Abstract

Black locust (Robinia pseudoacacia L.) is a tree species native to North America. The multipurpose tree is cultivated worldwide, but causes problems due to its partially invasive character. The application of nuclear microsatellite markers has many aims in population genetic studies. Here we introduce a very cost-effective method for combining the information of 14 nuclear microsatellite loci into two multiplex PCR sets as a contribution to greater standardisation and more comparable results.

Combined non-exclusion probabilities for clone identification using example populations are estimated at between 1.37*10^-5 and 1.67*10^-11, and for paternity analysis for 1.59*10^-4. The detected weak linkage between some microsatellite loci is not considered to be a substantial restriction to the reliability of the set of microsatellite loci into two multiplex PCR sets as a contribution to greater standardisation and more comparable results.

Optimisation of a multiplex PCR assay of nuclear microsatellite markers for population genetics and clone identification in Robinia pseudoacacia L.

By H. LIESEBACH1) and E. EWALD2)

(Received 11th October 2011)

1) Thünen Institute of Forest Genetics, Sieker Landstraße 2, D-22927 Großhansdorf. Phone ++49-4102-696158, Fax ++49-4102-696200. E-Mail: heike.liesebach@vti.bund.de
2) Thünen Institute of Forest Genetics, Eberswalder Chaussee 3A, D-15377 Waldsieversdorf.
markers in providing an appropriate method for fingerprinting and parentage analysis.

Key words: Black locust; SSRs; multiplexing; population genetics; clone identification; paternity analysis; linkage.

Introduction

The tree species black locust (Robinia pseudoacacia L., Fabaceae) is native to the south-eastern part of the United States and occurs in mountainous regions as an early colonizer of disturbed areas. Black locust has vegetative propagation ability via root suckers and can perform clonal structures in native stands, as was detected by studies with isozyme markers (McCaig et al., 1993; Chang et al., 1998). Black locust flowers are pollinated by insects, and seeds are dispersed by the wind. First mating system observations to describe the generative propagation were carried out by Sürles et al. (1990). Based on a multilocus isozyme study, they mainly found outcrossing mating, but with remarkable shares of inbreeding (average outcrossing 0.83, range 0.46–1.00).

Black locust, as a „multi-purpose tree,” is of increasing importance for forestry in many countries. It is cultivated for its durable timber, and for biomass and honey production. Numerous breeding efforts in many countries have been undertaken on the family and clonal level (Schrock, 1953; Keresztesi, 1983; Mebraitu and Hanover, 1989; Bongarten et al., 1992; Dini-Papanastasi, 2008), including the development of several methods for vegetative propagation (Naijoks et al., 1999; Redei et al., 2002). As an early colonizer, black locust is suitable for recultivation and slope plantings, but this disposition is also responsible for its partially invasive character. Black locust can become an invasive species in some regions, as ascertained, for example, in Germany (Böcker and Dirk, 1998; Böhmer et al., 2001) and Japan (Jung et al., 2009). In these countries, clonal structures have been observed in artificial stands (Hertel and Schneck, 2003; Jung et al., 2009).

The development of highly variable microsatellite markers by Lian and Hogetsu (2002, seven markers), Lian et al. (2004, three markers) and Mishima et al. (2009, eleven markers) followed the increasing interest in more detailed knowledge of genetic structures in natural and artificial, including invasive, black locust stands and their mating systems, as well as in breeding material. Two studies were published recently describing the application of only four microsatellite loci to recognize clonal structures in artificial populations in Japan (Jung et al., 2009; Kurokochi et al., 2010).

Guichoux et al. (2011) reviewed the current trends in microsatellite genotyping and emphasized the need for standardised protocols to enable more comparable results in joint projects. This issue is also under discussion for other species being bred and cultivated, i.e. poplar (Rathmacher et al., 2009), beech (Lefèvre et al., 2011), tomato and wheat (Vosman et al., 2001) or Pacific salmon (Moran et al., 2006; Ellis et al., 2011).

In this study, we developed a cost-effective multiplex PCR method with 14 loci for further population genetic studies, including mating system analysis and identification purposes, and we present first data on linkage and variation levels in a larger number of samples to give a dimension of non-exclusion probabilities for mating system studies. The presented multiplex sets were applied in two studies (Liesebach, 2012; Liesebach and Naijoks, 2012).

Materials and Methods

Plant material

The plant material originates from several samples, mainly from artificial stands in Germany. Among them are semi-natural populations which have not been managed for a long time; selected clones; material from actual breeding programs; open pollinated offspring from single clones, and samples from six planted stands from several seed sources located in Germany/Brandenburg. In addition to Robinia pseudoacacia, a few samples of other Robinia species were tested: an open pollinated offspring family from R. neomexicana A. Gray (37 individuals) and the clone ‘Casque Rouge’ (R. × margarettiae Ashe, R. hispida L. × R. pseudoacacia). The previous total sample size amounts to approximately 1300 individuals.

Microsatellite genotyping

The isolation of total DNA from fresh or frozen leaves or from seeds followed a modified CTAB protocol (Dumont et al., 1995). All 21 available loci were tested with the exception of Rops15, which was described as highly somatic instable (Lian et al., 2004), and RP102, RP211 and RP165, which were characterized by a low variation and a significant departure from Hardy-Weinberg equilibrium (Mishima et al., 2009).

PCRs for the remaining 17 microsatellite loci (Table 1) were carried out with fluorescent dye-labeled forward primers (delivered by biomers.net, www.biomers.net) using the “Multiplex PCR Kit” from Qiagen in accordance with the manufacturer’s instructions. PCRs were carried out in 15 µl reaction volume with 10–50 ng template DNA in the Multiplex PCR Master Mix containing HotStarTaq DNA Polymerase, buffer, dNTP mix and a final concentration of 3 mM MgCl₂. All primers were added with equal concentrations of 0.2 µM. PCRs were carried out in a Biometra TGradient and a Biometra UNOII thermocycler (Göttingen, Germany) with the following temperature profiles: activation step of 15 minutes at 95°C, 28 cycles (denaturation of 30 sec at 94°C, annealing 90 sec, and extension for 60 sec at 72°C) and a final extension of 30 min at 60°C. After optimisation, eight loci were analysed in Multiplex Set 1 at an annealing temperature of 56°C, six loci were analysed in Multiplex Set 2 at an annealing temperature of 63°C (Table 1) and the three loci Rops09, Rops10 and Rops18 were excluded (see below). Amplification products were detected with a Beckman Coulter CEQ 8000 capillary sequencer.

Data analysis

Population genetic parameters such as number of alleles per locus (A/L) and observed heterozygosity (Hₒ); as
well as combined non-exclusion probabilities for the identification of unrelated individuals, sib identity or the identification of the second parent when the first parent is known, were calculated with the software package CERVUS (MARSHALL et al., 1998; KALINOWSKI et al., 2007). The effective number of alleles $N_e$ was estimated by the reciprocal of expected homozygosity.

Homozygous genotypes of null alleles were assumed in the case of repeated missing amplification products in samples with other detectable loci. A further indication for null alleles in heterozygote genotypes is mismatching between mother and offspring individuals when mother and offspring exhibit different alleles in apparent homozygous genotypes. The software package Micro-

### Table 1. Overview of nuclear microsatellite loci applied on *Robinia* samples and their variation parameters.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Reference for primer development</th>
<th>Motif</th>
<th>PCR set</th>
<th>Dye used to label the forward primer</th>
<th>Range [bp]</th>
<th>Number of alleles in <em>R. pseudoacacia</em></th>
<th>Observed heterozygosity</th>
<th>Additional alleles in <em>R. neomexicana</em> and <em>R. hispida</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rops04</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>AC</td>
<td>1</td>
<td>BMN-6</td>
<td>107-112</td>
<td>4&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.3153</td>
<td>1</td>
</tr>
<tr>
<td>Rops05</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>AC</td>
<td>1</td>
<td>DY-751</td>
<td>115-156</td>
<td>14</td>
<td>0.8312</td>
<td>2</td>
</tr>
<tr>
<td>Rops06</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>GT</td>
<td>1</td>
<td>Cy5</td>
<td>117-147</td>
<td>10</td>
<td>0.6841</td>
<td></td>
</tr>
<tr>
<td>RP106</td>
<td>(Mishima et al., 2009)</td>
<td>GT</td>
<td>1</td>
<td>BMN-6</td>
<td>128-138</td>
<td>4</td>
<td>0.6321</td>
<td>1</td>
</tr>
<tr>
<td>RP01B</td>
<td>(Mishima et al., 2009)</td>
<td>CT</td>
<td>1</td>
<td>Cy5</td>
<td>155-175</td>
<td>9</td>
<td>0.8165</td>
<td>1</td>
</tr>
<tr>
<td>Rops10</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>T/AAT</td>
<td>1, test only</td>
<td>BMN-6</td>
<td>181-188</td>
<td>6&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.1659</td>
<td></td>
</tr>
<tr>
<td>RP150</td>
<td>(Mishima et al., 2009)</td>
<td>TC</td>
<td>1</td>
<td>BMN-6</td>
<td>175-202</td>
<td>12</td>
<td>0.8187</td>
<td></td>
</tr>
<tr>
<td>Rops08</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>CA</td>
<td>1</td>
<td>Cy5</td>
<td>193-207</td>
<td>7</td>
<td>0.6741</td>
<td></td>
</tr>
<tr>
<td>Rops16</td>
<td>(Lian et al., 2004)</td>
<td>CT</td>
<td>1</td>
<td>DY-751</td>
<td>194-220</td>
<td>13</td>
<td>0.8627</td>
<td>1</td>
</tr>
<tr>
<td>Rops09</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>TA/A</td>
<td>2, test only</td>
<td>BMN-6</td>
<td>77-127</td>
<td>10&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.0568</td>
<td></td>
</tr>
<tr>
<td>RP035</td>
<td>(Mishima et al., 2009)</td>
<td>TC</td>
<td>2</td>
<td>BMN-6</td>
<td>75-97</td>
<td>10&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.6805</td>
<td></td>
</tr>
<tr>
<td>RP032</td>
<td>(Mishima et al., 2009)</td>
<td>TG</td>
<td>2</td>
<td>DY-751</td>
<td>91-119</td>
<td>11&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.2913</td>
<td></td>
</tr>
<tr>
<td>Rops02</td>
<td>(Lian and Hogetsu, 2002)</td>
<td>AC/AT</td>
<td>2</td>
<td>Cy5</td>
<td>106-142</td>
<td>14&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.7866</td>
<td>2</td>
</tr>
<tr>
<td>RP109</td>
<td>(Mishima et al., 2009)</td>
<td>AG</td>
<td>2</td>
<td>BMN-6</td>
<td>123-145</td>
<td>11</td>
<td>0.7292</td>
<td>2</td>
</tr>
<tr>
<td>RP200</td>
<td>(Mishima et al., 2009)</td>
<td>AG</td>
<td>2</td>
<td>DY-751</td>
<td>136-182</td>
<td>19</td>
<td>0.7814</td>
<td>1</td>
</tr>
<tr>
<td>Rops18</td>
<td>(Lian et al., 2004)</td>
<td>AC</td>
<td>2, test only</td>
<td>DY-751</td>
<td>137-221</td>
<td>10</td>
<td>0.6837</td>
<td></td>
</tr>
<tr>
<td>RP206</td>
<td>(Mishima et al., 2009)</td>
<td>GT</td>
<td>2</td>
<td>Cy5</td>
<td>205-231</td>
<td>13&lt;sup&gt;★&lt;/sup&gt;</td>
<td>0.5799</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>★</sup> assumed null allele because of missing amplification products or mismatches between mother and offspring.
Checker (Van Oosterhout et al., 2004) was used to estimate null allele frequencies based on Hardy-Weinberg equilibrium in some example populations. One of the implemented methods (Brookfield, 1996) considers missing values as true homozygote genotypes of the null allele.

Linkage analyses were carried out for some two-locus combinations derived from available mother-offspring families. Haploid maternal genotypic data were generated from the offspring by subtracting the paternal alleles. Data were omitted in case of identity of maternal and offspring heterozygote genotype since the maternal allele cannot be determined. Recombination frequencies \( r \) with standard deviations were calculated as

\[
r = \frac{R}{N} \pm \frac{1 - R}{N} \sqrt{\frac{R}{N} \cdot \frac{(1 - R)}{N}}
\]

where \( N \) is the total sample size of two-locus combinations and \( R \) is the number of recombinants. Kosambi's mapping function was used to calculate map distances as

\[
d = \frac{1}{4} \ln \frac{1 + 2r}{1 - 2r}
\]

The classical LOD scores of linkage analysis was calculated as

\[
\text{LOD} = \log_{10} \left( \frac{(1 - R/N)^{1-K} \cdot (R/N)^{K}}{0.5^N} \right)
\]

as the decimal logarithm of the likelihood ratio of linkage between the two loci and of independent segregation. The common threshold is LOD \( \geq 3.0 \) for significant linkage as was suggested by Morton (1955), cited in Gerber and Rodolphe (1994).

### Results and Discussion

#### Variation

Initially, 17 microsatellite loci were analysed. After the first approximately 200 samples, three loci were excluded from further analysis to concentrate a maximum of information in a minimum of PCRs. Locus Rops09 was omitted because of its very low variation and a putative null allele (observed heterozygosity 0.057). Locus Rops10 has a rather mononucleotide pattern and was omitted because of a putative highly frequent null allele. About 70% of individuals did not reveal a peak at this locus. Locus Rops18 was excluded because of its wide range from 137 to 221 bp with no alleles between 152 and 209 bp, perhaps caused by a large indel overlaying the dinucleotide pattern. A tendency to allele dropout is suspected because large alleles do not amplify as efficiently as small alleles. Some extreme differences in peak height were observed in heterozygous genotypes combining alleles shorter than 152 bp and longer than 209 bp. The remaining 14 loci were combined in two multiplex PCR sets with 8 and 6 loci, respectively (Table 1).

All loci more or less fit to a dinucleotide base-pair periodicity with the exception of locus Rops05 with a dinucleotide repeat that exhibits an average of 2.2 nucleotides distance between adjacent alleles. This so-called size shift exists between the observed electrophoretic size and the expected repeat unit difference. Similar deviations from the expected periodicity were observed in humans (Amos et al., 2007), in poplar (Liesebach et al., 2010) and salmon (Ellis et al., 2011). Alleles of *R. neomexicana* also match to the *R. pseudoacacia* alleles, as was observed in an offspring family. However, we detected alleles apart from the ladder at five loci in the clone ‘Casque Rouge’, obviously originating from the *R. hispida* parent.

Even though these 14 loci are suitable for clone identification, one locus has to be excluded for population genetic studies, especially mating system and paternity analysis. Locus RP150 shows an indication for duplication. It reveals reproducible genotype patterns with 1 to 4 alleles per sample and partially shows behaviour like a tetraploid locus with a dosage effect, i.e., two small peaks and one large peak.

#### Linkage

Multilocus data evaluations, like the calculation of several non-exclusion probabilities, assume that the loci segregate independently. However, no serious bias is expected in multilocus calculations when loci are not so tightly linked (Marshall et al., 1998). Mapping distances of 10 cM or more are considered as an independent association of markers for fingerprinting and parentage analysis purposes (Slavov et al., 2004).

Here we present first results of linkage analysis in black locust in offspring families from two openly pollinated trees A and B. Forty five two-locus combinations could be tested from 10 heterozygote loci for each of the individuals A and B. Table 2 shows five significant linkages at the LOD score > 3 criterion. The suggested threshold value of 3 is very conservative, as was discussed by Gerber and Rodolphe (1994). Yet further two-locus combinations in this study have LOD scores below 1.6 and have to be considered as independent loci. Two pairs of linked loci Rops05 – RP035 and RP109 – RP200 were observed in both trees A and B, whereas RP106 – RP01B in tree B (Table 2) seems to be unlinked in tree A (LOD = 0.0098, recombination frequency 0.489).

#### Estimation of non-exclusion probabilities

Generally, non-exclusion probabilities should be as small as possible. This could be achieved by a high number of loci and a high level of variation. Nevertheless, it could be advisable to exclude single loci for paternity analysis in the case of high frequencies of null alleles in certain populations. Often they cannot contribute information because of missing amplification products, or their apparent homozygous genotype might be a true heterozygote with the null allele. In the six planted stands of black locust described here, maximum observed frequencies of null alleles per population are 18% for locus Rops04; 23% for RP032, and 29% for RP206. However, such loci can be used as additional information in paternity analysis to exclude candidate fathers in the case of more than one possible candidate.
Some examples for variation parameters and combined non-exclusion probabilities based on 13 microsatellite loci (locus RP150 was excluded because of duplication) were given for six populations to demonstrate the power of the presented set of highly variable microsatellite loci in black locust (Table 3). Calculations were carried out under the condition of unlinked loci. Taking into consideration the observed linkage, the combined non-exclusion probabilities might be slightly underestimated.

A realistic probability for identifying clones originating from vegetative propagation in populations or in breeding materials by their identical multilocus genotypes might be assumed between the estimates for identity by chance for unrelated individuals (on average 1.67E-11) and for sibs (on average 1.37E-05). This implies a highly reliable identification of clones including possible full sibs. The application of the set of microsatellite markers for parentage analysis is not restricted by linkage despite inexact non-exclusion probabilities. It allows a reliable identification of the pollinator among the candidate fathers or the exclusion of all candidates in natural or semi-natural populations with clonal structures or in seed orchards.

**Acknowledgements**

The authors thank Volker Schneck, Gisela Naufors, Jan Engel and Dirk Knoche for their assistance in the collection of *R. pseudoacacia* plant material and Roland Graeff for providing the *R. neomexicana* collection. Partial funding was provided by the Fachagentur Nachwachsende Rohstoffe e.V. (FNRL of the Federal Ministry of Food, Agriculture and Consumer Protection in Germany. The authors also thank two anonymous reviewers for their constructive contributions to improve the manuscript.
Microarray Analysis of Gene Expression in Triploid Black Poplar

By Baoquan Hu1), Bin Wang1), Chunguo Wang1), Wenqin Song1) and Chengbin Chen1),*)

(Received 19th November 2011)

Abstract

Triploidy is a widespread phenomenon in cultivated and natural breeding plants and it can confer some growth advantages. Here, we analyzed genome-wide gene expression in triploid Populus euramericana (black poplar) using the Affymetrix poplar microarray to detect any possible correlation between triploid vigor and a unique gene expression profile. Among the 38,400 transcripts that were detected in triploid poplar, 1,564 and 2,015 genes were up- or downregulated, respectively, compared with the diploid. The majority of the upregulated genes in the triploid were associated with carbon and nitrogen metabolism, especially lignin and secondary metabolism. Other genes upregulated in the triploid included genes involved in sugar transport, and brassinosteroid (BR) and auxin metabolism. Downregulated genes were mostly related to the assembly and biosynthesis of ribosomes and the nucleosome macro-molecular complex. The results suggested that BR and auxin levels were crucial in controlling sugar transport, photosynthesis and cell wall biosynthesis. Downregulated genes were associated with chromatin regulation in the triploid. The information from this analysis could provide an insight into the vigor of triploid poplar.

Key words: Gene expression; Populus; microarray; triploid.

Abbreviations: BR, brassinosteroid; cRNA, complementary RNA; ESTs, expressed-sequence tags; GO, gene ontology; IAA, indole-3-acetic acid; qRT-PCR quantitative real-time polymerase chain reaction; SEA, singular enrichment analysis.

Introduction

Polyplody is a widespread phenomenon in plants, and it is also a major mechanism driving speciation. Approximately 60–70% of flowering plants have undergone polyploidy during their ancestry (Blanc and Wolfe, 2004; Cui et al., 2006; Masterson, 1994), and it is estimated that 15% of angiosperm speciation has been achieved by polyploidization (Rieseberg and Willis, 2007). However, new polyploids often face reproductive isolation from their progenitors due to the triploid reproductive barrier, which is caused by the production of unreduced gametes (Kohler et al., 2010). Therefore, the plants that evolved by polyploidization remain in a triploid state. This speciation process has been defined as the triploid bridge, and is thought to play an important role in plant evolution (Yamauchi et al., 2004).

Triploidy is also found in natural plant groups, and these plants can have growth advantages. There exist triploid varieties in various species of poplar, including Populus tremula, Populus alba (Paulay, 1949), Populus balsamifera and Populus tremuloides (van Bulthoven et al., 1958). Compared with their diploid counterparts, the triploid trees are faster growing, have larger leaves and show greater vigor (Li et al., 2008). Several studies have examined triploid white poplar, which has good growth performance and other desirable properties (Zhang et