New Forests (2006) 32:179–196
DOI 10.1007/s11056-005-5961-7

A robust set of black walnut microsatellites for parentage and clonal identification

RODNEY L. ROBICHAUD[1,*], JEFFREY C. GLAUBITZ[1], OLIN E. RHODES Jr.[2] and KEITH WOESTE[3]

*[1]Department of Forestry and Natural Resources, Hardwood Tree Improvement and Regeneration Center (HTIRC), Purdue University, 715 West State St, Pfendler Hall, West Lafayette, IN 47907-2061, USA; [2]Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-2033, USA; [3]USDA Forest Service, Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-2033, USA; *Author for correspondence (e-mail: robichau@purdue.edu; phone: +1-765-494-9592; fax: +1-765-494-9461)*

Received 13 December 2004; accepted in revised form 2 December 2005

Key words: Allele sharing, Exclusion probability, Genetic diversity, *Juglans nigra*, nSSRs, Outcrossing rate

Abstract. We describe the development of a robust and powerful suite of 12 microsatellite marker loci for use in genetic investigations of black walnut and related species. These 12 loci were chosen from a set of 17 candidate loci used to genotype 222 trees sampled from a 38-year-old black walnut progeny test. The 222 genotypes represent a sampling from the broad geographic distribution of the species. Analysis of the samples using the 12 loci revealed an average expected heterozygosity of 0.83, a combined probability of identity of 3×10^{-19}, and a combined probability of exclusion for paternity analysis of >0.999. The 222 genotyped trees from the progeny test comprised 39 open-pollinated families, 29 of which (having at least five sampled progeny) were used to estimate the outcrossing rate for the progeny trial. The same 29 families were used to construct a Neighbor-Joining dendrogram based upon allele sharing between individuals. The multilocus estimate of the outcrossing rate was 100% (standard error of zero), higher than the 90% level found in previous studies at the embryo stage, suggesting that both artificial and natural selection against selfs may have occurred over the 38-year lifespan of the progeny trial. In the Neighbor-Joining dendrogram, the majority of the putative siblings grouped together in 21 out of the 29 families, showing that the microsatellites were able to discern most of the family structure in the dataset. Our results indicate that errors were sometimes committed during the establishment of the progeny test. This set of microsatellite loci clearly provides a powerful tool for future applications in black walnut.

Introduction

Black walnut (*Juglans nigra* L.) is one of the most valuable hardwood species in the Central Hardwoods Region, prized both for its quality veneer and nutritious nuts (Harlow et al. 1979). This species can be found throughout the eastern and central hardwood forests of the United States (Figure 1a) generally occurring as widely dispersed individuals or in small, spatially distinct groves (Rink et al. 1994). Unfortunately, black walnut was over-harvested during the first part of the 20th century, resulting in declining supplies of premium logs by the early

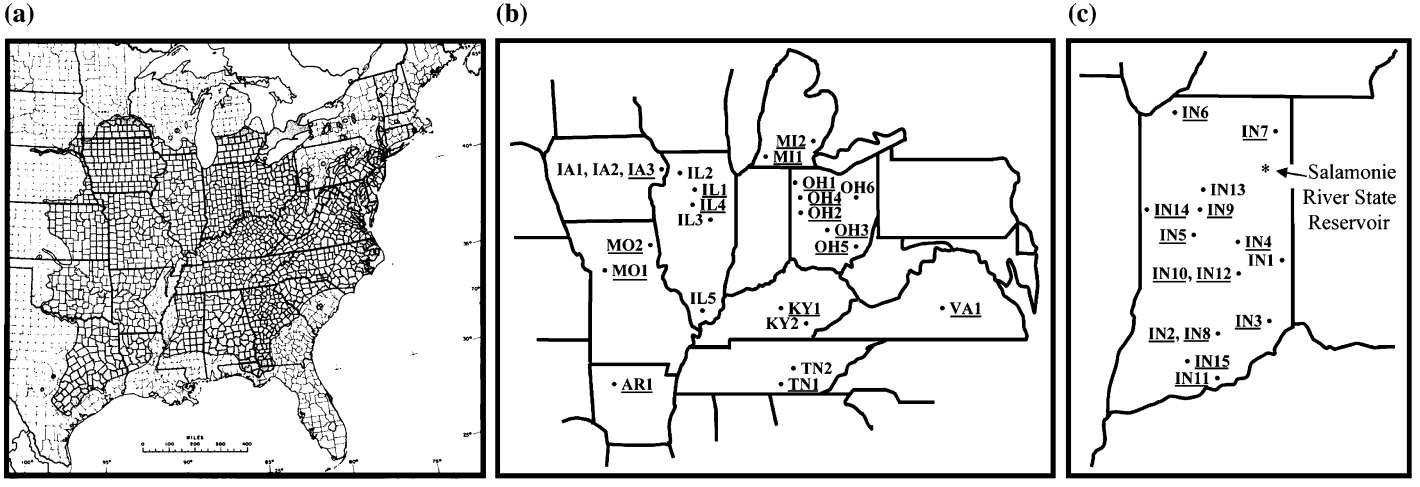

Figure 1. Locations of the source trees (maternal parents) of the 39 black walnut open pollinated families sampled from the Salamonie progeny test for this study. The family numbers of the 29 families (with at least five progeny sampled) that were used in the mating system analysis and for dendrogram construction are underlined. (a) Natural distribution of black walnut (after Williams 1990). (b) Source trees located outside of Indiana. (c) Locations of the Indiana source trees, as well as of the Salamonie progeny trial.

1960s (Beineke 1989). Intense harvesting in combination with extensive changes in land use practices may have resulted in significant losses of genetic diversity in this valuable hardwood species (Beineke 1974; Rink et al. 1994).

Although allozymes have been surveyed in black walnut for purposes such as the identification of walnut cultivars (Vyas et al. 2003), assessment of inbreeding and outcrossing rates (Rink et al. 1989, 1994; Busov et al. 2002), and estimation of genetic diversity within local populations (Rink et al. 1989, 1994; Busov et al. 2002), no genetic data of any kind are currently available with which to assess levels of genetic diversity in black walnut at a regional or range-wide scale. The advent of DNA-based markers, such as nuclear microsatellites (Weber and May 1989), has made it possible to rapidly assess the genetic diversity and population structure of species at various spatial scales with relative ease. Microsatellite markers, consisting of tandem repeats of simple (1–6 bases) sequence motifs, are ideal for population genetic studies because they are co-dominant, highly polymorphic, reproducible, and require little tissue. In recent years microsatellite loci have been used to estimate levels of genetic diversity, genetic structure, dispersal rates, and paternity in numerous tree species (e.g., Chase et al. 1996; Dow and Ashley 1996; Aldrich et al. 1998; Streiff et al. 1999; van der Schoot et al. 2000; Craft et al. 2002; Glaubitz et al. 2003; Tabbener and Cottrell 2003).

To take advantage of the powerful analytical opportunities provided by microsatellite markers (Parker et al. 1998; Luikart and England 1999), a panel of 30 nuclear microsatellites was developed by Woeste et al. (2002) for use in genetic investigations of black walnut. The purpose of our study was to further evaluate available microsatellites in black walnut, with the goal of obtaining a smaller, working set of reliable and highly informative markers for use in future population genetic studies. To accomplish this goal, we utilized the broad geographical sample of black walnut trees accessible within a 38-year old progeny trial in northern Indiana. In all, 222 individuals, representing 39 families from 10 states were sampled and genotyped. In addition to evaluating the robustness, level of polymorphism, and associated discrimination power of our set of microsatellites, the open-pollinated family structure of the progeny trial allowed us to estimate an outcrossing rate for black walnut as well as to test the power of the 12 microsatellites to distinguish half-sib families in a Neighbor-Joining dendrogram. The application and utility of these molecular markers for future black walnut studies is discussed.

Materials and methods

Plant material

Samples were taken from a 38-year-old progeny test of black walnuts planted at the Salamonie River State Reservoir, near Lagro, Indiana. The trees planted in this progeny test were grown from seed collected in the wild from

phenotypically superior black walnuts as part of a USDA Forest Service research study (Clausen 1983; Woeste 2002). Mature leaves were sampled from 222 trees in the progeny trial, representing 39 open-pollinated families from 10 states (Arkansas: 1 family; Illinois: 5; Indiana: 15; Iowa: 3; Kentucky: 2; Michigan: 2; Missouri: 2; Ohio: 6; Tennessee: 2; and Virginia: 1; (Figure 1b, c). The number of progeny sampled per family ranged from 1 to 10. Samples were collected during July and August of 2000 and 2001. Upon collection, samples were quickly placed on ice for transport and stored at −80 °C or freeze-dried until DNA extraction.

DNA extraction

Samples were prepared for DNA isolation by grinding about 100 mg of leaf tissue in a 2-ml microcentrifuge tube containing a 1/4 inch cylindrical ceramic bead (BIO 101-Savant, Carlsbad, CA) and 1.0 ml of CTAB extraction buffer (Lefort and Douglas 1999; modified with 2× PVP, 2× CTAB, and 2.0% ß-mercaptoethanol)[1]. Samples were then homogenized in an FP 120 Fastprep machine (BIO 101-Savant, Carlsbad, CA) by grinding the samples for 40 s and then cooling them on ice for about 1 min. This cycle was repeated two more times. After grinding, microcentrifuge tubes were placed into a 64 °C water bath for at least 30 min and periodically shaken by hand. Next, samples were centrifuged for 5 min at 12,500*g* using a tabletop centrifuge. DNA was isolated from 500 μl of the supernatant using an NA-2000 automated nucleic acid extractor (Autogen, Framingham, MA) employing a modification of Autogen's NA-2000 Plant DNA, V 1.01 DNA isolation protocol. In our modification, potassium acetate (Autogen reagent AG00317) was added first, followed by SDS/*N*-lauroyl sarcosine (Autogen reagent AG00212), and then chloroform (Autogen reagent AG00316). All the remaining reagents and protocols were the same as designated in the Autogen protocol. The extracted DNA was quantified using a FL 600 microplate fluorescence reader (Bio-Tek, Winooski, VT) and a Hoechst dye 33258 assay solution. The quantified DNA was diluted with 10 mM Tris–HCl (pH 8.0) and 1.0 mM EDTA (pH 8.0) buffer to a working stock concentration of 10 ng/μl prior to PCR amplification.

Initial locus selection and genotyping procedures

A total of 17 black walnut microsatellites, identified from a black walnut library (Woeste et al. 2002), were used to genotype all 222 sampled trees, with the goal of identifying a final working set of reliable and informative loci for

[1]The use of trade names is for the information and convenience of the reader and does not imply official endorsement or approval by the United States Department of Agriculture or the Forest Service of any product to the exclusion of others that may be suitable.

future studies in black walnut and related species. These 17 'candidate' loci were selected from the 30 loci described by Woeste et al. (2002) as well as from 18 additional unpublished, polymorphic loci. Selection of the 17 candidates from the 48 available loci was based upon an initial screen in which the same 10 black walnut trees utilized by Woeste et al. (2002) were genotyped, this time using fluorescently labeled primers instead of fluorescent dCTP. The forward primer for each microsatellite was modified with either a FAM, JOE, or TAMRA fluorescent label (MWG-Biotech, High Point, NC). The set of 17 candidate dinucleotide repeats were selected based upon degree of polymorphism among the 10 individuals, ease of genetic interpretation, and number of heterozygotes.

PCR amplification was performed using either Sigma ReadyMix Taq (PCR reaction mix with $MgCl_2$) or AmpliTaq Gold (Perkin-Elmer). PCR amplifications utilizing the Sigma ReadyMix Taq contained 10 ng of DNA template, 5.0 μl ReadyMix Taq PCR reaction mix with $MgCl_2$, 0.4 μl of 20 pmol/μl working primer stock, and 4 μl of nanopure, sterile water for a total volume of 10 μl. PCR amplifications utilizing AmpliTaq Gold Taq contained 10 ng of DNA template, 1.5 mM $MgCl_2$, 0.4 U AmpliTaq Gold, and 0.8 μM of each primer. All other components of the PCR mixture were as recommended by the manufacturer for a final 20 μl reaction volume. The PCR-amplification protocol was 50 cycles (AmpliTaq Gold Taq) and 30 cycles (Sigma ReadyMix Taq) of 92 °C for 30 s, 55 °C annealing temperature for 1 min, and 72 °C for 35 s run on either a PTC-100^{TM} or a PTC-200^{TM} Peltier Thermal Cycler (MJ Research, INC., San Francisco, CA). All primers were annealed at 55 °C except for WGA69, which was annealed at 50 °C.

In preparation for gel electrophoresis, 1.0 μl of the PCR product, 0.5 μl of CXR 400-bp Ladder Standard (Promega, Fitchburg Center, WI) and 1.5 μl of blue dextran loading solution (Promega, Fitchburg Center, WI) were combined, denatured for 2 min at 95 °C, and loaded onto CAL96 paper combs (The Gel Company, San Francisco, CA). Up to three loci (with different colored fluorescent tags) were run together in a single gel lane. Electrophoresis was in 5% polyacrylamide Long Ranger denaturing gels (BMA, Rockland, ME) at 3000 V, 60 mA, 200 W, 51 °C for 3 h using an ABI 377 XL automated DNA sequencer (Perkin-Elmer) with 36-cm plates and 0.2-mm spacers. The software programs GENESCAN v 3.1 and GENOTYPER v 2.5 were used to aid the assignment of genotypes.

For quality control purposes and to aid locus selection, 56 individuals, randomly selected from the total set of 222, were genotyped a second time (starting from the stock DNA) and independently scored, to test the reproducibility of genotype assignment. Reproducibility was scored for each of the 17 candidate loci as the percentage of the repeated genotypes that were identical both times; individuals with missing data (for either the original or the repeated genotype, or both) were not included in the calculation for the locus in question.

Data analysis

The software program GDA v 1.1 (Lewis and Zaykin 2001) was used to calculate the observed number of alleles per locus (allelic richness), the observed and expected heterozygosities, and the fixation index (*f*) based upon the entire sample of 222 trees. The unbiased probability of identity (P_{ID}), the probability that two randomly sampled, unrelated trees would have identical genotypes, was calculated according to Paetkau et al. (1998), based upon the final set of selected loci. We also calculated the probability that two randomly selected *full sibs* would have identical genotypes (the probability of identity of full-sibs, P_{IDsib}) using the formula provided by Waits et al. (2001). The exclusion probability provided by the selected loci for paternity analysis (P_{Epat}, the probability that a randomly sampled, *unrelated* male would be genetically excluded from being a pollen donor in a paternity analysis where the mother's genotype is known) was calculated according to Jamieson and Taylor (1997); Equations 1a and 4. The exclusion probability for parentage analysis (P_{Epar}, the probability that a randomly sampled, unrelated tree would be genetically excluded from being a parent of a younger tree) also was calculated according to Jamieson and Taylor 1997; Equations 2a and 4.

Multilocus and single-locus outcrossing rates under the mixed-mating model (Ritland and Jain 1981) were estimated using the program MLTR (Ritland 2002) for the set of 29 half-sib families with sample sizes of at least five individuals (Figure 1). All parameters of the model were estimated via the Expectation-Maximization method, allele frequencies in the pollen were assumed to be equivalent to those in ovules, and the program was allowed to infer the most likely maternal genotype for each half-sib family. Standard errors on the single-locus and multilocus outcrossing rate estimates were obtained via bootstrapping 10,000 times, with resampling performed over half-sib families.

The power of the selected microsatellite markers to partition the set of individual trees into half-sib families was tested by construction of a Neighbor-Joining dendrogram (Saitou and Nei 1987) based upon the allele sharing distance between individuals (D_{PS} = 1 – the proportion of shared alleles; Bowcock et al. 1994). Again, only those families consisting of 5 or more sampled individuals were used in the analysis ($n = 29$ families). The allele sharing distances were calculated between all pairwise combinations of individuals using the program MICROSAT (Minch et al. 1996). Based on these distances, the Neighbor-Joining dendrogram was constructed using PHYLIP (Felsenstein 1993).

Searches of the GenBank database were performed (via the National Center for Biotechnology Information internet site: http://www.ncbi.nlm.nih.gov) to see if the flanking sequences of any of the loci in our final working set had significant homology with any known genes. BLAST searches were conducted using both the flanking DNA sequences and their three possible amino-acid translations. The complete sequences of the corresponding cloned inserts were used in the searches (excluding the microsatellite repeats).

Results

Locus identification and characterization

Genotypes were obtained for most of the 222 individuals at each of the 17 candidate microsatellite loci, with the amount of missing data per locus ranging from 1 to 26 individuals (average of nine individuals). Five of the 17 candidate loci were dropped based upon the reproducibility test results, the prevalence of scoring ambiguities, and the likelihood that null alleles, alleles that fail to amplify in the PCR, were present. High frequency null alleles seemed to be present at two of the loci (WGA71 and WGA95), as indicated by large deficits of observed heterozygosity resulting in extremely high f estimates (f of 0.751 and 0.688, respectively). The other three loci (WGA2, WGA33 and WGA73) were removed because of scoring ambiguities that resulted in low reproducibility (61, 77 and 84% reproducibility, respectively). In the case of WGA2, scoring ambiguities were caused by the presence of numerous alleles differing in size by only one base, resulting in difficulty in resolving adjacent alleles. In the cases of WGA33 and WGA73, ambiguities resulted from the presence of extra peaks (in addition to the usual "stutter" bands) that made it difficult to identify the true alleles. Estimated reproducibility in the remaining 12 loci ranged from 89.5 to 99.1%, with an overall average of 96.0%.

The primer sequences and Genbank accession numbers for each of the remaining 12 loci are given in Table 1. Eight of the final 12 loci were previously reported (Woeste et al. 2002); we provide the primer sequences of all 12 here for the convenience of potential users. The flanking sequence of only 1 of the 12 selected microsatellites, WGA27, had significant homology with a known gene in the Genbank database (http://www.ncbi.nlm.nih.gov). A 116 nucleotide stretch of the cloned insert corresponding to WGA27 matched a Class III peroxidase gene from cotton (*Gossypium hirsutum*; Genbank accession number AF485267) with 87% homology, a score of 119 and an highly significant E-value of 4×10^{-24}.

Diverstiy estimates

The 12 microsatellite markers in the final working set displayed high levels of diversity based upon the complete sample of 222 trees from the 39 open-pollinated families. Allelic richness (number of alleles) at each locus ranged from 12 to 37, with an average of 19.9 (Table 2). Expected heterozygosities ranged from 0.651 to 0.958 across loci (average of 0.832) and observed heterozygosities ranged from 0.621 to 0.923 (average of 0.789); (Table 2). The fixation index (f) varied from 0.015 to 0.118 across individual loci, with an estimate of 0.052 based on all 12 loci, indicating only minor deviations from Hardy–Weinberg genotypic proportions in our geographically broad sample.

Table 1. Final working set of 12 microsatellites for studies in black walnut.

Locus	Accession number	Repeat array	Primer sequence (5′–3′)	Length (bp)[b]	Allele size range (bp)	Label	Locus set[c]
WGA06[a]	AY333949	$(AG)_4AA(AG)_{19}$ $AT(AG)_3$	F: CCATGAAACTTCATGCGTTG R: CATCCCAAGCGAAGGTTG	157	134–172	Joe	A
WGA24[a]	AY333950	$(T)_8N_{29}(CT)_{17}N_{24}(CT)_5$	F: TCCCCCTGAAATCTTCTCCT R: TTCTCGTGGTGCTTGTTGAG	242	222–248	Tamra	B
WGA27[a]	AY333951	$(GT)_3TT(GA)_{29}$	F: AACCCTACAACGCCTTGATG R: TGCTCAGGCTCCACTTCC	242	199–245	Tamra	C
WGA32[a]	AY333952	$(TC)_3CG(TC)_{19}$	F: CTCGGTAAGCCACACCAATT R: ACGGGCAGTGTATGCATGTA	176	163–217	Fam	A
WGA69[a]	AY333953	$(AG)_4N_6(AG)_{17}$	F: TTAGTTAGCAAACCCACCCG R: AGATGCACAGACCAACCCTC	182	164–188	Fam	D
WGA72[a]	AY333954	$(AG)_6AA(AG)_6(G)_{12}$	F: AAACCACCTAAAACCCTGCA R: ACCCATCCATGATCTTCCAA	151	135–159	Joe	B
WGA79[a]	AY333955	$(GA)_{10}$	F: CACTGTGGCACTGCTCATCT R: TTCGAGCTCTGGACCACC	206	192–226	Tamra	D
WGA82[a]	AY333956	$(CT)_{20}$	F: TGCCGACACTCCTCACTTC R: CGTGATGTACGACGGCTG	175	140–234	Fam	B
WGA86	AY352439	$(TC)_3TG(TC)_3$ $(CT)_4G(CT)_{11}(CA)_{16}$	F: ATGCCTCATCTCCATTCTGG R: TGAGTGGCAATCACAAGGAA	247	208–250	Tamra	A
WGA89	AY352440	$(TG)_9(GA)_{21}$	F: ACCCATCTTTCACGTGTGTG R: TGCCTAATTAGCAATTTCCA	215	179–233	Fam	C
WGA90	AY352441	$(CT)_4T(TC)_{14}$	F: CTTGTAATCGCCCTCTGCTC R: TACCTGCAACCCGTTACACA	157	142–178	Joe	C
WGA97	AY352442	$(GA)_{26}$	F: GGAGAGGAAAGGAATCCAAA R: TTGAACAAAAGGCCGTTTTC	180	149–189	Joe	D

[a]Locus previously published in Woeste et al. (2002).
[b]Expected length of the cloned and sequenced allele after PCR.
[c]Loci with the same letter were electrophoresed together.

Table 2. Genetic diversity parameters for the final working set of 12 black walnut microsatellite loci.

Locus	n[a]	A_R[b]	H_E[c]	H_O[d]	f[e]
WGA06	221	16	0.735	0.701	0.046
WGA24	211	15	0.862	0.848	0.015
WGA27	216	22	0.883	0.866	0.019
WGA32	212	28	0.930	0.901	0.031
WGA69	213	12	0.658	0.648	0.015
WGA72	221	12	0.651	0.624	0.041
WGA79	211	12	0.704	0.621	0.118
WGA82	196	37	0.958	0.923	0.036
WGA86	213	21	0.903	0.808	0.106
WGA89	214	26	0.926	0.879	0.051
WGA90	216	18	0.889	0.833	0.063
WGA97	214	20	0.890	0.813	0.086
Mean	213	19.9	0.832	0.789	0.052

[a]Number of genotypes obtained, out of a total of 222 trees from 39 half-sib families.
[b]Allelic richness, or the number of alleles/locus.
[c]Expected heterozygosity.
[d]Observed heterozygosity.
[e]Estimate of the fixation index (f) according to Weir and Cockerham (1984).

Probabilities of identity and exclusion

The probability that two unrelated individuals would share the same genotype (P_{ID}) is shown, for each locus individually and for the combined set of 12 loci, in Table 3. Also shown in this table are the probabilities that two full-sibs will have identical genotypes (P_{IDsib}), the probabilities that an unrelated male would be genetically excluded from being a possible father in a paternity analysis where the maternal genotype is known (P_{Epat}), and the probabilities that an unrelated tree would be genetically excluded from being a parent of a younger tree in a parentage analysis (P_{Epar}). Note that the ranking of loci for all four statistics corresponds perfectly to that for expected heterozygosity, reflecting the fact that they are all functions of this parameter. Loci with the highest expected heterozygosity values have the lowest probabilities of identity and the highest paternity exclusion probabilities. Across all 12 loci, P_{ID} and P_{IDsibs} are extraordinarily low (roughly one pair in 3×10^{18} for P_{ID} and one in 380,000 for P_{IDsibs}) and P_{Epat} and P_{Epar} are extraordinarily high (only one unrelated male in 11 million would falsely appear to be a father, and only one unrelated tree in 52,000 would falsely appear to be a parent).

However, it should be noted that the equations used to calculate these probabilities of identity and exclusion were all derived under the assumption that the allele frequencies are from a single population with genotypes in Hardy–Weinberg equilibrium proportions. Estimating allele frequencies from a geographically widespread sample of trees, rather than a single population,

Table 3. Discrimination power of the working set of 12 black walnut microsatellites.

Locus[a]	H_E	P_{Iunb}[b]	P_{Isib}[c]	P_{Epat}[d]	P_{Epar}[e]
WGA82	0.958	0.0031	0.273	0.910	0.834
WGA32	0.930	0.0086	0.289	0.854	0.746
WGA89	0.926	0.0097	0.291	0.846	0.733
WGA86	0.903	0.0162	0.304	0.802	0.669
WGA97	0.890	0.0205	0.312	0.778	0.636
WGA90	0.889	0.0209	0.312	0.776	0.633
WGA27	0.883	0.0220	0.316	0.768	0.624
WGA24	0.862	0.0326	0.329	0.722	0.563
WGA06	0.735	0.0842	0.405	0.562	0.373
WGA79	0.704	0.1251	0.431	0.482	0.306
WGA69	0.658	0.1391	0.458	0.459	0.271
WGA72	0.651	0.1573	0.466	0.429	0.252
Overall	0.832	2.95×10^{-19}	2.62×10^{-06}	>0.999	>0.999

[a]Loci are sorted in descending order by expected heterozygosity.
[b]Unbiased probability of genetic identity between two randomly sampled, unrelated individuals.
[c]Probability of genetic identity between two randomly sampled full sibs.
[d]Exclusion probability for paternity analysis (genotype of the maternal parent known).
[e]Exclusion probability for parentage analysis (both parents unknown).

would likely underestimate probabilities of identity and over overestimate probabilities of exclusion. However, our relatively low estimate of the total fixation index (f) across the 12 loci (0.052) indicates that the genotypic frequencies in our sample do not grossly deviate from Hardy–Weinberg proportions, so the above probabilities of identity and exclusion should at least give us a rough idea of how powerful this set of markers will be in studies carried out at finer spatial scales (e.g., parentage and paternity analysis).

Outcrossing rates

The maximum likelihood estimate of a generic multilocus outcrossing rate under the mixed-mating model for the 29 families with five or more progeny was 1.000 with a bootstrap standard error of zero. In various runs of the MLTR program, the multilocus outcrossing rate always converged to 1.000 after only a very small number of iterations of the Expectation-Maximization algorithm, no matter to what the initial values the parameters of the model were set. However, as the MLTR analysis assumes that the half-sib families are all drawn from a single population, rather than from a wide geographical area, which may indicate that these results are tentative at best. Even though these results may be tentative, what is evident, is that biologically, the species is extremely outcrossed.

In order to estimate the outcrossing rate under the mixed-mating model, the program MLTR first attempts to infer the most likely maternal genotype at each

locus for each family based upon the genotypes of the putative half-sib progeny array. In instances where the genotypes of a set of putative half-sibs are incompatible with having a common mother, the most likely maternal genotype is treated as missing data, and the corresponding family/locus combination is not used in the mating system analysis. Such genetic incompatibilities could occur because of errors committed during the establishment of the progeny test, sample collection errors, genotyping errors, or, in rare cases, mutation. The data set that we used in the MLTR analysis had a total of 348 family/locus combinations (29 families times 12 loci). Of these 348 'cells', there were 14 (4.0%) in which probable maternal genotypes could not be found because of genetic incompatibilities among the putative half-sibs. A majority of these problematic cells (13) were concentrated in three families, two from Indiana (IN14 and IN6) and one from Missouri (MO2). IN6 and MO2 both had five problematic cells, while IN14 had three. A fourth family, also from Missouri (MO1), contained only one problematic family/locus cell. We confirmed these results by repeating the genetic assays for all 14 problematic family/locus combinations, starting from our stock DNA samples. Given the concentration of incompatible genotypes in the three families, IN14, IN6 and MO2, errors in the progeny test would seem to be the best explanation in these cases. The single problem in MO1 appears to have resulted from a scoring ambiguity. It should be noted that for four of the families (IN2, IN5, IN8 and IN15) their maternal seed source trees were serendipitously sampled and genotyped in other ongoing experiments. When the true maternal genotypes were compared to the 'expected' maternal genotype generated by MLTR, the genotypes matched across all scored loci.

Family assignment

In the Neighbor-Joining dendrogram constructed based upon the degree of allele-sharing between individuals, there was a strong tendency for putative half-sibs within each family to cluster together. For this analysis, we included only those 29 families consisting of five or more individuals (196 individuals in total; Figure 2). In 8 of the 29 families, all putative half-sibs clustered together perfectly as a monophyletic group. Monophyletic clusters consisting of more than 50% of the individuals in a family were found for 18 of the 29 families. If we allow a maximum of one inclusion of an individual from outside the family within an otherwise monophyletic cluster, six more families can be added for a total of 24 out of 29. Hence, although the results of this analysis are far from the perfect, idealized result – where all individuals in each of 29 families would cluster together as monophyletic groups – it is clear that the set of 12 microsatellites are able to discern much of the family structure in the data set.

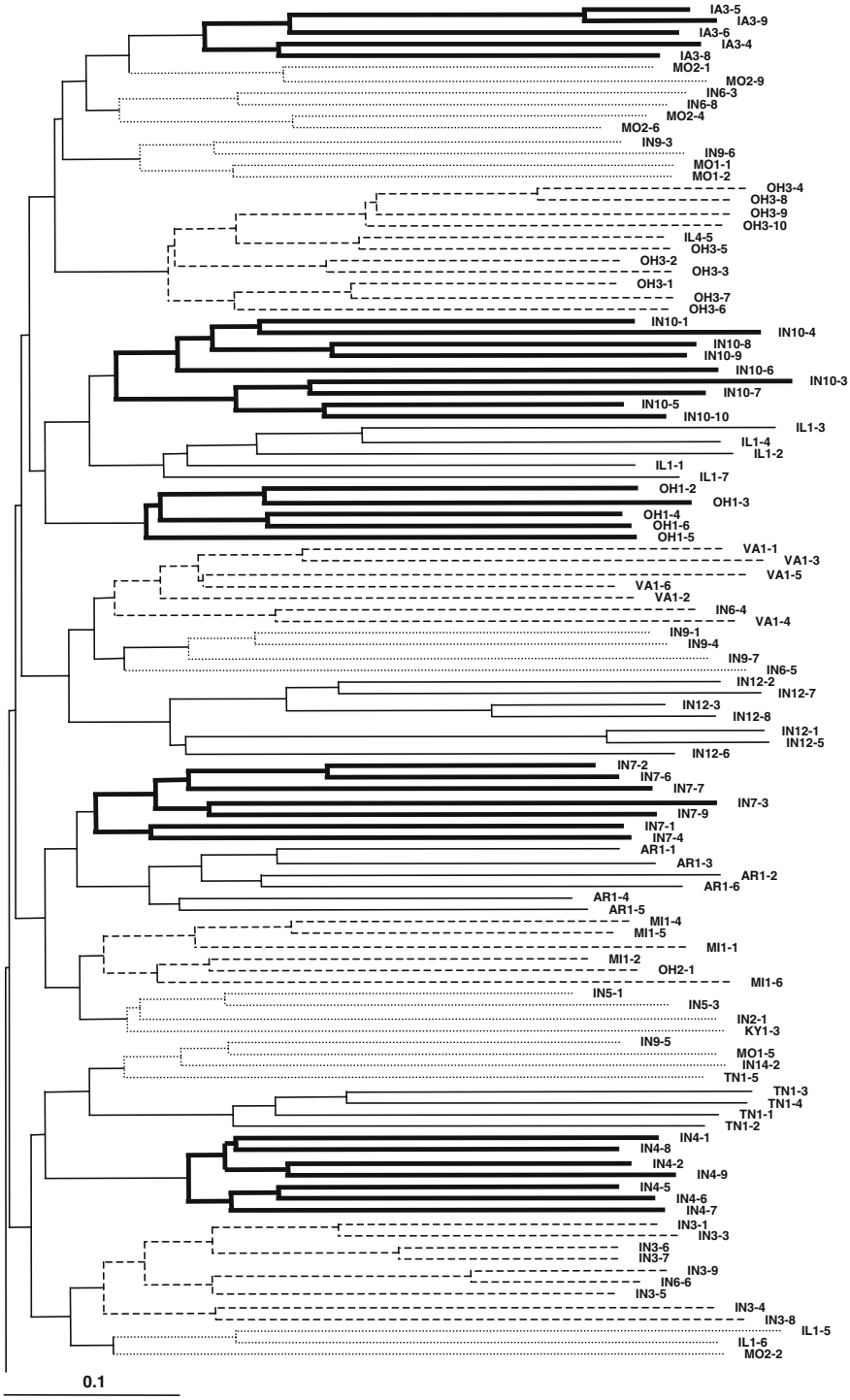
IA3-5
IA3-9
IA3-6
IA3-4
IA3-8
MO2-1
MO2-9
IN6-3
IN6-8
MO2-4
MO2-6
IN9-3
IN9-6
MO1-1
MO1-2
OH3-4
OH3-8
OH3-9
OH3-10
IL4-5
OH3-5
OH3-2
OH3-3
OH3-1
OH3-7
OH3-6
IN10-1
IN10-4
IN10-8
IN10-9
IN10-6
IN10-3
IN10-7
IN10-5
IN10-10
IL1-3
IL1-4
IL1-2
IL1-1
IL1-7
OH1-2
OH1-3
OH1-4
OH1-6
OH1-5
VA1-1
VA1-3
VA1-5
VA1-6
VA1-2
IN6-4
VA1-4
IN9-1
IN9-4
IN9-7
IN6-5
IN12-2
IN12-7
IN12-3
IN12-8
IN12-1
IN12-5
IN12-6
IN7-2
IN7-6
IN7-7
IN7-3
IN7-9
IN7-1
IN7-4
AR1-1
AR1-3
AR1-2
AR1-6
AR1-4
AR1-5
MI1-4
MI1-5
MI1-1
MI1-2
OH2-1
MI1-6
IN5-1
IN5-3
IN2-1
KY1-3
IN9-5
MO1-5
IN14-2
TN1-5
TN1-3
TN1-4
TN1-1
TN1-2
IN4-1
IN4-8
IN4-2
IN4-9
IN4-5
IN4-6
IN4-7
IN3-1
IN3-3
IN3-6
IN3-7
IN3-9
IN6-6
IN3-5
IN3-4
IN3-8
IL1-5
IL1-6
MO2-2
0.1

Figure 2. The unrooted Neighbor-Joining Dendrogram based upon the degree of allele-sharing between individuals. Only those families that had $n \geq 5$ individuals (29 families) were included in this analysis. The first two letters of each branch label gives the abbreviation of the state of origin, the first number indicates the family it came from, while the dashed number indicates the individual analyzed. Bold lines (▬) indicate all individuals from the half-sib family included within the group; Solid lines (——) means a majority of the individuals are within the group; Dashed lines (- - - -) all or majority of individuals are within the group with one inclusion; Dotted lines (.....) no clear grouping of the half-sib family.

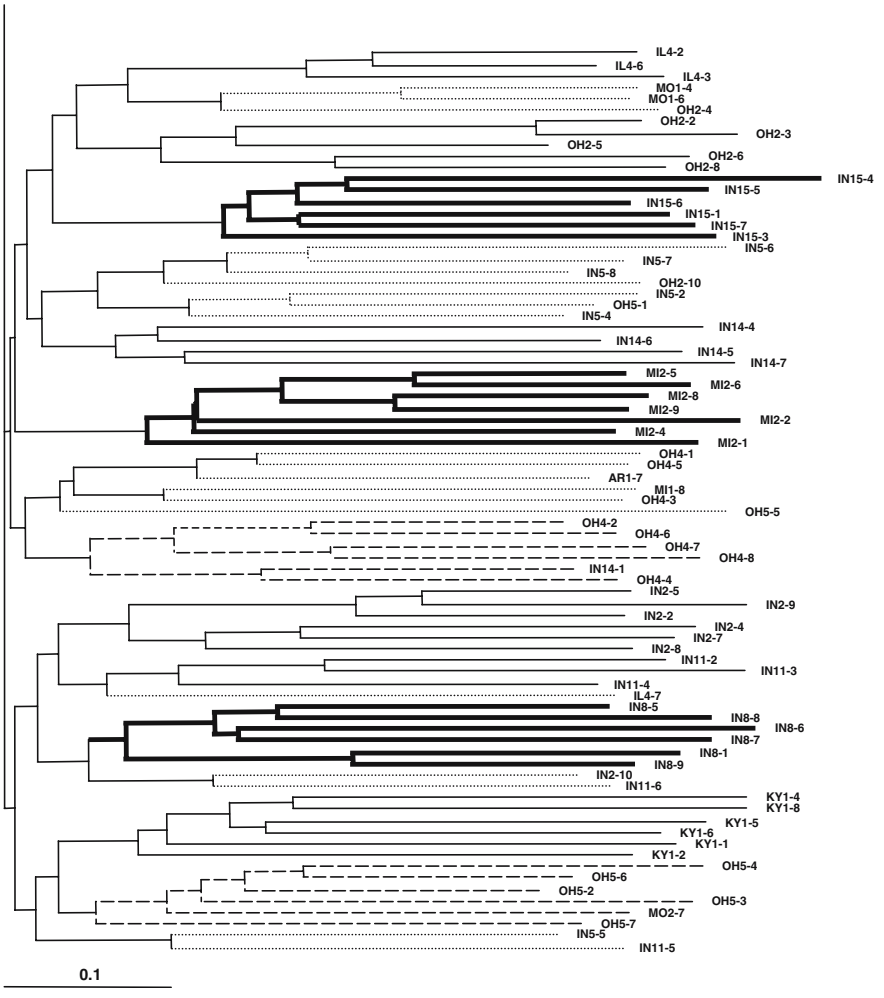

Figure 2. Continued

Discussion

It is not surprising that the 12 microsatellite loci displayed high levels of genetic diversity in our broad geographical sample, given that high levels of allozyme diversity have been documented for black walnut (Rink et al. 1989, 1994; Busov et al. 2002) and since degree of polymorphism was a key criterion in our locus selection process. The ranges of allelic richness and expected heterozygosity observed across the 12 loci are typical of those observed for microsatellites developed in other temperate, angiosperm tree species such as oaks (Dow and Ashley 1996; Lexer et al. 1999), black cherry (Downey and Iezzoni 2000), and black poplar (van der Schoot et al. 2000). High levels of allelic richness and expected heterozygosity in the black walnut markers resulted in extremely low probabilities of pairwise genetic identity (for genetic fingerprinting) and very high exclusion probabilities (for paternity or parentage analyses). Hence this set of markers undoubtedly provides a robust and powerful tool for use both in broad and fine scale population genetic applications (Parker et al. 1998). Broad scale applications include elucidating the genetic structure and evolutionary history of the species as a whole; fine scale applications include studies of pollen dispersal, seed dispersal, within-stand spatial genetic structure, and the genetic and ecological effects of fragmentation. We are currently engaged in all of these types of studies in black walnut using this set of markers.

Our estimate of the fixation index (f) across all 12 loci and for our whole sample of 222 trees was surprisingly low (0.052), indicating that, even over such a broadly sampled area, most of the genetic diversity resides within, rather than among populations. There are a number of factors that can cause positive or negative deviations of f from zero (Non-random mating, 'Wahlund effect', artificial or natural selection). Fundamentally, our estimated value of f is biased and only tentatively represents what the true value of f would be from each of the populations sampled within our study. We utilized the estimated f as a statistic to screen the original set of nSSRs for null alleles to produce a more reliable and powerful subset of microsatellites for future population studies. Our ongoing, broad-scale study using a more traditional sampling scheme (moderately-sized, 'random' samples of trees drawn from each of a large number of populations) will not only allow us to better estimate the fixation index for black walnut within its native range, but all three F-statistics, F_{IT}, F_{IS} and F_{ST} (Victory et al., in press).

We found no evidence of selfing based upon the collection of half-sib families from the Salamonie progeny test (estimated multilocus outcrossing rate of 1.000 with a standard error of zero). This contrasts somewhat with previous results in black walnut, based upon allozymes (Rink et al. 1989, 1994), where multilocus outcrossing rates of 0.905 and 0.880 were obtained, based on two different years of nut collection from 26 or 23 naturally occurring maternal trees in Jackson County, IL. These allozyme results indicate that, although black walnut is a predominantly outcrossing species, selfing does occur at a rate

of about 10%. We suggest two possible factors that together might explain the discrepancy between our results and the allozyme results. First, the use of a larger number of loci (12 microsatellites vs. 6–8 isozyme loci) with far higher levels of polymorphism should have vastly reduced the inflationary influence of biparental inbreeding on the multilocus selfing rate estimate. The greater the number of loci used, and the more polymorphic they are, the lesser degree to which multilocus outcrossing rate estimates are downwardly biased by biparental inbreeding (Ritland 2002). Second, since our outcrossing rate estimate was based upon arrays of 38-year-old progeny as opposed to viable embryos from mature nuts, there has been far more opportunity for selection – be it natural or artificial – to remove selfed individuals from our study population. Parentage studies that we are currently conducting in natural populations of black walnut will shed further light on this matter.

Divergence of the Neighbor-Joining dendrogram of individuals from the ideal result – where all individuals in each family cluster together as a monophyletic group – could have several causes: (1) limited power for 12 microsatellites to discern half-sib families based on allele-sharing; (2) errors in any phase of the microsatellite analysis, from sample collection to data set compilation; or (3) errors in the establishment of the progeny test, resulting in partial 'mixing' of the families. Power may be limited by the fact that any given pair of half-sibs are expected, on average, to share only one allele identical by descent at only half of their loci (Thompson 1975). More than 12 microsatellites may be required for a strong enough phylogenetic signal (i.e., a 'perfect' dendrogram) to emerge above the noise generated by chance sharing of alleles between unrelated individuals. As for genotyping errors, the reproducibility check that we performed as part of the locus selection process indicated that, albeit present, they are rare. Hence, the Neighbor-Joining results, together with the fact that three genetically incompatible half-sib groups were uncovered by the mating system (MLTR) analysis, suggest that errors were sometimes committed during the establishment of the Salamonie progeny test.

This finding of potential errors in the Salamonie progeny trial demonstrates that our working set of 12 microsatellite markers will be a powerful tool not only for population genetic studies, but also for more applied applications in the tree improvement and genetic management of black walnut. The establishment of large scale provenance or progeny trials of any species can be a daunting task when you take into consideration the large number of seed or scion wood that must be collected and catalogue from numerous sources, then germinated or grafted, and finally planted at numerous locations. Generally in these tree trials, years and sometimes decades must go by before the final data can be collected. Errors committed at any point along this process may dramatically diminish or skew the final results from these long-term experiments.

Given that errors committed during the breeding cycle have the potential to significantly reduce the genetic gain from tree improvement programs (Vaillancourt et al. 1998), it would seem imperative that microsatellites be utilized, where available, in quality control monitoring of breeding activities.

Paternity analysis based upon microsatellites also can be of great utility for analyses of seed orchard efficiency, resulting in improvements in seed orchard design (Adams et al. 1992), or even the use of polymix (or open) pollinations within a full-sib breeding strategy (Lambeth et al. 2001). Finally, given the extremely high value of black walnut veneer, there is demand for the ability to reliably genetically fingerprint individual trees to allow verification of clonal identity or, even, prosecution of timber theft and wood forensics (White et al. 2000; Deguilloux et al. 2002). The powerful and robust set of microsatellites presented here clearly fulfills this demand.

Acknowledgements

We would like to thank our research colleagues Brian Beheler, Ron Burns, and a number of undergraduate students who assisted in the early stages of the Salamonie black walnut project. We would like to extend our appreciation to the Indiana Department of Natural Resources for granting us permission to sample the black walnut provenance trial at the Salamonie River State Reservoir. Funding for this project was provided by the USDA Forest Service, Purdue University, the Indiana Hardwood Lumbermen Association, the National Hardwood Lumbermen Association and the van Eck Foundation.

References

Adams W.T., Birkes D.S. and Erickson V.J. 1992. Using genetic markers to measure gene flow and pollen dispersal in forest tree seed orchards. In: Wyatt R. (ed.), Ecology and Evolution of Plant Reproduction. Chapman and Hall, New York, pp. 37–61.

Aldrich P.R., Hamrick J.L., Chavarriaga P. and Kochert G. 1998. Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. Mol. Ecol. 7: 933–944.

Beineke W.F. 1974. Recent changes in the population structure of black walnut. In: Polk R.B. (ed.), Proceedings of the Eighth Central States Forest Tree Improvement Conference.October 11–13, 1972. School of Forestry, Fisheries and Wildlife, University of Missouri, Columbia, MO, pp. 43–46.

Beineke W.F. 1989. Twenty years of black walnut genetic improvement at Purdue University. North. J. Appl. Forest 6: 68–71.

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E., Kidd J.R. and Cavalli-Sforza L.L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455–457.

Busov V.B., Rink G. and Woeste K. 2002. Allozyme variation and mating system of black walnut (*Juglans nigra* L.) in the Central Hardwood Region of the United States. Forest Genet. 9: 319–326.

Chase M.R., Moller C., Kesseli R. and Bawa K.S. 1996. Distant gene flow in tropical trees. Nature 383: 398–399.

Clausen K.E. 1983. Age-age correlations in black walnut and white ash. In: Proceedings of the 7th North American Forest Biology Workshop: Physiology and Genetics of Intensive Culture. July 26–28, 1982. University of Kentucky, Lexington, Kentucky, pp. 113–117.

Craft K.J., Ashley M.V. and Koenig W.D. 2002. Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. Am. J. Bot. 89: 1792–1798.

Deguilloux M.F., Pemonge M.H. and Petit R.J. 2002. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. Phil. Roy. Soc. Lond. B Biol. 269: 1039–1046.

Dow B.D. and Ashley M.V. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol. Ecol. 5: 615–627.

Downey S. and Iezzoni A.L. 2000. Polymorphic DNA markers in black cherry (*Prunus serotina*) are identified using sequences from sweet cherry, peach, and sour cherry. J. Am. Soc. Hort. Sci. 125: 76–80.

Felsenstein J. 1993. PHYLIP (Phylogeny Inference Package). Version 3.5c. Department of Genetics, University of Washington, Seattle (Program available from: http://evolution.genetics.washington.edu/phylip.html).

Glaubitz J.C., Murrell J.C. and Moran G.F. 2003. Effects of native forest regeneration practices on genetic diversity in *Eucalyptus consideniana*. Theor. Appl. Genet. 107: 422–431.

Harlow W.M., Harrar E.S. and White F.M. 1979. Textbook of Dendrology, 6th ed. McGraw-Hill Book Company, New York, 510 pp.

Jamieson A. and Taylor St. C.S. 1997. Comparison of three probability formulae for parentage exclusion. Anim. Genet. 28: 397–400.

Lambeth C., Lee B.C., O'Malley D. and Wheeler N. 2001. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. Theor. Appl. Genet. 103: 930–943.

Lefort F. and Douglas G.C. 1999. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species *Acer*, *Fraxinus*, *Prunus* and *Quercus*. Ann. Forest Sci. 56: 259–263.

Lewis P.O. and Zaykin D. 2001. Genetic Data Analysis: Computer Program for the Analysis of Allelic Data. Version 1.1 (d16c). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html.

Lexer C., Heinze B., Steinkeller H., Kampfer S., Ziegenhagen B. and Glossl J. 1999. Microsatellite analysis of maternal half-sib families of *Quercus robur*, pendunculate oak: detection of seed contaminations and inference of the seed parents from the offspring. Theor. Appl. Genet. 99: 185–191.

Luikart G. and England P.R. 1999. Statistical analysis of microsatellite DNA data. Trends. Ecol. Evol. 14: 253–255.

Minch E., Ruiz-Linares A., Goldstein D., Feldman M. and Cavalli-Sforza L.L. 1996. Microsat (version 1.5b): A Computer Program for Calculating Various Statistics on Microsatellite Allele Data (Program available from: http://hpgl.stanford.edu/projects/microsat/).

Paetkau D., Waits L.P., Clarkson P.L., Craighead L., Vyse E., Ward R. and Strobeck C. 1998. Variation in genetic diversity across the range of North American brown bears. Conserv. Biol. 12: 418–429.

Parker P.G., Snow A.A., Schug M.D., Booton G.C. and Fuerst P.A. 1998. What molecules can tell us about populations: choosing and using a molecular marker. Ecology 79: 361–382.

Rink G., Carroll E.R. and Kung F.H. 1989. Estimation of *Juglans nigra* L. mating system parameters. Forest Sci. 35: 623–627.

Rink G., Zhang G., Jinghua Z., Kung F.H. and Carroll E.R. 1994. Mating parameters in *Juglans nigra* L. – seed orchard similar to natural population estimates. Silvae Genet. 43: 261–263.

Ritland K. 2002. Extensions of models for the estimation of mating systems using *n* independent loci. Heredity 88: 221–228.

Ritland K. and Jain S.K. 1981. A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. Heredity 47: 35–52.

Saitou N. and Nei M. 1987. The Neighbor-Joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Streiff R., Ducousso A., Lexer C., Steinkellner H., Gloessl J. and Kremer A. 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Mol. Ecol. 8: 831–841.

Tabbener H.E. and Cottrell J.E. 2003. The use of PCR based DNA markers to study the paternity of poplar seedlings. Forest Ecol. Manage. 179: 363–376.

Thompson E.A. 1975. The estimation of pairwise relationships. Ann. Hum. Genet. 39: 173–188.

Vaillancourt R.E., Skabo S. and Gore P.L. 1998. Fingerprinting for quality control in breeding and deployment. Aust. Forestry 61: 207–210.

van der Schoot J., Pospiskova M., Vosman B. and Smulders M.J.M. 2000. Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor. Appl. Genet. 101: 317–322.

Victory E., Glaubitz J., Rhodes O.E. and Woeste K. 2006. Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. Am. J. Bot. (In Press)

Vyas D., Sharma S.K. and Sharma D.R. 2003. Genetic structure of walnut genotypes using leaf isozymes as variability measure. Sci. Hort.-Amsterdam 97: 141–152.

Waits L.P., Luikart G. and Taberlet P. 2001. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. 10: 249–256.

Weber J.L. and May P.E. 1989. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–396.

Weir B.S. and Cockerham C.C. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38: 1358–1370.

White E., Hunter J., Dubetz G., Brost R., Bratton A., Edes S. and Sahota R. 2000. Microsatellite markers for individual tree genotyping: application in forest crime prosecutions. J. Chem. Technol. Biotechnol. 75: 923–926.

Williams R.D. 1990. *Juglans nigra* L.: black walnut. In: Burns R.M. and Honkala B.H. (eds), (tech. coords.), Silvics of North America: Vol. 2, Hardwoods. Agricultural Handbook 654. U.S. Department of Agriculture, Forest Service, Washington DC, pp. 391–399.

Woeste K.E. 2002. Heartwood production in a 35-year-old black walnut progeny test. Can. J. Forest Res. 32: 177–181.

Woeste K., Burns R., Rhodes O. and Michler C. 2002. Thirty polymorphic nuclear microsatellite loci from black walnut. J. Hered. 93: 58–60.