Galina Vasilyeva, Valeriya Vavilova, Kirill Ustyantsev, Igor Sukhikh, Alexander Blinov, Sergei Goroshkevich, Victor Sokolov

Genetic diversity of *Pinus sibirica*, *P. pumila* and their natural hybrids based on non-linked nuclear loci

Received: 13 February 2018; Accepted: 11 May 2018

**Abstract:** Frequent discordant phylogenies inferred from different loci, as well as the presence of sufficiently diverged gene variants within a single species isolate are indicative of potentially frequent non-monophyly in the genus *Pinus*. Interspecies hybridisation and incomplete lineage sorting have been suggested as possible explanations for the observed phylogenetic discrepancies. However, there is no direct evidence to support any of the proposed scenarios for the Eurasian five-needle pines. We used natural hybrids between *Pinus sibirica* and *P. pumila*, as well as their parental species, as a model to reproduce the scenario of non-monophyly in the subgenus *Strobus*. Three non-linked nuclear DNA loci (*LEA*, *AGP6* and *4CL*) were applied to detect introgressive alleles and to genetically discriminate the studied species. Comparative sequence analyses revealed two clusters of species-specific alleles for each of the markers, characteristic for either *P. sibirica* or *P. pumila*. No hybrid-specific alleles were found. We also found no hybrids with a genotype characteristic of only one of the parental species for all three loci. On average, the hybrids were characterised by an equal ratio of alleles from the *P. sibirica* and *P. pumila* clusters. We reveal that some trees of pure species originating from allopatric locations have non-specific loci that can be a result of genetic exchange between these species in the distant past or incomplete lineage sorting.

**Keywords:** hybridisation, Siberian stone pine, Siberian dwarf pine, genetic variation

**Addresses:**
G. Vasilyeva, S. Goroshkevich, Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch, Russian Academy of Sciences, 10/3, Akademichesky Ave., Tomsk, 634055 Russia, e-mail: galina_biology@mail.ru
V. Vavilova, K. Ustyantsev, I. Sukhikh, A. Blinov, Institute of Cytology and Genetics, Siberian Branch, Russian Academy of Sciences, 10 Prospekt Lavrentyeva, Novosibirsk, 630090 Russia
A. Blinov, V. Sokolov, Institute of Molecular and Cell Biology, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia
V. Sokolov, Vavilov Research Institute of Plant Industry, Russian Academy of Agricultural Sciences, 42-44, B. Morskaya Street, St. Petersburg, 190000 Russia; Department of Natural Sciences, Novosibirsk State University, 2 Pirogova Str., Novosibirsk, 630090 Russia
Introduction

Mechanisms of reproductive isolation are weak amongst pine species (Mirov, 1967). Therefore, the frequency of interspecies crosses under controlled pollination is higher than that under natural conditions (Critchfield, 1986; Lu et al., 2007). Five-needle pines (section Quinquefoliae, subgenus Strobus) are not exception to this rule. Siberian stone pine (Pinus sibirica Du Tour) and Siberian dwarf pine (P. pumila (Pall.) Regel) have the largest geographic distributions, with partial overlap among other five-needle pines. Although natural hybridisation between P. sibirica and P. pumila has been known for some time (Pozdnyakov, 1952), recent studies have begun to examine different aspects such as morphology, flow-er phenology, seed production, mating system and growth of hybrid seed progeny (Petrova et al., 2007; Goroshkevich et al., 2008; Petrova et al., 2008; Vasilyeva et al., 2010; Vasilyeva & Goroshkevich, 2012; Vasilyeva, 2014; Vasilyeva, 2017).

P. sibirica and P. pumila have different life-forms. Whereas P. sibirica is an upright tree, P. pumila is prostrate, with natural hybrids exhibiting intermediate morphology and life-form (Goroshkevich et al., 2008). It is important that the hybrids are fertile and crossed with both parental species (Vasilyeva & Goroshkevich, 2013). The seed efficiency of such hybrids is quite high and could have an impact on the genetic structure of the mixed population (Vasilyeva & Goroshkevich, 2012; Vasilyeva, 2014). Hence, hybridisation could have significant evolutionary consequences, such as the production of a novel species or deep introgression.

There are many gaps in the phylogeny of the genus Pinus, with non-monophyly, a consequence of ancient hybridisation or incomplete lineage sorting, particularly problematic in phylogenetic tree construction (Syring et al., 2007a,b; Tsutsui et al., 2009). The use of low-copy nuclear genes has been applied successfully in the study of pine phylogeny at low taxonomic levels (Syring et al., 2007a). Such multiple independent markers can provide more accurate data than cytoplasmic DNA regarding the phylogenetic relationships between closely related species (Syring et al., 2005). Several low-copy nuclear loci have been proposed for the phylogeny of Pinus species (Syring et al., 2005; Syring et al., 2007b; Mglinet et al., 2014). The aim of the study was to examine the genetic diversity of P. sibirica and P. pumila in the sympatric zone and beyond, as well as that of their natural hybrids based on non-linked nuclear loci, LEA (Late Embryogenesis Abundant (LEA)-like gene), 4CL (4-coumarate: CoA ligase) and AGP6 (Arabinogalactan-like protein 6).

Methods

Sample collection

Plant material (shoots with needles) was collected from the vegetative or seed progeny of species and hybrids of varying geographic origin (Table 1 in Supplementary material, Fig. 1). Both species and the hybrids were identified morphologically. The hybrids were of an intermediate growth habit and were thus identified as presumable F1. All trees were grown at the “Kedr” field station (scientific collection: 507474), situated 30 km south of Tomsk (56°13 N 84°51 E, 78 m above sea level) and managed by the Institute of Monitoring of Climatic and Ecological Systems SB RAS.

Amplification, T-vector cloning, and sequencing

Total DNA was isolated from fresh pine needles using a DNeasy Plant Mini Kit (QIAGEN) as specified
in the manufacturer’s protocol. The resulting total DNA was amplified via PCR using three pairs of primers, specific to the following partial nuclear gene sequences: **LEA** (intron fragment), **4CL** (the first exon and the following intron) and **AGP6** (coding part). PCR conditions and primer sequences for **AGP6** and **4CL** loci were as in Syring et al. (2005), and for **LEA**, as in Mglinets et al. (2014). PCR products were separated in 1.5% agarose gel and purified with a DNA purification kit (BioSilica, http://biosilica.ru/).

The PCR products (final concentration ~ 0.15 µg/µl) were ligated into pGEM-T Easy plasmid vector (Promega) in the presence of T4 DNA ligase (Promega, http://www.promega.com) as specified in the manufacturer’s protocol. The ligation products were transformed into *E. coli* competent cells via standard heat-shock transformation. Obtained white colonies were analysed for the presence of the inserted PCR products using PCR with universal M13 primers. The inserted PCR products were then sequenced automatically with an ABI PrISM 3100 Avant Genetic Analyzer (Applied Biosystems, USA) using a Big Dye terminator sequencing kit (Applied Biosystems, USA) at the SB RAS Genomics Core Facility (Novosibirsk, Russia, http://sequest.niboch.nsc.ru).

### Data analysis

The number of obtained genetic variants per marker, amount of polymorphic sites and mean number of nucleotide substitutions (or insertions/deletions in the case of **LEA**) per sequence per site (m) were calculated in DnaSP v. 5.10 (Librado & Rozas, 2009).

STRUCTURE v.2.3.4 software was used to infer the genetic structure of *P. sibirica* and *P. pumila* and to identify the species admixture in their hybrids (Hubisz et al., 2009). This particular program employs a Bayesian model-based clustering method; here the best K (number of clusters) value was identified by running the program for each K from 1 to 8 with 100000 burn-in cycles followed by 1000000 cycles of data collection (Evanno et al., 2005) based on the use of an admixture model.

Genetic differentiation of *P. sibirica*, *P. pumila* and their hybrids was estimated in Arlequin v. 3.5 (Excoffier & Lischer, 2010).

### Results

In total, 167 nucleotide sequences were obtained from **LEA**, **4CL** and **AGP6** nuclear loci corresponding to 13 *P. sibirica*, 11 *P. pumila*, and 12 *P. sibirica x P. pumila* hybrid trees. The obtained nucleotide sequences of **4CL** and **AGP6** loci were deposited in GenBank (Accession numbers: KT328511-KT328567 and KT447260-KT447315, respectively), while **LEA** nucleotide sequences were placed in the European Nucleotide Archive (Accession numbers: LN877971-LN878024). Six, 9 and 24 sequence variants were identified for **LEA**, **4CL** and **AGP6** loci, respectively. Six unique sequence variants (1 and 6 for **4CL** and **AGP6** loci, respectively) which were not found in the...
Genetic diversity of *Pinus sibirica*, *P. pumila* and their natural hybrids based on non-linked nuclear loci

Parental species were identified in hybrids (Table 1 in Supplementary material). The diversity of the **LEA** locus was the largest due to the presence/absence of deletions of 22 and 65 bp; in contrast, **4CL** and **AGP6** sequences varied only due to single nucleotide polymorphisms (Table 1).

Analysis of genetic structure indicated that the obtained data were better explained by assuming that all observed genetic variation was derived from two clusters (K=2). These two genetic clusters were species-specific, with 91% of *P. sibirica* genetic variation associated with the first cluster and 96% of *P. pumila* genetic diversity with the second cluster, respectively. Genetic variation of hybrids was represented as a mixture of 39% *P. sibirica*-specific and 61% *P. pumila*-specific genetic clusters. Average expected admixture levels (product of cluster frequencies multiplied by two) comprised 15%, 8% and 33% for *P. sibirica*, *P. pumila* and the hybrids, respectively (Table 2). No differences were recorded between individuals from the allopatric and sympatric zones of the studied species (Fig. 2).

Differentiation between *P. sibirica* and *P. pumila* (*F*$_{st}$ = 0.37) is greater than that between *P. sibirica* and hybrids (*F*$_{st}$ = 0.09) and between *P. pumila* and hybrids (*F*$_{st}$ = 0.12), supporting the intermediate position of the hybrids between the parental species (Table 3). The hybrids were also characterised by the highest pairwise intrapopulational difference, which is again in accordance with their mixed genetic pattern.

**Discussion**

Increased genetic diversity in plant hybrids is a well-known phenomenon (Zalapa et al., 2010 and others). *P. sibirica* x *P. pumila* hybrids are no exception,

![Fig. 2. STRUCTURE v.2.3.4 analysis of *P. sibirica*, *P. pumila* and their hybrids, assuming K=2. Cluster ratios indicated via light and dark grey colours. ID numbers on x-axis as in Table 1 in Supplementary material](image-url)
with the results obtained here in good agreement with those of earlier studies (Petrova et al., 2007, 2012; Vasilyeva & Semerikov, 2014).

Hybrids do not always have an equal ratio of alleles of the parental species. Here we found that whereas the hybrids from the Mount Alchanai region of southern Transbaikalia (sample ID 14–16) were more similar to P. pumila, those from northern Priabaikalia (sample ID 17–21) exhibited the reverse pattern and were more similar to P. sibirica. Established patterns of hybrids from distant geographic areas may be due to the duration of hybridisation, which leads to advanced hybrid generations. Hybrids from the northern Priabaikalia possessed pumila-specific mtDNA and sibirica-specific cpDNA (Watano et al., 2006); these hybrids were intermediate relative to their parental species, albeit slightly closer to P. sibirica in allozyme profile (Petrova et al., 2010).

The Mount Alchanai isolated population is located at the south-eastern boundary of the P. sibirica’s geographic distribution (Bobrinev et al., 2004), and hence at the most south-eastern point of the sympatric zone. Although P. sibirica and P. pumila in this region grow at different altitudes, their distributions overlap at 1500–1600 m a.s.l., where hybrids are abundant (Petrova et al., 2012). We found that pumila-specific alleles predominated in the hybrids, while allozyme analysis revealed that hybrids were intermediate (presumably F1) and only one hybrid with a prevalence of the P. sibirica alleles (presumably backcross) was found (Petrova et al., 2012). Such genetic heterogeneity suggests that interspecies hybridisation in the southern Transbaikalia is of considerable duration and includes at least several generations.

Most intermediate hybrids were found in southern Priabaikalia (Khamar-Daban Ridge, sample ID 22–26), but were slightly closer to P. pumila. According to AFLP markers, these hybrids are also intermediate, with the genetic distance to P. sibirica slightly less than that to P. pumila (Vasilyeva & Semerikov, 2014). Thus, the production of advanced generations of hybrids is also possible in the southern Priabaikalia.

We have thus far supposed that all the studied hybrids are F1, based on their intermediate growth habit. However, this intermediate growth habit does not comprise a single growth form but rather a number of possible morphological forms that reflect the contrast in life-form between the parental species. Perhaps, an intermediate growth habit can be retained in more advanced hybrid generations.

Using the non-linked nuclear loci LEA, 4CL and AGP6, the present study has revealed the genetic admixture in individual P. sibirica and P. pumila trees. Previous works have proposed to the use of the LEA gene to confirm the hybrid nature of morphologically intermediate individuals, since P. sibirica and P. pumila are well differentiated (Mglinets et al., 2014). However, we found that P. sibirica could be heterozygous at the locus in both the sympatric zone and beyond, while P. pumila could also be heterozygous beyond the sympatric zone. Therefore, the LEA locus is not a reliable marker for species differentiation and hybrid identification. Furthermore, the same could be said about the 4CL and AGP6 loci, with P. sibirica and P. pumila potentially heterozygous even outside the current sympatric zone. There are two possible explanations for the obtained results: an ancient genetic exchange and the retention of ancestral polymorphisms.

However, distinguishing introgression and retention of ancestral polymorphisms and incomplete lineage sorting is very difficult. In addition, it may be that both of these factors have contributed to the evolution of the species, as both phenomena are widespread in Pinus species (Syring et al., 2007b; Willyard et al., 2009). Therefore, further research is needed involving both more nuclear loci and increased population numbers representing different regions in the vast ranges of these species.

Conclusion

We have shown that genetic diversity in P. sibirica × P. pumila hybrids is higher than that in the parental species. We found shared genetic polymorphism in P. sibirica and P. pumila using non-linked nuclear loci. The obtained results suggest possible ancient hybridisation that took place in the distant past or incomplete lineage sorting.

Acknowledgements

This work was supported by the Russian Academy of Sciences (Program of Basic Research in State Academies, Theme 52.2.6 project No 0369-2016-0001), the Russian Foundation for Basic Research (Project No. 18-04-00833 a), the fundamental scientific research program (project No. 0310-2018-0010), and the Russian Government Budget (project No. 0324-2018-0017).

References


Evanno G, Regnaut S & Goudet J (2005) Detecting the number of clusters of individuals using the