Genetic timber tracking of Larix sp. in Eurasia

CÉLINE BLANC-JOLIVET1, YULAI YANBAEV2, BERND DEGEN1
1 Thünen Institute of Forest Genetics, Sieker Landstraße 2, 22927 Großhansdorf, Germany (celine.blanc-jolivet@thuenen.de, bernd.degen@thuenen.de)
2 Bashkirian State University, Z. Validi str.-32, 450076, Ufa, Russia (yanbaev ua@mail.ru)

Abstract

We sampled Larix sp. individuals over Eurasia and conducted genotyping at four chloroplast microsatellite and 253 newly developed SNP loci. Both marker types showed a very clear geographical pattern corresponding in Russia to different Larix species. The best SNP loci were selected to develop next sets of 76 and 13 loci for respectively cost-effective MassARRAY and SNaPShot genotyping. Genetic assignment tests provided theoretical success rates between 73 and 90% according to the marker set used. An analysis of timber samples showed varying amplification success rates, but demonstrated the feasibility of larch timber tracking with molecular markers at the regional level. We further discussed statistical improvements needed for forensic applications.

Key words: Larix sp., illegal logging, genetic timber tracking, genetic assignment, chloroplast microsatellites, Single Nucleotide Polymorphism

Introduction

With the entry in force of new laws for timber importation (EU Timber Regulation (EUTR), the US Lacey Act and the Australia Illegal Logging Prohibition Act prohibit), reliable information on the species and geographical origin are required to avoid placing on the market illegally harvested timber. The trade of larch timber (Larix sp.) exhibits a significant importance in the European market, and is harvested in natural forests of Europe and Russia. However, several Larix species occur in Eurasia, all showing contrasting geographical ranges (ARAKI et al. 2008, KHATAB et al. 2008, ORESHKOVA et al. 2013, POLEZHAEVA et al. 2010, SEMERIKOV and LASCOUX 1999, SEMERIKOV and LASCOUX 2003, SEMERIKOV et al. 2013). Since the anatomical timber identification of Larix species is not possible, other timber tracking methods are necessary to control the species declaration in the frame of timber market regulations. We present here the genetic variation of Larix sp. within Eurasia with chloroplastic microsatellite (cpSSRs) and newly developed SNP markers and show how this genetic variation can be used for the control of species and geographical origin.

Material and methods

Fresh cambium, buds or needles were collected from Larix sp trees ranging from the French Alps to the eastern cost if Russia, and a from a provenance trial located in Germany. DNA was extracted according to DUMOLIN et al. (1995). One Larix decidua and one Larix gmelinii tree were sampled from the collection at the Thünen Institute in Grosshansdorf and were used for RAD sequencing to detect putative SNP loci (RADseq, Baird et al. 2008, SLAVOV et al. 2014). 2094 individuals were genotyped at four cpSSRs (SEMERIKOV et al. 2013) on a capillary sequencer (Figure 1) and 1885 individuals were genotyped at 253 SNP loci (249 nuclear, 1 chloroplast and 3 mitochondrial) derived from the RADseq analysis and from MOSCA et al. (2012) on a MassARRAY platform (Figure 2). Cluster analysis was conducted to detect genetic groups (cpSSRs: SAMOVA (DUPANLOUP et al. (2002)), SNPs: STRUCTURE (PRITCHARD et al. 2000)). Since genotyping
of SNPs is more robust on timber material, we addressed the success rate of genetic assignment methods on the SNP dataset based on the results of the grouping of individuals according to the cluster analysis with self-assignment tests together with exclusion probability calculation (CORNUT et al. 1999, DEGEN et al. 2017, RANNALA and MOUNTAIN 1997). Most samples from West Russia were not included due to timing reasons.

**Figure 1:** Frequency of most common haplotypes observed in *Larix* sp. samples using four cpSSRs. The white colour in the pie charts represents the frequency of rare haplotypes.

**Figure 2:** Frequency of the genetic clusters and admixed individuals (“Hybrids”) in the sampled populations together with the corresponding *Larix* species.

We compared analysis methods and the complete set of loci with a reduced set of 76 loci optimised for informativeness. Finally, to address the feasibility to cost-effectively trace the origin of larch timber, we
genotyped three *L. decidua* timber samples with known origin and four blind timber samples at the reduced set of 76 loci. To address genotyping errors and decrease in DNA quality of timber, we included in the sampling needles or cambium from the logged tree and repeated each timber samples four times. Finally, we developed and optimised a SNaPShot Assay with highly informative 13 loci based on the primers used for MassARRAY analysis.

Amplification rate of timber samples ranged from 58 to 100 %, but genotyping errors among replicated samples were observed in 0 to 18 loci (Table 2). Most genotyping errors in timber samples were due to allele dropout at heterozygous loci. All blind samples were successfully assigned to their true population of origin, while two timber samples from Germany would have been misassigned to a population in Austria, albeit not very far from the true population. This result was not only due to the low amplification rate of the sample RP1, but also on the weakness of the reference data for genetic assignment at the population level. These results therefore suggest that genetic tracking should be done at a regional and not at a population level.

The SNaPShot assay was successfully tested on reference material from all *Larix* sp. regions of origin (Figure 3). Based on the reference data obtained with the MassARRAY genotyping, this set of 13 loci correctly assigned 90% of the samples to the most likely group, but here no exclusion probabilities were calculated. We nevertheless do not expect the assignment success rate to be substantially lower than at the other set of loci. The selection of loci showing strong differentiation among groups can therefore contribute to the development of cost-effective methods for timber tracking without strong decrease in statistical power.

### Table 1: Assignment success rates of the genetic assignment based on the Bayesian criteria (Rannala and Mountain, 1997) and on the nearest-neighbour approach (DEGEN et al. 2017)

<table>
<thead>
<tr>
<th></th>
<th>Bayesian criteria</th>
<th>Nearest-neighbour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All loci</td>
<td>Selected loci</td>
</tr>
<tr>
<td>Number of individuals</td>
<td>775</td>
<td>775</td>
</tr>
<tr>
<td>Assignment success (%)</td>
<td>76.0</td>
<td>83.6</td>
</tr>
</tbody>
</table>

### Table 2: Amplification rates and genotyping errors of timber samples together with genetic assignment results based on the Bayesian criteria (RANNALA and MOUNTAIN 1997) obtained from the genotyping of 76 loci

<table>
<thead>
<tr>
<th>Sample</th>
<th>Origin</th>
<th>Amplification success timber/fresh material [%]</th>
<th>Number of loci with mismatch</th>
<th>Assigned to</th>
<th>True Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>RP1</td>
<td>Germany</td>
<td>58</td>
<td>18</td>
<td>Austria</td>
<td>Germany</td>
</tr>
<tr>
<td>RP2</td>
<td>Germany</td>
<td>100</td>
<td>1</td>
<td>Austria</td>
<td>Germany</td>
</tr>
<tr>
<td>M1</td>
<td>France</td>
<td>75</td>
<td>17</td>
<td>France</td>
<td>France</td>
</tr>
<tr>
<td>1</td>
<td>?</td>
<td>100</td>
<td>0</td>
<td>Germany</td>
<td>Germany</td>
</tr>
<tr>
<td>2</td>
<td>?</td>
<td>100</td>
<td>0</td>
<td>Germany</td>
<td>Germany</td>
</tr>
<tr>
<td>3</td>
<td>?</td>
<td>100</td>
<td>4</td>
<td>Germany</td>
<td>Germany</td>
</tr>
<tr>
<td>4</td>
<td>?</td>
<td>75</td>
<td>15</td>
<td>France</td>
<td>France</td>
</tr>
</tbody>
</table>
Conclusions

The genotyping at newly developed SNP loci was very effective for cost-effective timber tracking purposes. However, average success rates are lower than 90%, which still hinders the use of this method for forensic timber identification. Further work is therefore needed. Besides the improvement of the marker set which would require further NGS sequencing of individuals from regions where success rates of our method is low, estimation of error rates might in some cases still forensically validate the presently described method. The calculation of likelihood ratios allows estimating the probability that the assignment test returned an erroneous result, and rely on the choice of a likelihood ratio threshold which minimise the probability or type 1 and type 2 errors. A preliminary analysis showed that likely ratio tests would be very helpful, but with limited interests to distinguish among L. gmelinii and L. olgensis.

Acknowledgements

We would like to thank Vivian Kuhlenkamp and Olca Cakmak for their work in the laboratory. Timber samples for the blind testing were provided by Barbara Fussi (Germany) and Pascal Trézalet (France). The Northwest German Forest Research Institute (Axel Noltensmeier) supported us with the sampling of trees from a European Larch provenance trial. The SNP genotyping was conducted at the Genomic and Sequencing Facility of Bordeaux.

References


