

References

- DEGEN, B. and M. FLADUNG (2008): Use of DNA-markers for tracing illegal logging. *In: Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October 8–9 2007, Bonn/Germany (Ed. BERND DEGEN). Landbauforschung, vTI Agriculture and Forestry Research, Sonderheft 321, Germany. Pp 6–14.*
- DEGUILLOUX, M. F., M. H. PEMONGE, L. BERTEL, A. KREMER and R. J. PETIT (2003): Checking the geographical origin of oak wood: molecular and statistical tools. *Mol Ecol* **12**: 1629–1636.
- DEGUILLOUX, M. F., M. H. PEMONGE and R. J. PETIT (2002): Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proc R Soc Lond B: Biological Sciences* **269**: 1039–1046.
- LEMES, M. R., C. W. DICK, C. NAVARRO, A. J. LOWE, S. CAVERS and R. GRIBEL (2010): Chloroplast DNA microsatellites reveal contrasting phylogeographic structure in mahogany (*Swietenia macrophylla* King, Meliaceae) from Amazonia and Central America. *Tropical Plant Biology* **3**: 40–49.
- DUMOLIN-LAPÈGUE, S., R. J. PETIT, L. GIELLY and P. TABERLET (1999): Amplification of DNA from ancient and modern oak wood. *Mol Ecol* **8**: 2137–2140.
- LOWE, A. J. (2008): Can we use DNA to identify the geographic origin of tropical timber? *In: Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October 8–9 2007, Bonn/Germany (Ed. BERND DEGEN). Landbauforschung, vTI Agriculture and Forestry Research, Sonderheft 321, Germany. Pp 15–19.*
- LOWE, A. J., S. A. HARRIS and P. ASHTON (2004): *Ecological Genetics: Design, Analysis and Application*. Blackwells, Oxford. 326 pp.
- LOWE, A., R. C. MUNRO, S. SAMUEL, and J. COTTRELL (2004): The utility and drawbacks of chloroplast DNA for identifying native British oak stands. *Forestry* **77**: 335–347.
- SPEIRS, A., G. MCCONNACHIE and A. J. LOWE (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 DNA Analysis in Honour of Tom Loy (Eds. HASLAM, ROBERTSON, CROWTHER, KIRKWOOD and NUGENT). Chapter 13 pp 165–179. Terra Australis*
- WEISING, K. and R. C. GARDNER (1999): A set of conserved PCR primers for the analysis of simple sequence repeat poly-morphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* **42**, 9–19.
- WONG, K. N., W. L. TAN and F. T. CHEW (2009): Identification and characterization of microsatellite loci in *Intsia palembanica* (Leguminosae), a valuable tropical timber species. *Molecular Ecology Resources* **9**, 360–364.
- ZAHNEN, J. (2008): Foreword from WWF-Germany. *In: Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins" October 8–9 2007, Bonn/Germany (Ed. BERND DEGEN). Landbauforschung, vTI Agriculture and Forestry Research, Sonderheft 321, Germany. Pp 5.*

Use of DNA-Fingerprints to Control The Origin of Forest Reproductive Material

By B. DEGEN^{1,*}, A. HÖLTKEN² and M. ROGGE³

(Received 28th May 2010)

Abstract

Well-adapted, high quality reproductive material is key to the success of forest plantations. Consequently in many countries the collection and trade of forest reproductive material is regulated. Paper documents are usu-

ally the only evidence for the origin of forest reproductive material. Certification schemes already established in Germany use genetic inventories to compare reference samples collected at different steps of the chain-of-custody. A new approach using DNA-fingerprints efficiently controls the origin of seed sources without these multiple reference samples. Only a sample of adult trees within the seed stands is needed. The control is directly made for each suspicious plant or a group of suspicious plants by use of multilocus genotype assignment. We made a field test with samples of adults and seedling from 5 registered seed stands of *Quercus robur* in Western Germany. Eight highly variable nuclear microsatellites were used to genotype each individual. We found in total 255 different alleles at all loci in the adult populations. The observed levels of genetic variation ($A_e = 9.18$), genetic differentiation ($\delta = 0.187$) and population fixation ($F_{ST} = 0.01$) were slightly higher than results

¹) Johann Heinrich von Thünen Institut, Institut für Forstgenetik, Sieker Landstrasse 2, D-22927, Grosshansdorf, Germany.

²) Plant Genetic Diagnostics GmbH, Sieker Landstrasse 2, D-22927, Grosshansdorf, Germany. E-Mail: hoeltken@holz.uni-hamburg.de

³) Landesbetrieb Wald und Holz Nordrhein-Westfalen, Obereimer 13, D-59821 Arnsberg, Germany. E-Mail: martin.rogge@wald-und-holz.nrw.de

*) Corresponding author: BERND DEGEN. Johann-Heinrich-von-Thünen-Institut, Institut für Forstgenetik, Sieker Landstrasse 2, D-22927, Grosshansdorf, Germany. E-Mail: bernd.degen@vti.bund.de

of similar studies. Individual and group assignment tests were performed with the Bayesian multi-locus approach. The proportion of correctly assigned seedlings was 65% for individuals with completely scored genotypes. In all 5 cases the groups of seedlings were assigned to the correct seed stand and an additional sample of seedlings from another stand could be successfully excluded with a probability test. The conclusion of the field study is that a large scale application of this new approach to control of the origin of forest reproductive material is feasible.

Key words: Certification, forest reproductive material, genotype assignment, nSSRs, seed, tracing origin, *Quercus robur*.

Introduction

Well-adapted, high quality reproductive material is key to the success of forest plantations. Thus the collection and trade of forest reproductive material is regulated in many countries by laws, regulations and certification to promote the use of reproductive material of known origin and genetic quality for consumer protection, and the use of material well adapted to a given region. More economic, efficient methods of identifying reproductive material are needed to comply with these regulations. These regulations include the Organisation for Economic Co-operation and Development's (OECD) *Scheme for the Certification of Forest Reproductive Material* at the international level, the Council Directive 1999/105/EC for the European Union, and within Germany, the Act on Forest Reproductive Material (FoVG), its executive regulations and the technical standards of the German Programme for the Endorsement of Forest Certification Schemes (PEFC).

In Germany forest seeds are mainly collected in registered stands of the category "selected" or "tested", and to a lesser degree in seed orchards. Paper documents are usually the only proof for material origin. Documents have been repeatedly falsified, resulting in the use of unsuitable seed sources for a given area, and reduced growth, poor form, increased mortality, or increased vulnerability to disease in forest plantations in Germany

and elsewhere (KONNERT and RUETZ, 2006). Thus there is an urgent need for an efficient control system using traits that cannot be manipulated. In the last years two certifications systems have been successfully established in Germany (BEHM and KONNERT, 2000; KONNERT, 2006). These systems use reference samples of forest reproductive material taken at different steps along the chain of custody. Gene markers are used to check the similarity between suspicious probes and reference samples. The systems work well, but they are time-consuming, involving many reference samples, and the statistical power of these approaches has never been tested.

In this paper we test a new approach using DNA-fingerprinting that efficiently controls the origin of seed source without multiple reference samples along the chain of custody. Only a sample of adult trees within the seed stands is needed. Multilocus genotype assignment of individuals or groups of suspicious plants will be used to control the origin of the material. This approach has already successfully been used to assign the geographic origin of plants and animals (HONJO et al., 2008; ODGEN, 2008; WASSER et al., 2004, 2007). We performed a field-study using DNA-fingerprints based on nuclear microsatellites to check the origin of forest reproductive material in five sessile oak seed stands in Western Germany.

Material and Methods

Plant material

For the field study five seed tree stands of the pedunculate oak, *Quercus robur*, 8 to 240 km apart were selected in North Rhine-Westphalia, Germany (NWR) (Fig. 1, Table 1). In each stand cambium from 56 to 89 adult trees was collected. A punch was cleaned with ethanol and then hammered into the bark of the target tree until stopped by the wood. The resultant bark plug had the cambial zone on the inner surface. These were kept on ice until further treatment in the lab. For further information on this method see (COLPAERT et al., 2005). Between 27 and 34 seedlings were also collected randomly within each stand.

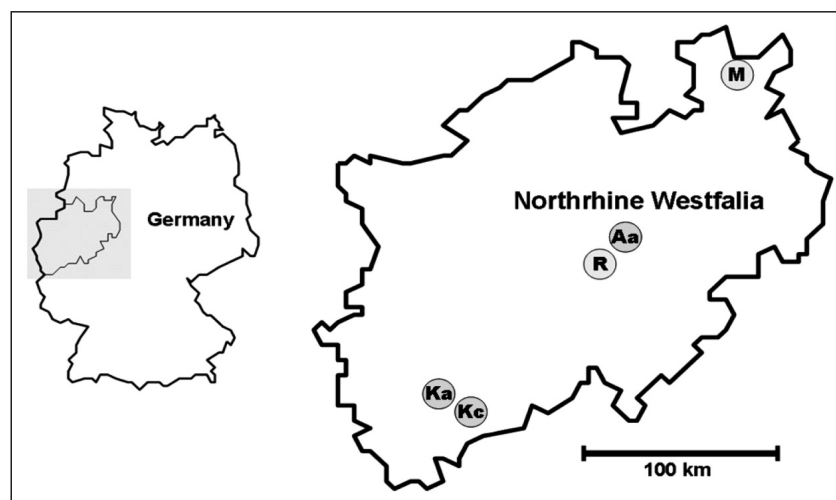


Figure 1. – Locations of the studied seed stands of *Quercus robur* (sessile oak) in North Rhine-Westphalia (Germany); Mindener Wald (M), Rumblebeck (R), Arnsberg (Aa), Kottenforst (Ka), Kottenforst (Kc).

Table 1. – Seed stands of *Quercus robur* sampled in North-Rhine Westphalia and sample sizes for adults and seedlings.

Stand	Size (ha)	Altitude (m)	Year of plantation	N adults	N seedlings
Mindener Wald (M)	9.6	46	1894	82	34
Rumbeck (R)	10.5	220-250	1874	89	33
Arnsberg (Aa)	5.5	185	1846	84	27
Kottenforst (Ka)	15.6	176	1905	56	36
Kottenforst (Kc)	10.9	151	1905	85	34

Genetic analyses

DNA-extraction: From the cambium/bark samples of the adult trees, a very thin disc of tissue was sliced off using a clean scalpel sterilised between individuals with 70% ethanol. After grinding the frozen plant material (cambium, leaf, and embryo tissue in liquid nitrogen) with a Mixer Mill (MM 300, Retsch), DNA was extracted using the protocol of (DUMOLIN et al., 1995) based on CTAB/dichloromethane. DNA concentrations were then measured and dilutions containing 10 ng/ μ L DNA were prepared and stored at -20°C .

SSR-Genotyping: Eight microsatellites were analysed for *Quercus robur*. A high throughput microsatellite genotyping method was applied using two sets of primer pairs. QrSET1 was comprised of four loci (QrZAG11, QrZAG96, QrZAG110, QrZAG112), and QrSET2 of four microsatellite loci (QrZAG7, QrZAG20, QrZAG65, QrZAG87) (KAMPFER et al., 1998; STEINKELLNER et al., 1997). The microsatellite loci were amplified by multiplexing all primers of each set. The cycling conditions consisted of a denaturation step of 5 min at 94°C ; followed by 35 cycles of 30 s at 94°C , 30 s at 50°C , 90 s at 65°C ; and a final extension of 10 min at 65°C after (LEPAIS et al., 2006, 2009), for further information on amplification conditions see Table 2). Amplified SSR fragments were analysed with an Amersham MegaBace 1000 capillary sequencer (GE Healthcare, Freiburg) and individual genotypes were determined with the Fragment Profiler software version 1.2.

Data analysis

Genetic variation and genetic differentiation

Locus allele frequencies (p_{ij}), the number of different alleles (A), observed heterozygosity (H_o), expected

heterozygosity (H_e), the effective number of alleles ($A_e = 1/(1-H_e)$), and the fixation index ($F = 1-(H_o/H_e)$) were calculated for each as described by (WEIR, 1990). WRIGHT's F_{st} (WEIR, 1990), the standardized $F_{st}^{(Hedrick)}$ of (HEDRICK, 2005), and delta (GREGORIUS et al., 2007) were computed as measures of fixation and genetic differentiation among populations.

Genetic distance (D_0) was computed to measure the genetic difference between pairs of populations (GREGORIUS, 1978):

$$D_0(i, j) = \frac{1}{2} \cdot \sum_{k=1}^n |p_{ik} - p_{jk}|,$$

where i and j represent two populations, n is the number of alleles, p_{ik} is the relative frequency of the k th allele in the i th population. Numerical tests based on Monte Carlo methods were used to estimate the significance of D_0 (MANLY, 1991). The genotypes of the two populations were pooled 1000 times, and from this joint group the two individual populations were reformed by random sampling without replacement, and the D_0 values were recalculated for both re-sampled populations (i and j). For each permutation, the genotypic structure of each population represented a random mixture of genotypes from both original populations. The probability of significance was estimated from the proportion of permutations with D_0 distances smaller than the observed.

Genotype assignment

Individual assignment tests and group assignment tests were performed using the Bayesian multilocus-approach (RANNALA and MOUNTAIN, 1997) in GeneClass2 (PIRY et al., 2004) and GDA_NT (DEGEN, unpublished).

Table 2. – Primer used to amplify the microsatellites for the 5 *Quercus robur* stands (LEPAIS et al., 2006, 2009).

Species	SSR-Set	Locus name	Primer μM	Allele size range (bp)	T_a ($^{\circ}\text{C}$)
<i>Quercus robur</i>	QrSet1	QrZAG11	0.35	241-303	50
		QrZAG96	0.12	138-184	50
		QrZAG112	0.12	70-136	50
		QpZAG110	0.30	193-251	50
	QrSet2	QrZAG7	0.12	106-154	50
		QrZAG20	0.30	158-204	50
		QrZAG65	0.50	248-338	50
		QrZAG87	0.05	102-186	50

The individuals of the adults were self-classified to the five sampled populations using the leave-one-out approach (self-assignment). The five samples of juveniles were assigned to reference populations once on an individual and once on a group basis. The assignment of an individual or group of individuals is based on a criterion relating the individual or group to each reference population, the best candidate population being the one with the highest probability. Thus a population is always designated to which the individual or group can be assigned, because there is always a most likely population in any reference set. However, the set of reference populations may not include the true population of origin of the control group (false positive assignment). Therefore, a measure of confidence that the test individual or test group truly belongs to a given population is needed (CORNUET et al., 1999). This can be achieved by comparing the likelihood value of the test group with the likelihood of sampled groups that belong to the population. More precisely, we need to locate the criterion value of the test individual or test group within the distribution of values for individuals or groups of the assigned population. If the test individual's or group's criterion is well outside the distribution, the individual or group does not belong to the population. Furthermore, the proportion of the distribution with values "worse" (lower probability/likelihood) than the tested individual's or group's value can be considered as a measure of the probability that this individual/group belongs to the population. As proposed by MARSHALL et al. (1998), we used the difference in LOD scores between the most-likely reference population and the next most-likely reference population as a test criterion for paternity analysis. In cases where the true population of origin is among the reference samples, this difference is clearly larger compared to the situation if the true population is not part of the reference samples.

Results

Field-study

Diversity/differentiation

We found in total 255 different alleles at all loci. The mean number of alleles (A) per adult population and

locus ranged from 19.75 to 23.12, the mean effective number of alleles (A_e) ranged from 8.81 to 9.23, and the values for the observed heterozygosity (H_o) ranged from 0.75 to 0.80 (Table 3). For the adults the fixation-index was close to 0, indicating that the level of heterozygosity corresponds to expectations under Hardy-Weinberg. The genetic differentiation and the fixation between all samples of adults was moderate ($\Delta_{Gregorius} = 0.187$; $F_{ST} = 0.008$, $F_{ST(Hedrick)} = 0.091$).

The gene pool distances (D_0 Gregorius) varied between 0.200 and 0.265 and most of them were statistically significant (Table 4). Only among the two populations from Kottenforst (Ka, Kc) and between Kottenforst (Ka) and Mindener Wald (M) did the genetic distance have a significance level below 0.9. Spatial and genetic distances were not correlated for the five populations.

Genotype assignment

Individual self-assignment tests were performed with the Bayesian allele frequency estimation method with the leave-one-out option. Adult individuals were self-classified to the 5 reference populations. Accuracy of individual assignment to reference population was, with 37% correctly assigned individuals, relatively low. The individual assignment of the seedlings was correct in 56% of all cases. If we considered only seedlings with completely scored genotypes at all 8 loci the accuracy increased to 65%. The group-wise assignment was correct in all 5 cases. The probabilities of exclusion for the correctly assigned reference populations varied between 26% and 99%. The seedling samples from the outside seed stand got assigned to the population Rumbeck but the exclusion probability was 100%. Thus it was typed as a "wrong positive" assignment.

Table 4. – Gene pool distances (D_0 Gregorius) among the 5 adult populations (above diagonal) and probability of significance (below diagonal).

	M	R	Aa	Ka	Kc
M	-	0.200	0.263	0.218	0.205
R	0.967	-	0.254	0.219	0.204
Aa	1.000	1.000	-	0.265	0.217
Ka	0.889	0.914	1.000	-	0.227
Kc	0.923	0.955	0.989	0.888	-

Table 3. – Estimates of genetic parameters for adults of the five populations of *Quercus robur*: A is the mean number of alleles per locus; A_e is the mean effective number of alleles; H_o is the mean observed heterozygosity; H_e is the mean expected heterozygosity; F is the average fixation index. The Gregorius genetic differentiation index ($\Delta_{Gregorius}$); Wright's F_{ST} -statistic (F_{ST}); and the standardized Hedrick F_{ST} ($F_{ST(Hedrick)}$) are measures of genetic differentiation and fixation based on all groups of sampled adults and seedlings.

Parameter	Population					Mean
	M	R	Aa	Ka	Kc	
A	20.25	22.62	20.75	19.75	23.12	21.30
A_e	8.93	9.31	9.65	8.81	9.23	9.18
H_o	0.78	0.80	0.79	0.75	0.79	0.78
F	0.01	0.00	0.02	0.06	0.03	0.02
$\Delta_{Gregorius}$	--	--	--	--	--	0.18
F_{ST}	--	--	--	--	--	0.01
$F_{ST(Hedrick)}$	--	--	--	--	--	0.09

Discussion

The five pedunculate oak populations studied in Western Germany had slightly higher levels of genetic variation ($A_e = 9.18$), genetic differentiation ($\delta = 0.18$) and population fixation ($F_{ST} = 0.008$) than found in similar studies in oaks using nuclear microsatellites. For four pedunculate oak populations in Northern Germany, (GREGORIUS et al., 2007) found mean values of $A_e = 8.2$, $\delta = 0.14$, and $F_{ST} = 0.006$ based on six microsatellite loci. But MUIR et al. (2004) observed the same average fixation of 0.008 among *Quercus petraea* populations in Ireland.

Significant genetic differentiation is required for all types of genotype assignment. As indicated by the significant gene pool distances among samples (Table 4), significant differentiation could be found in most pair-wise comparisons of our field study. For these oak populations, genetic and geographic distances were not significantly correlated. This may be due to the common practise of growing oaks from seeds or saplings from various seed sources in the past. In our case, the origin of reproductive material and not post-glacial re-colonisation caused the observed spatial genetic structure (KOENIG et al., 2002).

Our field study showed that 8 microsatellites would be sufficient for highly reliable group-wise assignment of seedlings to their population of origin. Using the Bayesian approach implemented in GeneClass2 and GDA_NT, the group-wise assignment of offspring to the adult populations was correct in all 5 cases. This underlines the power of the proposed approach. In our field study 65% of all oak seedlings with a completely scored genotype were assigned correctly. This means for a highly accurate assignment test of single individuals more than 8 microsatellites would be required.

Multilocus-genotype assignment based on nuclear or chloroplast microsatellites and Single Nucleotide Polymorphism (SNP) has been used to trace the origins of fish (ODGEN, 2008), cattle (ORRU et al., 2009), endangered plant species (HONJO et al., 2008) and humans (XING et al., 2009). NARUM et al. (2008) applied a combination of microsatellites and SNPs to assign the origin of salmon. In their study the individual assignment test accuracy was higher for microsatellites (73.1%) than SNPs (66.6%), and pooling all 50 markers provided the highest accuracy (83.2%). The same approach using 16 nuclear microsatellites has been used by (WASSER et al., 2004, 2007) to assign the geographic origin of African elephant DNA in order to control the ivory trade. For the elephant study an individual assignment accuracy of 85% was observed.

Our field study shows that highly variable nuclear microsatellites and genotype assignment are feasible for controlling of the origin of forest reproductive material on a large scale. The group-wise assignment approach is completely feasible in the trade of forest reproductive material. The traded seeds or plants are accompanied by documents with information on the seed stand or seed orchard of origin. Thus a group of individuals can always be collected for a control. In case of undeclared or illegal mixtures of seed material the assignment of groups to a single seed source will be not successful. Assignment of individuals would make it possible to identify false documentation.

Acknowledgement

This project was financially supported by the forests service of North-Rhine Westphalia and the EU Network of Excellence "EvolTree." We are thankful to VIVIAN KUHLENKAMP for her excellent laboratory work. We thank SHEILA WARD and ARMIN KÖNIG for their helpful comments on former versions of the manuscript.

References

- BEHM, A. and M. KONNERT (2000): Proposal for a seed certification scheme. *Dendrobiology* **47**, 105–108.
- COLPAERT, N., S. CAVERS, E. BANDOU, H. CARON, G. GHEYSEN and A. J. LOWE (2005): Sampling Tissue for DNA Analysis of Trees: Trunk Cambium as an Alternative to Canopy leaves. *Silvae Genet.* **54**, 265–269.
- CORNUET, J. M., S. PIRY, G. LUIKART, A. ESTOUP and M. SOLIGNAC (1999): New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**: 1989–2000.
- DUMOLIN, S., B. DEMESURE and R. J. PETIT (1995): Inheritance of chloroplast and mitochondrial genomes in pediculate oak investigated with an efficient PCR method. *Theor. Appl. Genetics* **91**, 1253–1256.
- GREGORIUS, H.-R., B. DEGEN and A. KOENIG (2007): Problems in the analysis of genetic differentiation among populations – A case study in *Quercus robur*. *Silvae Genet.* **56**, 190–199.
- GREGORIUS, H.-R. (1978): The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. *Math. Bioscience* **41**, 253–271.
- HEDRICK, P. W. (2005): A standardized genetic differentiation measure. *Evolution* **59**, 1633–1638.
- HONJO, M., S. UENO, Y. TSUMURA, T. HANDA, I. WASHITANI and R. OHSAWA (2008): Tracing the origins of stocks of the endangered species *Primula sieboldii* using nuclear microsatellites and chloroplast DNA. *Conservation Genetics* **9**, 1139–1147.
- KAMPFER, S., C. LEXER, J. GLÖSSL and H. STEINKELLNER (1998): Characterization of $(GA)_n$ microsatellite loci from *Quercus robur*. *Hereditas* **129**, 183–186.
- KOENIG, A. O., B. ZIEGENHAGEN, B. C. V. DAM, U. M. CSAIKL, E. COART, B. DEGEN, K. BURG, S. M. G. D. VRIES and R. J. PETIT (2002): Chloroplast DNA variation of oaks in western Central Europe and genetic consequences of human influences. *For. Ecol. Manage.* **156**, 147–166.
- KONNERT, M. and W. RUETZ (2006): Genetic aspects of artificial regeneration of Douglas-fir (*Pseudotsuga menziesii*) in Bavaria. *Eur. J. For. Res.* **125**, 261–270.
- KONNERT, M. (2006): Success with origin evidence using isoenzym and DNA analyses. *AFZ/Wald, Allg. Forst Z. Waldwirtsch. Umweltvorsorge* **61**.
- LEPAIS, O., V. LÉGER and S. GERBER (2006): Short Note: High Throughput Microsatellite Genotyping in Oak Species. *Silvae Genet.* **55**, 238–240.
- LEPAIS, O., J. PETIT, E. GUICHOUX, J. E. LAVABRE, F. ALBERTO, A. KREMER and S. GERBER (2009): Species relative abundance and direction of introgression in oaks. *Mol. Ecol.* **18**, 2228–2242.
- MANLY, B. F. J. (1991): *Randomization and Monte Carlo Methods in Biology*. Chapman and Hall, London.
- MARSHALL, T. C., J. SLATE, L. E. B. KRUK and J. M. PEMBERTON (1998): Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* **7**: 639–655.

- MUIR, G., A. J. LOWE, C. C. FLEMING and C. VOGL (2004): High nuclear genetic diversity, high levels of outcrossing and low differentiation among remnant populations of *Quercus petraea* at the margin of its range in Ireland. *Ann. Bot.* **93**, 691–697.
- NARUM, S. R., M. BANKS, T. D. BEACHAM, M. R. BELLINGER, M. R. CAMPBELL, J. DEKONING, A. ELZ, C. M. GUTHRIE, C. KOZFKAY, K. M. MILLER, P. MORAN, R. PHILLIPS, L. W. SEEB, C. T. SMITH, K. WARHEIT, S. F. YOUNG and J. C. GARZA (2008): Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Mol. Ecol.* **17**, 3464–3477.
- OGDEN, R. (2008): Fisheries forensics: the use of DNA tools for improving compliance, traceability and enforcement in the fishing industry. *Fish and Fisheries* **9**, 462–472.
- ORRU, L., G. CATILLO, F. NAPOLITANO, G. DE MATTEIS, M. C. SCATA, F. SIGNORELLI and B. MOIOLI (2009): Characterization of a SNPs panel for meat traceability in six cattle breeds. *Food Control* **20**, 856–860.
- PIRY, S., A. ALAPETITE, J. M. CORNUET, D. PAETKAU, L. BAUDOIN and A. ESTOUP (2004): GENECLASS2: A software for genetic assignment and first-generation migrant detection. *J. Hered.* **95**, 536–539.
- RANNALA, B. and J. L. MOUNTAIN (1997): Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9197–9201.
- STEINKELLNER, H., S. FLUCH, E. TURETSCHKEK, C. LEXER, R. STREIFF, A. KREMER, K. BURG and J. GLÖSSL (1997): Identification and characterization of (GA/CT)_n microsatellite loci from *Quercus petraea*. *Plant Mol. Biol.* **33**, 1093–1096.
- WASSER, S. K., C. MAILAND, R. BOOTH, B. MUTAYOBA, E. KISAMO, B. CLARK and M. STEPHENS (2007): Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4228–4233.
- WASSER, S. K., A. M. SHEDLOCK, K. COMSTOCK, E. A. OSTRANDER, B. MUTAYOBA and M. STEPHENS (2004): Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14847–14852.
- WEIR, B. S. (1990): *Genetic Data Analysis*. Sinauer Associates, Sunderland, MA.
- XING, J. C., W. S. WATKINS, D. J. WITHERSPOON, Y. H. ZHANG, S. L. GUTHERY, R. THARA, B. J. MOWRY, K. BULAYEVA, R. B. WEISS and L. B. JORDE (2009): Fine-scaled human genetic structure revealed by SNP microarrays. *Genome Research* **19**, 815–825.

Population Structure and Molecular Characterization of Nigerian Field Genebank Collections of Cacao, *Theobroma cacao* L.

By P. O. AIKPOKPODION^{1,2,4,*}, M. KOLESNIKOVA-ALLEN², V. O. ADETIMIRIN³, M. J. GUILTINAN⁴, A. B. ESKES⁵, J.-C. MOTAMAYOR⁶ and R. J. SCHNELL⁷

(Received 2nd October 2008)

Abstract

Inadequate knowledge of the population structure and diversity present often hamper the efficient use of germplasm collections. Using a high through-put system, twelve microsatellite loci were used to analyze genetic diversity and population structure in a national

field genebank repository of 243 cacao accessions grouped into 11 populations based on their known sources. Based on multi-locus profiles, the Bayesian method was used for individual assignment to verify membership in each population, determine mislabeling and ancestry of some important accessions used in breeding program. A total of 218 alleles was revealed with a mean number of 18.2 alleles per locus. Gene diversity ($H_e = 0.70$) and allelic richness (4.34 alleles per locus) were highest in the F₁ hybrid population. Differential mating system was suggested as responsible for the observed deficit and excess of heterozygotes observed among the populations. Analysis of molecular variance showed that within-population variance accounted for 63.0% of the total variance while the rest 37% was accounted for by the among-population variance. Cluster dendrogram based on UPGMA revealed two main subsets. The first group was made up of the Amelonado/Trinitario ancestry and the other of Nanay/Parinari ancestry. We found that Nanay and Parinari populations were the major source of Upper Amazon genes utilized while a large proportion of genetic diversity in the field genebank remained under-utilized in development of improved cultivars released to

¹) Plant Breeding Division, Cocoa Research Institute of Nigeria PMB 5244 Ibadan Nigeria.

²) Central Biotechnology Laboratory, International Institute of Tropical Agriculture PMB 5320 Ibadan Nigeria.

³) Department of Agronomy, University of Ibadan, Nigeria.

⁴) Department of Horticulture, 422 Life Sciences Building, Pennsylvania State University, University Park, PA 16802, USA.

⁵) Bioersivity International, Parc Scientifique Agropolis II, 34397 Montpellier Cedex 5, France.

⁶) Mars Incorporated, Subtropical Horticulture Research Station, USDA-ARS, 13601 Old Cutler Road, FL33158.

⁷) Subtropical Horticulture Research Station, USDA/ARS, 13601 Old Cutler Road, FL33158.

* Corresponding author's current address: Department of Genetics and Biotechnology, University of Calabar, PMB 1115, Calabar, Cross River State, Nigeria. E-Mail: paikpokpodion@yahoo.com