Genetic timber tracking of *Larix* sp. in Eurasia

CÉLINE BLANC-JOLIVET[1], YULAI YANBAEV[2], BERND DEGEN[1]

[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 developped 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 adressed 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 (CORNUET 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.

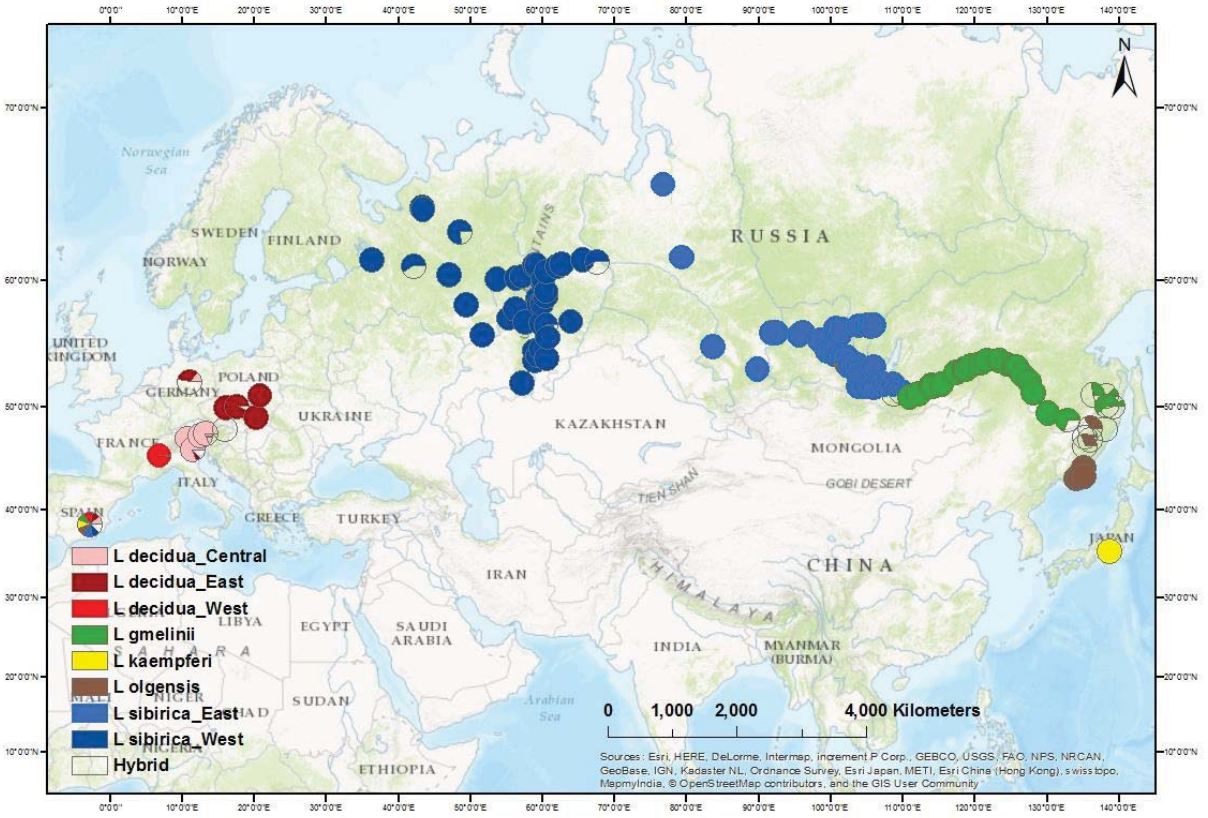

Figure 2: Frequency of the genetic clusters and admixed individuals ("Hybrids") in the sampled populations together with the corresponding *Larix* species.

We compared analysis methods and the complete set of loci with a reduced set of 76 loci optimised for informativeness. Finally, to adress the feasilbility 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 adress 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 developped 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)

	Bayesian criteria		Nearest-neighbour	
	All loci	Selected loci	All loci	Selected loci
Number of individuals	775	775	914	913
Assignment success (%)	76.0	83.6	78.3	73.8

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

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

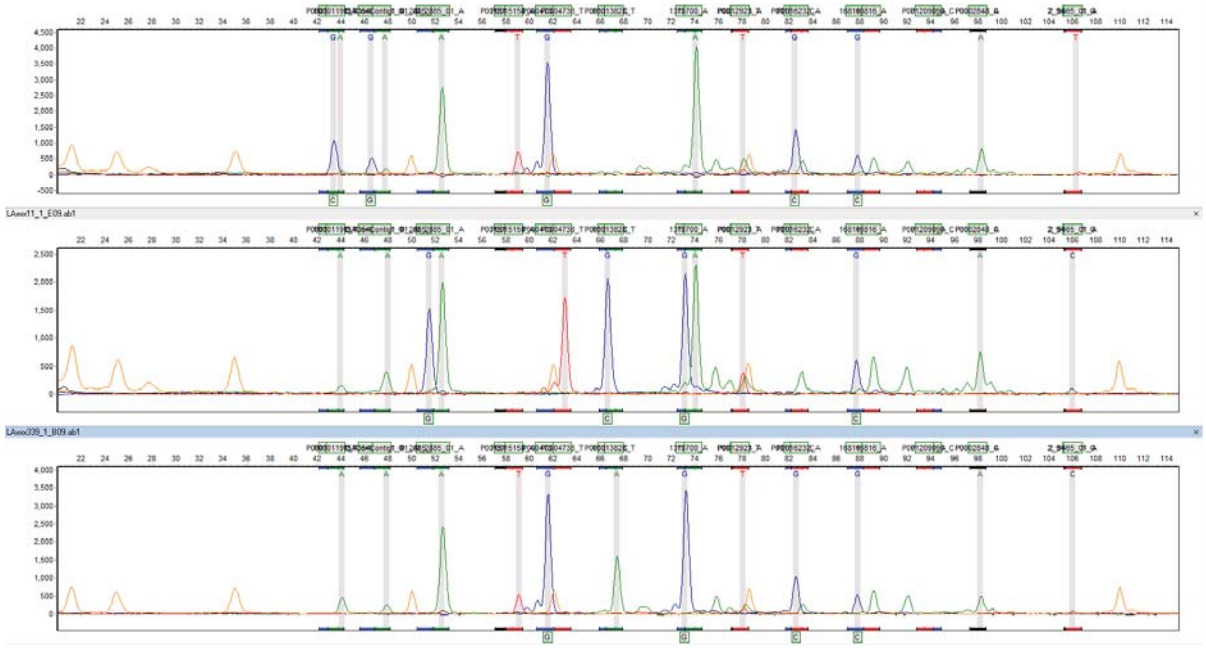

Figure 3: SNaPShot analysis at a combination of 13 loci.

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

ARAKI NHT, KHATAB IA, HEMAMALI KKGU, INOMATA N, WANG X-R, SZMIDT AE (2008) Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes. Tree Genet. Genomes 4: 611-623.

BAIRD NA, ETTER PD, ATWOOD TS, CURREY MC, SHIVER AL, LEWIS ZA et al. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. Plos One 3: 10.

CORNUET JM, PIRY S, LUIKART G, ESTOUP A, SOLIGNAC M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics 153: 1989-2000.

DEGEN B, BLANC-JOLIVET C, STIERAND K, GILLET E (2017) A nearest neighbour approach by genetic distance to the assignment of individuals to geographic origin. Forensic Sci. Int.-Gen. 27: 132-141.

DUMOLIN S, DEMESURE B, PETIT RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor. Appl. Genet. 91: 1253-1256.

DUPANLOUP I, SCHNEIDER S, EXCOFFIER L (2002) A simulated annealing approach to define the genetic structure of populations. Molecular Ecology 11(12): 2571-81.

KHATAB IA, ISHIYAMA H, INOMATA N, WANG X-R, SZMIDT AE (2008) Phylogeography of Eurasian Larix species inferred from nucleotide variation in two nuclear genes. Genes Genet. Sys. 83: 55-66.

MOSCA E, ECKERT AJ, LIECHTY JD, WEGRZYN JL, LA PORTA N, VENDRAMIN GG et al. (2012) Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. Evol. Appl. 5: 762-775.

ORESHKOVA NV, BELOKON MM, JAMIYANSUREN S (2013) Genetic diversity, population structure, and differentiation of Siberian larch, Gmelin larch, and Cajander larch on SSR-marker data. Russ. J. Genet. 49: 178-186.

POLEZHAEVA MA, LASCOUX M, SEMERIKOV VL (2010) Cytoplasmic DNA variation and biogeography of *Larix* Mill. in Northeast Asia. Mol. Ecol. 19, 1239-1252.

PRITCHARD JK, STEPHENS M, DONNELLY P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

RANNALA B, MOUNTAIN JL (1997) Detecting immigration by using multilocus genotypes. P. Natl. Acad. Sci. USA 94: 9197-9201.

SEMERIKOV VL, LASCOUX M (1999*)* Genetic relationship among Eurasian and American *Larix* species based on allozymes. Heredity 83: 62-70.

SEMERIKOV VL, LASCOUX M (2003) Nuclear and cytoplasmic variation within and between Eurasian *Larix* (*Pinaceae*) species. Am. J. Bot. 90: 1113-1123.

SEMERIKOV VL, SEMERIKOVA SA, POLEZHAEVA MA, KOSINTSEV PA, LASCOUX M (2013) Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. Mol. Ecol. 22: 4958-4971.

SLAVOV GT, NIPPER R, ROBSON P, FARRAR K, ALLISON GG, BOSCH M, et al. (2014) Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass Miscanthus sinensis. New Phytol. 201: 1227-1239.