

Microsatellite analysis for differentiation and identification of the source tree of *Fagus orientalis* Lipsky

Parvin Salehi Shanjani^{1*}, Giovanni Giuseppe Vendramin², Mohsen Calagari¹

¹Research Institute of Forests and Rangelands, P.O. Box 13185-116, Tehran, I.R. Iran ²Institute of Plant Genetics, Consiglio Nazionale delle Ricerche (CNR), via Madonna del Piano, I-50019 Sesto Fiorentino, Firenze, Italy

Abstract

The present study describes approaches for the identification of individual beech trees using maternal tissues from their seeds or fruits. Four microsatellite markers were used for genetic analysis of seedlots from *Fagus orientalis* Lipsky, a highly out-crossing tree species. Seeds from 11 single-tree harvests belonging to one population, (7 seeds from each), as well as non-paranchymatic maternal tissues of the seeds (single woody pericarps) were genotyped. No prior information about the genotypes of the mother trees was available for the seedlot samples. Two methods for detection of mother tree genotypes were adopted; (1) analysis of woody pericarps and their applicability for a direct molecular identification of the mother trees; and (2) analysis of maternal half-sib families and usefulness contaminations and inference of the seed parents from the offspring. A comparison of the multi-locus genotypes from pericarps with those of the inferred from the offspring revealed absolute identity. Analysis of the half-sib families revealed that the maternal genotypes can be inferred from offspring genotypes due to codominant Mendelian inheritance of the microsatellites. Differences among the seeds from the same mother tree suggested that paternal genotypes and gene flow have more influence on differentiation. The results of this research show that microsatellite analysis is a suitable means to monitor the number of trees included in commercial seedlot samples and detect seed contaminations.

Keywords: *Fagus orientalis* Lipsky; Hyrcanian forests; Iran; Microsatellite; Half-sib; Pericarp.

INTRODUCTION

Conservation of forest genetic resources is, increasingly, an issue of major national and international concern. Iranian forestry services also have clear policies to restore large areas of degraded land to native forest. However, knowledge needed for the success of these ambitious programs is still inadequate, and considerable further research is required. One problem for any forest restoration program is selection of trees from which to obtain seeds. The Convention on Biological Biodiversity (Rio de Janeiro, Brazil, 1992) emphasized the importance of maintaining intraspecific genetic diversity and evolutionary potential. Consequently, adaptability and maintenance of a broad genetic base must be considered. Genetic variation in a founding population is essential, particularly if restored areas are far from pollen sources, or if the restored area is likely to become a seed source itself. Collection of seeds from few individual seed trees can result in a low effective population size by narrowing the genetic base, inbreeding depression and a decrease in the adaptive evolutionary potential of the population (Blakesley *et al.*, 2002; Barrett and Kohn, 1991). General guidelines for seed collection which consider the capture of biodiversity have been published in a number of texts (eg Schmidt, 2000; Guarino *et al.*, 1995). For provenance or progeny trials, it is recommended to sample 10-20 individuals, which may be increased to 25-50 individuals per population for *ex situ* conservation purposes (Guarino *et al.*, 1995).

*Correspondence to: Parvin Salehi Shanjani, M.Sc.
Tel: +98 2144195901-5, Fax: +98 21 44196575
E-mail: psalehi@rifra-ac.ir

However, these measures run contrary to economical considerations during seed harvest, as gathering seed from only a few trees bearing abundant fruit is less costly.

Microsatellites are highly polymorphic and codominantly inherited, and therefore have a high potential to resolve genetic relatedness (Blouin *et al.*, 1996). Therefore they are the ideal genetic marker to confirm or reject the half-sib family structure within the samples and infer the number of different maternal trees (seed parents) involved (Lexer *et al.*, 1999). The multilocus genotype of the seed pericarp is unambiguously identified with the maternal tree of the seed due to the fact that the pericarp tissue in beech is of maternal origin (Schopmeyer, 1974).

In order to test the applicability of microsatellite markers to directly identify mother trees, this investigation involved a study of the economically and ecologically important forest tree species *Fagus orientalis* Lipsky. The tissues studied were single woody pericarps of the seed (Fig. 1), which were known to be of purely maternal origin. However in some situations there is no available material from the mother tree. Therefore, the other objective of this investigation was to validate the methods of data analysis that require no prior knowledge of the possible parental genotypes. In order to approach these problems a maternal model half-sib family that included the genotype of the mother trees was used to examine 11 seedlot samples.

MATERIALS AND METHODS

Plant material: This study was conducted in the Kordkuy forest (Part of the Hyrcanian forest and located in the Golestan Province, North of Iran) that con-

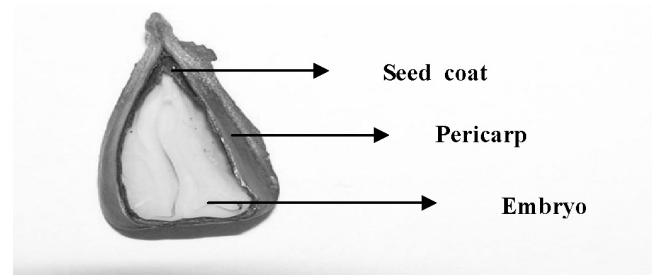


Figure 1. Schematic representation of beech seed.

sisted of mixed beech forest (*F. orientalis*), with an elevation 600 m from sea level. At this site, seeds from 11 mother trees (randomly chosen and separated by at least 30 m) were collected (7 seeds from each mother tree).

The seed pericarp was split open and separated by hand from the seed content (embryo with no endosperm, Figure 1), and collected for later grinding. We do not expect contamination of the pericarp tissue with remains of the seed content, as the whole embryo is separated from the pericarp by a thin seed coat that is cleanly extracted when the pericarp is split open.

DNA extraction and microsatellite genotyping:

DNA was isolated from the embryo and pericarp (100 mg as starting material) using the Nucleospin plant kit (Macherey Nagel, Germany). Four microsatellites (FS1-15, FS1-03, FS1-11 and FS3-04) from Pastorelli *et al.* (2003) were amplified according to the following temperature profile: 5 min of denaturation at 95°C followed by 30 cycles involving 1 min of denaturation at 95°C, 1 min of annealing (Table 1), 1 min of extension at 72°C, with a final extension step of 8 min at 72°C. The PCR was performed in volume of 25 µl using 10 ng of template DNA, 10x Amersham reaction buffer (500 mM KCl, 15 mM MgCl₂ and 100 mM Tris-HCl,

Table 1. Characteristics of the 4 polymorphic nuclear microsatellite markers used to analyse genetic characteristics of seedlot harvests.

Locus	Repeat	Annealing temp. °C	MgCl ₂ concen. mM	No. alleles in <i>Fagus orientalis</i> from Europe	N _a	N _e	Ho	He	Fis	Fit	Fst
FS1-15	(GA) ₂₆	60	2.5	12	3.82	2.53	0.64	0.57	-0.129	0.104	0.206
FS1-03	(GA) ₁₈	60	1.5	12	4.18	2.79	0.72	0.59	-0.21	-0.008	0.167
FS1-11	(GA) ₁₅	63	2.5	9	3.54	2.43	0.43	0.53	-0.167	0.090	0.22
FS3-04	(GCT) ₅ (GTT) ₃ (GCT) ₆	60	1.5	4	1.91	2.36	0.27	0.21	-0.284	-0.059	0.175
Mean					3.36	2.28	0.52	0.48	-0.197	0.032	0.192

N_a: mean number of alleles per locus; N_e: effective number of alleles; Ho: average observed heterozygosity; He: average expected heterozygosity; Fis: Fixation index; Fit: Genetic differentiation among all individuals; Fst: Genetic differentiation among populations.

pH 9.0), MgCl₂ concentration as in Table 1, 0.2 mM dNTPs (Amersham), 0.4 μM of each primer and, 1U of *Taq* DNA polymerase (Amersham). The success of the amplification was confirmed on a 1.4% (w/v) agarose gel. Amplified fragments were then multiplexed by size (mixed two by two) standards (50, 100, 150, 200, 250 and 300 bps) and added to each mix before loading onto a ReproGel Long Read acrylamide gels (Amersham). The gel was run on an automated sequencing machine (Alf Express, Amersham) at 1500V, 60 mA and, 30 W at 55°C. The results of the run were then analyzed with Fragment Manager 1.2 (Amersham).

Data analysis: Amplification reactions from all individuals (embryo/pericarp) were scored (based on diploid genotypes) and the following statistics of genetic variation within seeds/trees were computed as averages over loci using the GENALEX 6 software (Peakal and Smouse, 2006): mean number of alleles per locus (N_a), effective number of alleles (N_e), average observed heterozygosity (H_o) and average expected heterozygosity (H_e) were computed according to Nei (1978). An estimator of Wright's F-Statistics (F_{is} , F_{it} and F_{st} values) was calculated to assess population differentiation (Wright, 1951, 1931). The significance of the F_{st} values was tested using 999 random permutations of the data matrix, a procedure that is not dependent on Hardy-Weinberg equilibrium. Analysis of molecular variance (AMOVA) (Schneider *et al.*, 2000) was used to partition the genetic variation among trees, and among seeds within trees. The significance of each variance component was tested with permutation tests (Excoffier *et al.*, 1992). Principal coordinate analysis (PCoA) based on pairwise F_{st} values was performed to identify the patterns of genetic relationships of seeds and trees (Gower, 1966).

RESULTS

Genotyping of the pericarp tissue for the identification of the source tree: We successfully extracted DNA from the woody pericarp tissue of individual beech seeds and analysed the pericarp using 4 microsatellite loci, FS1-15, FS3-04, FS1-11 and FS1-03. Because pericarp tissue is of maternal origin, its multilocus genotype should be identical to that of the mother tree. This was confirmed for individual mother

trees by comparing the genotypes of both pericarps and embryos of seeds collected from several pure single tree harvests. As expected, the embryo genotypes of seeds collected from the same tree were variable (due to their parental origin), however, the pericarp genotypes were identical to each other (due to maternal origin). The genotyping of seed pericarp thus provides a convenient tool for the unambiguous assignment of maternity for dispersed seeds. Furthermore, our results revealed that a PCR amplification that is routinely used on beech DNA from paranchymatic tissues was successful for pericarps (non-paranchymatic tissue) as well.

Genetic analysis of the maternal half-sib families: Eleven maternal half-sib families (each seven seeds) of *F. orientalis* were genotyped at 4 microsatellite loci. Number of alleles ranged from 1 to 6 per locus with an average of 3.36 alleles per locus. The genotypes of the offsprings were used in identifying reliable and simple methods (1) to confirm or reject the hypothesis of the half-sibling relationship among embryos supplied as single tree harvests; (2) to reconstruct the maternal genotypes of single tree harvests.

For inference of the seed mother from the offspring, when one allele was detected in each offspring individual, it suggested that the mother is homozygous at the analyzed loci. All other alleles at these loci occurred at much lower frequencies and would therefore not fit a segregation ratio of 1:1 as expected for the maternal alleles under a simple codominant inheritance hypothesis. Therefore all other alleles could be unequivocally assigned to the pollen contribution. When more than one allele was presented in each offspring individual, it indicated that the mother tree is heterozygous at the analyzed loci. In each case only one possible pair of alleles constituting the maternal genotype was identified. All other alleles could be excluded on the basis of co-dominant mendelian inheritance: 1) an allele can only be of maternal origin if there is a second allele that is present whenever the first one is absent; (2) an allele that is present in a homozygous state in at least one offspring individual must be present in the mother tree; (3) if two different homozygotes are found among the offsprings then no other allele can be of maternal origin (Fig. 2A and B).

Unrelated individuals within each of seedlot harvests were detected among the half-sibs (Figure 2B, seed NO. 2 not have any alleles from the mother tree).

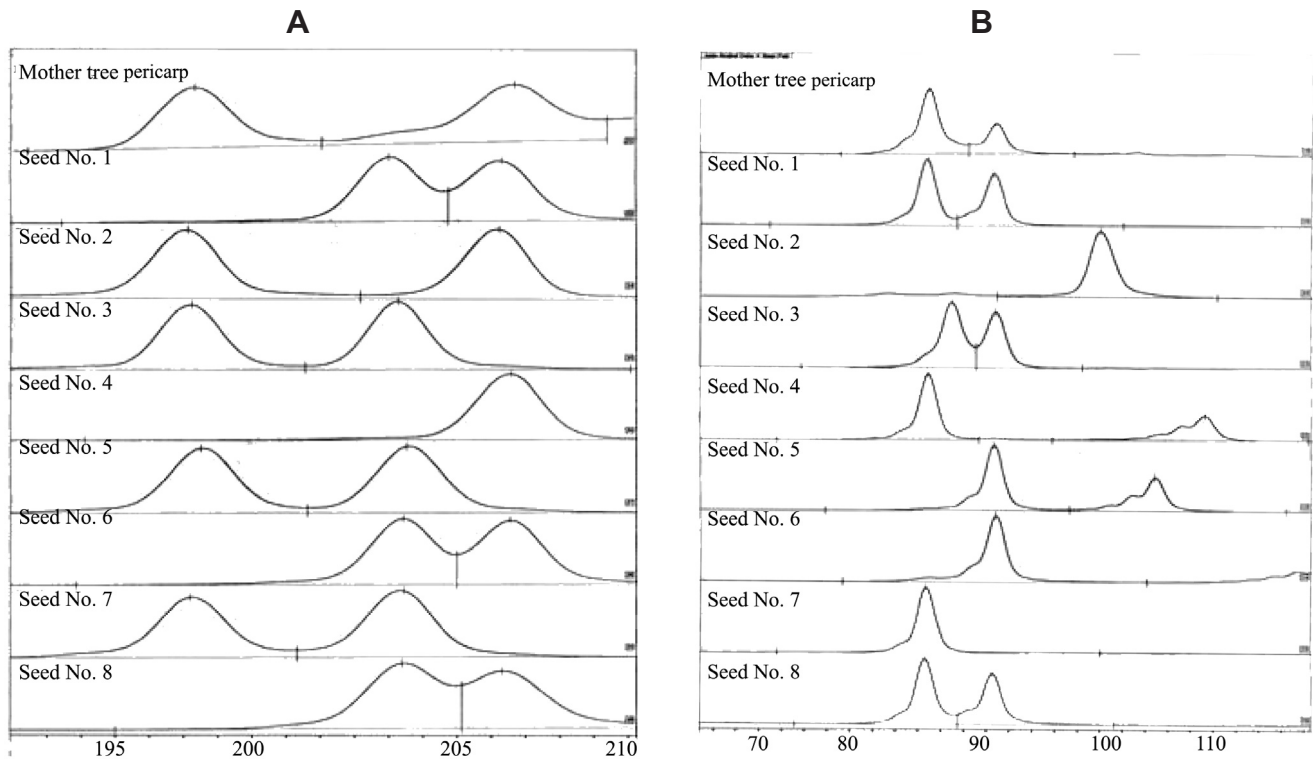


Figure 2. A: Nuclear SSR alleles at locus FS3-04 displayed in the fluoroform modus of the FRAGMENT MANAGER software on a Amersham ALF automated sequencer. B: Nuclear SSR alleles at locus FS1-15 displayed in the fluoroform modus of the FRAGMENT MANAGER software on a Amersham ALF automated sequencer.

These seeds were removed from the dataset. Successful reconstruction of the maternal genotype was confirmed for each locus by comparison with the microsatellite genotype inferred from the analysis of the seed pericarp (an example for microsatellite locus of FS3-04 is reported in Figure 2A). Homozygous offspring genotypes were only observed for alleles of true maternal origin, strongly suggesting that false homozygotes carrying a null-allele were not present among the offspring. The absence of null-alleles is an essential prerequisite and is also very likely for the loci under study, since none of them showed null-alleles in studies of controlled crosses.

Differentiation between single tree harvests: We genotyped a total of 77 seeds (embryo) from 11 trees, seven seeds from each tree. In this population, we determined a total of 77 distinct multilocus genotypes, each seed in the population showing a unique multilocus genotype. Gene diversity was 0.48 according to the mean expected heterozygosity and a mean effective number of 2.28 alleles per locus (Table 1).

Wright's fixation index (F_{is}) measures the devia-

tion from a panmictic genotype distribution and shows the correlation between similar alleles within individuals of one population (Brown and Weir, 1983; Barrett and Shore, 1990). In all the loci, the mean values of F_{is} were negative showing heterozygote excess and no obvious selfing among trees (Table 1).

Differentiation between the pure single tree harvests, was estimated using F_{st} according to Weir and Cockerham (1984). F_{st} between the ($n=11$) harvests was 0.19, indicating significant differentiation (Table 1). As 11 seedlot samples originate from one population, the strong genetic differentiation suggests that they indeed represent a distinct family structure. It may be expected, however, that differentiation decreases as more half-sib families of one population are analysed because this would result in an increased probability for some of the seed parents to act also as pollen donors. Inbreeding in the overall seeds, as measured by the overall inbreeding coefficient, was not important ($F_{it} = 0.032$)

AMOVA. revealed variation among trees which variation accounted for 28%, and within trees accounting for 72% of the total (Table 2). The variation for

Table 2. Analysis of molecular variance (AMOVA).

Source	df	Sum of squares	Mean of squares	Variance components	variation (%)	Prob.
Among trees	10	70.922	7.092	0.737	28%	
Among seeds within trees	66	127.429	1.931	1.931	72%	0.001
Total	76	198.351	9.023	2.668		

both sources was significant ($P = 0.01$).

In order to determine the number of different mother trees present in the samples we calculated pairwise F_{st} between the 11 mother trees. A PcoA was conducted on the pairwise F_{st} matrix and the first three principal coordinates were plotted (Fig. 3). Additionally, 7 individual seeds of one of the single tree harvests (tree A) were analyzed as 7 different trees in order to simulate 7 single tree harvested from the same mother tree. F_{st} was calculated between the 17 samples (10 mother trees A-K, and 7 seeds of tree A) to obtain a pairwise F_{st} matrix, PcoA was conducted on the pairwise F_{st} matrix (Fig. 3 below). Differentiation between the 11 single tree harvests appeared to be pronounced, while, in contrast the expected differentiation among the 7 samples from the same mother tree was also strong. The first three prin-

cipal coordinates accounted for 37, 18 and 16% of total variance, respectively. This analysis was repeated with all other trees culminating in the same result showing strong differentiation among samples from same the mother tree.

DISCUSSION

Exact identification of source trees with pericarp:

The pericarp tissue from *F. orientalis* seeds can be used to identify the maternal source tree when the pericarp genotypes of seeds collected from the same tree were identical to each other, however, embryo genotypes were variable. This ‘hypervariability’ of the embryos confirmed previous results by Streiff *et al.* (1999, 1998; Degen *et al.*, 1999). Thus, it was legiti-

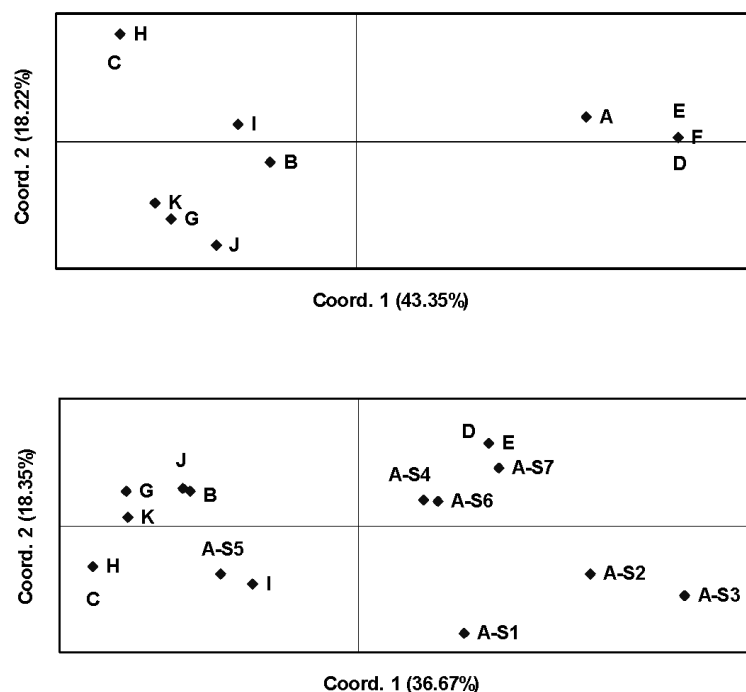


Figure 3. Two-dimensional scatter plot of principal coordinate analysis based on first and second principal coordinates of pairwise F_{st} between eleven mother trees (above: A-K) and 18 samples including 11 mother tree (A-K) and 7 seeds from the same mother tree A (below: A-S1 to A-S7).

mate to unambiguously attribute the analysed pericarps to the respective mother tree without any assumption regarding the mating system or recombination frequencies of the SSR loci. This is expected on the basis of the anatomical origin of beech seeds, with no endosperm. By comparing the genotypes of leaves and tissues obtained from the seed progenies (endocarp/pericarp and embryo) *Prunus mahaleb* (Godoy and Jordano, 2001), *Quercus robur* and *Abies alba* Mill (Ziegenhagen *et al.*, 2003) confirmed that the genotype of the endocarp/pericarp tissue is identical to that of the mother tree. Thus, this method of direct genotyping can be easily combined with regular sampling schemes of seed rain using seed traps (Harms *et al.*, 2000; Kollmann and Goetze, 1997), to assess patterns of seed dispersal at the landscape level.

The approach described in this study can be applied to a variety of species that typically show a woody pericarp, although care should be taken in determining the anatomical origin of the tissue analysed. When assessing species with complex fruit structures, such as arillate seeds, etc., a preliminary comparison with other maternal tissues can be undertaken to assure the reliability of using a particular tissue to compare with the maternal genotype. The method can also be used with wind-dispersed species that typically show ancillary structures such as wings, pappus, etc., that are presumably of maternal origin. The approach of this study can be used to estimate relative female fertilities, the diversity of trees contributing seeds to particular landscape patches, especially in relatively small populations. Some of these parameters can be estimated even with incomplete genotyping of the adults in larger populations (Slate *et al.*, 2000). For instance, an exclusion approach requiring a limited genotyping effort can be used to test whether a subset of candidate trees in the neighborhood of a given seed sampling point are the source for the sampled seeds.

Inference of seed parents from the offspring: In principle the maternal alleles can be inferred from the maternal half-sib families of *F. orientalis* using highly polymorphic microsatellites. No prior information about the mother tree is necessary. The parental alleles can be assigned to the pollen donor on the basis of codominant mendelian inheritance. Hence, this method has the potential to determine the purity of single tree harvests and to infer their maternal genotypes. Using this method, it is possible to detect the unrelated seeds.

The detection of unrelated individuals within single tree harvests may be explained by the fact that the seeds of some trees are generally collected from the ground, as is usually done for commercial seedlots. The collection of seeds directly from the ground takes into consideration the possibility of dispersal by animals such as glis glis. Furthermore, in the specific case of this investigation, the harvest site was located on a steep slope, making it likely that seeds were dispersed by gravity. Our results show that such contaminations, whatever the cause may be, can be detected with microsatellites. These observations are consistent with a half-sibling relationship. Similar observations were made by Lexer *et al.* (1999) in *Quercus robur* and Dow and Ashley (1996) in parentage studies of *Q. Macrocarpa*.

Lexer *et al.* (1999) suggested to calculate pairwise F_{st} between single tree harvests in order to determine the number of different mother trees present in the samples. According to his work on *Quercus* the maternal genotypes have a strong effect on differentiation because the maternal alleles are present in each of the offspring individuals. In other words he believes that identical mother trees create low pairwise F_{st} values, while different mother trees cause higher pairwise F_{st} values. Contrary to his results, this study found strong differentiation among the samples from the same mother tree, suggesting that maternal genotypes do not have a strong effect on differentiation. Although the maternal alleles are present in each of the offspring individuals, the results of this investigation suggested that paternal genotypes and gene flow have also a strong influence on differentiation and identical mother trees do not create low pairwise F_{st} values. In contrast to what has been found by Lexer *et al.* (1999), who suggest that pairwise F_{st} may be a suitable first check for the number of different mother trees included in commercial seed harvests.

Acknowledgements

This project was supported by the International Plant Genetic Resource Institute, IPGRI (Research Grant # D06C fellowships).

References

- Barrett SCH, Kohn JR (1991). Genetic and evolutionary consequences of small population sizes in plants: implications for conservation. In: *Genetics and Conservation of Rare Plants*, Oxford University Press, eds., Falk DA, Holsinger KE, New

- York, PP. 3-30.
- Barrett SCH, Shore JS (1990). Isozyme variation in colonizing plants. In: *Isozymes in plant biology*, Chapman and Hall, eds., Soltis DE, Soltis PS, London, PP. 1-280.
- Blakesley D, Hardwick K, Elliott S (2002). Research needs for restoring tropical forests in Southeast Asia for wildlife conservation: framework species selection and seed propagation. *New Forests* 24: 165-174
- Blouin MS, Parsons M, Lacaille V, Lotz S (1996). Use of microsatellite loci to classify individuals by relatedness. *Mol Ecol* 5: 393-401.
- Brown AHD, Weir BS (1983). Measuring genetic variability in plant populations. In: *Isozymes in plant genetics and breeding*, part A, Elsevier, eds., Tanksley SD, Orton L. Amstredam, PP: 219-239.
- Degen B, Streiff R, Ziegenhagen B (1999). Comparative study of genetic variation and differentiation of two pedunculate oak (*Quercus robur*) stands using microsatellite and allozyme loci. *Heredity* 83: 597-603.
- Dow BD, Ashley MV (1996). Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol Ecol* 5: 615-627.
- Excoffier L, Smouse P, Quattro J (1992). Analysis of molecular variance inferred from metric distance among DNA restrictiondata. *Genetics* 131: 471-491.
- Godoy JA, Jordano P (2001). Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. *Mol Ecol* 10: 2275-2283.
- Gower JC (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53: 325-338.
- Guarino L, Ramanath Rao V, Reid R (1995). *Collecting Plant Genetic Diversity. Technical Guidelines*. CAB International, Wallingford, UK, pp. 31-63.
- Harms KE, Wright SJ, Calderón O, Hernández A, Herre EA (2000). Pervasive density-dependent recruitment enhances seedling diversity in a tropical forest. *Nature* 404: 493-495.
- Kollmann J, Goetze D (1997). Notes on seed traps in terrestrial plant communities. *Flora* 192: 1-10.
- Lexer C, Heinze B, Steinkellner H, Kamfer S (1999). Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: detection of seed contaminations and inference of the seed parents from the offspring. *Theo Appl Genet* 99: 185-191.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York. PP.1-512.
- Pastorelli R, Smulders MJM, Westende WPC van 't; Vosman B, Giannini R, Vettori C, Vendramin GG (2003). Characterization of microsatellite markers in *Fagus sylvatica* L. and *Fagus orientalis* Lipsky. *Mol Ecol Notes* 3: 76-78.
- Peakal R, Smouse PE. (2006). GenA1Ex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288-295.
- Schmidt L (2000). *Guide to handling of tropical and subtropical forest seed*. Danida Forest Tree Centre, Denmark. PP.1-511.
- Schneider S, Roessli D, Excoffier L (2000). *Arlequin, Version 2.000: A software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva. PP.1-111.
- Schopmeyer CS (1974). *Seeds of Woody Plants in the United States*. U.S. Department of Agriculture, Forest Service, Washington Publisher, PP. 1-401.
- Slate J, Marshall T, Pemberton J (2000). A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Mol Ecol* 9: 801-808.
- Streiff R, Ducouso A, Lexer C, Steinkellner H, Glossl J, Kremer A (1999). Pollen dispersal inferred from paternity analysis in a mixed stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Mol Ecol* 8: 831-841.
- Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glossl J, Kremer A (1998). Within-population genetic structure in *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Mol Ecol* 7: 317-328.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Wright, S (1931). Evolution in mendelian populations. *Genetics* 16: 97-159.
- Wright, S (1951). The genetical structure of populations. *Annal Eugenetics* 15: 323-354.
- Ziegenhagen B, Liepelt S, Kuhlenskamp V, Fladung M (2003). Molecular identification of individual oak and fir trees from maternal tissues of their fruits or seeds. *Trees* 17: 345-350.