Selection of nuclear microsatellite loci for specific identification of *Larix gmélinii* Ruhr. and comparison of genetic profiles of Larix to solve agricultural problems

E Amyaga*, S Nifontov

Branch of Roslesozashchita Federal State-Funded Institution—Forest Protection Center of Khabarovsk Krai, 4 Volochaevskaya Street, Khabarovsk 600000, Russian Federation

*Corresponding email: amyagaen@rcfh.ru

**Abstract.** Preservation and study of forest plant species diversity is one of the fundamental challenges of modern botany, genetics, and dendrology. Genetic structure determines variability and controls adaptation mechanisms, letting every population adapt to environmental conditions. The most topical issue is genetic study of valuable forest species to ensure their integrity, specific identification, and control over wood origin. One of such environmentally and economically important coniferous species in our country is larch (*Larix*).

Currently, there is a problem to use azonal larch species for reforestation in Russia, particularly in the Russian Far East. In this regard, selection of genetic markers that let us detect differences between *Larix gmélinii* Ruhr. and *Larix sibirica* Ledeb. and their subspecies is a crucial and relevant task.

**1. Introduction**

In last decades, thanks to development of forest genetics, forest species natural population reproduction and recovery programs have been implemented, and genetic variation of economic and endangered species has been characterized. Development of DNA polymorphism detection methods and work with genetic markers allows monitoring of genetic diversity dynamics, gene constellation, study of kin connections between taxa, genetic mapping, history analysis of natural areas formation, as well as search for loci that control specific indicators that are typical for a certain species. Information on genetic diversity and population structure of major forest-forming species is necessary to organize measures to protect and rationally use biological resources [1]. According to the Manual of Forestry Seed Collection and Storage of the Russian Federation, as well as in accordance with Article 65 of the Forest Code of the Russian Federation, ‘genetic heterogeneity of wood species within the borders of vast areas require strict ordering of harvesting and usage of seeds of forest-forming species with due allowance for hereditary properties and habitat conditions’.

In the article presented here, the objects of study are *Larix gmélinii* (*Larix gmélinii*) and *Larix sibirica*. Larch (*Larix*) is the most widespread, valuable, and economic species across the globe. Its growing area in Russia is massive, including Ural, West Siberia, Altai, Sayan Mountains, Far East, China, northwest of Mongolia [2]. Its timber, thanks to its strength and resistance to decay, is used in shipping industry, construction of hydraulic structures, pulp production. However, genetics of not all larch species are studied in equal measure. There is enough information on *Larix sibirica* genetics in literary sources [3-5]; a set of nuclear microsatellite loci is determined that allows a detailed exploration of intraspecific genetic differentiation. Genetic processes in *Larix gmélinii* populations are not studied, and there are a number of problems related to identification of subspecies and hybrid forms, determination of genetic potential and sustainability of populations. Therefore, the main goal of research is the selection of nuclear microsatellite loci for marking *Larix gmélinii* subspecies and populations, as well as the comparison of genetic profiles of larch to solve agricultural problems.
2. Materials and methods
For the study, fir needles are selected from three populations of *Larix gmélinii* in the Khabarovsk Krai (Sukpai forestry, Sukpai district forestry) and two populations of *Larix gmélinii* from the Jewish Autonomous Oblast (Birobidzhan forestry, Gorodskoe district forestry; Kuldur forestry, Birakan district forestry).

For the purpose of larch species identification, reference samples of *Larix gmélinii* seeds (batch from the Khabarovsk Krai) and *Lárix sibírica* seeds (batch from the Republic of Khakassia) were selected, and seedlings and seeds from forest nurseries of the Khabarovsk Krai were analyzed against admixtures of azonal larch species.

DNA extraction from fir needles and seeds was conducted using a modified CTAB method [6].

To determine genetic traits and analyze population polymorphism of *Larix gmélinii*, nuclear microsatellite loci were selected that had been selected for Japanese larch—bcLK group [7]. Main characteristics of PCR conditions and alleles are presented in Table 1.

**Table 1.** List of examined nuclear microsatellite loci for analysis of *Larix gmélinii* polymorphism.

| Locus   | Number of alleles | Fragment size | Annealing temperature | Literary source |
|---------|-------------------|---------------|------------------------|-----------------|
| bcLK056 | 21                | 142-196       | Touchdown 63 – 53°C    | [6]             |
| bcLK066 | 8                 | 143-157       | Touchdown 63 – 53°C    | [6]             |
| bcLK224 | 8                 | 128-148       | Touchdown 63 – 53°C    | [6]             |
| bcLK232 | 8                 | 133-149       | Touchdown 63 – 53°C    | [6]             |
| bcLK260 | 9                 | 94-110        | Touchdown 63 – 53°C    | [6]             |
| bcLK235 | 21                | 172-220       | 58°C                   | [6]             |
| UACTy6  | 23                | 214-262       | 58°C                   | [2]             |

Nuclear microsatellite loci [8], which variability was analyzed to compare *Larix gmélinii* and *Lárix sibírica* samples, are presented in Table 2.

PCR is performed using a kit of GenPak™ PCR Core lyophilized ready-made reaction mixtures (0.5 ml) made by Laboratory Isogen LLC [9, 10].

Visualization of PCR products is made by a vertical electrophoresis method in 5% polyacrylamide gel. The molecular mass of an amplicon is determined by its electrophoretic mobility [11]. As a standard length marker, DNA of pBR322, processed with Hpa II restriction, was used.

The size of amplicons was determined in Photo-Capt. The genotypes obtained were analyzed in GenAlEx 6.2 [9].

**Table 2.** Characteristics of nuclear microsatellite loci selected to compare *Larix gmélinii* and *Lárix sibírica* [12].

| Locus     | Nucleotide sequence                  | Annealing temperature | Product size |
|-----------|--------------------------------------|-----------------------|--------------|
| Ls_152449 | FW: CGACAAACACAGTCACATTCATC          | Touchdown 60 – 51°C   | 179          |
| Ls_951631 | FW: GAAACATCGTGACTCTTTGTA           | Touchdown 60 – 51°C   | 150          |
| Ls_254200 | FW: TTGTAAATGCACTCTTAACTCCA         | Touchdown 60 – 51°C   | 252          |
3. Results and discussion

In the context of comparative analysis and determination of Larix gmélinii genetic traits, analyses were made for 7 loci, with samples of fir needles from three populations in the Khabarovsk Krai and the Jewish Autonomous Oblast. Allelic polymorphism is typical for all analyzed loci. The number of allelic variants of bclk 066 locus, when working with Far Eastern larch species, is 10, which exceeds the number mentioned in literary source [10]. For bclLK235 and UACLy6 loci, the number of allelic variants is less for Larix gmélinii, which is 5 and 8 correspondingly, instead of 21 and 23 alleged for Japanese larch [7, 8].

The largest number of allelic variants is achieved by bclk 232 (13-15 alleles), bclk 056 loci (11-13 alleles). Other loci have a comparably lower polymorphism. The number of alleles and the most often seen variants are shown in graphs.

Although the identified nuclear microsatellite loci were common, their allelic composition was different in three populations. The maximum allelic variation — 61 alleles — is noted in the Larix gmélinii population from the Jewish Autonomous Oblast (Birobidzhan/Gorodskoe), while the least one — 55 alleles — is in the population from the Khabarovsk Krai (Sukpai/Sukpai).

Table 3 shows values of key genetic variability indicators calculated for 7 loci in three populations. The value of the Wright fixation index (characterizing the individual, subpopulation, and population levels of the population’s genetic structure) is maximum for the population from the Khabarovsk Krai (Sukpai/Sukpai) (F=0.553) and minimum for the population from the Jewish Autonomous Oblast (Birobidzhan/Gorodskoe) (F=0.502). Comparably, the difference between the indicators is within the limits of insignificant differences. The comparison of the observed and the expected heterozygosis showed that in all three populations there is a deficit in heterozygous genotypes. Probably, it can be explained by a low count of populations and self-pollination that causes a high degree of inbreeding [3-5].

| Populations          | N    | Na   | Ne  | Ho  | He  | F    |
|----------------------|------|------|-----|-----|-----|------|
| Sukpai, Sukpai       | 50   | 11.286 | 6.109 | 0.377 | 0.801 | 0.553 |
| Birobidzhan, Gorodskoe | 50  | 10.000 | 6.055 | 0.409 | 0.798 | 0.502 |
| Kuldur, Birakan      | 50   | 11.143 | 6.435 | 0.403 | 0.818 | 0.522 |

where, N — selection, Na — average number of alleles per locus, Ne — effective number of alleles per locus, Ho — observed heterozygosis, He — expected heterozygosis, F — fixation index.

The values of key genetic polymorphism indicators obtained as a result of the study are indicative of a quite high level of Larix gmélinii genetic diversity in the analyzed regions and coherent with the results of study of other Larix sp. as pertaining to the loci studies [6, 7, 11].

The values of specimen inbreeding coefficients with regard to the Fis population, specimen inbreeding with regard to the Fix species, and the population inbreeding with regard to the Fix species, calculated for every analyzed Larix gmélinii locus, are presented in Table 4.

| Locus | Fis  | Fst  | Fix  |
|-------|------|------|------|
| bclk_235 | 0.710 | 0.717 | 0.025 |
| bclk_260 | 0.879 | 0.886 | 0.057 |
| bclk_066 | 0.139 | 0.182 | 0.050 |
| bclk_056 | 0.778 | 0.806 | 0.125 |
| bclk_224 | 0.164 | 0.106 | 0.050 |
| bclk_232 | 0.831 | 0.863 | 0.186 |
The $F_{st}$ coefficient varies from 0.139 (bcll_066) to 0.879 (bcll_260), being 0.573 on average. A positive average value $F_{st}$ indicates a 57% lack of heterozygous genotypes. The $F_{st}$ coefficient also has a positive value, with the average one equal to 0.507, which indicates a 58% lack of the species’ heterozygotes in the explored part of the area in general. As for the $F_{it}$ value, which reflects the population subdivision level, it is found out that 92% genetic variability defined in three populations is accomplished within the population, and only 8% ($F_{it}=0.08$) is divided between the populations. Highly polymorphic loci bcll_232 ($F_{it}=0.186$) and bcll_056 ($F_{it}=0.125$) are of the highest importance for the analysis of the interpopulation variability component, and bcll_235 ($F_{it}=0.025$) is of the lowest importance.

While selecting nuclear microsatellite loci to compare genetic profiles of *Larix gmélinii* and *Lárix sibirica*, as well as their subspecies, 19 loci were analyzed. For specific identification, only 5 out of them were selected, with different values for different *Larix* sp. [9, 10].

**Table 5.** Characteristics of nuclear microsatellite loci selected for specific identification of *Larix*.

| Locus     | Indicators of nuclear microsatellite loci | *Larix gmélinii* and its subspecies | *Lárix sibirica* and its subspecies |
|-----------|------------------------------------------|-------------------------------------|-------------------------------------|
| UACLly6   | 214-262                                  | 180-215                             |
| bclLk232  | 133-149                                  | 151-164                             |
| Ls_1524449| unstable spectra with a big number of 0 alleles | highly polymorphic, with 9 allelic variants |
| Ls_951631 | a wide scatter of the variety of allelic variants and, probably, having two overlapping activity zones, which makes genotyping much more difficult | polymorphism |
| Ls_254200 | weak polymorphism (or monomorphism)      | prominent polymorphism, but with a big number of 0 alleles and unstable amplification |

The analysis results for the samples of larch seeds and seedlings received in the laboratory are described in Table 6.

**Table 6.** Analyzed samplings and their characteristics.

| Analyzed samples | Collection area                      | Result of nuclear microsatellite loci study | Species and genus on grounds of laboratory study |
|------------------|-------------------------------------|--------------------------------------------|-------------------------------------------------|
| 50 seeds         | Khabarovsk Krai, Soviet forestry, Kolpino district forestry | UACLly6, bclLk232, Ls_1524449, Ls_951631, Ls_254200 | correspondence to data from Table 5 *Larix gmélinii* |
| 50 seeds         | Republic of Khakassia, Tuim forestry, Kommunar district forestry | UACLly6, bclLk232, Ls_1524449, Ls_951631, Ls_254200 | correspondence to data from Table 5 *Lárix sibirica* |
At the initial stage, the work was conducted with control batches of seeds of Larix gmélinii (50 samples from the Khabarovsk Krai, Soviet forestry, Kolpino district forestry) and Larix sibírica (50 samples from the Republic of Khakassia, Tuim forestry, Kommunar district forestry) of the known origin. In order to confirm the genus and the species, UACLy6 and bcLK232 loci were analyzed, and the values were obtained that became ground-breaking for further diagnostics of seed and seedling batches, which were received in the laboratory to confirm and determine generic and specific assignment. For UACLy6 locus, the value of the Larix gmélinii fragment size was 214-262 bps and one of Larix sibírica was 180-215 bps. As for bcLK232 locus, the values are 151-164 bps and 133-149 bps for Larix gmélinii and Larix sibírica correspondingly (Table 6, Figure 1, 2).

The information obtained as a result of laboratory work with Larix samples for Ls_1524449, s_951631, Ls_254200 loci fully correspond to data from the paper by E I Bondar [12] (Figure 3-5).
Figure 5. Electropherogram of Ls_254200 locus (values for Larix gmélinii and Lárix sibírica)

In all 5 studied batches, Larix gmélinii seed and seedling samples were not found, but the analysis of 50 seeds from the nursery garden in the Khabarovsk Krai failed to clearly determine the species, since, as for UACLy6, bcLK232, and Ls_254200 loci, the values were typical for Larix gmélinii and its closely related species, while for Ls_1524449 and Ls_951631 loci the data corresponded to values that are typical for Lárix sibírica.

This research paper is part of papers of the Russian Forest Protection Center on creation of a united genetic database of major forest-forming species.

4. Conclusions
1. As a result of studying 3 larch populations from the Khabarovsk Krai and the Jewish Autonomous Oblast, their interpopulation variability was evaluated for the first time, based on the analysis of nuclear microsatellite loci. Significant differences in allele frequencies among the samplings were noted, as well as a prominent lack of heterozygotes in all three populations, which confirms a high degree of inbreeding. Unique alleles are identified for all three populations for 7 analyzed loci, which determine specificities of the population genetics and define its belonging to a certain region. The information obtained on the fragment lengths for the loci studies and the number of allelic variants will be filed in the united database of population and genetic characteristics of major forest-forming species of the Russian Federation.

2. Loci and primers thereto are selected, by which it is possible to identify stark differences between Larix gmélinii and Lárix sibírica and their subspecies, as well as to compare their genetic profiles. The values of DNA fragment lengths are determined, distinctive polymorphism criteria for 5 selected loci are noted. It is found out that among the studies seed and seedling samples there are no azonal species of Lárix sibírica and its subspecies, but one of the batches (seeds from a nursery garden in the Khabarovsk Krai), when genetically evaluated for nuclear microsatellite loci UACLy6, bcLk 232, and Ls 254200, corresponded to values that are typical for Larix gmélinii and its closely related species and, for Ls_1524449 and Ls_951631 loci, to the typical data for Lárix sibírica and its closely related species. Consequently, in all probability, there is a hybrid variant in the batch. The work on it is in progress, the methodology is being enhanced, and there is a search for new specific microsatellite markers.

Acknowledgments
For the provided information and assistance in the study, the authors express their gratitude to N V Oreshkova, Ph. D. in Biology, Research Associate of the Laboratory of Forest Genetics and Selection of the Federal State Budgetary Institution of Science, Sukachev Institute of Forest of the Siberian Division of the Russian Academy of Sciences; E I Bondar, Master from the Federal State Autonomous Educational Institution of Higher Education, Siberian Federal University, Institute of Fundamental Biology and Biotechnology, Specialized Department of Protection and Modern Technologies of Forest Monitoring.
References

[1] Khasa D 2006 Contrasting microsatellite variation between sub-alpine and western larch, two closely related species with different distribution patterns Mol. Eco. pp 3907-18

[2] Bousquet J and Simon L 1990. DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction Can. J. For. Res. pp 254–257

[3] Dylis N 1981 Listvennitsa (Larch) (Moscow: Lesn. prom-st’ (Forest industry) p 96

[4] Hamrick J 2004 Response of forest trees to global environmental changes For. Eco. and Manag. pp 323-335

[5] Isoda K and Watanabe A 2006 Isolation and characterization of microsatellite loci from Larix kaempferi Mol. Eco.y pp 664-666

[6] Oreshkova N 2009 Population and genetic parameters of Larix gmelinii in the Eastern Transbaikal [in Russian with English summary] Vestnik Tomskogo gosudarstvennogo universiteta pp 193-198

[7] Peakall R and Smouse P 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research pp 288–295

[8] Oreshkova N 2012 Variability of nuclear microsatellite loci of Dahurian larch (Larix gmelinii (Rupr.) Rupr.) and Kamchatka larch (Larix kamtchatica (Rupr.) Carr) [in Russian with English summary] Khvoine borealnoi zony pp 145-151

[9] Politov D 2007 Genetics of populations and evolitional relationships of pine species of Northern Eurasia: synopsis of a thesis by Doctor of Biology Science, PhD thesis, Moscow University

[10] Taheri S; Lee Abdullah T; Yusop M R; Hanafi M M; Sahebi M; Azizi P and Shamshiri R R Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants 2018 Molecules 23 p 399

[11] Oreshkova N 2008 Allozyme polymorphism of ferments of Siberian larch (Larix sibirica Ledeb.) and Dahurian larch (Larix cajanderi Mayr) [in Russian with English summary] Khvoine boreal. zony pp 160-167

[12] Bondar E 2018 Development of microsatellite markers of Siberian larch (Larix sibirica Ledeb.) on the basis of whole genome de novo sequencing [in Russian with English summary] Krasnoyarsk p 65