Use of Microsatellite Markers in an American Beech (Fagus grandifolia) Population and Paternity Testing

By J. L. Koch1), D. W. Carey1) and M. E. Mason2)

(Received 22nd October 2008)

Abstract

Cross-species amplification of six microsatellite markers from European beech (Fagus sylvatica Linn) and nine markers from Japanese beech (Fagus crenata Blume) was tested in American beech (Fagus grandifolia Ehrh.). Three microsatellites from each species were successfully adapted for use in American beech and were found to be highly polymorphic, with 4-22 alleles at each locus and an expected heterozygosity value of 0.291 to 0.913. Twenty-five trees (including two clonal clusters) from a mature stand were sampled and genotyped to compute population statistics. No linkage disequilibrium between pairs of loci was detected, and the marker loci indicated that the population is at Hardy-Weinberg equilibrium. The markers were also used to genotype two full-sibling families consisting of a combination of American beech and Japanese beech.
bined total of 99 individuals and were found to contain sufficient genetic information to assign paternity using a maximum likelihood method.

Key words: beech bark disease, microsatellite markers, paternity testing, population genetics.

Introduction

Simple sequence repeats (SSRs, also called microsatellites) are becoming the markers of choice because of their high level of polymorphism, high information content, co-dominant inheritance, reproducibility, locus specificity, and transferability in linkage mapping (Gianfranceschi et al., 1998). For a number of tree species, SSRs have been used to infer paternity and estimate gene flow. In each case, only a few SSR loci (4-8) were needed to achieve high exclusionary power (Dow and Ashley, 1998; Jones et al., 2008; Lian et al., 2001; Stieff, 1999). Microsatellite markers are also the markers of choice in studies aimed at assessing genetic diversity and gene conservation in tree species such as white pine (Marquardt et al., 2007), black poplar (Storme et al., 2004) and black walnut (Victory et al., 1998; Jones et al., 2008; Lian et al., 2001; Stieff, 1999). Because development of microsatellite markers is expensive and time consuming, and heterologous species often share marker loci, SSR primers often are adapted from closely related species. Previous work has shown successful transferability among members of the Fagaceae family within the same genus such as in Quercus (Steinkeller et al., 1997; Aldrich et al., 2003) and Castanea (Kubisiak and Roberts, 2006; Wang, 2008). In Fagus, transferability between the closely related species F. crenata and F. japonica (Tanaka et al., 1999) and F. sylvatica and F. orientalis (Pastorelli et al., 2003) has also been reported. However, there are no previously published reports of transferability between Fagus grandifolia and other members of the genus Fagus.

Here we report the successful amplification of heterologous primers from European beech (Fagus sylvatica) and Japanese beech (Fagus crenata) in American beech (Fagus grandifolia). The loci have been evaluated for their level of polymorphism and power of discrimination in Fagus grandifolia. In addition, the markers were used to compute preliminary population statistics and to confirm paternity in full-sib families.

Materials and Methods

Study area and plant material

Breeding experiments and population sampling were carried out at Ludington State Park, Ludington, Michigan, USA. Twenty-three trees within a half-mile radius and two clusters of trees of root-sprout origin within 1.25 miles were sampled for DNA marker work (Ludington population). Four of the trees that were not part of the clusters were selected as parents (1504, 1505, 1506 and 1510). Controlled cross-pollinations and seed germination were carried out as described previously (Koch and Carey, 2004). The half-sib progeny used for analysis included three open-pollinated families (1504, 1506 and 1510) consisting of a total of forty-five progeny (N=17, 17 and 12, respectively). The full-sib progeny population consisted of two cross-pollinated families, one with fifty-three progeny (1506 x 1504) and the other having forty-six progeny (1505 x 1504). Seedlings were raised in containers in the greenhouse prior to sampling.

DNA extraction

DNA extraction was carried out using a modified version of the method published by Doyle and Dickson (1987). For SSR analysis, DNA was extracted from either dormant buds or leaf tissue. Tissues were kept stored at −80°C until used. The extraction buffer contained 3% hexadecyltrimethylammonium bromide (CTAB), 200 mM Tris pH 8.0, 30 mM ethylenediamine tetraacetic acid (EDTA), 1% polyvinylpyrrolidone, 1% polyethylene glycol 8000, 1.4 M NaCl, 10 mM dithiothreitol, 1% polyvinylpolypyrrolidone, and 2% beta-mercaptoethanol. The dithiothreitol, polyvinylpolypyrrolidone and beta-mercaptoethanol were added just prior to use. Tissue was added to prechilled buffer at a ratio of 1:10 (by volume) and ground in a size 23 Kontes Duall homogenizer (Kimble Chase LLC, Vineland, NJ). The homogenate was heated to 37°C and treated with RNase A for 30 minutes followed by extraction with chloroform-isooamyl alcohol (24:1). The supernatant was then transferred to a new tube and Proteinase K (Sigma) was added to a final concentration of 0.05 mg/ml. The samples were incubated at 70°C for 30 minutes. A 25:24:1 phenol:chloroform:isooamyl alcohol extraction was performed followed by a second chloroform-isooamyl alcohol extraction. Supernatants were chilled to 4°C and debris and carbohydrates were removed by centrifugation at 10,000 g for 5 minutes at 4°C. DNA was then precipitated from the supernatant with 0.6 volumes of isopropanol. Purified DNA was resuspended in TE buffer and stored at −20°C until use. All DNA samples were quantified using a Hoescht dye 33258 fluorometric assay and diluted to a working stock concentration of 10 ng/μl prior to polymerase chain reaction (PCR) amplification.

Microsatellite analysis

Previously published PCR primer sequences for nine nuclear microsatellite loci in F. crenata (Tanaka et al., 1999) and for seven nuclear SSR microsatellite loci in F. sylvatica (Pastorelli et al., 2003) were screened in F. grandifolia. Three of the primer pairs from F. sylvatica and three of the primer pairs from F. crenata each amplified polymorphic loci in F. grandifolia. However, despite attempts to optimize the amplification conditions, the primers from F. crenata resulted in amplification of a high number of nonspecific bands. To improve the specificity of these primers, the PCR fragments were cloned using the pGEM®-T cloning kit (Promega Corp., Madison, WI) so the loci could be sequenced and primers based on F. grandifolia sequences could be developed. Sequencing of the cloned inserts was performed by the Plant-Microbe Genomics Facility at the Ohio State University using the 3730 DNA analyzer and Big Dye Terminator™ cycle-sequencing kit (Applied Biosystems, Foster City, CA). Both forward and reverse sequences were obtained by using T7 and SP6 primers. New
primers were designed, internal to the original primers, based on the *F. grandifolia* sequences obtained. In the case of primer FG-7, not enough non-repeat *F. grandifolia* sequence was obtained, so this primer consists of the *F. crenata* sequence with three *F. grandifolia* specific nucleotides added to 3-prime end (Table 1).

Amplification reactions were carried out in 96 well plates with primers from Integrated DNA Technologies, Inc (Coralville, IA). One primer of each pair was labeled at the 5-prime end with Cy 5. Reactions using FG5, FG7, FS3-04, and FS4-46 contained 2 mM MgCl₂, 0.4 mM dNTPs, 0.4 µM of each primer, 0.8 units of Vent (exo-) DNA polymerase (New England BioLabs, Ipswich, MA) and 10 ng of DNA template in a total volume of 20 µl. Reactions using FS1-15 and FG11-2 were the same except that one unit of Taq DNA polymerase (Promega Corp., Madison, WI) was used instead of Vent (exo-) DNA polymerase. In addition, the concentration of dNTPs was reduced to 0.2 mM and the primer concentration to 0.2 µM. The reactions were placed in an Eppendorf Mastercycler gradient thermocycler (Hamburg, Germany) and the cycling profile used consisted of a denaturation step for 5 minutes at 95°C, then 30 cycles of 95°C for 5 seconds, 92°C for 25 seconds, and 15 seconds at the Tm of the primer (see Table 1) followed by 72°C for 1 minute. A final extension step at 72°C was carried out for 7 minutes. When Taq DNA polymerase was used, the initial denaturing step was reduced to 2 minutes. The resulting reaction products were separated on denaturing acrylamide gels consisting of 6% acrylamide and 7 M urea. A size marker was generated by sequencing a previously characterized clone with a 5-prime Cy 5 labeled T7 primer and the SequiTherm Excel II DNA sequencing kit (Epitecte Biotechnologies, Madison, WI). Acrylamide gels were visualized using a Molecular Dynamics Storm 860 imaging system (GE Healthcare Bio-sciences Corp., Piscataway, NJ).

**Table 1.** Microsatellite primer sequences, amplification conditions, and size of observed alleles in American beech.

| Microsatellite locus | Primer sequences 5’-3’* | Repeat | Anneal. temp. (°C) | No. of alleles* | Observed allele size range (bp) |
|---------------------|-------------------------|--------|-------------------|----------------|-------------------------------|
| FG5                 | ACTGGGACAAAAAAACAAAAAACA AATGCAATCTCAGGGG AAATAGTATGGTTATGTC ACACCTAGTGTGCTG AG            | (AG)32 | 54                | 16             | 209-257                       |
| FG7                 | GAAAGCCTTGTGCTGAG        | (GA)32 | 54                | 23             | 58-92                         |
| FG11-2              | ATATCCACAAAGTAAAGG       | (AG)7  | 62                | 11             | 250-278                       |
| FS1-15              | ACACGACAAAATAACTCATAC    | (GA)36 | 54                | 21             | 66-120                        |
| FS3-04              | in Pastorelli et al., Mol. Ecol. Notes 2003 (3): 76-78. | (GCT)5(GTT)5 | 62 | 4 | 169-183 | |
| FS4-46              | in Pastorelli et al., Mol. Ecol. Notes 2003 (3): 76-78. | (TGA)23 | 62 | 8 | 194-454 |

Multi-locus avg. 13.8

* determined by screening 25 mature individuals from the Ludington population plus the 45 open-pollinated progeny.

alleles were included in the data set when the segregation pattern could be determined. In the case of FS4-46, the homozygous condition could not be distinguished from the presence of null alleles so individuals showing only one band were excluded for that locus. Individual loci and multi-locus exclusion probabilities were determined by subtracting the nonexclusion probabilities calculated by CERVUS 3.0 from one. Polymorphic information content (PIC) was also calculated for each locus and averaged across all loci. Loci were assessed for deviation from Hardy-Weinberg equilibrium using HW-QuickCheck (Kalinski, 2006). HW-QuickCheck uses Monte Carlo randomization to generate p-values, which is appropriate for the microsatellite data because of the large number of singleton and rare alleles present.

**Data analysis of microsatellites**

The software program CERVUS 3.0 (Marshall et al., 1998; Kalinowski et al., 2007) was used to calculate allele frequency, observed and expected heterozygosity, and nonexclusion probability for each locus. Null alleles were included in the data set when the segregation pattern could be determined. In the case of FS4-46, the homozygous condition could not be distinguished from the presence of null alleles so individuals showing only one band were excluded for that locus. Individual loci and multi-locus exclusion probabilities were determined by subtracting the nonexclusion probabilities calculated by CERVUS 3.0 from one. Polymorphic information content (PIC) was also calculated for each locus and averaged across all loci. Loci were assessed for deviation from Hardy-Weinberg equilibrium using HW-QuickCheck (Kalinski, 2006). HW-QuickCheck uses Monte Carlo randomization to generate p-values, which is appropriate for the microsatellite data because of the large number of singleton and rare alleles present.

GEPHOP 4.0 (Rousset, 2008) was used to calculate linkage disequilibrium and a Bonferroni correction for multiple comparisons was applied (Rice, 1989).

Paternity was assigned to individuals in cross-pollinated families by first estimating the number of potential pollen donors using the open-pollinated progeny. CERVUS 3.0, which relies on a maximum likelihood approach, was used to identify all candidate fathers for the open-pollinated progeny (N=46) given the maternal genotype (1504, 1506 or 1510). An individual was considered a candidate father if a positive LOD score (log of likelihood ratio) was attained. LOD scores are obtained from the natural log of the overall likelihood ratio and were based on the following simulation parameters: 10,000 simulated offspring, 200 candidate fathers, 12.5% of candidate fathers sampled (N= 25), 1% error rate in genotyping, and a minimum of two loci typed. Once the number of potential pollen donors was determined, that number was used to run a second progeny simulation to assign paternity to the control-pollinated progeny. This second simulation used 40 candidate fathers (as determined by the first simulation in open-pollinated progeny) and a sampling rate of 55% of the 40 candidate fathers identified were included in the

DOI:10.1515/sg-2010-0008

edited by Thünen Institute of Forest Genetics
the 25 sampled trees). All other parameters were the same as in the first simulation, except this time self-fertilization was allowed as a possibility. The most likely male parent, including possible self-fertilization, was assigned based on the highest calculated LOD score. Confidence level of the assigned parent is based on the distribution of LOD scores for known parents among the simulated genotypes.

Results

Characterization of microsatellites in Fagus grandifolia

Three out of six primer pairs from *F. sylvatica* (Pastorelli et al., 2003) and three out of nine primer pairs from *F. crenata* (Tanaka et al., 1999) successfully amplified a single, polymorphic locus in *F. grandifolia*. Due to the production of nonspecific bands using the *F. crenata* primers, *F. grandifolia* specific primers were designed and their sequences are listed in Table 1. No such modifications were required for *F. sylvatica* primers. A total of 71 individuals were screened, including the open-pollinated progeny (N=45) and the Ludington population (N=25), using the three *F. grandifolia* (FG) primer pairs and the three *F. sylvatica* (FS) primer pairs. Allelic richness (the number of alleles) was determined at each locus and ranged from four to 22, with an average of 13.8 (Table 1). The total number of alleles identified across all six loci was 87, and 16 of these occurred only once.

![Image](https://example.com/figure1.png)

*Figure 1.* – Allele frequency distribution for six microsatellite loci in twenty-five individuals (Ludington population) of *F. grandifolia*. 

DOI:10.1515/sg-2010-0008
edited by Thünen Institute of Forest Genetics
Table 2 – Ludington population (N = 25) statistics for six microsatellite loci.

| Microsatellite locus | Hs⁰ | Hsᵇ | PICᶜ | HWᵣᵈ | Pₑ⁰ᵉᵗ | Pₑᵇₑᵗ |
|---------------------|-----|-----|------|------|--------|--------|
| FG5                 | .920| .932| .907 | .51  | .824   | .700   |
| FG7                 | .970| .934| .908 | .18  | .829   | .708   |
| FG11-2              | .720| .634| .562 | .21  | .371   | .216   |
| FS1-15              | .275| .958| .913 | .18  | .836   | .719   |
| FS3-04              | .320| .291| .273 | .51  | .156   | .042   |
| FS4-46              | 1.00| .874| .814 | .22  | .672   | .501   |

*a* observed heterozygosity.  
*b* expected heterozygosity.  
*c* polymorphic information content.  
*d* probability that loci deviates from Hardy-Weinberg equilibrium.  
*exclusion probability for paternity analysis (genotype of maternal parent known).*  
*exclusion probability for parentage analysis (genotype of both parents unknown).*  
*incompletely genotyped individuals due to the presence of null alleles, were excluded.*

All six microsatellite markers were used to obtain population statistics by screening 23 mature trees of seedling origin and two additional genotypes of root- or stump-sprout origin from a beech stand in Ludington State Park (Ludington population). Allele frequencies at each locus are reported in Figure 1, illustrating the variation found in these loci. Expected heterozygosities ranged from 0.291 to 0.938 with an average of 0.767 and observed heterozygosities ranged from 0.320 to 1.0 with an average of 0.784 (Table 2). The analysis of these markers in full-sib progeny revealed the presence of null alleles at the FS 4-46 locus, which is consistent with observations reported in F. sylvatica (PASTORELLI et al., 2003). No loci deviated significantly from the expected Hardy-Weinberg equilibrium (Table 2). Additionally, pairwise comparisons among all loci were performed and no significant linkage disequilibrium was detected (overall α = .05). The polymorphic information content (PIC) is reported as an indication of informativeness of the loci as well as usefulness in linkage mapping. A PIC greater than 0.7 is considered best for linkage mapping (HEARNE et al., 1992). The PIC values ranged from 0.273 to 0.907 with an average of 0.723.

Exclusion probabilities for each locus and across all loci are reported in Table 2. The probability that an unrelated male would be genetically excluded from being a possible father in a paternity analysis where the maternal genotype is known is reported as Pₑᵉᵗ. The probability that an unrelated tree would be genetically excluded from being a parent of a younger tree in a parentage analysis is reported as Pₑᵇₑᵗ. This confirms that these alleles contain sufficient information to distinguish parentage, although we chose to conduct the more robust maximum likelihood-based paternity analysis (MARSHALL et al., 1998).

Paternity analysis

CERVUS 3.0 uses a maximum likelihood approach to assign likelihoods to candidate fathers and allows for null alleles, errors in genotyping, and incomplete sampling of candidate fathers (KALINOWSKI et al., 2007; MARSHALL et al., 1998). The program is based on the assumption that the allele frequencies are in Hardy-Weinberg equilibrium. The analysis of the distribution of alleles within the Ludington State Park population sample of F. grandifolia provides evidence that the microsatellite loci tested are in Hardy-Weinberg equilibrium (Table 2) and therefore can be used by CERVUS 3.0 to perform paternity analysis.

To estimate numbers of candidate fathers or potential pollen contaminants in the control-pollinated crosses, paternity analysis was performed using CERVUS 3.0 on the 46 open-pollinated individuals. The 22 additional mature trees sampled were entered as candidate fathers as well as the three parent trees (1504, 1505 and 1506) for a total of 25 candidate fathers sampled from an estimated 200 trees in the stand. CERVUS 3.0 identified between one and six possible fathers per individual and 22 of the 25 sampled trees were identified as possible fathers. An additional 18 unsampled trees were identified as possible fathers for a total of 40 candidate fathers.

Based on these results, a second simulation was performed on the controlled-cross progeny using 40 as the estimated number of candidate fathers, of which 22 had a known genotype. This new simulation was used to conduct the paternity analysis on the cross-pollinated progeny. In all cases, CERVUS assigned the actual pollen parent, 1504, as the most likely male parent (including the mother tree as a possible male parent) based on LOD scores that ranged from 5.84 to 9.11. The paternity assignment, based on the genotypes of the known mother (1505 or 1506), the most likely father, and the individual progeny being tested, was significant with a p-value less than 0.021 for all individuals and a p-value less than 0.010 for 76% of the progeny (p-values are based upon the distribution of LOD scores in the simulation).

Discussion

Microsatellite markers from F. sylvatica and F. crenata were adapted for use in F. grandifolia. Those originating from F. sylvatica were transferred directly, while for the F. crenata markers it was necessary to clone and sequence amplification products to develop F. grandifolia sequence specific primers. Three out of six (50%) of the microsatellite loci from F. sylvatica were transferred to F. grandifolia and three out of nine (33.3%) of the F. crenata ones were transferred, in contrast to the 100% transferability reported between the closely related species F. sylvatica and F. orientalis (PASTORELLI et al., 2003) and F. crenata and F. japonica (TANAKA et al., 1999). These findings support the idea that the proportion of microsatellites that can be transferred between species decreases with increasing genetic distance (STEINKELLNER et al., 1997; WANG, 2008). The allelic richness, or number of alleles per locus, increased in F. grandifolia for FG7, FG11-12, FS1-15 and FS3-4 compared to the number in their respective species of origin.
FG7 increased from 11 alleles in *F. japonica* to twenty-three in *F. grandifolia*. One locus, FG5, maintained the same number of alleles between species, and another, FS4-46, decreased from 10 alleles in *F. sylvatica* to eight in *F. grandifolia*. A number of rare alleles were detected, which is consistent with beech being a wind-pollinated species with a high degree of outcrossing. High outcrossing rates have been reported both in *F. grandifolia* (Kitamura et al., 1998) and in *F. sylvatica* (Rossi et al., 1996). The informativeness of these loci is apparent when looking at the number of alleles and the various allele frequencies (Fig. 1). The high numbers of alleles and high values for expected and observed heterozygosity of these markers resulted in high exclusion probabilities and a high polymorphic information content, indicating that these loci can be a useful part of genetic diversity analysis, paternity analysis, and linkage mapping in *F. grandifolia*.

In addition, because there was no significant deviation from Hardy-Weinberg equilibrium or evidence of linkage disequilibrium, the loci are not likely to be closely linked and are therefore appropriate for use in maximum likelihood-based paternity analysis. Further support that two of the loci, FS3-04 and FS4-46, are not linked is found on the *F. sylvatica* linkage map, where they are located on two separate linkage groups (Scalfi et al., 2004). The exclusion probability given all six loci in *F. grandifolia* is reported as 0.999 for paternity analysis and 0.990 for parentage analysis (Table 2). This indicates that the amount of information and heterozygosity captured across these six loci is sufficient to allow efficient paternity analysis for confirmation of the full-sib families. In fact, when these six microsatellite markers were used for paternity analysis using the maximum likelihood approach, the father tree (1504) was identified as the most likely father, significant at greater than 97% confidence for all of the control-pollinated progeny. In all cases, the pollen parent was selected as the most likely parent over the possibility of self-fertilization. These data support that the controlled cross-pollinations carried out in the field (Koch and Carey, 2004) did not result in any self-fertilization events despite the fact that the flowers could not be emasculated prior to placement of the pollination bags. Furthermore, the data indicate no evidence that any pollen contamination occurred during hand pollinations. Screening for additional informative markers is necessary to continue to increase the available tools for paternity testing, pollen flow analyses and other genetic studies.

Acknowledgements

The authors gratefully acknowledge the technical assistance of Jany Chan, Bill Sickinger and Donna Wilburn in carrying out this work. We also would like to thank T. Poland, K. Woeste, C. Dana Nelson and an unidentified reviewer for commenting on an earlier version of this manuscript. This work was partially funded through a grant from the Special Technology Development Program of Forest Health Monitoring, USDA Forest Service.

References

Aldrich, P. R., M. Jagtap, C. H. Mihuler and J. Romero-Severson (2003): Amplification of North American red oak microsatellite markers in European white oaks and Chinese chestnuts. Silvae Genet. 52: 3–4.

Doyle, J. J. and E. E. Dickson (1987): Preservation of plant samples for DNA restriction endonuclease analysis. Taxon 36: 715–722.

Dow, B. D. and M. V. Ashley (1998): High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. J. of Heredity 89: 62–70.

Gianfranceschi, L., N. Seglias, R. Tarchini, M. Komijanč and C. Gessler (1998): Simple sequence repeats for the genetic analysis of apple. Theor. Appl. Genet. 96: 1069–1076.

Hearne, C. M., S. Ghosh and J. A. Todd (1992): Microsatellites for linkage analysis of genetic traits. Trends Genet. 8: 288–94.

Jones, M. E., M. Shepherd, R. Henry and A. Delves (2008): Pollen flow in *Eucalyptus grandis* determined by paternity analysis using microsatellite markers. Tree Gen. Gen. 4: 37–47.

Kalinowski, S. T., M. L. Taper and T. C. Marshall (2007): Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16: 1099–1106.

Kalinowski, S. T. (2006): HW-QUICKCHECK: an easy-to-use computer program for checking genotypes for agreement with Hardy-Weinberg expectations. Mol. Ecol. Notes 6: 974–979.

Kitamura, K., J. O'Neil, D. F. Whigham and S. Kawano (1998): Demographic genetic analyses of the American beech (*Fagus grandifolia* Ehrh.). Genetic variations of seed populations in Maryland. Plant Species Biol. 13: 147–154.

Koch, J. L. and D. W. Carey (2004): Control cross-pollinations with American beech trees that are resistant to beech bark disease. In: Proceedings of the 14th Central Hardwoods Forest Conference. Gen. Tech. Rep. NE-316. USDA Forest Service, Northeastern Research Station, Newtown Square, PA, 358–364 p.

Kubitsch, T. L. and J. H. Roberts (2006): Genetic structure of American chestnut populations based on neutral DNA markers. In: Steiner, K. C. and J. E. Carlsson, eds. 2006. Restoration of American Chestnut to Forest Lands-Proceedings of a conference and workshop. Natural Resources Report NPS/NCR/CUE/NRR-2006/001. National Park Service, Washington, D.C. 109–122.

Lian, C., M. Miwa and T. Hogetsu (2001): Outcrossing and paternity analysis of *Pinus densiflora* (Japanese red pine) by microsatellite polymorphism. Heredity 87: 88–98.

Marshall, T. C., J. Slate, L. E. B. Kruuk and J. M. Peterkin (1998): Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655.

Marquardt, P. E., C. S. Echt, B. K. Epffer and D. M. Purban (2007): Genetic structure, diversity, and inbreeding of eastern white pine under different management conditions. Can. J. For. Res. 37: 2652–2662.

Pastorelli, R., M. J. M. Smulders, W. P. C. van’t Westende, B. Vosman, R. Giannini, C. Vettori and G. G. Vendramin (2003): Characterization of microsatellite markers in *Fagus sylvatica* L. and *Fagus orientalis* Lipsky. Mol. Ecol. Notes 3: 76–78.

Rice, W. R. (1989): Analyzing tables of statistical tests. Evolution 43: 223–225.
Isolation, Characterization and Phylogenetic Analysis of Nucleotide Binding Site-encoding Disease-resistance Gene Analogues from European Aspen (Populus tremula)

By Yong Zhang1,2,4, Shougong Zhang3, Liwang Qi3, Tao Zhang1, Chungen Wang2, Chengbin Chen2 and Wenqin Song2,5

(Received 5th February 2009)

Abstract

The majority of verified plant disease resistance genes (R genes) isolated to date was of the nucleotide binding site-leucine rich repeat (NBS-LRR) class. The conservation between different NBS-LRR R genes opens the avenue for the use of PCR based strategies in isolating and cloning other R gene family members or analogs (resistance gene analogue, RGA) using degenerate primers for these conserved regions. In this study, to better understand the R gene in European aspen (Populus tremula), a perennial tree, we used degenerate primers to amplify RGA sequences from European aspen. Cloning and sequence characterization identified 37 European aspen RGAs, which could be phylogenetically classified into seven subfamilies. Deduced amino acid sequences of European aspen RGAs showed strong identity, ranging from 30.41 to 46.63%, to toll inter-leukin receptor (TIR) R gene subfamily. BLAST searches with reference to the genomic sequence of P. trichocarpa found 209 highly homologous regions distributed in 28 genomic loci, suggesting the abundance and divergence of NBS-encoding R genes in European aspen genome. Although, numerous studies have reported that plant R genes are under diversifying selection for specificity to evolving pathogens, non-synonymous to synonymous nucleotide substitution (dN/dS) ratio were <1 for NBS domains of European aspen RGA, showing the evidence of purifying selection in this perennial tree. In further analysis, many intergenic exchanges were also detected among these RGAs, indicating a probable role in homogenising NBS domains. The present study permits insights into the origin, diversification, evolution and function of NBS-LRR R genes in perennial species like European aspen and will be useful for further R gene isolation and exploitation.

Key words: Populus tremula, disease resistance gene, nucleotide binding site, resistance gene analogue, phylogenetic analysis.