin of Jatoba (*Hymenaea courbaril *L.) Timber.* Journal of Heredity 109 (5): 543–52. doi:10.1093/jhered/esy017

Chen Y., Dai X., Hou J., Guan H., Wang Y., Li Y., Yin T. 2016 - *DNA fingerprinting of oil camellia cultivars with SSR markers.* Tree Genetics & Genomes 12 (1): 7. doi:10.1007/s11295-015-0966-7

Cornuet J.-M., Piry S., Luikart G., Estoup A., Solignac M. 1999 - *New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals.* Genetics 153 (4): 1989–2000. doi:10.1093/genetics/153.4.1989

Degen B., Ward S. E., Lemes M. R., Navarro C., Cavers S., Sebbenn A. M. 2013 - *Verifying the geographic origin of mahogany (*Swietenia macrophylla *King) with DNA-fingerprints.* Forensic Science International: Genetics 7 (1): 55–62. doi:10.1016/j.fsigen.2012.06.003

Degen B., Blanc-Jolivet C., Stierand K., Gillet E. 2017 - *A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic*

origin. Forensic Science International: Genetics 27: 132–41. doi:10.1016/j.fsigen.2016.12.011

Degen B., Caron H., Bandou E., Maggia L., Chevallier M. H., Leveau A., Kremer A. 2001 - *Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs.* Heredity 87 (4): 497–507. doi:10.1046/j.1365-2540.2001.00942.x

Deguilloux M., Pemonge M., Petit R. 2002 - *Novel perspectives in wood certification and forensics: dry wood as a source of DNA.* Proceedings of the Royal Society of London. Series B: Biological Sciences 269 (1495): 1039–46. doi:10.1098/rspb.2002.1982

Dormontt E. E., Boner M., Braun B., Breulmann G., Degen B., Espinoza E., Gardner S., Guillery P., Hermanson J. C., Koch G., Lee S. L., Kanashiro M., Rimbawanto A., Thomas D., Wiedenhoeft A. C., Yin Y., Zahnen J., Lowe A. J. 2015 - *Forensic timber identification: It's time to integrate disciplines to combat illegal logging.* Biological Conservation 191: 790–98. doi:10.1016/j.biocon.2015.06.038

Dumolin-Lapegue S., Pemonge M.-H., Gielly L., Taberlet P., Petit R. J. 1999 - *Amplification of oak DNA from ancient and modern wood.* Molecular Ecology 8 (12): 2137–40. doi:10.1046/j.1365-294x.1999.00788.x

Earl D. A., vonHoldt B. M. 2012 - *STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method.* Conservation Genetics Resources 4 (2): 359–61. doi:10.1007/s12686-011-9548-7

Emiliani G., Paffetti D., Vettori C., Giannini R. 2004 - *Geographic distribution of genetic variability of* Fagus sylvatica *L. in Southern Italian populations.* Forest Genetics 11 (3): 231–37.

Evanno G., Regnaut S., Goudet J. 2005 - *Detecting the number of clusters of individuals using the software structure: a simulation study.* Molecular Ecology 14 (8): 2611–20. doi:10.1111/j.1365-294X.2005.02553.x

Excoffier L., Lischer H. E. L. 2010 - *Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.* Molecular Ecology Resources 10 (3): 564–567. doi:10.1111/j.1755-0998.2010.02847.x

Finch K. N., Cronn R. C., Ayala Richter M. C., Blanc-Jolivet C., Correa Guerrero M. C., de Stefano Beltrán L., García-Dávila C. R., Honorio Coronado E. N., Palacios-Ramos S., Paredes-Villanueva K., Jones F. A. 2020 - *Predicting the geographic origin of Spanish Cedar (*Cedrela odorata *L.) based on DNA variation.* Conservation Genetics 21 (4): 625–639. doi:10.1007/s10592-020-01282-6

Finkeldey R., Leinemann L., Gailing O. 2010 - *Molecular genetic tools to infer the origin of forest plants and wood.* Applied Microbiology and Biotechnology 85 (5): 1251–58. doi:10.1007/s00253-009-2328-6

Hewitt G. M. 2004 - *Genetic consequences of climatic oscillations in the Quaternary.* Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359 (1442): 183–195. doi:10.1098/rstb.2003.1388

Honorio Coronado E. N., Blanc-Jolivet C., Mader M., García-Dávila C. R., Aldana Gomero D., del Castillo Torres D., Flores Llampazo G., Hidalgo Pizango G., Sebbenn A. M., Meyer-Sand B. R. v, Paredes-Villanueva K., Tysklind N., Troispoux V., Massot M., Carvalho C., de Lima H. C., Cardoso D., Degen B. 2020 - *SNP Markers as a Successful Molecular Tool for Assessing Species Identity and Geographic Origin of Trees in the Economically Important South American*

Legume Genus Dipteryx. Journal of Heredity 111 (4): 346–56.

doi:10.1093/jhered/esaa011

Jakobsson M., Rosenberg N. A. 2007 - *CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure.* Bioinformatics 23 (14): 1801–06.

doi:10.1093/bioinformatics/btm233

Jiao L., Liu X., Jiang X., Yin Y. 2015 - *Extraction and amplification of DNA from aged and archaeological* Populus euphratica *wood for species identification.* Holzforschung 69 (8): 925–931. doi:10.1515/hf-2014-0224

Jiao L., Yin Y., Cheng Y., Jiang X. 2014 - *DNA barcoding for identification of the endangered species* Aquilaria sinensis*: comparison of data from heated or aged wood samples.* Holzforschung 68 (4): 487–94. doi:10.1515/hf-2013-0129

Jolivet C., Degen B. 2012 - *Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon.* Forensic Science International: Genetics 6 (4): 487–493.

doi:10.1016/j.fsigen.2011.11.002

Lawson S., MacFaul L. 2010 - *Illegal Logging and Related Trade: Indicators of the global response*. London. 154 p.

Lowe A. J., Cross H. B. 2011 - *The Application of DNA methods to Timber Tracking and Origin Verificat ion.* IAWA Journal 32 (2): 251–62. doi:10.1163/22941932-90000055

Lowe A. J., Wong K.-N., Tiong Y.-S., Iyerh S., Chew F.-T. 2010 - *A DNA Method to Verify the Integrity of Timber Supply Chains; Confirming the Legal Sourcing of Merbau Timber From Logging Concession to Sawmill.* Silvae Genetica 59 (1–6): 263–68. doi:10.1515/sg-2010-0037

Magri D., Vendramin G. G., Comps B., Dupanloup I., Geburek T., Gömöry D., Lataowa M., Litt T., Paule L., Roure J. M., Tantau I., van der Knaap W. O., Petit R. J., de Beaulieu J. 2006 - *A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences.* New Phytologist 171 (1): 199–221. doi:10.1111/j.1469-8137.2006.01740.x

Nei M. 1973 - *Analysis of Gene Diversity in Subdivided Populations.* Proceedings of the National Academy of Sciences 70 (12): 3321–3323. doi:10.1073/pnas.70.12.3321

Nithaniyal S., Newmaster S. G., Ragupathy S., Krishnamoorthy D., Vassou S. L., Parani M. 2014 - *DNA Barcode Authentication of Wood Samples of Threatened and Commercial Timber Trees within the Tropical Dry Evergreen Forest of India.* PLoS ONE 9 (9): e107669. doi:10.1371/journal.pone.0107669

Ogden R., Linacre A. 2015 - *Wildlife forensic science: A review of genetic geographic origin assignment.* Forensic Science International: Genetics 18: 152–159. doi:10.1016/j.fsigen.2015.02.008

Pastorelli R., Smulders M. J. M., VAN'T Westende W. P. C., Vosman B., Giannini R., Vettori C., Vendramin G. G. 2003 - *Characterization of microsatellite markers in* Fagus sylvatica *L. and* Fagus orientalis *Lipsky.* Molecular Ecology Notes 3 (1): 76–78. doi:10.1046/j.1471-8286.2003.00355.x

Peakall R., Smouse P. E. 2006 - *genalex 6: genetic analysis in Excel. Population genetic software for teaching and research.* Molecular Ecology Notes *6* (1): 288–295. doi:10.1111/j.1471-8286.2005.01155.x

Peakall R., Smouse P. E. 2012 - *GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update.* Bioinformatics 28 (19): 2537–2539. doi:10.1093/bioinformatics/bts460

Petit R. J., Aguinagalde I., de Beaulieu J.-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Müller-Starck G., Demesure-Musch B., Palmé A., Martín J. P., Rendell S., Vendramin G. G. 2003 - *Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity.* Science 300 (5625): 1563–65. doi:10.1126/science.1083264

Piry S., Alapetite A., Cornuet J.-M., Paetkau D., Baudouin L., Estoup A. 2004 - *GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection.* Journal of Heredity 95 (6): 536–39. doi:10.1093/jhered/esh074

Pritchard J. K., Stephens M., Donnelly P. 2000 - *Inference of Population Structure Using Multilocus Genotype Data.* Genetics 155 (2): 945–959. doi:10.1093/genetics/155.2.945

Rachmayanti Y., Leinemann L., Gailing O., Finkeldey R. 2006 - *Extraction, amplification and characterization of wood DNA from dipterocarpaceae.* Plant Molecular Biology Reporter 24 (1): 45–55. doi:10.1007/BF02914045

Rachmayanti Y., Leinemann L., Gailing O., Finkeldey R. 2009 - *DNA from processed and unprocessed wood: Factors influencing the isolation success.*

Forensic Science International: Genetics 3 (3): 185–92.

doi:10.1016/j.fsigen.2009.01.002

Rannala B., Mountain J. L. 1997 - *Detecting immigration by using multilocus genotypes.* Proceedings of the National Academy of Sciences 94 (17): 9197–201.

doi:10.1073/pnas.94.17.9197

Rosenberg N. A. 2003 - *distruct: a program for the graphical display of population structure.* Molecular Ecology Notes 4 (1): 137–138. doi:10.1046/j.1471-8286.2003.00566.x

Ryman N., Palm S. 2006 - *POWSIM: a computer program for assessing statistical power when testing for genetic differentiation.* Molecular Ecology Notes 6 (3): 600–02. doi:10.1111/j.1471-8286.2006.01378.x

Schroeder H., Cronn R., Yanbaev Y., Jennings T., Mader M., Degen B., Kersten B. 2016 - *Development of Molecular Markers for Determining Continental Origin of Wood from White Oaks (*Quercus *L. sect. Quercus).* PLOS ONE 11 (6): e0158221.

doi:10.1371/journal.pone.0158221

Sebastiani F., Carnevale S., Vendramin G. G. 2004 - *A new set of mono- and dinucleotide chloroplast microsatellites in Fagaceae.* Molecular Ecology Notes 4 (2): 259–261. doi:10.1111/j.1471-8286.2004.00635.x

Shabrina H., Siregar U. J., Matra D. D., Kamiya K., Siregar I. Z. 2019 - *Short Communication: DNA extraction from stored wood of* Falcataria moluccana *suitable for barcoding analysis.* Biodiversitas Journal of Biological Diversity 20 (6). doi:10.13057/biodiv/d200635

Speirs A. K., McConnachie G., Lowe A. J. 2009 - *Chloroplast DNA from 16th Century Waterlogged Oak in a Marine Environment: Initial Steps in Sourcing the Mary Rose Timbers.* In "Archaeological Science Under a Microscope: Studies in Residue and Ancient DNA Analysis in Honour of Thomas H. Loy". ANU Press, M. Haslam, G. Robertson, A. Crowther, S. Nugent, & L. Kirkwood (Eds.).

Tani N., Tsumura Y., Sato H. 2003 - *Nuclear gene sequences and DNA variation of* Cryptomeria japonica *samples from the postglacial period.* Molecular Ecology 12 (4): 859–868. doi:10.1046/j.1365-294X.2003.01779.x

Tereba A., Woodward S., Konecka A., Borys M., Nowakowska J. A. 2017 - *Analysis of DNA profiles of ash (*Fraxinus excelsior *L.) to provide evidence of*

illegal logging. Wood Science and Technology 51 (6): 1377–1387. doi:10.1007/s00226-017-0942-5

Tnah L. H., Lee S. L., Ng K. K. S., Faridah Q.-Z., Faridah-Hanum I. 2010 - *Forensic DNA profiling of tropical timber species in Peninsular Malaysia.* Forest Ecology and Management 259 (8): 1436–1446. doi:10.1016/j.foreco.2010.01.017

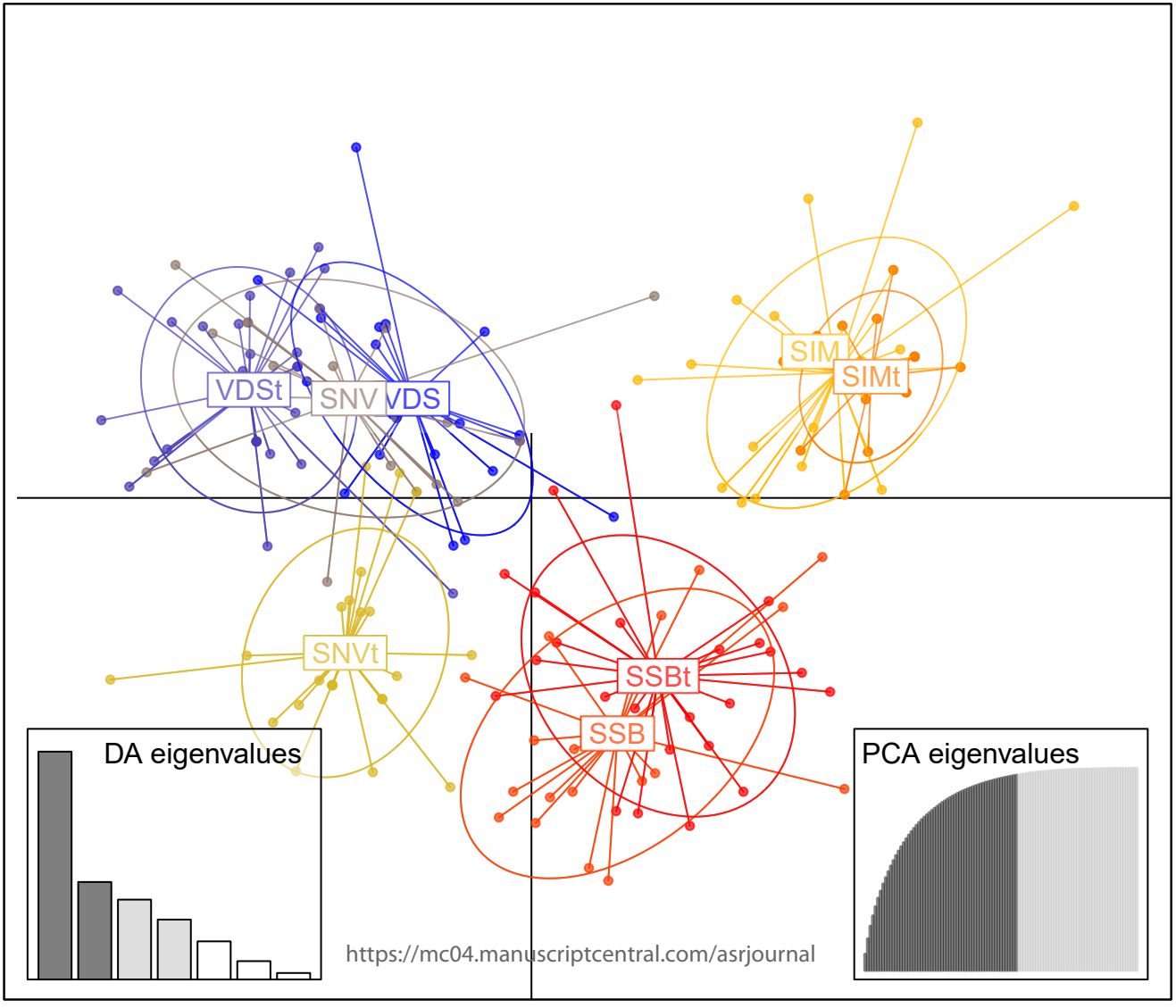
Tnah L. H., Lee S. L., Ng K. K. S., Tani N., Bhassu S., Othman R. Y. 2009 - *Geographical traceability of an important tropical timber (*Neobalanocarpus heimii*) inferred from chloroplast DNA.* Forest Ecology and Management 258 (9): 1918–1923. doi:10.1016/j.foreco.2009.07.029

Vettori C., Vendramin G. G., Anzidei M., Pastorelli R., Paffetti D., Giannini R. 2004 - *Geographic distribution of chloroplast variation in Italian populations of beech (*Fagus sylvatica *L.).* Theoretical and Applied Genetics 109 (1): 1–9. doi:10.1007/s00122-004-1609-9

Vlam M., de Groot G. A., Boom A., Copini P., Laros I., Veldhuijzen K., Zakamdi D., Zuidema P. A. 2018 - *Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature.* Biological Conservation 220: 262–271. doi:10.1016/j.biocon.2018.01.031

Weising K., Gardner R. C. 1999 - *A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms.* Genome 42 (1): 9–19. doi:10.1139/g98-104


SNV
VDS
SIM
SSB
0
200
400 km
N


VDSt
SNV
VDS
SIM
SIMt
SNVt
SSBt
SSB
DA eigenvalues
PCA eigenvalues
https://mc04.manuscriptcentral.com/asrjournal

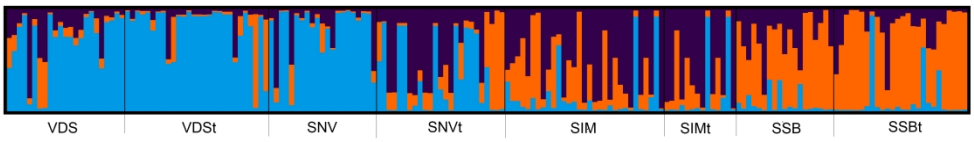


Figure 3 - Assignment per individual based on Bayesian clustering of microsatellite data using STRUCTURE. Individuals are represented by vertical bars. Colours per individual represent percentage assignment to each of the three assumed genetic clusters.

Additional material

Materials and Methods

Genetic diversity and genetic differentiation

To assess genetic diversity at the population level (reference tree populations), within- and among-population genetic parameters were evaluated by different approaches. Expected heterozygosity (HE) (Nei 1973) were calculated using GeneAlEx v6.5 (Peakall & Smouse 2006, Peakall & Smouse 2012; Tab. S1). As an assessment of allelic richness through measuring allele frequencies must consider variation in population sizes, allelic richness (AR) and the number of private alleles (PA) were computed using the rarefaction method with HP-RARE software (Kalinowski 2005; Tab. S1). The inbreeding coefficient (FIS) for each population was obtained by computing a hierarchical AMOVA using Arlequin v3.5.2.2 (Excoffier & Lischer 2010), and its statistical significance was determined by a non-parametric approach using 1000 permutations (Tab. S1; Tab. S2). Population differentiation (FST) was calculated in GeneAlEx v6.5 (Peakall & Smouse 2006, Peakall & Smouse 2012; Tab. S3).

Tab. S1 - Genetic indices computed on reference samples

label	N	AR	PA	HE	FIS
VDS	23	10.3	1.1	0.76	0.20
SNV	21	8.3	1.0	0.73	0.14
SIM	31	10.1	1.3	0.80	0.22
SSB	34	10.8	2.4	0.77	0.20

N = Total number of sampled individuals; AR = allelic richness; PA = number of private alleles with rarefaction; HE = expected unbiased heterozygosity; F = fixation index for each sampled reference population

Tab. S2 – Analysis of molecular variance (AMOVA) of the four reference populations

Source of variation	Df	Variance components	% variation	F-statistic	
Among populations	3	0.035	2.3	Fst	0.03
Among individuals within populations	105	0.292	19.1	Fis	0.19
Within individuals	109	1.206	78.6	Fit	0.21
Total	217	2.431	100		

Tab. S3 – Population pairwise FST. *** indicate significant differences (p<0.05)

	VDS	SNV	SIM	SSB
VDS	-			
SNV	0.003	-		
SIM	0.032 ***	0.046 ***	-	
SSB	0.048 ***	0.068 ***	0.037 ***	-

Tables and captions

Table 1 - Nuclear microsatellite markers used in this study.

Locus	bp	NA	NE	*f*na	HE	F
FS1-15	69–131	22	7.3	0.068	0.89	0.02
FS3-04	192–207	4	1.9	0.002	0.47	0.02
FCM5	260–346	23	9.1	0.147	0.93	0.03
FS1-03	68–120	21	5.0	0.147	0.82	0.01
FS1-11	92–126	12	4.4	0.166	0.81	0.07
FS4-46	170–311	24	6.8	0.109	0.90	0.05

Base pairs (bp), numbers of alleles (Na), effective numbers of alleles (Ne), frequency of null allele (*f*na), expected heterozygosity (He) and fixation index (F)

Table 2 - POWSIM power simulation results for the 6 microsatellite loci indicating proportion of statistical significance ($p<0.05$) to detect population structure among *Fagus sylvatica* locations. t = generations to drift before sampling.

t	FST	Power probability
0	0	0.051
50	0.03	1

Table 3 – Values of A) membership probabilities calculated in GeneClass2 and B) membership probabilities (Standard Deviation) obtained by AssignPop via Monte Carlo assignment, trained using all loci and a set of individuals from source populations.

A)	VDS	SNV	SIM	SSB
VDSt	42	16	11	31
SNVt	34	8	27	31
SIMt	11	2	72	15
SSBt	16	5	30	49

B)

VDSt	0.48 (0.11)	0.22 (0.11)	0.15 (0.12)	0.15 (0.11)
SNVt	0.24 (0.13)	0.28 (0.13)	0.26 (0.14)	0.22 (0.12)
SIMt	0.18 (0.10)	0.09 (0.10)	0.54 (0.12)	0.19 (0.13)
SSBt	0.18 (0.15)	0.09 (0.09)	0.27 (0.18)	0.46 (0.19)