

## Functional characteristics of EST-SSR markers available for Scots pine

Gulyaeva E.N.<sup>1</sup>, Tarelkina T.V.<sup>2</sup>, Galibina N.A.<sup>2</sup>

<sup>1</sup>*Department of Multidisciplinary Scientific Research, Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk, Russia*

<sup>2</sup>*Forest Research Institute of Karelian Research Centre Russian Academy of Sciences, Petrozavodsk, Russia*

**Abstract.** EST-derived simple sequence repeat markers (EST-SSRs) are important tools for studies on genetic diversity, phylogeny, evolution, comparative genomics, QTL analysis, and gene-based associations. We have searched the literature for known EST-SSRs used for Scots pine (*Pinus sylvestris* L.) – one of the world’s major forest species. Then, 91 of 102 EST-SSRs suggested for Scots pine studies were manually aligned against the reference genome of *Pinus taeda* L. as well as available genes of *P. sylvestris*. For 83 EST-SSRs, genome location and consensus putative functions of the associated genes were identified through conservative domain analysis (CDD), functional analysis of known homologs in terms of Gene Ontology annotations, and KEGG pathway analysis. Many of the markers were located in untranslated regions (mostly in 3’UTR), as well as in coding sequences of Scots and loblolly pine genes. For eight markers whose EST sequences were known no genes could be identified in any of the species. Seven of these markers were located in *P. taeda* scaffold regions carrying no genes in the current genome assembly (v.1.0). The results can be used in the future to improve the choice of markers for population genetic research, studies of adaptive traits, and QTL mapping of *P. sylvestris*, as well as other pine species.

**Key words:** EST-SSR, *Pinus sylvestris*, marker-gene association, marker genome location, functional annotation.

### INTRODUCTION

Scots pine *Pinus sylvestris* L. (*Pinaceae*) is one of the world’s key coniferous forest species, having high economical and ecological importance [1, 2]. Different Scots pine organs and tissues are used in sawmilling, wood processing, chemical and paper-making industries, medicine and cosmetology [2, 3]. The species also plays a major role in reforestation [3–6].

Scots pine distribution range covers over 14000 km, spanning from 8°W in Spain to 141 E in Russia [3, 7], so the species grows under diverse soil and climate conditions. This diversity of ecotopes gives rise to a wide spectrum of phenotypic and genotypic variation in this species [1, 8, 9]. Knowledge of the characteristics of Scots pine intrazonal forms and ecotypes is the background for designing selective breeding programs to enhance economic values [10–12]. Meanwhile, selection for superior *P. sylvestris* genotypes featuring high timber quality, high growth rates and/or resistance to abiotic and biotic stress impacts is commonly based on phenotype combined with biochemical markers, while identification of target traits using DNA markers is so far limited [10, 13].

Pine genomes are large (10000–40000 Mbp), with a lot of repetitive and noncoding DNA (introns, intergenic regions), making it difficult to implement whole-genome sequencing projects [14–16]. Scots pine is a diploid species ( $2n = 24$ ), with provisionally estimated genome size of 22520 Mbp [17]. In spite of the long history of *P. sylvestris* research, the reference genome for this species is so far missing [18]. That is why DNA genotyping of

Scots pine has long been primarily performed using RAPD and AFLP markers, for which no prior knowledge of the genome sequence is needed [19–27]. These markers, however, have poor reproducibility and are not suitable for reliable identification of loci within species or orthologous loci between species, which could provide a reliable basis for finding genes associated with the variation of the investigated traits [28–30].

In view of the above, it would be interesting to perform transcriptome sequencing, compile expressed sequence tag (EST) libraries, and use them to produce molecular markers, such as SSRs (Simple Sequence Repeats; microsatellite) and SNPs (Single Nucleotide Polymorphism). The high polymorphism of these markers turns them into a reliable tool for estimating genetic similarity between individuals or closely related taxa [31, 32]. Besides, EST-SSRs and EST-SNPs can reveal changes in the coding and untranslated (5'UTR and 3'UTR) regions of the genome, potentially contributing to the understanding of the genetic control of many important traits [33–35]. Among them, SSR markers remain more efficient and resource-saving, especially in genetic diversity studies. They are neutral, multi-allelic and co-dominant in nature, and permit genetic relatedness to be determined with higher probability using fewer markers [36, 37]. SNP markers are predominantly biallelic, and their allele frequencies are often distorted, resulting in reduced heterozygosity compared to SSR markers [31, 37]. In addition, EST-SSR markers have a higher transferability among closely related species [38–40], highlighting evolutionary processes and enlarging the pool of available markers for the entire genus [41–43].

As of now, several papers have been published describing EST-SSR markers for Scots pine [41, 44–46], but information about their location and functions in the Scots pine genome is fragmentary and scarce. In this study, we searched the literature for all known EST-SSR markers available for Scots pine, and aligned them against the *Pinus taeda* L. reference genome and known *P. sylvestris* gene sequences, thus identifying their exact locations and functions for *P. sylvestris*. The results can be used for improving choice of EST-SSR markers and for in-depth data analysis in future studies.

## MATERIAL AND METHODS

Having searched the literature, we identified EST-SSR markers specially developed or tested for *P. sylvestris* (Supplementary Table S1). Then, based on the NCBI accession numbers indicated by the authors of the articles, the EST nucleotide sequences, which were used to create the markers, were obtained. Where the NCBI accession number for an EST was not provided, we performed BLAST search against the EST database for the respective *Pinus* species using the nucleotide sequence containing the forward primer sequence, SSR, and the reverse complementary sequence of the reverse primer as the query. The selection criterion among BLAST search results was 100 % alignment of the forward- and reverse primer sequences, and coincidence between potential product length and the expected length indicated by the authors [41]. Where the EST sequence containing the EST-SSR marker had been entered in NCBI before publication of the research results, the DNA library used to create the EST-SSR was indicated in Supplementary Table S1.

The search for EST-SSR-associated genes was based on published *P. taeda* (Congenie.org (<https://congenie.org/>)) [47] and *P. sylvestris* (GymnoPlaza (<https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/>)) genomes. The search for *P. taeda* genes was also done in GymnoPlaza if search against Congenie database yielded no matches. Further in the text, PITA in the gene IDs refers to the genes of *P. taeda* from the Congenie database, PTA and PSY refers to the genes of *P. taeda* and *P. sylvestris* from the GymnoPlaza database, respectively. The BLAST search query was the nucleotide sequence corresponding to the marker (i.e. located between primers). The test for filtering out false positives was 100 % alignment of the nucleotide sequences of the marker primers and the

gene. Although no NCBI accession number could be retrieved for SsrPt\_ctg5167, we managed to align this marker onto Scots pine genome using forward and reverse primer.

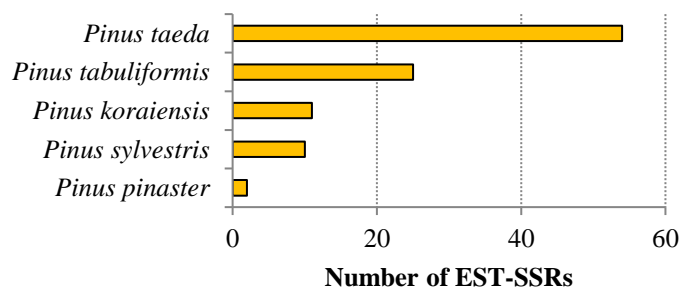
The marker's location within the gene was determined through alignment of the sequences of the marker, the gene, and the corresponding scaffold/isotig in MEGAX program [48]. The percent identity of the nucleotide sequences of the marker and the corresponding gene region was determined using EMBOSS Needle tool ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)).

The function of EST-SSR-associated genes were predicted using Blast2GO program [49] with E-value =  $10^{-6}$ . Where several nucleotide sequences were identified for an EST-SSR marker, identical GO results for several sequences were merged. The nucleotide sequences were also compared against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [50], where they were grouped into functional categories. Conserved domains in the putative proteins were predicted using the National Centre for Biotechnology Information (NCBI) resource (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) [51]. The percent identity of the amino acid sequences of the two pine species proteins was determined using EMBOSS Needle tool ([https://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](https://www.ebi.ac.uk/Tools/psa/emboss_needle/)).

## RESULTS

### Analysis of the literature, search for known Scots pine EST-SSR markers and their sequences

We analyzed 74 sources published between 2000 and 2021. Analysis of the data available in the literature revealed 102 EST-SSR markers, of which ten were developed from libraries of *P. sylvestris* nucleotide sequences [44] and 92 markers were derived from cDNA libraries for other *Pinus* species and tested for *P. sylvestris* [41,45, 46; 52–54] (fig. 1, Supplementary Table S1). EST-SSRs used in *P. sylvestris* studies were based on the libraries of DNA isolated from different tissues (Supplementary Table S1). Most of the markers were developed using libraries of DNA from xylem of different ages and states. Other sources of DNA libraries for EST-SSRs development were needles and male cones, root and branch xylem, roots, and stem apical meristem. A majority of known markers are widely used in populational and phylogenetic studies of *P. sylvestris* across Europe and Asia (see references in Supplementary Table S1).



**Fig. 1.** Species used to create DNA libraries and develop EST-SSR markers for *Pinus sylvestris*.

For 56 EST-SSR markers, NCBI accession numbers of the nucleotide sequences were indicated in the articles [41, 45, 46, 52, 53]. Yet, NCBI accession numbers of the nucleotide sequences were provided only for six of the 41 EST-SSRs customized for *P. sylvestris* by Chagné et al. [41]. Knowing that the markers suggested by Chagné et al. [41] are often used in *P. sylvestris* studies (Supplementary Table S1), we performed BLAST search against *Pinus pinaster* Aiton and *P. taeda* EST sequences deposited in NCBI and thus managed to identify the accession numbers for another 34 nucleotide sequences containing markers.

Eventually, we identified the NCBI accession numbers of 143 nucleotide sequences, which contained 90 markers, among which 72 markers were located in one nucleotide sequence and 18 markers were aligned with several nucleotide sequences (Supplementary Table S1).

**Identification of *P. taeda* and *P. sylvestris* genes associated with EST-SSRs**

Having aligned the sequences onto *P. taeda* genome, we link 126 nucleotide sequences containing 76 EST-SSR markers with 76 genes (Supplementary Table S3). The markers were located in different regions of *P. taeda* genes, mostly in the CDS and CDS/3'UTR regions (fig. 2). Another 14 sequences containing 12 EST-SSRs were aligned with 12 gene-free scaffold regions.

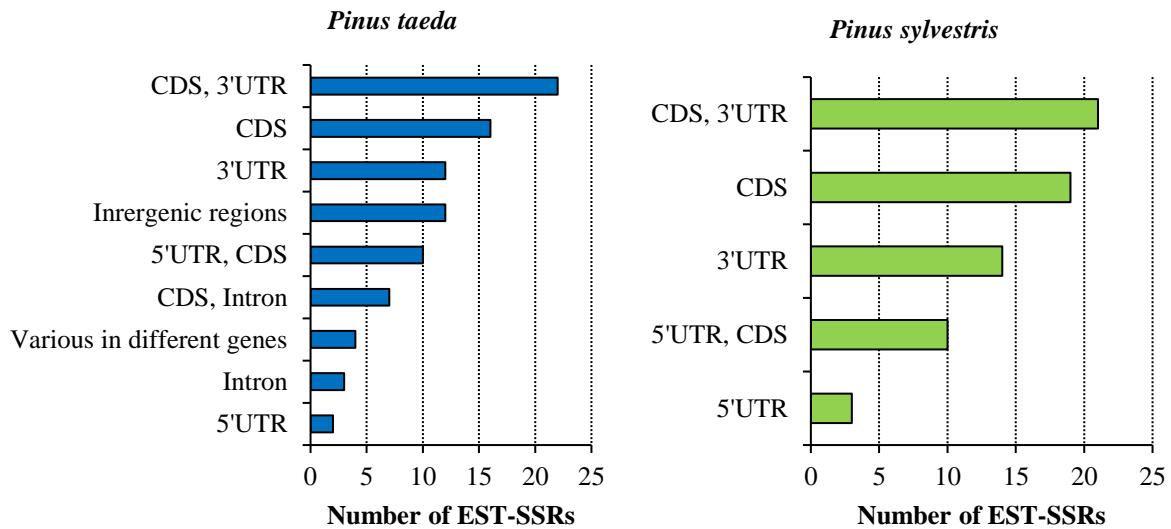


Fig. 2. Location of markers in *P. taeda* and *P. sylvestris* genes.

Alignment of the markers onto *P. sylvestris* contigs/isotigs deposited in the GymnoPLAZA database permitted matching 114 nucleotide sequences containing 67 EST-SSR markers with 56 Scots pine genes (Supplementary Table S2). The location did not always coincide with the location of these markers in *P. taeda* genome, but the distribution of the markers among gene functional regions was generally the same (fig. 2).

Table 1. List of Scots pine redundant EST-SSR markers

Redundant markers with the same orientation	Redundant markers with the reverse orientation
SsrPt_ctg2300 / SsrPt_ctg64	PtTX2146 / RPtest9 / PtSIFG_3145
SsrPt_ctg4487a/SsrPt_ctg4487b/ SsrPt_ctg7141	SsrPt_ctg8064 / RPtest15
SsrPt_ctg4698 / lw_isotig07383	SsrPt_ctg7024 / psyl36
SsrPt_ctg7444 / LOP3	SsrPt_ctg7867 / lw_isotig00080
SsrPt_ctg5167 / SsrPt_ctg5333	psyl16 / psyl17
RPtest1 /LOP6	psyl18 / psyl19
lw_isotig00080 / lw_isotig00081	

Analysis of the sequences revealed that some of the markers produced by different research teams were redundant pairs, derived from the same gene (tab. 1). Some of these markers aligned with the same gene region but with the reverse orientation.

## Identification of the functions of genes associated with EST-SSRs

BLAST search against the NCBI nr database with E-value threshold at  $10e^{-6}$  retrieved homologs with the predicted function for 65 *P. taeda* genes and 44 *P. sylvestris* genes containing EST-SSR markers (Supplementary Table S4, Supplementary Table S5). We also detected homologs with unknown function for five *P. taeda* genes and three *P. sylvestris* genes. For 6 *P. taeda* genes and 9 *P. sylvestris* genes no homologs were detected in the NCBI database (Supplementary Table S4, Supplementary Table S5). Functions of predicted homologs matched the results from the analysis of conserved domains in *P. taeda* and *P. sylvestris* genes.

Functional classification using standardized GO terms enabled the establishment of associations between 69 *P. taeda* genes containing 66 markers and 150 GO annotations of different levels, belonging to one of the three functional categories: Biological process (61 GO terms), Molecular function (64 GO terms), and Cellular component (25 GO terms) (Supplementary Table S5). The same analysis was performed to 48 *P. sylvestris* genes associated with 54 markers. As a result, 121 GO annotations of different levels were obtained. The terms were arranged in three functional categories as follows: Biological process (48 GO terms), Molecular function (50 GO terms), and Cellular component (23 GO terms) (Supplementary Table S4). The assigned GO terms were identical for genes containing the same marker in two pine species (tab. 2, Supplementary Table S4, Supplementary Table S5). For a number of markers, the GO terms coincided (Supplementary Table S4, Supplementary Table S5).

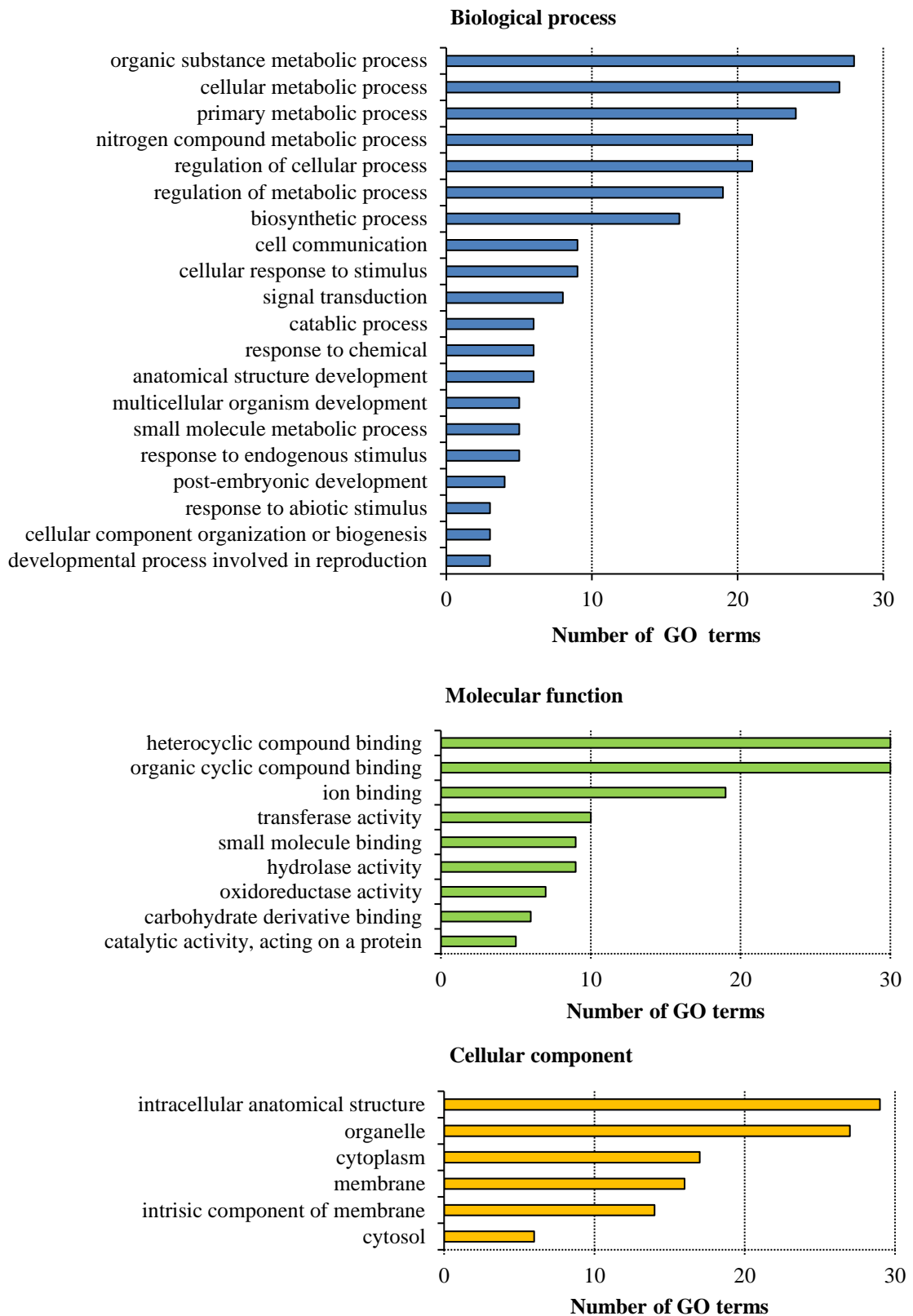
Functions at level 3 in the Biological process category were grouped into 20 subcategories, the most common ones being Organic substance biosynthetic process (GO:0071704), Cellular metabolic process (GO:0044237), Primary metabolic process (GO:0044238), and Nitrogen compound metabolic process (GO:0006807) (fig. 3). In the Molecular function category (level 2), the EST-SSR containing genes fell into three subcategories: Binding (GO:0005488), Catalytic activity (GO:0003824), and Transcription regulator activity (GO:0140110). At level 3, this category was subdivided into 9 subcategories, among which Heterocyclic compound binding (GO:1901363), Organic cyclic compound binding (GO:0097159), and Ion binding (GO:0043167) were the most common (fig. 3). When grouped according to the Cellular component category (fig. 3c) at level 3, genes were placed into one of six subcategories, and 27 % and 25 % of them belonged to the subcategories Intracellular anatomical structure (GO:0005622) and Organelle (GO:0043226), respectively (fig. 3). GO terms at higher and lower levels for each of the three basic GO categories are given in supplementary3.

KEGG pathway analysis allowed to associate 22 of the 91 EST-SSR markers with 15 different pathways belonging to four major functional categories:

Metabolism: Carbohydrate metabolism (9101), Biosynthesis of other secondary metabolites (9110), Nucleotide metabolism (9104), Energy metabolism (9102),

- Metabolism of terpenoids and polyketides (9109), Metabolism of cofactors and vitamins (9108), Metabolism of other amino acids (9106);
- Genetic Information Processing: Translation (9122), Replication and repair (9124);
- Cellular process: Cell growth and death (9143), Transport and catabolism (9141);
- Environmental information processing: Signal transduction (9132).

Another 15 markers were assigned to the Brite Hierarchies group, in which 14 markers belonged to the category Protein families: genetic information processing: Transcription factors (3000), Spliceosome (3041), Chromosome and associated proteins (3036), Membrane trafficking (4131). One marker was assigned to the category Protein families: metabolism: Glycosyltransferases (1003).



**Fig. 3** Distribution of 3<sup>rd</sup> level GO terms, associated with EST-SSR containing genes, in Biological process, Molecular function, and Cellular component categories.

A comparative analysis of 65 protein sequences encoded by *P. taeda* and *P. sylvestris* genes associated with the same marker shown that protein pairs shared over 35 % of identical amino acids. According to the results of conserved domains analysis and prediction of functional homologs, gene pairs of two pine species associated with the same marker have

also been suggested to perform a similar function. We compared protein sequences for another 4 pairs of genes containing the markers LOP6, psyl36, SsrPt\_ctg17601, and SsrPt\_ctg7024, whose function in the genome was identified only for one of the studied pine species. Their protein sequences also shared more than 37 % of identical amino acids, suggesting they performed similar functions in *P. taeda* and *P. sylvestris* [55]. We can thus infer that the 20 markers detected only in *P. taeda* genome will occur in homologous sequences in *P. sylvestris* genome and, hence, perform a similar function there. Eventually, having compared the results from the analysis of the conservative domain, functions of predicted homologs, and KEGG pathway analysis of *P. taeda* and *P. sylvestris* genes containing EST-SSRs, we suggested a consensus function for 83 EST-SSRs (tab. 2).

**Table 2.** Potential functions of the 91 EST-SSR-associated genes of *P. sylvestris* and *P. taeda*

Locus	<i>Pinus sylvestris</i> gene ID	<i>Pinus taeda</i> gene ID	Potential function according to Blast2GO analysis	Functional ortholog ( <i>ko</i> and <i>k</i> numbers), Functional category according to KEGG pathway analysis
PtTX2146	NP	PITA_000026471	aspartyl protease APCB1	
SsPp_cn524	PSY00010973	PITA_000001867	galacturonosyltransferase 13	K20867, Glycosyltransferases
SsrPt_AW010960	NP	NP	NA	
SsrPt_AW225917	NP	NP	NA	
SsrPt_AW981772	NP	PITA_000031044	transcription termination factor	
SsrPt_BF049767	NP	NP	NA	
SsrPt_ctg1376	NP	PITA_000000887	3-isopropylmalate dehydratase large subunit, chloroplastic	K01703, Carbohydrate metabolism
SsrPt_ctg1525	NP	NP	NA	
SsrPt_ctg17601	PSY00008240	PITA_000037826	starch-binding domain	
SsrPt_ctg18103	PSY00026391 PSY00024530	PITA_000025493 PITA_000086735 PITA_000047867 PITA_000047868 PITA_000051460 PITA_000072543 PITA_000030491	olee1-like protein	
SsrPt_ctg2300	PSY00014290	PTA00008090	cold shock protein 2-like	K09250, Transcriptionfactors
SsrPt_ctg275	NP	NP	NA	
SsrPt_ctg3021	PSY00012130	PITA_000031795	GRF1-interacting factor 1	
SsrPt_ctg3089	NP	NP	NA	
SsrPt_ctg3754	NP	PITA_000016257	DELLA protein GAI-like	K14494, Signaltransduction
SsrPt_ctg4363	NP	PITA_000068716	Protein FAM70A	
SsrPt_ctg4487a	PSY00005414	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	
SsrPt_ctg4487b	PSY00005414	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	
SsrPt_ctg4698	PSY00006544	PITA_000017622	protein RCC2 homolog isoform X1	
SsrPt_ctg5167	PSY00010492	NP	probable transcription factor PosF21	
SsrPt_ctg5333	PSY00010492	PTA00009525	probable transcription factor PosF21	
SsrPt_ctg64	PSY00014290	PTA00008090	cold shock protein 2-like	K09250, Transcriptionfactors
SsrPt_ctg7024	PSY00003770	PITA_000023332	mitogen-activated protein kinase 15 isoform X1	K04371, Signaltransduction
SsrPt_ctg7081	PSY00020303	NP	tubulin beta chain	K07375, Transport andcatabolism

Locus	<i>Pinus sylvestris</i> gene ID	<i>Pinus taeda</i> gene ID	Potential function according to Blast2GO analysis	Functional ortholog ( <i>ko</i> and <i>k</i> numbers), Functional category according to KEGG pathway analysis
SsrPt_ctg7141	PSY00005414	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	
SsrPt_ctg7170	PSY00002648	PITA_000027567	polyadenylate-binding protein RBP47-like isoform X1	
SsrPt_ctg7425	PSY00000824	PITA_000004120	UBP1-associated protein 2C	K1274, Transcription
SsrPt_ctg7444	PSY00000153	PITA_000012096	unknown	
SsrPt_ctg7731	PSY00017742	PITA_000091704	glucan endo-1,3-beta-glucosidase	
SsrPt_ctg7824	PSY00014046	PITA_000076250	Fiber protein fb34	
SsrPt_ctg7867	PSY00001445	PITA_000007834	cold shock protein 2	K09250, Transcription factors
SsrPt_ctg8064	PSY00002483	PITA_000008378	RNA-binding protein	K13195, Spliceosome
SsrPt_ctg865	NP	NP	NA	
SsrPt_ctg9249	PSY00009266	PITA_000010076	class III HD-Zip protein	K09338, Transcription factors
RPtest1	PSY00008891	PITA_000070971 PITA_000006918	putative histone-lysine N-methyltransferase SUVR2-like	
RPtest5	PSY00003137	PITA_000007179	B2 protein	
RPtest6	PSY00019692	NP	AT-hook motif nuclear-localized protein 10-like	
RPtest9	NP	PITA_000026471	aspartyl protease APCB1	
RPtest11	PSY00011770	PITA_000052634 PITA_000085253 PITA_000013530	N-terminal acetyltransferase A complex catalytic subunit NAA10	K20791, Chromosome and associated proteins
RP11est13	PSY00025105	PITA_000065673	E3 ubiquitin-protein ligase	K11982, Membrane trafficking Ubiquitin system
RPtest15	PSY00002483	PITA_000008378	RNA-binding protein	K13195, Spliceosome
RPtest16	PSY00014450	NP	NA	
LOP1	PSY00018563	PITA_000008107 PITA_000008108	CBL-interacting protein kinase 5	K07198, Signal transduction
LOP3	PSY00000153	PITA_000012096	unknown	
LOP5	PSY00023395	NP	chaperone protein dnaJ C76, chloroplastic	
LOP6	PSY00008891	PITA_000070971	putative histone-lysine N-methyltransferase SUVR2-like	
LOP8	PSY00004341	PITA_000044661	unknown	
LOP9	NP	PTA00066643	Chitinase-like protein 1	
LOP11	PSY00023490	PITA000023770	Lactoylglutathione lyase	K01759, Carbohydrate metabolism
LOP12	NP	PITA_000035518	putative TIR-NBS-LRR protein	
PtSIFG_1102	PSY00008719	PTA00050445	unknown	
PtSIFG_0737	NP	PITA_000017323	protein G1-like4	
PpSIFG_3145	NP	PITA_000026471	aspartyl protease APCB1	
PtSIFG_5015	NP	NP	NA	
PtSIFG_6044	PSY00004353	PITA_000008712	pleiotropic drug resistance protein 1-like	
PtSIFG_6065	PSY00027604	NP	(R)-mandelonitrile lyase-like	K08248, Metabolism of other amino acids
psyl2	PSY00005261	PITA_000017503	oligoridylate-binding protein 1-like	K13201, Messenger RNA biogenesis; Spliceosome
psyl16	PSY00018010	PITA_000010550	patellin-3-like protein	K19996, Membranetrafficking
psyl17	PSY00018010	PITA_000010550	patellin-3-like protein	K19996, Membranetrafficking
psyl18	PSY00016934	PITA_000064129	polyadenylate-binding protein 3	K13126, Translation



<b>Locus</b>	<b><i>Pinus sylvestris</i> gene ID</b>	<b><i>Pinus taeda</i> gene ID</b>	<b>Potential function according to Blast2GO analysis</b>	<b>Functional ortholog (<i>ko</i> and <i>k</i> numbers), Functional category according to KEGG pathway analysis</b>
psyl19	PSY00016934	PITA_000086838 PITA_000041571	polyadenylate-binding protein 3	K13126, Translation
psyl25	PSY00000191	PITA_000008677	ribulose biphosphate carboxylase small subunit	K01602, Carbohydrate metabolism
psyl36	PSY00003770	PITA_000023332	mitogen-activated protein kinase 15 isoform X1	K04371, Signaltransduction
psyl42	NP	PITA_000007965	WD repeat-containing protein 48	K15361, Replicationandrepair
psyl44	PSY00013146	PITA_000017051	thiazole biosynthetic enzyme	K03146, Metabolism of cofactors and vitamins
psyl57	NP	PITA_000010946	dicarboxylate transporter 1, chloroplastic	
lw_isotig00080	PSY00026757	PITA_000007834 PITA_000009956	cold shock protein 2-like	K09250, Transcriptionfactors
lw_isotig00081	PSY00026757	PITA_000009956	cold shock protein 2-like	K09250, Transcriptionfactors
lw_isotig00542	PSY00008008	PTA00082932	glycine dehydrogenase (decarboxylating), mitochondrial	K00281, Carbohydrate metabolism
lw_isotig01420	PSY00020425	PITA_000023899	pimaradiene oxidase 1	K05280, Biosynthesis of other secondary metabolites
lw_isotig02138	PSY00016229	PITA_000020633	E3 ubiquitin-protein ligase At1g12760-like	
lw_isotig02347	PSY00016640	NP	transcriptional corepressor SEUSS-like isoform X1	
lw_isotig02842	PSY00012746	PITA_000016857	WD-40 repeat-containing protein MSI1	K10752, Cell growth and death
lw_isotig03088	NP	PTA00015290	Poly A polymerase	K00974, Translation
lw_isotig04195	NP	PITA_000001889 PITA_000034486	Late embryogenesis abundant protein	
lw_isotig04204	PSY00005987	PITA_000008362	probable endopeptidase p60	
lw_isotig04306	PSY00011561	PTA00014456	protein FATTY ACID EXPORT 2, chloroplastic	
lw_isotig04600	PSY00011465	PTA00083895	trihelix transcription factor ASIL2	
lw_isotig04931	PSY00018521	PITA_000025726	dihydropyrimidinase	K01464, Nucleotide metabolism
lw_isotig05123	NP	PTA00083918	DUF674 domain-containing protein	
lw_isotig06215	PSY00022080	PTA00008461	Transcription factor MYB44	K09422, Transcription factors
lw_isotig06440	PSY00018317	PITA_000080433	zinc finger protein GIS3-like	
lw_isotig07383	PSY00006544	PITA_000017622	protein RCC2 homolog isoform X1	
lw_isotig10603	PSY00011044	PITA_000023436	proline-rich receptor-like protein kinase PERK1	
lw_isotig11166	PSY00030466	PITA_000059576	basic blue protein-like	
lw_isotig12667	PSY00012407	PTA00083219	adenyl-sulfate kinase 3 isoform X1	K00860, Energy metabolism
lw_isotig17679	NP	PITA_000014378	protein NETWORKED 1D	
lw_isotig20215	PSY00031065	PTA00083968	indole-3-acetaldehyde oxidase	K09842, Metabolism of terpenoids and polyketides
lw_isotig21953	PSY00005415	PTA00006979	protein trichome birefringence-like 6	
lw_isotig26230	PSY00034638	PITA_000038492	protein RSI-1	
lw_isotig27940	PSY00003211	PITA_000016972	auxin-responsive protein IAA9 isoform X3	K14484, Signaltransduction

NP – not predicted, NA – not available

## DISCUSSION

Polymorphous EST-SSR markers are important tools in the various genetic studies [56–62]. As a rule, functional annotation of EST-SSR markers is based on the search of homologous sequences in genetic databases. This approach, however, is not always effective for markers of conifers because there exist specific genes and gene families for which there are no homologous sequences in angiosperms [63–66]. In view of this, we manually aligned EST-SSR markers available for *P. sylvestris* onto *P. taeda* reference genome and onto Scots pine gene sequences deposited in the GymnoPLAZA database. Owing to this approach, we managed to align 83 EST-SSR markers onto *P. taeda* and *P. sylvestris* genes, and 23 of the markers were identified only for one of the species (tab. 2). For eight of the markers with known containing EST sequences, no gene matches were identified in either of the species. Of these, seven markers were located in *P. taeda* scaffold regions containing no genes in the current genome assembly version (v.1.0). Analysis of the literature showed that apart from the studies in which those 8 markers were developed, *P. sylvestris* studies have employed only two markers, SsrPt\_ctg1525 and PtSIFG\_5015 (Supplementary Table S1). In these studies, amplification for the marker SsrPt\_ctg1525 was absent at all in 11 pine species, while the marker PtSIFG\_5015 produced clear and reproducible results [67]. Markers SsrPt\_ctg3089 and SsrPt\_ctg865 were successfully amplified in two other pine species, *Pinus canariensis* C. Sm. [68] and *Pinus nigra* Arn. [69]. Proceeding from the above, one of the possible reasons for the failure to align these markers onto genes is challenges with gene assembly and annotation which are known for conifer mega-genomes [18, 47]

Since EST-SSR markers are developed using EST libraries produced by different experiments, one of the problems associated with this type of markers is the high proportion of redundant markers [39]. We identified 13 pairs/triplets of redundant markers among the 102 EST-SSRs used in *P. sylvestris* studies. This is not surprising considering that available for Scots pine EST-SSRs were proposed relying on EST libraries developed for different tissues of 5 *Pinus* species (Supplementary Table S1). Researchers usually try to avoid using redundant markers, realizing this can give rise to redundant information in genetic maps, inadequate estimation of the genetic diversity, and issues during multiplex PCR [53, 70]. On the other hand, detection of redundant markers, especially ones reversely oriented relative to one another within the same gene region, by conventional methods can be problematic. E.g., among the redundant markers we have identified, 4 EST-SSR pairs were used by research teams within the same study: PtTX2146/PtSIFG\_3145 [71], lw\_isotig00080/lw\_isotig00081 [72], psy116/psy117 [73-78], psy118/psy119 [73, 79].

Another major concern about using EST-SSRs is their level of polymorphism, which was reported to be lower in comparison with genomic SSRs [39, 80, 81]. Furthermore, microsatellite repeats can be located both in coding as well in non-coding gene regions, with higher polymorphism in the 5' and 3' ends [39, 82–84]. A large number of markers in our study were located in the 5'UTR and 3'UTR in genes of Scots and loblolly pines. More specifically, the markers for which a greater number of alleles per locus has been previously demonstrated were usually located in the 3'UTR, whereas markers with fewer alleles per locus were located in CDS. Examples of marker pairs with high/low number of alleles per locus are LOP5 (3'UTR)/LOP3 (CDS) [45, 86], psy116 (3'UTR)/psy144 (CDS) [44, 73, 87], lw\_isotig21953 (CDS/3'UTR)/lw\_isotig04306 (CDS) [46, 72, 88]. The lower polymorphism of the markers located in CDS has to do with the type of SSR. Among the microsatellites used for SSR genotyping, trinucleotide repeats prevail in the plant genome [88, 89], but dinucleotide repeats are more polymorphic, which improves the estimation accuracy [90]. We know from the literature that mutations in microsatellites with repeat length divisible by three (tri- and hexanucleotide repeats) do not shift the reading frame. Mutations in non-trimeric microsatellites (i.e. mono-, di-, tetra-, and pentanucleotide repeats) can shift the reading frame if they happen in the coding DNA, and selection against these mutations

reduces their chances of getting fixed [91, 92]. Hence, information about the marker's location can be useful for the choice of markers with a potentially higher number of alleles per locus.

Information about marker location can also be of interest in terms of the marker's involvement in gene regulation [32, 38, 39]. E.g., markers located in the 5'UTR can influence gene expression [93–97]. It has been demonstrated that in some cases the regulatory effect of SSRs was due to the coincidence of their sequences with sequences of *cis*-regulatory elements, such as TATA-box [98], GAGA-element [99], or regulatory elements involved in light and salicylic acid responses [100]. SSRs in the 3'UTR can cause transcription slippage [38, 39, 101].

Alignment of EST-SSR markers used in Scots pine studies onto *P. taeda* and *P. sylvestris* genes helped reveal the putative functions of SSR-containing genes and the processes in which these genes are probably involved. Functional characterization of Scots pine EST-SSRs by means of GO annotation demonstrated the presence of all basic functional categories. This is in agreement with previous studies of EST sequences in conifers, such as *Pinus taeda* [64], *Pinus sylvestris* [102], *Pinus pinaster* [103], *Pinus tabulaeformis* Carr.[104], *Pinus koraiensis* Siebold et Zucc. [105], and *Pinus canariensis* [106]. Comparative analysis of the studies mentioned above shows that ESTs in pine species were similarly distributed among GO categories, and not a single functional group was overrepresented, indicating there were no preferences in the location of microsatellites in terms of gene function. Hence, EST-SSRs appear to be a convenient and cheap gene mapping tool.

Numerous studies for mapping commercially valuable traits have demonstrated the expediency of using associative mapping in *Pinus* species selection [20, 26, 107–111]. However, mass-scope introduction of marker-assisted selection in the breeding of conifers, especially *Pinaceae*, is so far inhibited by a lack of molecular markers, specifically from coding regions of the genome, hindering identification of gene-trait associations. New data on the associations of Scots pine EST-SSRs with the functions of the containing genes can be useful when looking for candidate genes contributing to the expression of commercially valuable traits.

E.g., in *Pinus* species such as *Pinus radiata* D.Don, *P. sylvestris*, *P. taeda*, *P. pinaster*, QTL mapping is now mainly focused on the genes that directly influence timber quality – one of the key criteria in forestry [107–110, 112–114]. A large number of ESTs have been sequenced and analyzed in pine to detect the genes associated with wood formation and timber quality characteristics [115–120]. In this study, we identified several markers located in the genes potentially associated with timber quality in *P. sylvestris*. The marker LOP9 is located in the gene coding for chitinase-like protein 1, which belongs to the glycosyl hydrolases (GHs) family and has a role in cellulose synthesis in primary and secondary cell walls in both herbaceous and woody plants [121–124]. Besides, chitinases constitute the largest group of enzymes with diverse roles in plant defense. Especially interesting is their activity during callus formation and growth in *P. sylvestris*. Because chitinases are activated in plant tissues primarily in response to microbial presence, they can be used as markers of culture infection [125]. The marker lw\_isotig21953 is located in the gene coding for a member of the TRICHOME BIREFRINGENCE-like (TBL) family, which has been shown to impact cellulose biosynthesis as well as pathogen resistance and freezing tolerance [126, 127]. The gene encoding the fiber protein fb34 (associated with SsrPt\_ctg7824 in the current study) was first identified as one of the genes up-regulated during secondary cell wall formation in cotton fibers [128]. Later, this gene was shown to be active in wood-forming tissues of poplar [129] and suggested to be potentially involved in monolignol transport [130]. Members of the MYB transcription factor (TF) family (marker lw\_isotig06215) regulate multiple processes, including lignin biosynthesis in *Arabidopsis* Heynh [131]. In conifers, various MYB TF were shown to participate in wood formation [106, 132–136]. It is also interesting to study the member of Class III HD-Zip TFs (marker SsrPt\_ctg9249). Members of this family participate

in secondary growth regulation in different woody plants [137, 138]. A correlation was shown to exist between HD ZIP III TFs and cellulose content in *P. pinaster* wood [109] as well as wood density in *P. radiata* [120].

We also found that some of the markers were associated with genes potentially involved in the adaptation of plants to abiotic stresses (SsrPt\_ctg5333, LOP1, psyl2, psyl36, lw\_isotig00080, lw\_isotig02842, lw\_isotig04195, and others) and in defense against various pathogens and herbivores (PtTX2146, SsrPt\_ctg4698, SsrPt\_ctg7731, LOP12, PtSIFG\_6044, lw\_isotig01420, lw\_isotig11166). Functions of the genes associated with the indicated markers have been experimentally demonstrated for several conifers. Thus, different LEA proteins (marker lw\_isotig04195) of *P. tabuliformis* expressed in *Escherichia coli* demonstrated a protective function under heat and salt stresses [139]. Genes of the family *TIR-NBS-LRR* (marker LOP12) were associated with resistance to different pathogens in several *Pinus* species [140]. Members of the pleiotropic drug resistance (PDR) family (marker PtSIFG\_6044) were shown to play a key role in tree defense against spruce budworm [141] and spruce shoot weevil [142] in two spruce species, as well as against pitch canker disease in loblolly pine [143].

To conclude, 91 of 102 EST-SSRs available for *P. sylvestris* studies were manually aligned against the reference genome of *P. taeda* as well as available genes of *P. sylvestris*. For 83 EST-SSRs, genome location and putative functions of the associated genes were identified through different bioinformatics methods. The results can be useful for more adequate choice of markers for population genetics research, studies of adaptive traits, and QTL mapping of *P. sylvestris*, as well as other pine species.

#### ACKNOWLEDGMENTS

The study was carried out under state assignment to the Department for Multidisciplinary Scientific Research, Karelian Research Centre of the Russian Academy of Sciences.

## SUPPLEMENTARY MATERIAL

Table S1. Characteristics of 102 *Pinus sylvestris* EST-SSRs

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
PtTX2146	AF143963	F:CCTGGGGATTGGATTGGTATTTG R:ATATTTTCCTTGCCCCTTCCAGACA	(GCT) <sub>4</sub> GCC (GCT) <sub>7</sub> GCC (GCT) <sub>8</sub>	55	181 193 211 220	Locus PtTX2146 derived from a total genomic library (needles) amplified the same marker as EST-SSR locus RPtest9 [1] ( <i>Pinus taeda</i> )	<i>Pinus taeda</i> , <i>P. ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>P. sylvestris</i> [45, 144]	[71, 79, 85, 145–169]
SsPp_cn524	BX255732	F:CGATTGTTTTGCCTTTTAAGC R:AAATATGGCGGGGTGTGC	(AG) <sub>14</sub>	50	156	Differentiating xylem wood (adult, <i>Pinus pinaster</i> ) BioSample: SAMN00172245	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> [41]	[67]
SsrPt_AW010960	AW010960	F: ATCGACTAGGCATCAGGTGG R: TCCTCGTAGCCAGCTTTTA	(AT) <sub>9</sub>	49	225	Shoot tips ( <i>Pinus taeda</i> ) BioSample: SAMN00156911	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_AW225917	AW225917	F: TGCATTGAAAAATACAGCGG R: ATTATGTACGAGGCCCCACA	(AT) <sub>9</sub>	49	198	Shoot tips ( <i>Pinus taeda</i> ) BioSample: SAMN00156911	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_AW981772	AW981772	F: GATCCTGTTCTCCTCCTCC R: CCTGGACAGAAACAGCAACA	(CCT) <sub>4</sub>	49	266	Pollen cone ( <i>Pinus taeda</i> ) BioSample: SAMN00156910	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_BF049767	BF049767	F: TTTTGGGTCGTAGGAACCTG R: TAAAACGGGTGTCTTCCGG	(AG) <sub>22</sub>	51	227	Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> [41]	-
SsrPt_ctg1376	CD024501 CD024819	F: CGATATTATGGATTTTGCTTGTA R: AAATGCATGCCAAACTAAATAC	(AT) <sub>20</sub>	60-50	145	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> [41]	[72, 149–152, 162, 163, 170–175]
SsrPt_ctg1525	BQ655604	F: TTGAAACCATATAAGCAATGCC R: AGGACCTGGGTAAGGAGGC	(AGG) <sub>7</sub>	60-50	173	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
SsrPt_ctg17601	AW985278	F: CGCCATTAATATGCCTACCG R: ATCTCTGCGCTGCTTGAAGT	(AAG) <sub>9</sub>	54	225	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg18103	BE187051	F: CCTGGATTCATTTGTGGCTAA R: CATGCCAACTCTTGCATTG	(AT) <sub>10</sub>	60	184	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg2300	-	F: CACTTTGCGAGAGACTGCAC R: ACGCTGAAGGAAATCGAGAA	(CCG) <sub>6</sub>	49	173	-	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
SsrPt_ctg275	AW869955	F: ACGGAGATATATTGCTGGCG R:AAAGAATAACGTGAAACAAACCC	(AT) <sub>16</sub>	60-50	137	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> [41]	-
SsrPt_ctg3021	-	F: CTCAGATTCCTCCAAATGCG R:CATGCAACATATGCAAACCG	(AGC) <sub>14</sub>	60-50	234	-	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
SsrPt_ctg3089	BQ291150	F: CTTTCTTCACGTTGGACTTCTT R:TTAGCCATGGAGAGTGCAGA	(AT) <sub>17</sub>	45	482	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg3754	BQ701798	F: TCTTGGGTTCTGGAGTGG R:GCTGTTGCTGTTGTTCTTGG	(AGC) <sub>6</sub>	60-50	421	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg4363	AI812743	F: TAATAATTCAAGCCACCCCG R:AGCAGGCTAATAACAACACGC	(AT) <sub>10</sub>	60-50	100	Differentiating xylem wood (adult, <i>Pinus taeda</i> ) BioSample: SAMN00156479	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67, 72, 149–152, 162, 163, 170–184]
	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372							
	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774							
SsrPt_ctg4487a	BQ702547 BG040578 BG040033 BG039427 BF517866 BF517086 BF060592	F: TCTGCTGTGTGGACAAACCT R:TTCTTGGCTCAAAATCTCGG	(CCG) <sub>5</sub>	60-50	155	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>(Pinus pinaster, Pinus taeda, Pinus radiata, Pinus sylvestris, Pinus halepensis</i> [41])	-
	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108							
	Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033							
SsrPt_ctg4487b	BG039427 BQ701415	F:ATGACGCATTATCAGGGGAA R:TTGCACAGAAAGCAGGTTTG	(CCG) <sub>10</sub>	45	254	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> [41]	[148, 154]

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
SsrPt_ctg4698	BG275886	F:CGAAAAGGTGGTCTGATGG R:TTTTCCGCTGGATTTACCAC	(ATC) <sub>10</sub>	49	246	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[76, 167]
SsrPt_ctg5167	-	F:TGCAGAGAGATTCGATGGG R:ATTTGGTTTGGTTGCTGGC	(AAC) <sub>7</sub>	60–50	293	-	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
SsrPt_ctg5333	BQ700216	F:GAAGGAGTCGGCGATAACAG R:GGGAATTCGACCTGTGAAGA	(AGC) <sub>7</sub>	49	163	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> [41]	-
	BF610405					Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774		
SsrPt_ctg64	-	F:GGAAGCTGTTACAAGTGC GG R:ATCGAGAAGAGAGGAAGGGC	(CCG) <sub>7</sub>	60-50	284	-	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[148, 154, 165]
SsrPt_ctg7024	BQ695936	F:GGGAATTCTGAAAGACAAGGG R:AACCTACCCATCGAGAGCCCC	(AAG) <sub>7</sub>	60-50	277	Xylem planings wood (secondary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00163209	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> [41]	-
SsrPt_ctg7081	BE187193	F:GTCATCCACGTTTCATTGGC R:TCACAACCTGACCAAACTGCC	(AAG) <sub>7</sub>	60-50	442	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> [41]	-
SsrPt_ctg7141	-	F:GAATGACGCATTATCAGGGG R:TCACCTTTCTACCTCTGCC	(CCG) <sub>8</sub>	45	381	-	<i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg7170	BQ696574	F:GGTTTTTCGATTTCTGAGGC R:AACAGGTGTGCAAATAGCCC	(AGC) <sub>5</sub>	60-50	385	Xylem planings wood (secondary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00163209	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> [41]	-
	BG039619 BF778892					Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774		
SsrPt_ctg7425	CF385379	F:AATAAGACCCCAGAGGAGCC R:GACGTCTTTACCAAATCGC	(AAG) <sub>6</sub>	60-50	384	Roots ( <i>Pinus taeda</i> ) BioSample: SAMN00173690	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> [41]	-
	CF394315					Roots ( <i>Pinus taeda</i> ) BioSample: SAMN00173693		

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
SsrPt_ctg7444	BQ196808 BQ107124 BM367065	F:TCTTCACCATCGGTTTCTCC R:TGGATCTGTACCTCCTCATC	(AT) <sub>10</sub>	58	285	Primary xylem wood (late wood, <i>Pinus taeda</i> ) BioSample: SAMN00164483	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774							
SsrPt_ctg7731	BF517779	F:AGTGGTGAAGGGTCCATCTG R:GCATAACACAAAAGCCAGCA	(AT) <sub>12</sub>	51	217	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
	BE451949					Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033		
SsrPt_ctg7824	BQ703132	F:TGACCTGTCTTGTGAGACGC R:TTTTGAAACAGATTGCAGCC	(AT) <sub>12</sub>	60-50	501	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> [41]	-
SsrPt_ctg7867	BQ701488 BG275803 BG040533 BG040463 BF777863	F:GGTCGTGGAGGAGGTAGGG R:ACTGATAACAGCTGCCCCC	(CCG) <sub>6</sub>	45	154	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
	BE643974					Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033		
SsrPt_ctg8064	-	F:GAACGTGGTTATGGCGGTAG R:TCGTGGCAACTATCTCCTCC	(ACC) <sub>6</sub>	50	147	-	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
SsrPt_ctg865	BQ699190 BQ699199	F:TTTCAGAAGCTCCCGATTG R:CTTGTGGACATGGTTAATGAAG	(AT) <sub>15</sub>	45	232	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
SsrPt_ctg9249	BE761900	F:CTGCTCCCTCAGCTCTTCC R:AGACGTCACTGCCATTACCC	(AAG) <sub>7</sub>	55	156	Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[167]



Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
RPtest1	BF517869 BF517853	F:GATCGTTATTCCTCCTGCCA R:TTCGATATCCTCCCTGCTTG	(AAT) <sub>7</sub>	50	125	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
RPtest5	BQ697205	F:ACAACAATAATAACGGGGGC R:ACGCTTTAGATCCTCCTGCA	(AAC) <sub>6</sub>	55	197	Xylem planings wood (secondary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00163209	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67]
	Compression xylem wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00160033							
RPtest6	BG040975 BF518209	F:AGGATTCCAACAGCATCACC R:CTGAACATGAAGCGCAGTGT	(TGC) <sub>5</sub>	55	147	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
RPtest9	CF393810	F:CCAGACAACCCAAATGAAGG R:GCCTGCTATCGAATCCAGAA	(AGC) <sub>10</sub>	51	289	Roots ( <i>Pinus taeda</i> ) BioSample: SAMN00173693	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[185]
RPtest11	CF391802	F:AGGATGCCTATGATATGCGC R:AACCATAACAAAAGCGGTCCG	(ATC) <sub>7</sub>	56	213	Roots ( <i>Pinus taeda</i> ) BioSample: SAMN00173692	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[67, 177, 185]
	Xylem planings wood (secondary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00163209							
RP1test13	AA739656	F:GATTTTTTCAGGAAGACCCCC R:TGTAAGGCACAAGCCCTCTT	(CTG) <sub>5</sub>	51	277	Xylem branches ( <i>Pinus taeda</i> ) BioSample: SAMN00155658	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus canariensis</i> [41]	-
RPtest15	BG275136	F:GAACGTGGTTATGGCGGTAG R:CCAGGGACAGTTACCAGCAT	(ACC) <sub>6</sub>	56	246	Xylem side wood (juvenile, <i>Pinus taeda</i> ) BioSample: SAMN00161774	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus halepensis</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	[185]
RPtest16	AA739818	F:CAGAAATGGCGTCCAAATTC R:ACCCCACTTATATCCCCAGC	(AGT) <sub>5</sub>	56	132	Xylem branches ( <i>Pinus taeda</i> ) BioSample: SAMN00155658	<i>Pinus pinaster</i> , <i>Pinus taeda</i> , <i>Pinus radiata</i> , <i>Pinus sylvestris</i> , <i>Pinus pinea</i> , <i>Pinus canariensis</i> [41]	-
LOP1	AI812473	F:GGCTAATGGCCGGCCAGTGCT R:GCGATTACAGGGTTGCAGCCT	(TA) <sub>10</sub>	55	158, 162	Differentiating xylem wood (adult, <i>Pinus taeda</i> ) BioSample: SAMN00156479	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	[71, 85, 148–152, 154, 155, 157, 162–166, 168]

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
LOP3	AA556662	F:GTCTCCAGCCAGTTCACCTGC R:CAGTGGATCTGTCACCTCCTC	(TA) <sub>9</sub>	48	209, 213, 217	Immature xylem wood ( <i>Pinus taeda</i> ) BioSample: SAMN00155556	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	[85, 148, 154, 166–168]
LOP5	AW758812	F:AGCCGTAAGCTATCTTGTG R:GGCATACTACATTTAATAA	(TA) <sub>33</sub>	45	166–190	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	[71, 85]
LOP6	AA556221	F:AGTTTATCCATGCTGCACAG R:ACCTAAAGCCCAATATCCACA	(AAT) <sub>7</sub>	55	220–290	Immature xylem wood ( <i>Pinus taeda</i> ) BioSample: SAMN00155556	<i>Pinus taeda</i> [45]	[171]
LOP8	AI725303	F:TATCCACCAGAAGGGCATC R:CGGGAGCTTTAATGATCTTGA	(CCT) <sub>6</sub>	50	367	Xylem branches ( <i>Pinus taeda</i> ) BioSample: SAMN00155658	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	-
LOP9	AI813163	F:GGATTCTCGTTGTGGCTGG R:TTGCCTTGCACATAATATCT	(GGC) <sub>6</sub>	55	135	Differentiating xylem wood (adult, <i>Pinus taeda</i> ) BioSample: SAMN00156479	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	-
LOP11	AA739689	F:CCAGAAGGCTATAGTACAC R:CAACAATACAAGTAGCAATAC	(TA) <sub>2</sub> T (TA) <sub>12</sub>	45	235	Xylem branches ( <i>Pinus taeda</i> ) BioSample: SAMN00155658	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	-
LOP12	AW888197	F:AGGACAGTCCTTACTGCCCAAR:CAT GTTTTCCCATGGTTTTCC	(TA) <sub>26</sub>	45	156, 160	Xylem normal wood ( <i>Pinus taeda</i> ) BioSample: SAMN00157372	<i>Pinus taeda</i> , <i>Pinus ponderosa</i> , <i>Pinus contorta</i> ssp. <i>latifolia</i> , <i>Pinus sylvestris</i> [45]	-
PtSIFG_1102	BQ700433	F:ACGGAGATATATTGCAGGCG R:AAAGAATAACCTGAAACAAACCC	(TA) <sub>16</sub>	60	119	Xylem root (primary) (transitional, <i>Pinus taeda</i> ) BioSample: SAMN00170108	<i>Pinus taeda</i> [53]	[71]
PtSIFG_0737	CF476612	F:GCAAGGGGAATTGCTTATGA R:GGGATCGCATCAGCTGTAAT	(CAG) <sub>6</sub> (CAGCAT) <sub>6</sub>	60	429	Roots ( <i>Pinus taeda</i> ) BioSample: SAMN00173780	<i>Pinus taeda</i> [53]	[71]
PpSIFG_3145	BX254108	F:TGTATATTGCGCCTGGTGGT R:ATCAAATCCAGAATCAGGCG	(GAG) <sub>5</sub> (CAG) <sub>5</sub> (GCA) <sub>6</sub>	60	372	Differentiating xylem wood (adult, <i>Pinus pinaster</i> ) BioSample: SAMN00172245	<i>Pinus taeda</i> [53]	[71]
PtSIFG_5015	CZ894980	F:AAGTCAAAGGTCAAATCATTATGC R:AAACAAGCCACACAAACACG	(AAT) <sub>8</sub>	55	388–405	Methylation filtered library ( <i>Pinus taeda</i> ) BioSample: SAMN00183208	<i>Pinus taeda</i> [52]	[67]
PtSIFG_6044	CO164176	F:TGATGGGTGCCAGTAACAAA R:AATTGACCAATCGTGCCTCT	(TA) <sub>8</sub>	55	235–248	Roots of 1-year-old cuttings that were rooted ( <i>Pinus taeda</i> ) BioSample: SAMN00175137	<i>Pinus taeda</i> [52]	[67]
PtSIFG_6065	CO365046	F: TATTCCGGTAGTCTGGCACC R:GAAGGAAGCAGCTTTTGCAC	(AT) <sub>8</sub>	55	204–222	Roots of 1-year-old cuttings that were rooted ( <i>Pinus taeda</i> ) BioSample: SAMN00175162	<i>Pinus taeda</i> [52]	[67]

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
psyl2	HQ113935	F:TTGCTTTTGCAGAACATTTCG R:GTCCTGCAGGCAATCAAAAT	(GCT) <sub>5</sub>	55	199-211	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73, 76, 79, 86, 157, 164, 169, 186–191]
psyl16	HQ113936	F:GCTCTGCCCATGCTATCACT R:TGATGCTACCCAATGAGGTG	(AT) <sub>7</sub>	55	202-210	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–79, 86, 169, 178, 186–189, 192]
psyl17	HQ113937	F:TGGTCTGCAAATCAATCGAA R:GGGTAGGAATGCAAGTTAGGC	(TA) <sub>7</sub>	55	219-251	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–78, 157, 168, 169, 182, 193, 194]
psyl18	HQ113938	F:ACTACCTGGCATTTCGTCCTG R:GGATCTGGTCCATTTCGTGT	(GCA) <sub>7</sub>	55	297-306	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73, 79, 86, 164, 169, 186–189, 192, 195]
psyl19	HQ113939	F:GGCTGTAATTGGCACAGGTT R:CGAGGTGGTACACAGCAACA	(GCT) <sub>7</sub>	55	315-324	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–75, 79, 87, 169, 196]
psyl25	HQ113940	F:CAGCACGCGTTCTTTGTATC R:ACCGTTGCTCGTTGTCTCT	(GCA) <sub>5</sub>	55	214-244	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73, 79, 86, 164, 169, 186–191, 195]
psyl36	HQ113941	F:TATCATCGAGACCCCAAAA R:GAAAGGCGAAAGCAAAAGTG	(GTC) <sub>7</sub>	55	245-257	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–76, 79, 164, 190, 191, 193, 195]
psyl42	HQ113942	F:CAACTTCAGCCTTGAACAA R:CGACTTCATTGGAACACCA	(TC) <sub>9</sub>	55	171-179	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–79, 86, 164, 169, 178, 186–191, 194, 195]
psyl44	HQ113943	F:TCCAAGTTCGGTTCCTTGTG R:GACACGATGGATTCCCTGAT	(CGG) <sub>5</sub>	55	166-175	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73, 76, 78, 79, 164, 169, 187, 190, 191, 193–195]
psyl57	HQ113944	F:CCCCACATCTCTACAGTCCAA R:TGCTCTTGGATTGTTGTCTG	(ACC) <sub>7</sub>	55	187-202	Needles ( <i>Pinus sylvestris</i> ) Project EU-EVOLTREE	<i>Pinus sylvestris</i> [44]	[73–76, 78, 79, 86, 87, 164, 169, 182, 186–192, 195, 196]
lw_isotig00080	KF501196	F:CGGGCAAAATGACCGAAG R:TGGAGGAGGTAGAGGGGG	(CCG) <sub>6</sub>	55	177	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 182]
lw_isotig00081	KF501197	F:TGCGGAAGGCGTGAGTAG R:TGGAGGAGGTAGAGGGGG	(CCG) <sub>6</sub>	58	290	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
lw_isotig00542	KF501186	F:AACAGGAGCATATCAATCAA R:GTGGCATTCTACAAGCAATT	(T) <sub>40</sub>	55	257	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig01420	KF501198	F:TCCGTGACCCTATTACGT R:CGATTAGTTGCTTGCCCTT	(CTG) <sub>5</sub>	50	174	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 182, 197]
lw_isotig02138	KF501199	F:ATGCATCTTGTCTCTCT R:TTCCTGATTACACTCCC	(AG) <sub>6</sub>	42	124	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig02347	KF501200	F:CTCGTCTTCTTGTCCGC R:GCTATTGCTCCACTTGCC	(TG) <sub>7</sub>	50	198	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig02842	KF501203	F:GTGATGGTGTGGTGGCTGTA R:TCCTTTGTGGGAGATTGGTG	(AGA) <sub>5</sub>	55	229	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [72]	[72]
lw_isotig03088	KF501201	F:CATTTGGTTGACTTTGT R:TTGTAGTGAGATCTGTGC	(GA) <sub>6</sub>	45	235	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig04195	KF501204	F:GAGATCACCGAAACAACAAA R:TACAAGTCCCAGCAAAACAAT	(GAG) <sub>5</sub>	55	189	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169]
lw_isotig04204	KF501187	F:CTCCGTTTGGGTTGTGTTG R:ATCCTTGCCGCCAGATTTGT	(CGGCT) <sub>5</sub>	55	230	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 197]
lw_isotig04306	KF501205	F:GCCATTTTTTCTTCTCTCTCT R:GGTCGGTTCTGAATTTCTAA	(TCC) <sub>7</sub>	55	196	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 87, 169, 194, 196]
lw_isotig04600	KF501188	F:TCAGGGAAAATGTAGGAAAATG R:AATCTGTTGTGTGGGACTTGA	(CAG) <sub>10</sub>	55	305	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig04931	KF501202	F:TAGACCTCATCACAACACT R:ACAAAAACGAATACAAAT	(AC) <sub>6</sub>	40	132	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig05123	KF501206	F:TGTGCGTATAGGAGGTGGAG R:ATGAAAGGTGACAAAGCGGT	(GAG) <sub>6</sub>	55	166	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 182, 197]
lw_isotig06215	KF501207	F:TCAGGTGCTTACCCTTTTC R:TGGCAGCTATTCCAGTCTTT	(CAA) <sub>5</sub>	55	275	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig06440	KF501189	F:GGGACAAGGCATCG R:TGGAGACTTCGGGTGC	(AGGTTG) <sub>5</sub> (AGGCTG) <sub>6</sub>	55	298	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
lw_isotig07383	KF501190	F:CAAACAAAAACAGTCTGCA R:ATCGTCATCATCATCGTCAC	(GAT) <sub>8</sub>	55	191	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 182, 197]
lw_isotig10603	KF501191	F:CAAAATCGTCTACTTCTCCCC R:CAAAGCAAAGAAGTCCAACGA	(CAG) <sub>7</sub>	55	196	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 197]
lw_isotig11166	KF501208	F:ACACACACTGAGCTCCAATTT R:AGTCCCACCTCTGCTGATACA	(TA) <sub>7</sub>	55	137	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 182, 197]
lw_isotig12667	KF501209	F:CCAAGGTGAAAAGGAAATGA R:TTCTGACAGGGAGCGACTGA	(CA) <sub>6</sub>	55	199	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig17679	KF501192	F: TTGTTTGCCACATTGTTGCC R:CAAACCACCGTGCTTCTAA	(TTAA) <sub>5</sub>	55	277	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72]
lw_isotig20215	KF501210	F:AGAGGTGATCGCAGTCAAAGA R:TCAAAAAGACCAAACCGTAG	(TA) <sub>7</sub>	55	186	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 197]
lw_isotig21953	KF501193	F:ATGGTGTGTTTGAAGCGGAA R:ATTGCAGCCACTGGTGTCTT	(ATGGG) <sub>7</sub>	55	208	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 87, 169, 182, 194, 196, 197]
lw_isotig26230	KF501194	F:GGGCATTACATAAACACGGG R:TGCCCTTGAGCATTGATTA	(TA) <sub>10</sub>	55	260	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[72, 169, 182, 197]
lw_isotig27940	KF501195	F:GCAGGCAACAACAAAAGTGACA R:AGCAATCGAGTGCCAAATCTTC	(TGGA) <sub>5</sub>	55	231	Needle and male cones (Sexual mature, <i>Pinus tabuliformis</i> ) BioSample: SAMN10845670	<i>Pinus sylvestris</i> var. <i>mongolica</i> [46]	[87, 72, 169, 194, 196]
NFPK-40	-	F:TCGCTCTCTTCTTGACCACA R:CCGCTACTTCATCAGGGTTC	(TGA) <sub>6</sub>	60	196	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-32	-	F:AAATGGACGAAGTTGGATGG R:CTCAGTGTCTTCAGGCAGGA	(GCT) <sub>6</sub>	59	197	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-53	-	F:TGGAGATGCAGCAGATTAGG R:CTGCACACAGGATGTCACAA	(ATG) <sub>6</sub>	59	197	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-117	-	F:GCCCAATGGATGTGTCTCTT R:TCGGCCTGCAATTAGTCTCT	(TC) <sub>12</sub>	60	203	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus sylvestris</i> [54]	-

Locus	NCBI accession	Primer sequences (5'–3')	Repeat motif	T <sub>a</sub> (°C)	Allele size (bp)	Tissue for DNA library (developmental stage, species)	Identification in pine species (key reference)	Marker citation for <i>Pinus sylvestris</i>
NFPK-72	-	F:ATCACCGCTGCCTTTCAGTA R:TCACTTCCCAATCAATTCC	(ATA) <sub>8</sub>	60	204	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-43	-	F:ATGCAGGGTTTGAATACAG R:AATACGAGCACCGCGTTATC	(GAG) <sub>6</sub>	60	227	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-38	-	F:TGATGGTGTGGTGAGGGTTA R:AGCGTGGGAGGAGTGTGTAG	(AAG) <sub>6</sub>	60	229	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-150	-	F:AAATAACGGGGCTGTGTGTC R:ACGGATGTTGTAATCCCAA	(GA) <sub>6</sub>	60	241	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-179	-	F:CCAAGCCAGGTAAGGCACTA R:TGGACAAGGAGATGAGACA	(CA) <sub>10</sub>	60	246	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-175	-	F:AAGGTCACGGCGTTCATTAC R:CCTGTGACCTCAACTGGGAT	(GA) <sub>6</sub>	60	259	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-
NFPK-184	-	F:AAGTCTCCACTGCATCAACCTT R:TGTCTCCCAACTTCTGCTT	(TC) <sub>8</sub>	60	275	Needles (Seedling, <i>Pinus koraiensis</i> ) BioSample: SAMN10617534	<i>Pinus koraiensis</i> , <i>Pinus sibirica</i> , <i>Pinus pumila</i> , <i>Pinus wallichiana</i> , <i>Pinus parviflora</i> , <i>Pinus bungeana</i> , <i>Pinus tabuliformis</i> , <i>Pinus sylvestris</i> [54]	-

**Table S2.** List of *Pinus sylvestris* genes associated with EST-SSRs of *P. sylvestris*.

NP - not predicted

Locus	GenBank accession number	<i>Pinus sylvestris</i> gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
PtTX2146	AF143963	NP	NP				
SsPp_cn524	BX255732	PSY00010973	isotig42993:269-973	+	5'UTR/CDS	929-972/973-1072	86.5
SsrPt_AW010960	AW010960	NP	NP				
SsrPt_AW225917	AW225917	NP	NP				
SsrPt_AW981772	AW981772	NP	NP				
SsrPt_BF049767	BF049767	NP	NP				
SsrPt_ctg1376	CD024501	NP	NP				
SsrPt_ctg1376	CD024819	NP	NP				
SsrPt_ctg1525	BQ655604	NP	NP				
SsrPt_ctg17601	AW985278	PSY00008240	isotig38539:923-1720	+	CDS	1285-1509	96.9
SsrPt_ctg18103	BE187051	PSY00026391	isotig68644:261-740	+	3'UTR	937-1134	83.4
SsrPt_ctg18103	BE187051	PSY00024530	isotig65588:160-639	+	3'UTR	836-1025	79.9
SsrPt_ctg2300	JQ021072	PSY00014290	isotig50849:90-527	+	CDS/3'UTR	405-527/528-574	93.6
SsrPt_ctg275	AW869955	NP	NP				
SsrPt_ctg3021	CV135003	PSY00012130	isotig45153:530-1255	+	CDS	768-995	94.5
SsrPt_ctg3021	DT625916	PSY00012130	isotig45153:530-1255	+	CDS	768-995	94.5
SsrPt_ctg3021	CF478541	PSY00012130	isotig45153:530-1255	+	CDS	768-995	94.5
SsrPt_ctg3089	BQ291150	NP	NP				
SsrPt_ctg3754	BQ701798	NP	NP				
SsrPt_ctg4363	AI812743	NP	NP				
SsrPt_ctg4363	AW736917	NP	NP				
SsrPt_ctg4363	BQ702924	NP	NP				
SsrPt_ctg4487a	BQ702547	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	96.8
SsrPt_ctg4487a	BG040578	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	97.5
SsrPt_ctg4487a	BG040033	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	97.5
SsrPt_ctg4487a	BG039427	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	97.5
SsrPt_ctg4487a	BF517866	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	96.8

Locus	GenBank accession number	Pinus sylvestris gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
SsrPt_ctg4487a	BF517086	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	96.8
SsrPt_ctg4487a	BF060592	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	94.9
SsrPt_ctg4487a	BQ700883	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	97.5
SsrPt_ctg4487a	BM492929	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	96.8
SsrPt_ctg4487a	BE458136	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-940/784	96.8
SsrPt_ctg4487b	BG039427	PSY00005414	isotig28724:320-784	-	CDS	537-784	90.9
SsrPt_ctg4487b	BQ701415	PSY00005414	isotig28724:320-784	-	CDS	537-784	89.0
SsrPt_ctg4698	BG275886	PSY00006544	isotig33050:1465-1893	+	CDS	1560-1809	93.6
SsrPt_ctg5167	unknown	PSY00010492	isotig42207:345-2087	+	CDS	1687-1979	-
SsrPt_ctg5333	BQ700216	PSY00010492	isotig42207:345-2087	+	CDS	348-504	95.1
SsrPt_ctg5333	BF610405	PSY00010492	isotig42207:345-2087	+	CDS	348-504	95.1
SsrPt_ctg64	CV147495	PSY00014290	isotig50849:90-527	+	CDS/3'UTR	281-527/528-562	83.7
SsrPt_ctg64	CV144063	PSY00014290	isotig50849:90-527	+	CDS/3'UTR	281-527/528-562	84.4
SsrPt_ctg64	DT625112	PSY00014290	isotig50849:90-527	+	CDS/3'UTR	281-527/528-562	83.7
SsrPt_ctg7024	BQ695936	PSY00003770	isotig21292:478-1161	+	CDS/3'UTR	930-1161/1162-1207	96.6
SsrPt_ctg7081	BE187193	PSY00020303	isotig60441:381-1721	-	CDS/3'UTR	381-633/256-380	96.6
SsrPt_ctg7141	BV728811	PSY00005414	isotig28724:320-784	-	5'UTR/CDS	785-786/403-784	91.7
SsrPt_ctg7170	BQ696574	PSY00002648	isotig13512:534-1829	-	5'UTR/CDS	1830-1873/1490-1829	98.7
SsrPt_ctg7170	BG039619	PSY00002648	isotig13512:534-1829	-	5'UTR/CDS	1830-1873/1490-1829	97.9
SsrPt_ctg7170	BF778892	PSY00002648	isotig13512:534-1829	-	5'UTR/CDS	1830-1873/1490-1829	98.7
SsrPt_ctg7425	CF385379	PSY00000824	isotig06879:715-1842	+	CDS	736-1119	97.7
SsrPt_ctg7425	CF394315	PSY00000824	isotig06879:715-1842	+	CDS	736-1119	97.4
SsrPt_ctg7444	BQ196808	PSY00000153	118-2-51-H01.r.1:1-648	+	CDS	367-638	85.6
SsrPt_ctg7444	BQ107124	PSY00000153	118-2-51-H01.r.1:1-648	+	CDS	367-638	85.9
SsrPt_ctg7444	BM367065	PSY00000153	118-2-51-H01.r.1:1-648	+	CDS	367-638	85.6
SsrPt_ctg7444	BF518214	PSY00000153	118-2-51-H01.r.1:1-648	+	CDS	367-638	86.2
SsrPt_ctg7731	BF517779	PSY00017742	isotig58027:487-1944	+	5'UTR/CDS	331-486/487-543	92.4
SsrPt_ctg7731	BE451949	PSY00017742	isotig58027:487-1944	+	5'UTR/CDS	331-486/487-543	92.4
SsrPt_ctg7824	BQ703132	PSY00014046	isotig50027:353-880	-	CDS/3'UTR	353-494/1-352	90.9
SsrPt_ctg7867	BQ701488	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	74.2
SsrPt_ctg7867	BG275803	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	73.6



Locus	GenBank accession number	Pinus sylvestris gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
SsrPt_ctg7867	BG040533	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	73.6
SsrPt_ctg7867	BG040463	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	73.6
SsrPt_ctg7867	BF777863	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	73.6
SsrPt_ctg7867	BE643974	PSY00001445	isotig102325:78-566	+	CDS/3'UTR	426-566/567-568	73.6
SsrPt_ctg8064	CV148734	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV148725	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV148543	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV148254	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV148092	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV147527	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV146957	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV146737	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV146675	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV146556	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV146193	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV145664	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	CV145578	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	DR181307	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	DR092258	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	DR069750	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	DR068842	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg8064	DR048459	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-853	95.3
SsrPt_ctg865	BQ699190	NP	NP				
SsrPt_ctg865	BQ699199	NP	NP				
SsrPt_ctg9249	BE761900	PSY00009266	isotig40387:2378-2830	-	5'UTR/CDS	2831-2973/2816-2830	95.6
RPtest1	BF517869	PSY00008891	isotig39897:766-3753	-	3'UTR	344-468	92.0
RPtest1	BF517853	PSY00008891	isotig39897:766-3753	-	3'UTR	344-468	86.4
RPtest5	BQ697205	PSY00003137	isotig17309:594-1424	-	CDS	942-1141	74.2
RPtest5	BE997187	PSY00003137	isotig17309:594-1424	-	CDS	942-1141	70.8
RPtest5	BE643880	PSY00003137	isotig17309:594-1424	-	CDS	942-1141	72.2

Locus	GenBank accession number	Pinus sylvestris gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
RPtest6	BG040975	PSY00019692	isotig59809:462-1625	+	CDS	524-670	100.0
RPtest6	BF518209	PSY00019692	isotig59809:462-1625	+	CDS	524-670	100.0
RPtest9	CF393810	NP	NP				
RPtest11	CF391802	PSY00011770	isotig4442:769-1344	-	CDS/3'UTR	769-920/708-768	96.7
RPtest11	BI202760	PSY00011770	isotig4442:769-1344	-	CDS/3'UTR	769-920/708-768	97.2
RPt11est13	AA739656	PSY00025105	isotig66460:506-1204	-	CDS/3'UTR	506-640/364-505	96.4
RPtest15	BG275136	PSY00002483	isotig11881:200-844	+	CDS/3'UTR	707-844/845-951	95.6
RPtest16	AA739818	PSY00014450	isotig51443:196-663	+	CDS	281-414	94.0
LOP1	AI812473	PSY00018563	isotig58753:615-1898	+	CDS/3'UTR	1529-1898/1899-1910	95.5
LOP3	AA556662	PSY00000153	118-2-51-H01.r.1:1-648	+	CDS	435-642	81.9
LOP5	AW758812	PSY00023395	isotig63967:168-1271	+	3'UTR	1300-1534	59.3
LOP6	AA556221	PSY00008891	isotig39897:766-3753	-	3'UTR	271-498	95.2
LOP8	AI725303	PSY00004341	isotig24276:3-350	+	3'UTR	453-659	65.7
LOP9	AI813163	NP	NP				
LOP11	AA739689	PSY00023490	isotig64088:321-1190	+	3'UTR	1194-1428	87.9
LOP12	AW888197	NP	NP				
PtSIFG_1102	BQ700433	PSY00008719	isotig39566:335-640	+	5'UTR	180-311	55.6
PtSIFG_0737	CF476612	NP	NP				
PpSIFG_3145	BX254108	NP	NP				
PtSIFG_5015	CZ894980	NP	NP				
PtSIFG_6044	CO164176	PSY00004353	isotig24357:349-4695	-	3'UTR	51-299	85.2
PtSIFG_6065	CO365046	PSY00027604	isotig70975:361-1041	-	CDS/3'UTR	361-362/152-360	90.8
psyl2	HQ113935	PSY00005261	isotig28291:1094-1984	+	CDS/3'UTR	1875-1984/1985-2086	98.6
psyl16	HQ113936	PSY00018010	isotig58249:576-1496	-	3'UTR	367-567	100.0
psyl17	HQ113937	PSY00018010	isotig58249:576-1496	-	3'UTR	247-468	100.0
psyl18	HQ113938	PSY00016934	isotig57336:287-2248	+	CDS	1636-1845	97.7
psyl19	HQ113939	PSY00016934	isotig57336:287-2248	+	CDS	1736-1960	100.0
psyl25	HQ113940	PSY00000191	Cl#29_1740_18:1095-1670	-	5'UTR/CDS	1671-1692/1473-1670	98.6
psyl36	HQ113941	PSY00003770	isotig21292:478-1161	+	CDS/3'UTR	950-1161/1162-1202	100.0

Locus	GenBank accession number	Pinus sylvestris gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
psyl42	HQ113942	NP	NP				
psyl44	HQ113943	PSY00013146	isotig47431:360-1460	-	CDS	1027-1379	100.0
psyl57	HQ113944	NP	NP				
lw_isotig00080	KF501196	PSY00026757	isotig69313:87-539	+	CDS	338-514	99.4
lw_isotig00081	KF501197	PSY00026757	isotig69313:87-539	+	CDS/3'UTR	338-539/540-627	97.9
lw_isotig00542	KF501186	PSY00008008	isotig38196:81-3197	+	3'UTR	3515-3820	80.4
lw_isotig01420	KF501198	PSY00020425	isotig60567:1360-1923	-	5'UTR/CDS	1924-1925/1782-1923	82.8
lw_isotig02138	KF501199	PSY00016229	isotig56763:1078-2463	+	CDS/3'UTR	2348-2463/2464-2471	99.2
lw_isotig02347	KF501200	PSY00016640	isotig57098:2-2437	+	3'UTR	2530-2781	99.6
lw_isotig02842	KF501203	PSY00012746	isotig46483:455-1654	-	5'UTR/CDS	1655-1725/1500-1654	98.7
lw_isotig03088	KF501201	NP	NP				
lw_isotig04195	KF501204	NP	NP				
lw_isotig04204	KF501187	PSY00005987	isotig30656:578-1024	+	5'UTR	299-534	70.7
lw_isotig04306	KF501205	PSY00011561	isotig44019:976-1800	-	CDS	1298-1496	92.2
lw_isotig04600	KF501188	PSY00011465	isotig43845:480-1664	-	5'UTR/CDS	1665-1807/1512-1664	94.2
lw_isotig04931	KF501202	PSY00018521	isotig58716:127-1716	+	3'UTR	1853-1986	97.8
lw_isotig05123	KF501206	NP	NP				
lw_isotig06215	KF501207	PSY00022080	isotig62302:472-1542	-	CDS/3'UTR	472-709/432-471	98.9
lw_isotig06440	KF501189	PSY00018317	isotig58527:503-1960	+	CDS/3'UTR	1615-1960/1961-1992	98.2
lw_isotig07383	KF501190	PSY00006544	isotig33050:1465-1893	+	CDS	1522-1719	96.5
lw_isotig10603	KF501191	PSY00011044	isotig43111:455-2419	-	CDS	2028-2226	98.5
lw_isotig11166	KF501208	PSY00030466	isotig78457:423-893	-	CDS/3'UTR	423-424/292-422	56.1
lw_isotig12667	KF501209	PSY00012407	isotig45721:725-1738	+	5'UTR	234-433	98.5
lw_isotig17679	KF501192	NP	NP				
lw_isotig20215	KF501210	PSY00031065	isotig80097:1-552	+	3'UTR	600-787	98.9
lw_isotig21953	KF501193	PSY00005415	isotig28727:318-971	-	CDS/3'UTR	389-971/318-388	80.0
lw_isotig26230	KF501194	PSY00034638	isotig92892:2-403	+	CDS/3'UTR	395-403/404-643	92.7
lw_isotig27940	KF501195	PSY00003211	isotig17895:393-1301	+	3'UTR	1738-1984	92.3

**Table S3.** List of *Pinus taeda* genes associated with EST-SSRs available for Scots pine

NP - not predicted

Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
PtTX2146	AF143963	PITA_000026471	tscaffold2717:130233-179614	+	CDS	130428-130602	89.6
SsPp_cn524	BX255732	PITA_000001867	tscaffold934:486337-504350	+	5'UTR/CDS	486529-486573/486574-486672	85.9
SsrPt_AW010960	AW010960	NP	scaffold652460.2:191756-192535		NP	156441-156669	95.2
SsrPt_AW225917	AW225917	NP	tscaffold2587:1-1003788		NP	666204-666397	97.5
SsrPt_AW981772	AW981772	PITA_000031044	scaffold702685.2:82973-83909	+	CDS/Intron	83171-83449/83450-83487	84.2
SsrPt_BF049767	BF049767	NP	NP				
SsrPt_ctg1376	CD024501	PITA_000000887	tscaffold1619:2204779-2299325	+	3'UTR	2299570-2299725	76.9
SsrPt_ctg1376	CD024819	PITA_000000887	tscaffold1619:2204779-2299325	+	3'UTR	2299570-2299725	76.9
SsrPt_ctg1525	BQ655604	NP	tscaffold464:1-1643423		NP	55236-55705	97.7
SsrPt_ctg17601	AW985278	PITA_000037826	scaffold605875:29557-33185	+	CDS	30576-30800	97.8
SsrPt_ctg18103	BE187051	PITA_000025493	tscaffold5146:241165-242512	-	3'UTR	241201-241369	91.3
SsrPt_ctg18103	BE187051	PITA_000086735	scaffold572318:20221-21576	-	3'UTR	20257-20441	84.5
SsrPt_ctg18103	BE187051	PITA_000047867	scaffold23967.1:25138-26414	+	3'UTR	26273-26414	85.1
SsrPt_ctg18103	BE187051	PITA_000047868	scaffold23967.1:98370-99596	+	3'UTR	99446-99596	86.2
SsrPt_ctg18103	BE187051	PITA_000051460	scaffold692212:69637-70852	+	3'UTR	70707-70852	85.8
SsrPt_ctg18103	BE187051	PITA_000072543	scaffold93005:38194-39458	-	3'UTR	38194-38345	81.2
SsrPt_ctg18103	BE187051	PITA_000030491	scaffold808138:10414-11725	+	3'UTR	11574-11725	83.1
SsrPt_ctg2300	JQ021072	PTA00008090	scaffold501460.1:146871-147533	+	CDS/3'UTR	147405-147533/147534-147580	98.3
SsrPt_ctg275	AW869955	NP	scaffold885361:1-63699				78.8
SsrPt_ctg3021	CV135003	PITA_000031795	tscaffold9139:83281-146328	+	Intron/CDS	145302-145306/145307-145538	99.2
SsrPt_ctg3021	DT625916	PITA_000031795	tscaffold9139:83281-146328	+	Intron/CDS	145302-145306/145307-145538	99.2
SsrPt_ctg3021	CF478541	PITA_000031795	tscaffold9139:83281-146328	+	Intron/CDS	145302-145306/145307-145538	99.2
SsrPt_ctg3089	BQ291150	NP	scaffold896629:1-46925				88.0
SsrPt_ctg3754	BQ701798	PITA_000016257	tscaffold5544:46493-48634	-	5'UTR/CDS	48635-48900/48480-48634	98.6
SsrPt_ctg4363	AI812743	PITA_000068716	scaffold56036:50303-51756	-	3'UTR	50538-50633	95.0
SsrPt_ctg4363	AW736917	PITA_000068716	scaffold56036:50303-51756	-	3'UTR	50538-50633	94.1
SsrPt_ctg4363	BQ702924	PITA_000068716	scaffold56036:50303-51756	-	3'UTR	50538-50633	95.0

Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
SsrPt_ctg4487a	BQ702547	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	99.4
SsrPt_ctg4487a	BG040578	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	100.0
SsrPt_ctg4487a	BG040033	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	100.0
SsrPt_ctg4487a	BG039427	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	100.0
SsrPt_ctg4487a	BF517866	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	99.4
SsrPt_ctg4487a	BF517086	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	99.4
SsrPt_ctg4487a	BF060592	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	97.4
SsrPt_ctg4487a	BQ700883	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	100.0
SsrPt_ctg4487a	BM492929	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	99.4
SsrPt_ctg4487a	BE458136	PITA_000029961	scaffold463697:28233-34820	-	Intron/CDS	29128-29243/29089-29127	99.4
SsrPt_ctg4487b	BG039427	PITA_000029961	scaffold463697:28233-34820	-	CDS	28845-29089	95.7
SsrPt_ctg4487b	BQ701415	PITA_000029961	scaffold463697:28233-34820	-	CDS	28845-29089	94.1
SsrPt_ctg4698	BG275886	PITA_000017622	tscaffold7099:63111-67113	-	CDS	63452-63697	98.0
SsrPt_ctg5167	unknown	NP	NP				
SsrPt_ctg5333	BQ700216	PTA00009525	tscaffold8498:59439-61351	-	CDS	61186-61348	100.0
SsrPt_ctg5333	BF610405	PTA00009525	tscaffold8498:59439-61351	-	CDS	61186-61348	100.0
SsrPt_ctg64	CV147495	PTA00008090	scaffold501460.1:146871-147533	+	CDS/3'UTR	147283-147533/147534-147569	98.6
SsrPt_ctg64	CV144063	PTA00008090	scaffold501460.1:146871-147533	+	CDS/3'UTR	147283-147533/147534-147569	98.6
SsrPt_ctg64	DT625112	PTA00008090	scaffold501460.1:146871-147533	+	CDS/3'UTR	147283-147533/147534-147569	98.6
SsrPt_ctg7024	BQ695936	PITA_000023332	tscaffold3287:65555-104502	-	CDS/3'UTR	65995-66226/65949-65994	99.3
SsrPt_ctg7081	BE187193	NP	scaffold398783:1-309391				84.4
SsrPt_ctg7141	BV728811	PITA_000029961	scaffold463697:28233-34820	-	CDS/Intron/CDS	28806-29091/28705-28805/28610-28704	77.6
SsrPt_ctg7170	BQ696574	PITA_000027567	tscaffold5297:152922-233184	+	5'UTR/CDS	153116-153160/153161-153500	100.0
SsrPt_ctg7170	BG039619	PITA_000027567	tscaffold5297:152922-233184	+	5'UTR/CDS	153116-153160/153161-153500	99.2
SsrPt_ctg7170	BF778892	PITA_000027567	tscaffold5297:152922-233184	+	5'UTR/CDS	153116-153160/153161-153500	100.0
SsrPt_ctg7425	CF385379	PITA_000004120	tscaffold2162:112628-118544	+	CDS	112887-113270	99.7
SsrPt_ctg7425	CF394315	PITA_000004120	tscaffold2162:112628-118544	+	CDS	112887-113270	99.5
SsrPt_ctg7444	BQ196808	PITA_000012096	tscaffold7249:434167-436159	+	CDS/3'UTR	435708-435890/435891-435993	97.9
SsrPt_ctg7444	BQ107124	PITA_000012096	tscaffold7249:434167-436159	+	CDS/3'UTR	435708-435890/435891-435993	98.3

Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
SsrPt_ctg7444	BM367065	PITA_000012096	tscaffold7249:434167-436159	+	CDS/3'UTR	435708-435890/435891-435993	97.6
SsrPt_ctg7444	BF518214	PITA_000012096	tscaffold7249:434167-436159	+	CDS/3'UTR	435708-435890/435891-435993	97.9
SsrPt_ctg7731	BF517779	PITA_000091704	scaffold886273.1:5846-11254	+	5'UTR/CDS	5942-6101/6102-6158	95.1
SsrPt_ctg7731	BE451949	PITA_000091704	scaffold886273.1:5846-11254	+	5'UTR/CDS	5942-6101/6102-6158	95.1
SsrPt_ctg7824	BQ703132	PITA_000076250	scaffold891409:6243-7852	+	CDS/3'UTR	7359-7500/7501-7866	96.3
SsrPt_ctg7867	BQ701488	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	98.7
SsrPt_ctg7867	BG275803	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	100.0
SsrPt_ctg7867	BG040533	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	100.0
SsrPt_ctg7867	BG040463	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	100.0
SsrPt_ctg7867	BF777863	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	100.0
SsrPt_ctg7867	BE643974	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299417-299570	100.0
SsrPt_ctg8064	CV148734	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV148725	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV148543	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV148254	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV148092	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV147527	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV146957	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV146737	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV146675	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV146556	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV146193	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV145664	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	CV145578	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	DR181307	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	DR092258	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	DR069750	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	DR068842	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg8064	DR048459	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285387	100.0
SsrPt_ctg865	BQ699190	NP	scaffold40707				92.7

Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
SsrPt_ctg865	BQ699199	NP	scaffold40707				92.7
SsrPt_ctg9249	BE761900	PITA_000010076	tscaffold2087:94332-152939	-	5'UTR/CDS	152940-153080/152925-152939	98.1
RPtest1	BF517869	PITA_000070971	scaffold863061:37061-41395	+	CDS	40931-41055	96.0
RPtest1	BF517869	PITA_000006918	tscaffold3252:353148-364676	-	CDS/Intron/CDS	362987-363012/362954-362986/362890-362953	85.6
RPtest1	BF517853	PITA_000070971	scaffold863061:37061-41395	+	CDS	40931-41055	90.4
RPtest1	BF517853	PITA_000006918	tscaffold3252:353148-364676	-	CDS/Intron/CDS	362987-363012/362954-362986/362890-362953	80.8
RPtest5	BQ697205	PITA_000007179	tscaffold1149:676817-682707	+	CDS	677823-678019	71.4
RPtest5	BE997187	PITA_000007179	tscaffold1149:676817-682707	+	CDS	677823-678019	68.1
RPtest5	BE643880	PITA_000007179	tscaffold1149:676817-682707	+	CDS	677823-678019	69.5
RPtest6	BG040975	NP	scaffold489083: 1-299329				100.0
RPtest6	BF518209	NP	scaffold489083: 1-299329				100.0
RPtest9	CF393810	PITA_000026471	tscaffold2717:130233-179614	+	CDS	130447-130719	95.6
RPtest11	CF391802	PITA_000052634	scaffold59926.1:57681-58076	+	CDS/3'UTR	57925-58076/58077-58137	95.8
RPtest11	CF391802	PITA_000085253	scaffold737600:2371-2766	-	CDS/3'UTR	2371-2522/2310-2370	95.8
RPtest11	CF391802	PITA_000013530	tscaffold8323:496444-496839	+	CDS/3'UTR	496688-496839/496840-496900	97.2
RPtest11	BI202760	PITA_000052634	scaffold59926.1:57681-58076	+	CDS/3'UTR	57925-58076/58077-58137	96.2
RPtest11	BI202760	PITA_000085253	scaffold737600:2371-2766	-	CDS/3'UTR	2371-2522/2310-2370	96.2
RPtest11	BI202760	PITA_000013530	tscaffold8323:496444-496839	+	CDS/3'UTR	496688-496839/496840-496900	97.7
RPtest13	AA739656	PITA_000065673	scaffold7694:10422-12292	-	CDS/3'UTR	10993-11127/10851-10992	94.9
RPtest15	CF391220	PITA_000008378	tscaffold2564:284513-285787	+	CDS/3'UTR	285241-285378/285379-285486	100.0
RPtest16	AA739818	NP	scaffold568940:1-356899				94.8
LOP1	AI812473	PITA_000008107	tscaffold702:125040-162886	+	CDS/3'UTR	162153-162612/162613-162624	94.9
LOP1	AI812473	PITA_000008108	tscaffold702: 233620-234915	+	CDS/3'UTR	234546-234915/234916-234927	91.6
LOP3	AA556662	PITA_000012096	tscaffold7249:434167-436159	+	CDS/3'UTR	435776-435890/435891-435996	96.4
LOP5	AW758812	NP	scaffold3633.2:1-185090				75.7
LOP6	AA556221	PITA_000070971	scaffold863061:37061-41395	+	CDS/3'UTR	40901-41066/41067-41128	97.4
LOP8	AI725303	PITA_000044661	scaffold904057:80942-81535	+	CDS/3'UTR	81401-81535/81536-81609	72.1
LOP9	AI813163	PTA00066643	C32322558:2799-5248	+	CDS	2821-2958	97.8
LOP11	AA739689	PITA_000023770	tscaffold21:83186-167212	+	3'UTR	166917-167180	88.7

Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
LOP12	AW888197	PITA_000035518	scaffold701498.2:777-37567	+	3'UTR	37620-37778	80.1
PtSIFG_1102	BQ700433	NP	scaffold885361:1-63699				95.2
PtSIFG_1102	BQ700433	PTA00050445	scaffold895022:14751-15418	-	5'UTR	15442-15573	61.6
PtSIFG_0737	CF476612	PITA_000017323	scaffold225811:286750-287481	-	CDS/3'UTR	286750-287007/286579-286749	100.0
PpSIFG_3145	BX254108	PITA_000026471	tsc scaffold2717:130233-179614	+	5'UTR/CDS	130320-130385/130386-130712	93.6
PtSIFG_5015	CZ894980	NP	C27750010				
PtSIFG_6044	CO164176	PITA_000008712	tscaffold8432:288348-297321	-	3'UTR	288387-288621	81.1
PtSIFG_6065	CO365046	NP	tscaffold2414:1-223447				77.1
psyl2	HQ113935	PITA_000017503	tscaffold4148:279914-285934	-	CDS/3'UTR	280153-280262/280054-280152	95.3
psyl16	HQ113936	PITA_000010550	tscaffold4278:587042-589507	+	3'UTR	588976-589181	95.6
psyl17	HQ113937	PITA_000010550	tscaffold4278:587042-589507	+	3'UTR	589076-589301	94.7
psyl18	HQ113938	PITA_000064129	scaffold847237:25-831	+	CDS/Intron	297-403/404-497	92.6
psyl19	HQ113939	PITA_000086838	scaffold835269:23558-24346	-	Intron/CDS	23896-24001/23895-23777	97.8
psyl19	HQ113939	PITA_000041571	scaffold458326:63544-64497	+	CDS	63985-64209	96.4
psyl25	HQ113940	PITA_000008677	tscaffold2670:368130-369520	+	5'UTR/CDS	368269-368290/368291-368491	81.1
psyl36	HQ113941	PITA_000023332	tscaffold3287:65555-104502	-	CDS/3'UTR	65995-66206/65954-64994	99.2
psyl42	HQ113942	NP	tscaffold1683:1-902503				87.9
psyl42	HQ113942	PITA_000007965	tscaffold2193:111273-184939	+	Intron	119217-119397	72.0
psyl44	HQ113943	PITA_000017051	tscaffold8715:106632-110579	+	CDS	106774-106946	97.1
psyl57	HQ113944	PITA_000010946	tscaffold4287:302033-306896	+	CDS	302090-302280	90.5
lw_isotig00080	KF501196	PITA_000007834	tscaffold5017:299018-299620	+	CDS	299422-299595	92.3
lw_isotig00080	KF501196	PITA_000009956	scaffold501460.1:146871-147533	+	CDS	147335-147508	88.4
lw_isotig00081	KF501197	PITA_000009956	scaffold501460.1:146871-147533	+	CDS/3'UTR	147335-147533/147534-147621	88.8
lw_isotig00542	KF501186	PTA00082932	tscaffold16:105685-743656	-	3'UTR	105034-105312	82.4
lw_isotig01420	KF501198	PITA_000023899	tscaffold6848:267284-268261	+	5'UTR/CDS	267249-267283/267284-267422	99.4
lw_isotig02138	KF501199	PITA_000020633	tscaffold5747:199738-238934	-	CDS/3'UTR	200393-200508/200385-200392	99.1
lw_isotig02347	KF501200	NP	NP				
lw_isotig02842	KF501203	PITA_000016857	tscaffold8246:70583-70978	+	5'UTR/CDS	70509-70582/70583-70737	99.1
lw_isotig03088	KF501201	PTA00015290	scaffold609698:7916-209563	-	Intron	86587-86827	94.2
lw_isotig04195	KF501204	PITA_000001889	tscaffold2175:1756736-1757977	-	5'UTR/CDS	1757978-1758055/1757870-1757977	92.6



Locus	GenBank accession number	Pinus taeda gene ID	Genomic Sequence	Strand (gene)	Gene region	Start-end positions of marker within corresponding gene region	% identity of marker and corresponding gene region
lw_isotig04195	KF501204	PITA_000034486	scaffold587116:88897-89798	+	5'UTR/CDS	88819-88896/88897-89004	84.7
lw_isotig04204	KF501187	PITA_000008362	tscaffold2655:632355-636821	+	5'UTR	632531-632761	76.3
lw_isotig04306	KF501205	PTA00014456	scaffold802134:180296-264396	-	CDS	263930-264131	93.2
lw_isotig04600	KF501188	PTA00083895	scaffold5983:39891-217039	+	Intron/CDS	215709-215968/215969-216007	90.8
lw_isotig04931	KF501202	PITA_000025726	tscaffold5404:18473-113604	+	3'UTR	113566-113732	71.3
lw_isotig05123	KF501206	PTA00083918	tscaffold6082:163677-284369	-	Intron	283934-284099	97.6
lw_isotig06215	KF501207	PTA00008461	C32573268:542327-543394	+	CDS/3'UTR	543160-543394/543395-543432	96.4
lw_isotig06440	KF501189	PITA_000080433	scaffold63108.1:17038-19097	+	Intron/CDS/ Intron	18150-18202/18203-18478/18479-18536	97.7
lw_isotig07383	KF501190	PITA_000017622	tscaffold7099:63111-67113	-	CDS	63545-63735	92.3
lw_isotig10603	KF501191	PITA_000023436	tscaffold3281:62057-67715	-	CDS	67040-67226	89.4
lw_isotig11166	KF501208	PITA_000059576	scaffold596801:49948-50602	+	CDS/3'UTR	50494/50495-50629	94.9
lw_isotig12667	KF501209	PTA00083219	tscaffold1096 :49329-249162	+	5'UTR	48834-49033	96.2
lw_isotig17679	KF501192	PITA_000014378	tscaffold7457:409736-461083	-	Intron	457599-457875	96.0
lw_isotig20215	KF501210	PTA00083968	tscaffold6257:5961-365375	-	3'UTR	5728-5912	93.5
lw_isotig21953	KF501193	PTA00006979	C32573532:675332-678188	-	CDS/3'UTR	675332-675402/675204-675331	76.3
lw_isotig26230	KF501194	PITA_000038492	tscaffold8841:72708-73953	+	CDS/3'UTR	73512-73566/73567-73756	90.0
lw_isotig27940	KF501195	PITA_000016972	tscaffold8713:402087-404324	-	3'UTR	402184-402378	78.5

**Table S4.** Results of Blast2GO analysis and analysis of conserved domains performed for the *Pinus sylvestris* genes associated with EST-SSRs

NA - not available, NP - not predicted

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
PtTX2146	NP	NA	-	-	NA	-	NA	NA	NA
SsPp_cn524	PSY00010973	galacturonosyltransferase 13	probable galacturonosyltransferase 14 isoform X2 [ <i>Durio zibethinus</i> ], XP_022774547.1	4.30e-75	PLN02870 super family (cl30824)	8.73e-109	pollen development (GO:0009555); pollen tube growth (GO:0009860); cell wall pectin biosynthetic process (GO:0052325);	polygalacturonate 4-alpha-galacturonosyltransferase activity (GO:0047262);	Golgi apparatus (GO:0005794); membrane (GO:0016020); pollen tube (GO:0090406);
SsrPt_AW010960	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_AW225917	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_AW981772	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_BF049767	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg1376	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg1525	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg17601	PSY00008240	unknown	unknown [ <i>Picea sitchensis</i> ], ABK23821.1	1.39e-45	NA	-	NA	NA	NA
SsrPt_ctg18103	PSY00026391	olee1-like protein	extensin-like protein [ <i>Pinus taeda</i> ], QDZ06028.1	3.49e-106	Pollen_Ole_e_I Superfamily (cl03128)	2.26e-12	NA	NA	extracellular space (GO:0005615);
	PSY00024530	olee1-like protein	extensin-like protein [ <i>Pinus taeda</i> ], QDZ06028.1	1.29e-91	Pollen_Ole_e_I Superfamily (cl03128)	1.79e-14	NA	NA	extracellular space (GO:0005615);
SsrPt_ctg2300	PSY00014290	cold shock protein 2-like	putative glycine-rich protein [ <i>Picea glauca</i> ], ABA54143.1	7.57e-67	CSP_CDS (cd04458)	5.69e-26	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
SsrPt_ctg275	NP	NA	-	-	NA	-	NA	NA	NA

Locus	<i>Pinus sylvestris</i> gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg3021	PSY00012130	GRF1-interacting factor 1	unknown [ <i>Picea sitchensis</i> ], ABK22130.1	2.83e-109	SSXT (pfam05030)	3.83e-19	cell population proliferation (GO:0008283); positive regulation of transcription by RNA polymerase II (GO:0045944); leaf development (GO:0048366);	transcription coactivator activity (GO:0003713);	nucleus (GO:0005634);
SstPt_ctg3089	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg3754	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg4363	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg4487a	PSY00005414	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	1.15e-50	zf-FLZ (pfam04570)	9.44e-22	NA	NA	NA
SsrPt_ctg4487b	PSY00005414	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	1.15e-50	zf-FLZ (pfam04570)	9.44e-22	NA	NA	NA
SsrPt_ctg4698	PSY00006544	PREDICTED: protein RCC2-like	PREDICTED: protein RCC2-like [ <i>Nelumbo nucifera</i> ], XP_010271434.1	1.86e-10	NA	-	NA	NA	NA
SstPt_ctg5167	PSY00010492	probable transcription factor PosF21	unknown [ <i>Picea sitchensis</i> ], ABR17748.1	0	bZIP_plant_RF 2 (cd14703)	9.40e-24	regulation of transcription, DNA-templated (GO:0006355);	DNA-binding transcription factor activity (GO:0003700);	NA
SsrPt_ctg5333	PSY00010492	probable transcription factor PosF21	unknown [ <i>Picea sitchensis</i> ], ABR17748.1	0	bZIP_plant_RF 2 (cd14703)	9.40e-24	regulation of transcription, DNA-templated (GO:0006355);	DNA-binding transcription factor activity (GO:0003700);	NA

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg64	PSY00014290	cold shock protein 2-like	putative glycine-rich protein [ <i>Picea glauca</i> ], ABA54143.1	7.56e-67	CSP_CDS (cd04458)	5.69e-26	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
SsrPt_ctg7024	PSY00003770	unknown	unknown [ <i>Picea sitchensis</i> ], ABK24150.1, ABK26366.1	4.42E-60	NA	-	NA	NA	NA
SsrPt_ctg7081	PSY00020303	tubulin beta chain	tubulin beta chain-like [ <i>Nymphaea colorata</i> ], XP_031494355.1	0	NA	-	microtubule cytoskeleton organization (GO:0000226); mitotic cell cycle (GO:0000278);	GTPase activity (GO:0003924); structural constituent of cytoskeleton (GO:0005200); GTP binding (GO:0005525);	cytoplasm (GO:0005737); microtubule (GO:0005874);
SsrPt_ctg7141	PSY00005414	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	1.15e-50	zf-FLZ (pfam04570)	9.44e-22	NA	NA	NA
SsrPt_ctg7170	PSY00002648	polyadenylate-binding protein RBP47-like isoform X1	unknown [ <i>Picea sitchensis</i> ], ACN39849.1	0	RRM2_SECp43_like (cd12345) RRM1_SECp43_like (cd12344)	1.70e-45 9.53e-43	NA	mRNA binding (GO:0003729);	cytosol (GO:0005829);
SsrPt_ctg7425	PSY00000824	UBP1-associated protein 2C	unknown [ <i>Picea sitchensis</i> ], ABR18205.1	1.14e-128	RRM_SF super family (cl117169)	6.63e-22	NA	RNA binding (GO:0003723);	nucleus (GO:0005634); membrane (GO:0016020);
SsrPt_ctg7444	PSY00000153	unknown	unknown [ <i>Picea sitchensis</i> ], ABK21134.1, ABR16779.1	1.22e-45	NA	-	NA	NA	NA
SsrPt_ctg7731	PSY00017742	glucan endo-1,3-beta-glucosidase	Glucan endo-1,3-beta-glucosidase 7 [ <i>Pinus taeda</i> ], AZA14776.1	0	Glyco_hydro Superfamily (cl 23725) X8 (smart00768)	1.11e-89 1.32e-42	carbohydrate metabolic process (GO:0005975);	glucan endo-1,3-beta-D-glucosidase activity (GO:0042973);	anchored component of plasma membrane (GO:0046658);

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg7824	PSY00014046	Fiber protein fb34	unknown [ <i>Picea sitchensis</i> ], ABK21553.1, ABK22633.1, ACN40982.1	9.82e-93	DUF1218 (pfam06749)	4.63e-25	NA	NA	integral component of membrane (GO:0016021);
SsrPt_ctg7867	PSY00001445	cold shock protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	4.80e-51	CSD (pfam00313) zfCCHC (pfam00098)	1.09e-25 3.77e-05	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
SsrPt_ctg8064	PSY00002483	RNA-binding protein	glycine-rich RNA-binding protein [ <i>Picea glauca</i> ], AAD28176.1	5.26e-55	RRM2_NsCP33_like (cd21608)	3.21e-37	NA	mRNA binding (GO:0003729);	nucleus (GO:0005634);
SsrPt_ctg865	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg9249	PSY00009266	class III HD-Zip protein	class III HD-Zip protein HDZ32 [ <i>Pinus taeda</i> ], ABG73246.1	1.93e-79	Homeobox (pfam00046)	3.69e-14	NA	DNA binding (GO:0003677); lipid binding (GO:0008289);	nucleus (GO:0005634);
RPtest1	PSY00008891	putative histone-lysine N-methyltransferase SUV2-like	uncharacterized protein LOC105638411 [ <i>Jatropha curcas</i> ], XP_012077593.1, XP_012077594.1, XP_020536692.1	6.08e-26	DUF3527 Superfamily (cl13476)	3.94e-54	translation (GO:0006412);	structural constituent of ribosome (GO:0003735);	ribosome (GO:0005840);
RPtest5	PSY00003137	B2 protein	unknown [ <i>Picea sitchensis</i> ], ABK22365.1	2.16e-167	DCD (smart00767)	2.10e-75	NA	NA	NA
RPtest6	PSY00019692	AT-hook motif nuclear-localized protein 10-like	unknown [ <i>Picea sitchensis</i> ], ADE76363.1	1.66e-67	DUF296 (pfam03479)	2.91e-27	NA	DNA binding (GO:0003677); minor groove of adenine-thymine-rich DNA binding (GO:0003680);	nucleus (GO:0005634)
RPtest9	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
RPtest11	PSY00011770	N-terminal acetyltransferase A complex catalytic subunit NAA10	unknown [ <i>Picea sitchensis</i> ], ABK23715.1	9.17e-117	RimI (COG0456)	4.02e-30	response to water deprivation (GO:0009414); embryo development ending in seed dormancy (GO:0009793); N-terminal peptidyl-serine acetylation (GO:0017198); N-terminal peptidyl-glutamic acid acetylation (GO:0018002);	mRNA binding (GO:0003729); peptidyl-serine-N-acetyltransferase activity (GO:1990189); peptide-glutamate-N-acetyltransferase activity (GO:1990190);	cytosol (GO:0005829); NatA complex (GO:0031415);
RP11est13	PSY00025105	E3 ubiquitin-protein ligase	unknown [ <i>Picea sitchensis</i> ], ADE77028.1	1.40e-134	RING-H2_PATM-RING (cd16454)	2.40e-18	NA	NA	NA
RPtest15	PSY00002483	RNA-binding protein	glycine-rich RNA-binding protein [ <i>Picea glauca</i> ], AAD28176.1	5.33e-55	RRM2_NsCP33_like (cd21608)	3.21e-37	NA	mRNA binding (GO:0003729);	nucleus (GO:0005634);
RPtest16	PSY00014450	NA	-	-	NA	-	NA	NA	NA
LOP1	PSY00018563	CBL-interacting protein kinase 5	unknown [ <i>Picea sitchensis</i> ], ABK24914.1	0	STKc_SnRK3 (cd14663); CIPK_C (cd12195);	2.26e-176 7.17e-55	protein phosphorylation (GO:0006468); signal transduction (GO:0007165); response to abscisic acid (GO:0009737);	ATP binding (GO:0005524); protein serine kinase activity (GO:0106310); protein threonine kinase activity (GO:0106311);	integral component of membrane (GO:0016021);
LOP3	PSY00000153	unknown	unknown [ <i>Picea sitchensis</i> ], ABK21134.1, ABR16779.1	1.19e-45	NA	-	NA	NA	NA

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
LOP5	PSY00023395	chaperone protein dnaJ C76,chloroplastic	unknown [ <i>Picea sitchensis</i> ] ABK23437.1	0	DnaJ (pfam00226) Fer4_13 (pfam13370)	4.25e-20 2.14e-17	NA	protein binding (GO:0005515); metal ion binding (GO:0046872); iron-sulfur cluster binding (GO:0051536);	chloroplast (GO:0009507);
LOP6	PSY00008891	putative histone-lysine N-methyltransferase SUV2-like	uncharacterized protein LOC105638411 [ <i>Jatropha curcas</i> ] XP_012077593.1 XP_012077594.1 XP_020536692.1	6.05e-26	DUF3527 Superfamily (cl13476)	3.94e-54	translation (GO:0006412);	structural constituent of ribosome (GO:0003735);	ribosome (GO:0005840);
LOP8	PSY00004341	NA	-	-	NA	-	NA	NA	NA
LOP9	NP	NA	-	-	NA	-	NA	NA	NA
LOP11	PSY00023490	Lactoylglutathione lyase	unknown [ <i>Picea sitchensis</i> ] ABK22263.1	0	PLN02300 (PLN02300)	0,00	methylglyoxal catabolic process to D-lactate via S-lactoylglutathione (GO:0019243);	lactoylglutathione lyase activity (GO:0004462); metal ion binding (GO:0046872); dioxygenase activity (GO:0051213)	cytoplasm (GO:0005737);
LOP12	NP	NA	-	-	NA	-	NA	NA	NA
PtSIFG_1102	PSY00008719	unknown	unknown [ <i>Picea sitchensis</i> ] ADE76399.1	9.42e-8	NA	-	NA	NA	integral component of membrane (GO:0016021);
PtSIFG_0737	NP	NA	--	-	NA	-	NA	NA	NA
PpSIFG_3145	NP	NA	-	-	NA	-	NA	NA	NA
PtSIFG_5015	NP	NA	-	-	NA	-	NA	NA	NA
PtSIFG_6044	PSY00004353	pleiotropic drug resistance protein 1-like	PREDICTED: pleiotropic drug resistance protein 1-like [ <i>Nelumbo nucifera</i> ] XP_010273226.1	0	PLN03140 super family	0,00	transmembrane transport (GO:0055085);	ATP binding (GO:0005524); ATPase-coupled transmembrane transporter activity (GO:0042626);	chloroplast (GO:0009507); integral component of membrane (GO:0016021);
PtSIFG_6065	PSY00027604	(R)-mandelonitrile lyase-like	unknown [ <i>Picea sitchensis</i> ] ABK22756.1	1.78e-137	PLN02785 super family (cl33566)	1.21e-64	NA	oxidoreductase activity, acting on CH-OH group of donors (GO:0016614); flavin adenine	NA

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
								dinucleotide binding (GO:0050660);	
psyl2	PSY00005261	oligouridylylate-binding protein 1-like	unknown [ <i>Picea sitchensis</i> ] ABR17996.1	2.72e-122	RRM3_PUB1 (cd12622)	1.63e-42	NA	mRNA binding (GO:0003729);	NA
psyl16	PSY00018010	patellin-3-like protein	unknown [ <i>Picea sitchensis</i> ] ABR18053.1 ABR18173.1	5.24e-158	SEC14 (cd00170)	4.95e-34	NA	lipid binding (GO:0008289);	NA
psyl17	PSY00018010	patellin-3-like protein	unknown [ <i>Picea sitchensis</i> ] ABR18053.1 ABR18173.1	5.24e-158	SEC14 (cd00170)	4.95e-34	NA	lipid binding (GO:0008289);	NA
psyl18	PSY00016934	polyadenylate-binding protein 3	polyadenylate-binding protein 2, partial [ <i>Amborella trichopoda</i> ] XP_006842170.2	0	PABP-1234 (TIGR01628)	0,00	nuclear-transcribed mRNA poly(A) tail shortening (GO:0000289); translational initiation (GO:0006413);	mRNA 3'-UTR binding (GO:0003730); poly(A) binding (GO:0008143); poly(U) RNA binding (GO:0008266);	nucleus (GO:0005634); cytosol (GO:0005829); ribonucleoprotein complex (GO:1990904);
psyl19	PSY00016934	polyadenylate-binding protein 3	polyadenylate-binding protein 2, partial [ <i>Amborella trichopoda</i> ] XP_006842170.2	0	PABP-1234 (TIGR01628)	0,00	nuclear-transcribed mRNA poly(A) tail shortening (GO:0000289); translational initiation (GO:0006413);	mRNA 3'-UTR binding (GO:0003730); poly(A) binding (GO:0008143); poly(U) RNA binding (GO:0008266);	nucleus (GO:0005634); cytosol (GO:0005829); ribonucleoprotein complex (GO:1990904);



Locus	<i>Pinus sylvestris</i> gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
psyl25	PSY00000191	ribulose biphosphate carboxylase small subunit	RecName: Full=Ribulose biphosphate carboxylase small chain, chloroplastic; Short=RuBisCO small subunit; Flags: Precursor [ <i>Pinus thunbergii</i> ] P10053.1	1.92e-112	RbcS super family (cl29940)	4.18e-92	photorespiration (GO:0009853); reductive pentose-phosphate cycle (GO:0019253);	monooxygenase activity (GO:0004497); ribulose-biphosphate carboxylase activity (GO:0016984);	chloroplast (GO:0009507);
psyl36	PSY00003770	unknown	unknown [ <i>Picea sitchensis</i> ] ABK24150.1 ABK26366.1	4.48e-60	NA	-	regulation of gene expression (GO:0010468); phosphorylation (GO:0016310); intracellular signal transduction (GO:0035556);	MAP kinase activity (GO:0004707);	nucleus (GO:0005634); cytoplasm (GO:0005737);
psyl42	NP	NA	-	-	NA	-	NA	NA	NA
psyl44	PSY00013146	thiazole biosynthetic enzyme	unknown [ <i>Picea sitchensis</i> ] ABK23376.1	0	PLN02661 (PLN02661)	0,00	thiamine biosynthetic process (GO:0009228); thiazole biosynthetic process (GO:0052837);	iron ion binding (GO:0005506); transferase activity, transferring pentosyl groups (GO:0016763);	cytosol (GO:0005829); chloroplast stroma (GO:0009570);
psyl57	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig00080	PSY00026757	cold shock protein 2-like	putative glycine-rich protein [ <i>Picea glauca</i> ] ABA54143.1	3.24e-70	CSD (pfam00313)	2.79e-26	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270)	NA
lw_isotig00081	PSY00026757	cold shock protein 2-like	putative glycine-rich protein [ <i>Picea glauca</i> ] ABA54143.1	3.24e-70	CSD (pfam00313)	2.79e-26	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270)	NA

Locus	<i>Pinus sylvestris</i> gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
lw_isotig00542	PSY00008008	glycine dehydrogenase (decarboxylating), mitochondrial	glycine dehydrogenase (decarboxylating), mitochondrial [ <i>Elaeis guineensis</i> ] XP_010942763.1	0	PLN02414 (PLN02414)	0,00	glycine decarboxylation via glycine cleavage system (GO:0019464);	glycine dehydrogenase (decarboxylating) activity (GO:0004375); glycine binding (GO:0016594); pyridoxalphosphate binding (GO:0030170);	mitochondrion (GO:0005739); glycine cleavage complex (GO:0005960); chloroplast envelope (GO:0009941); apoplast (GO:0048046);
lw_isotig01420	PSY00020425	pimaradiene oxidase 1	CYP782B7 [ <i>Taxus chinensis</i> ] ATG29936.1	4.48e-43	p450 super family (cl12078)	4.66e-25	paclitaxel biosynthetic process (GO:0042617);	monooxygenase activity (GO:0004497); iron ion binding (GO:0005506); oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (GO:0016705); heme binding (GO:0020037);	NA
lw_isotig02138	PSY00016229	E3 ubiquitin-protein ligase At1g12760-like	E3 ubiquitin-protein ligase At1g63170 isoform X1 [ <i>Selaginella moellendorffii</i> ] XP_024531855.1	8.40e-143	RING_Ubox super family (cl17238)	1.06e-16	NA	NA	integral component of membrane (GO:0016021);
lw_isotig02347	PSY00016640	transcriptional corepressor SEUSS-like isoform X1	hypothetical protein [ <i>Adiantum capillus-veneris</i> ] MBC9828157.1	1.32e-162	LIM_bind (pfam01803)	1.25e-56	negative regulation of transcription by RNA polymerase II (GO:0000122); positive regulation of transcription by RNA polymerase II (GO:0045944);	RNA polymerase II activating transcription factor binding (GO:0001102);	nucleus (GO:0005634); transcription regulator complex (GO:0005667); membrane (GO:0016020);

Locus	<i>Pinus sylvestris</i> gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
lw_isotig02842	PSY00012746	WD-40 repeat-containing protein MSII	unknown [ <i>Picea sitchensis</i> ] ADE76527.1	0	WD40 super family (cl29593) CAF1C_H4-bd (pfam12265)	1.35e-37 1.65e-31	regulation of gene expression by genetic imprinting (GO:0006349); pollen development (GO:0009555); embryo development ending in seed dormancy (GO:0009793); regulation of flower development (GO:0009909); trichome differentiation (GO:0010026); seed coat development (GO:0010214); heterochromatin assembly (GO:0031507); positive regulation of cell cycle (GO:0045787); leaf development (GO:0048366); regulation of genetic imprinting (GO:2000653);	protein binding (GO:0005515); metal ion binding (GO:0046872); iron-sulfur cluster binding (GO:0051536);	chromatin silencing complex (GO:0005677);
lw_isotig03088	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig04195	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig04204	PSY00005987	probable endopeptidase p60	unknown [ <i>Picea sitchensis</i> ] ABK21059.1 ABR17490.1	1.84e-68	LysM (pfam01476)	1.37e-13	NA	NA	integral component of membrane (GO:0016021);
lw_isotig04306	PSY00011561	protein FATTY ACID EXPORT 2, chloroplastic	unknown [ <i>Picea sitchensis</i> ] ABR17562.1	1.83e-80	Tmemb_14 (pfam03647)	1.01e-21	NA	NA	integral component of membrane (GO:0016021);

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
lw_isotig04600	PSY00011465	trihelix transcription factor ASIL2	hypothetical protein AMTR_s00019p00163540 [ <i>Amborella trichopoda</i> ] ERN07191.1	5.78e-105	Myb_DNA-bind_4 (pfam13837)	7.88e-16	NA	transcription regulatory region sequence-specific DNA binding (GO:0000976);	nucleus (GO:0005634);
lw_isotig04931	PSY00018521	dihydropyrimidinase	unknown [ <i>Picea sitchensis</i> ] ABK24502.1	0	PLN02942 (PLN02942)	0.00	uracil catabolic process (GO:0006212); cellular response to nitrogen levels (GO:0043562);	dihydropyrimidinase activity (GO:0004157);	endoplasmic reticulum (GO:0005783); Golgi apparatus (GO:0005794);
lw_isotig05123	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig06215	PSY00022080	Transcription factor MYB44	R2R3-MYB transcription factor MYB6 [ <i>Picea glauca</i> ] ABQ51222.1	0	PLN03091 super family (cl33633) PRK13850 super family (cl29733)	5.14e-26 2.15e-03	regulation of transcription, DNA-templated (GO:0006355); regulation of transcription by RNA polymerase II (GO:0006357)	RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978); DNA-binding transcription factor activity, RNA polymerase II-specific (GO:0000981);	nucleus (GO:0005634);
lw_isotig06440	PSY00018317	zinc finger protein GIS3-like	hypothetical protein PHAVU_007G029500g [ <i>Phaseolus vulgaris</i> ] XP_007142933.1ESW14927.1	1.35e-10	NA	-	cytokinin-activated signaling pathway (GO:0009736); gibberellic acid mediated signaling pathway (GO:0009740); trichome differentiation (GO:0010026);	transcription regulatory region sequence-specific DNA binding (GO:0000976); DNA-binding transcription factor activity (GO:0003700);	nucleus (GO:0005634);
lw_isotig07383	PSY00006544	NA	-	-	NA	-	NA	NA	NA

Locus	<i>Pinus sylvestris</i> gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
lw_isotig10603	PSY00011044	proline-rich receptor-like protein kinase PERK1	unknown [ <i>Picea sitchensis</i> ] ACN40429.1	0	STKc_IRAK (cd14066)	2.09e-105	protein phosphorylation (GO:0006468);	ATP binding (GO:0005524); protein serine kinase activity (GO:0106310); protein threonine kinase activity (GO:0106311);	integral component of membrane (GO:0016021);
lw_isotig11166	PSY00030466	basic blue protein-like	APRN [ <i>Pinus tabulaeformis</i> ] AJP06237.1	4.99e-72	Plantacyanin (cd11013)	4.37e-46	electron transport chain (GO:0022900);	electron transfer activity (GO:0009055);	anchored component of plasma membrane (GO:0046658);
lw_isotig12667	PSY00012407	adenylyl-sulfate kinase 3 isoform X1	unknown [ <i>Picea sitchensis</i> ] ABR16196.1	0	CysC (COG0529)	1.17e-104	sulfate assimilation (GO:0000103); phosphorylation (GO:0016310); hydrogen sulfide biosynthetic process (GO:0070814);	adenylylsulfate kinase activity (GO:0004020); ATP binding (GO:0005524);	cytosol (GO:0005829); chloroplast (GO:0009507);
lw_isotig17679	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig20215	PSY00031065	indole-3-acetaldehyde oxidase	unnamed protein product [ <i>Triticum turgidum subsp. durum</i> ]VAH30502.1	1.98e-79	PLN00192 super family (cl33422)	1.56e-106	electron transport chain (GO:0022900);	iron ion binding (GO:0005506); electron transfer activity (GO:0009055); 2 iron, 2 sulfur cluster binding (GO:0051537); FAD binding (GO:0071949);	NA
lw_isotig21953	PSY00005415	protein trichome birefringence-like 6	unknown [ <i>Picea sitchensis</i> ] ADE76095.1	8.96e-87	GNH_hydrolase super family (cl01053)	2.10e-70	NA	NA	cellular anatomical entity (GO:0110165);
lw_isotig26230	PSY00034638	protein RSI-1	gasa5 like protein [ <i>Pinus pinaster</i> ] CBL95259.1	2.33e-73	GASA (pfam02704)	9.46e-35	NA	NA	cellular anatomical entity (GO:0110165);

Locus	Pinus sylvestris gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCB)	Biological Process	Molecular Function	Cellular Component
lw_isotig27940	PSY00003211	auxin-responsive protein IAA9 isoform X3	auxin-induced protein 2 [ <i>Pinus massoniana</i> ] ALI16894.1	0	AUX_IAA (pfam02309)	3.35e-101	regulation of transcription, DNA-templated (GO:0006355); auxin-activated signaling pathway (GO:0009734);	protein binding (GO:0005515);	nucleus (GO:0005634);

**Table S5.** Results of Blast2GO analysis and analysis of conserved domains performed for the *Pinus taeda* genes associated with EST-SSRs

NA - not available, NP - not predicted

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
PtTX2146	PITA_000026471	aspartyl protease APCB1	aspartyl protease APCB1 [ <i>Nymphaea colorata</i> ], XP_031492729.1	0	Pepsin_retropepsin_like Superfamily (c111403)	2.31e-141	proteolysis (GO:0006508);	hydrolase activity (GO:0016787); aspartic-type endopeptidase activity (GO:0004190);	NA
SsPp_cn524	PITA_000001867	probable galacturonosyltransferase 13 isoform X1	probable galacturonosyltransferase 14 isoform X1 [ <i>Durio zibethinus</i> ], XP_022774546.1	0	PLN02870 Superfamily (c130824)	0	pectin biosynthetic process (GO:0045489); cell wall organization (GO:0071555);	polygalacturonate 4-alpha-galacturonosyltransferase activity (GO:0047262);	Golgi membrane (GO:0000139); integral component of membrane (GO:0016021);
SsrPt_AW010960	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_AW225917	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_AW981772	PITA_000031044	NA	-	-	PRK12678 Superfamily (c136163)	4.41e-05	NA	NA	NA
SsrPt_BF049767	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg1376	PITA_000000887	3-isopropylmalate dehydratase large subunit, chloroplastic	unknown [ <i>Picea sitchensis</i> ], ACN40409.1	4.04e-129	Aconitase super family (c100285)	1.07e-76	leucine biosynthetic process (GO:0009098);	3-isopropylmalate dehydratase activity (GO:0003861); metal ion binding (GO:0046872); 4 iron, 4 sulfur cluster binding (GO:0051539);	NA
SsrPt_ctg1525	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg17601	PITA_000037826	uncharacterized protein LOC109718409	uncharacterized protein LOC109718409 [ <i>Ananas comosus</i> ], XP_020100234.1	4.01e-29	CBM20 Superfamily (cl15347)	3.83e-23	NA	carbohydrate binding (GO:0030246); starch binding (GO:2001070);	NA
SsrPt_ctg18103	PITA_000025493	olee1-like protein	extensin-like protein [ <i>Pinus taeda</i> ], QDZ06028.1	1.87e-114	Pollen_Ole_e_I Superfamily (cl03128)	3.30e-14	NA	NA	extracellular space (GO:0005615);
	PITA_000086735	olee1-like protein		5.96e-72	Pollen_Ole_e_I Superfamily (cl03128)	5.33e-09	NA	NA	extracellular space (GO:0005615);
	PITA_000047867	olee1-like protein		3.77e-106	Pollen_Ole_e_I Superfamily (cl03128)	7.39e-15	NA	NA	extracellular space (GO:0005615);
	PITA_000047868	olee1-like protein		2.28e-106	Pollen_Ole_e_I Superfamily (cl03128)	1.66e-14	NA	NA	extracellular space (GO:0005615);
	PITA_000051460	olee1-like protein		5.51e-101	Pollen_Ole_e_I Superfamily (cl03128)	2.08e-12	NA	NA	extracellular space (GO:0005615);
	PITA_000072543	olee1-like protein		8.44e-108	Pollen_Ole_e_I Superfamily (cl03128)	1.18e-14	NA	NA	extracellular space (GO:0005615);
	PITA_000030491	olee1-like protein		7.52e-107	Pollen_Ole_e_I Superfamily (cl03128)	3.96e-14	NA	NA	extracellular space (GO:0005615);
SsrPt_ctg2300	PTA00008090	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	8.19e-63	CSD (pfam00313)	1.33e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
SsrPt_ctg275	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg3021	PITA_000031795	GRF1-interacting factor 1	unknown [ <i>Picea sitchensis</i> ], ABK22130.1	3.68e-110	SSXT (pfam05030)	5.13e-19	cell population proliferation (GO:0008283); positive regulation of transcription by RNA polymerase II (GO:0045944); leaf development (GO:0048366);	transcription coactivator activity (GO:0003713);	nucleus (GO:0005634);
SstPt_ctg3089	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg3754	PITA_000016257	DELTA protein GAI-like	DPL [ <i>Pinus tabulaeformis</i> ], AHW42471.1	0	GRAS (pfam03514), DELTA (pfam12041)	1.09e-169, 8.20e-29	regulation of transcription, DNA-templated (GO:0006355); response to ethylene (GO:0009723); response to abscisic acid (GO:0009737); salicylic acid mediated signaling pathway (GO:0009863); jasmonic acid mediated signaling pathway (GO:0009867); negative regulation of gibberellic acid mediated signaling pathway (GO:0009938); negative regulation of seed germination (GO:0010187); hyperosmotic salinity response (GO:0042538); regulation of seed dormancy process (GO:2000033); regulation of reactive oxygen species metabolic process (GO:2000377);	DNA-binding transcription factor activity (GO:0003700); transcription coregulator activity (GO:0003712); sequence-specific DNA binding (GO:0043565);	nucleus (GO:0005634);



Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg4363	PITA_000068716	Protein FAM70A	uncharacterized protein LOC109944079 [ <i>Zea mays</i> ], XP_020404174.1	1.88e-14	NA	-	NA	NA	NA
SsrPt_ctg4487a	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	4.95e-48	zf-FLZ (pfam04570)	2.71e-20	NA	NA	NA
SsrPt_ctg4487b	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	4.95e-48	zf-FLZ (pfam04570)	2.71e-20	NA	NA	NA
SsrPt_ctg4698	PITA_000017622	protein RCC2homolog isoform X1	hypothetical protein ZOSMA_1G02490 [ <i>Zostera marina</i> ], KMZ70272.1	3.65e-10	NA	-	NA	DNA binding (GO:0003677);	NA
SstPt_ctg5167	NP	NA	-	-	NA	-	NA	NA	NA
SsrPt_ctg5333	PTA00009525	probable transcription factor PosF21	unknown [ <i>Picea sitchensis</i> ], ABR17748.1	0	bZIP_plant_RF2 (cd14703)	2.24e-24	regulation of transcription, DNA-templated (GO:0006355);	DNA-binding transcription factor activity (GO:0003700);	NA
SsrPt_ctg64	PTA00008090	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	8.19e-63	CSD (pfam00313)	1.33e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
SsrPt_ctg7024	PITA_000023332	mitogen-activated protein kinase 15 isoform X1	unknown [ <i>Picea sitchensis</i> ], ABK24488.1	9.03e-163	PKc_like Superfamily (cl21453)	2.04e-98	regulation of gene expression (GO:0010468); phosphorylation (GO:0016310); intracellular signal transduction (GO:0035556);	MAP kinase activity (GO:0004707);	nucleus (GO:0005634); cytoplasm (GO:0005737);
SsrPt_ctg7081	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg7141	PITA_000029961	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase-like isoform 1	unknown [ <i>Picea sitchensis</i> ], ADE77804.1	4.95e-48	zf-FLZ (pfam04570)	2.71e-20	NA	NA	NA
SsrPt_ctg7170	PITA_000027567	polyadenylate-binding protein RBP47-like	Lipase 1 [ <i>Pinus tabulaeformis</i> ], AJP06316.1	0	RRM2_SECp43_like (cd12345), RRM1_SECp43_like (cd12344)	3.48e-44, 3.57e-41	lipid metabolic process (GO:0006629);	mRNA binding (GO:0003729);	nucleus (GO:0005634); cytosol (GO:0005829);
SsrPt_ctg7425	PITA_000004120	UBP1-associated protein 2C	unknown [ <i>Picea sitchensis</i> ], ABR18205.1	2.30e-130	RRM2_NsCP33_like (cd21608)	4.43e-23	NA	RNA binding (GO:0003723);	nucleus (GO:0005634); membrane (GO:0016020);
SsrPt_ctg7444	PITA_000012096	Unknown	unknown [ <i>Picea sitchensis</i> ], ABK21134.1, ABR16779.1	1.74e-51	NA	-	NA	NA	NA
SsrPt_ctg7731	PITA_000091704	glucan endo-1,3-beta-glucosidase	Glucan endo-1,3-beta-glucosidase 7 [ <i>Pinus taeda</i> ], AZA14776.1	3.39e-65	Glyco_hydroSuperfamily (cl23725)	1.07e-18	carbohydrate metabolic process (GO:0005975);	glucan endo-1,3-beta-D-glucosidase activity (GO:0042973);	anchored component of plasma membrane (GO:0046658);
SsrPt_ctg7824	PITA_000076250	Fiber protein fb34	unknown [ <i>Picea sitchensis</i> ], ABK21553.1, ABK22633.1, ACN40982.1	1.81e-91	DUF1218 Superfamily (cl06009)	2.03e-25	NA	NA	integral component of membrane (GO:0016021);
SsrPt_ctg7867	PITA_000007834	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	4.37e-73	CSD (pfam00313)	2.62e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	mitochondrion (GO:0005739);
SsrPt_ctg8064	PITA_000008378	RNA-binding protein	glycine-rich RNA-binding protein [ <i>Picea glauca</i> ], AAD28176.1	5.59e-55	RRM2_NsCP33_like (cd21608)	1.14e-36	NA	mRNA binding (GO:0003729);	nucleus (GO:0005634);
SsrPt_ctg865	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
SsrPt_ctg9249	PITA_000010076	class III HD-Zip protein	Homeobox-leucine zipper proteinATHB-15 isoform D [ <i>Glycine soja</i> ], RZC00893.1	0	START_ArGLABRA2_1 like (cd08875), MEKHLA (pfam08670), Homeobox (pfam00046), bZIP (cd14686)	4.12e-79, 1.06e-64, 2.35e-16, 4.46e-06	NA	DNA binding (GO:0003677); lipid binding (GO:0008289);	nucleus (GO:0005634);
RPtest1	PITA_000070971	hypothetical protein F0562_000559	hypothetical protein F0562_000559 [ <i>Nyssa sinensis</i> ], KAA8548875.1	5.91e-40	DUF3527 Superfamily (c113476)	4.04e-72	NA	NA	NA
	PITA_000006918	PREDICTED: uncharacterized proteinLOC104600962	PREDICTED: uncharacterized protein LOC104600962 [ <i>Nelumbo nucifera</i> ], XP_010262438.1	4.37e-36	DUF3527 Superfamily (c113476)	1.35e-66	translation (GO:0006412);	structural constituent of ribosome (GO:0003735);	ribosome (GO:0005840);
RPtest5	PITA_000007179	B2 protein	unknown[ <i>Picea sitchensis</i> ], ABK22365.1	8.10e-172	DCD (smart00767)	7.72e-76	NA	NA	NA
RPtest6	NP	NA	-	-	NA	-	NA	NA	NA
RPtest9	PITA_000026471	aspartyl protease APCB1	aspartyl protease APCB1 [ <i>Nymphaea colorata</i> ], XP_031492729.1	0	Pepsin_retropepsin_like Superfamily (c111403)	2.31e-141	proteolysis (GO:0006508);	hydrolase activity (GO:0016787); aspartic-type endopeptidase activity (GO:0004190);	NA
RPtest11	PITA_000052634	N-terminal acetyltransferase A complex catalytic subunit NAA10-like	unknown [ <i>Picea sitchensis</i> ], ABK23715.1	1.21e-58	RimI (COG0456)	6.09e-16	response to water deprivation (GO:0009414); embryo development ending in seed dormancy (GO:0009793); N-terminal peptidyl-serine acetylation (GO:0017198); N-terminal peptidyl-glutamic acid acetylation (GO:0018002);	mRNA binding (GO:0003729); peptide-serine-N-acetyltransferase activity (GO:1990189); peptide-glutamate-N-acetyltransferase activity (GO:1990190);	cytosol (GO:0005829); NatA complex (GO:0031415);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
	PITA_000085253	N-terminal acetyltransferase A complex catalytic subunit NAA10	unknown [ <i>Picea sitchensis</i> ], ABK23715.1	4.56e-51	RimI (COG0456)	2.19e-19	response to water deprivation (GO:0009414); embryo development ending in seed dormancy (GO:0009793); N-terminal peptidyl-serine acetylation (GO:0017198); N-terminal peptidyl-glutamic acid acetylation (GO:0018002);	mRNA binding (GO:0003729); peptide-serine-N-acetyltransferase activity (GO:1990189); peptide-glutamate-N-acetyltransferase activity (GO:1990190);	cytosol (GO:0005829); NatA complex (GO:0031415);
	PITA_000013530	N-terminal acetyltransferase A complex catalytic subunit NAA10	unknown [ <i>Picea sitchensis</i> ], ABK23715.1	1.97e-49	RimI (COG0456)	1.52e-17	response to water deprivation (GO:0009414); embryo development ending in seed dormancy (GO:0009793); N-terminal peptidyl-serine acetylation (GO:0017198); N-terminal peptidyl-glutamic acid acetylation (GO:0018002);	mRNA binding (GO:0003729); peptide-serine-N-acetyltransferase activity (GO:1990189); peptide-glutamate-N-acetyltransferase activity (GO:1990190);	cytosol (GO:0005829); NatA complex (GO:0031415);
RP11est13	PITA_000065673	E3 ubiquitin-protein ligase	unknown [ <i>Picea sitchensis</i> ], ADE77028.1	1.10e-126	RING-H2_PA-TM-RING (cd16454)	1.09e-17	ubiquitin-dependent protein catabolic process (GO:0006511); protein ubiquitination (GO:0016567);	ubiquitin protein ligase activity (GO:0061630);	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
RPtest15	PITA_000008378	RNA-binding protein	glycine-rich RNA-binding protein [ <i>Picea glauca</i> ], AAD28176.1	5.67e-55	RRM2_NsCP33_like (cd21608)	1.14e-36	NA	mRNA binding (GO:0003729);	nucleus (GO:0005634);
RPtest16	NP	NA	-	-	NA	-	NA	NA	NA
LOP1	PITA_000008107	CBL-interacting protein kinase 5	unknown [ <i>Picea sitchensis</i> ], ABK24914.1	0	STKc_SnRK3 (cd14663), CIPK_C (cd12195)	1.06e-177, 2.23e-52	protein phosphorylation (GO:0006468); signal transduction (GO:0007165); response to abscisic acid (GO:0009737);	ATP binding (GO:0005524); protein serine kinase activity (GO:0106310); protein threonine kinase activity (GO:0106311);	integral component of membrane (GO:0016021);
	PITA_000008108	protein kinase C	unknown [ <i>Picea sitchensis</i> ], ABK24914.1	0	STKc_SnRK3 (cd14663), NAF (pfam03822), CIPK_C (cd12195);	1.32e-176, 1.30e-14, 2.39e-17	protein phosphorylation (GO:0006468); signal transduction (GO:0007165);	ATP binding (GO:0005524); protein serine kinase activity (GO:0106310); protein threonine kinase activity (GO:0106311);	integral component of membrane (GO:0016021);
LOP3	PITA_000012096	unknown	unknown [ <i>Picea sitchensis</i> ], ABK21134.1, ABR16779.1	1.71e-51	NA	-	NA	NA	NA
LOP5	NP	NA	-	-	NA	-	NA	NA	NA
LOP6	PITA_000070971	hypothetical protein F0562_000559	hypothetical protein F0562_000559 [ <i>Nyssa sinensis</i> ], KAA8548875.1	5.88e-40	DUF3527 Superfamily (cl13476)	4.04e-72	NA	NA	NA
LOP8	PITA_000044661	unknown	unknown [ <i>Picea sitchensis</i> ], ABR17498.1	8.45e-5	NA	-	NA	NA	membrane (GO:0016020); integral component of membrane (GO:0016021);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
LOP9	PTA00066643	Chitinase-like protein 1	class II chitinase [ <i>Pinus banksiana</i> ], QEL09515.1	0	chitinase_GH19 (cd00325)	9.85e-80	carbohydrate metabolic process (GO:0005975); chitin catabolic process (GO:0006032); cell wall macromolecule catabolic process (GO:0016998);	chitinase activity (GO:0004568);	NA
LOP11	PITA000023770	lactoylglutathione lyase-like	unknown [ <i>Picea sitchensis</i> ], ABK22263.1	0	PLN02300 (PLN02300)	0	methylglyoxal catabolic process to D-lactate via S-lactoylglutathione (GO:0019243);	lactoylglutathione lyase activity (GO:0004462); metal ion binding (GO:0046872); dioxygenase activity (GO:0051213);	cytoplasm (GO:0005737);
LOP12	PITA_000035518	putative TIR-NBS-LRR protein	putative TIR-NBS-LRR protein [ <i>Pinus monticola</i> ], ADW94527.1	0	PLN03210 super family (cl33662)	7.74e-62	signal transduction (GO:0007165);	ADP binding (GO:0043531); NAD(P)+ nucleosidase activity (GO:0050135); NAD+ nucleotidase, cyclic ADP-ribose generating (GO:0061809);	NA
PtSIFG_1102	PTA00050445	unknown	unknown [ <i>Picea sitchensis</i> ], ABK24072.1, ABK26657.1	9.42e-8	NA	-	NA	NA	integral component of membrane (GO:0016021);
PtSIFG_0737	PITA_000017323	protein G1-like4	unknown [ <i>Picea sitchensis</i> ], ABR18275.1	6.18e-110	DUF640 Superfamily (cl04800)	2.30e-86	mRNA transcription (GO:0009299); response to light stimulus (GO:0009416);	NA	nucleus (GO:0005634);
PpSIFG_3145	PITA_000026471	aspartyl protease APCB1	aspartyl protease APCB1 [ <i>Nymphaea colorata</i> ], XP_031492729.1	0	Pepsin_retropepsin_like Superfamily (cl11403)	2.31e-141	proteolysis (GO:0006508);	hydrolase activity (GO:0016787); aspartic-type endopeptidase activity (GO:0004190);	NA
PtSIFG_5015	NP	NA	-	-	NA	-	NA	NA	NA

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
PtSIFG_6044	PITA_000008712	pleiotropic drug resistance protein 1-like	PREDICTED: pleiotropic drug resistance protein 1-like [ <i>Nelumbo nucifera</i> ], XP_010273226.1	0	PLN03140 Superfamily (cl33646)	0	transmembrane transport (GO:0055085);	ATP binding (GO:0005524); ATPase-coupled transmembrane transporter activity (GO:0042626);	chloroplast (GO:0009507); integral component of membrane (GO:0016021);
PtSIFG_6065	NP	NA	-	-	NA	-	NA	NA	NA
psyl2	PITA_000017503	oligouridylate-binding protein 1-like	unknown [ <i>Picea sitchensis</i> ], ABR17996.1	0	RRM2_PUB1 (cd12619), RRM3_TIA1_like (cd12354)	8.42e-53, 6.72e-41	NA	mRNA binding (GO:0003729);	NA
psyl16	PITA_000010550	patellin-3-like protein	unknown [ <i>Picea sitchensis</i> ], ABR18053.1, ABR18173.1	0	SEC14 (cd00170), CRAL_TRIO_N (smart01100)	1.87e-34, 1.23e-07	NA	lipid binding (GO:0008289);	NA
psyl17	PITA_000010550	patellin-3-like protein	unknown [ <i>Picea sitchensis</i> ], ABR18053.1, ABR18173.1	0	SEC14 (cd00170), CRAL_TRIO_N (smart01100)	1.87e-34, 1.23e-07	NA	lipid binding (GO:0008289);	NA
psyl18	PITA_000064129	polyadenylate-binding protein 3	hypothetical protein RHSIM_Rhsim06G0093100 [ <i>Rhododendron simsii</i> ], KAF7140837	3.62e-61	PABP-1234 Superfamily (cl31127)	4.56e-49	nuclear-transcribed mRNA poly(A) tail shortening (GO:0000289); translational initiation (GO:0006413);	mRNA 3'-UTR binding (GO:0003730); poly(A) binding (GO:0008143); poly(U) RNA binding (GO:0008266);	nucleus (GO:0005634); cytosol (GO:0005829); ribonucleoprotein complex (GO:1990904);
psyl19	PITA_000086838	polyadenylate-binding protein 2-like	hypothetical protein CISIN_1g006282mg [ <i>Citrus sinensis</i> ], KDO84108.1	5.81e-52	PABP-1234 super family (cl31127)	1.35e-41	nuclear-transcribed mRNA poly(A) tail shortening (GO:0000289); translational initiation (GO:0006413);	mRNA 3'-UTR binding (GO:0003730); poly(A) binding (GO:0008143); poly(U) RNA binding (GO:0008266);	nucleus (GO:0005634); cytosol (GO:0005829); ribonucleoprotein complex (GO:1990904);
	PITA_000041571	polyadenylate-binding protein 3-like	hypothetical protein AMTR_s00078p00149860 [ <i>Amborella trichopoda</i> ], ERN03845.1	9.54e-136	PABP-1234 super family (cl31127)	1.71e-73	nuclear-transcribed mRNA poly(A) tail shortening (GO:0000289); translational initiation (GO:0006413);	mRNA 3'-UTR binding (GO:0003730); poly(A) binding (GO:0008143); poly(U) RNA binding (GO:0008266);	nucleus (GO:0005634); cytosol (GO:0005829); ribonucleoprotein complex (GO:1990904);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
psyl25	PITA_000008677	ribulose biphosphate carboxylase small subunit	RecName: Full=Ribulose biphosphate carboxylase small chain, chloroplastic; Short=RuBisCO small subunit; Flags: Precursor [ <i>Pinus thunbergii</i> ], P10053.1	1.85e-107	RbcS Superfamily (cl29940)	3.81e-92	photorespiration (GO:0009853); reductive pentose-phosphate cycle (GO:0019253);	monooxygenase activity (GO:0004497); ribulose-biphosphate carboxylase activity (GO:0016984);	chloroplast (GO:0009507);
psyl36	PITA_000023332	mitogen-activated protein kinase 15 isoform X1	unknown [ <i>Picea sitchensis</i> ], ABK24488.1	9.03e-163	PKc_like super family (cl21453)	2.04e-98	regulation of gene expression (GO:0010468); phosphorylation (GO:0016310); intracellular signal transduction (GO:0035556);	MAP kinase activity (GO:0004707);	nucleus (GO:0005634); cytoplasm (GO:0005737);
psyl42	PITA_000007965	WD repeat-containing protein 48	WD repeat-containing protein 48 [ <i>Amborella trichopoda</i> ], XP_006854552.1, XP_011627121.1, XP_020529314.1, XP_020529315.1	1.68e-145	WD40 super family (cl29593)	1.08e-27	protein phosphorylation (GO:0006468); lateral root formation (GO:0010311); positive regulation of protein deubiquitination (GO:1903003);	myosin heavy chain kinase activity (GO:0016905); ubiquitin binding (GO:0043130);	NA
psyl44	PITA_000017051	thiazole biosynthetic enzyme	unknown [ <i>Picea sitchensis</i> ], ABK23376.1	0	PLN02661 Superfamily (cl30472)	0	thiamine biosynthetic process (GO:0009228); thiazole biosynthetic process (GO:0052837);	iron ion binding (GO:0005506); transferase activity, transferring pentosyl groups (GO:0016763);	cytosol (GO:0005829); chloroplast stroma (GO:0009570);



Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
psyl57	PITA_000010946	dicarboxylate transporter 1, chloroplastic	unknown [ <i>Picea sitchensis</i> ], ABK24751.1	0	Dass (TIGR00785)	2.90e-151	transmembrane transport (GO:0055085);	transmembrane transporter activity (GO:0022857);	chloroplast inner membrane (GO:0009706); integral component of membrane (GO:0016021);
lw_isotig00080	PITA_000007834	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	4.37e-73	CSD (pfam00313)	2.62e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	mitochondrion (GO:0005739);
	PITA_000009956	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	8.31e-63	CSD (pfam00313)	1.33e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
lw_isotig00081	PITA_000009956	glycine-rich protein 2	unknown [ <i>Picea sitchensis</i> ], ABK22299.1, ABR17619.1, ACN40273.1	8.31e-63	CSD (pfam00313)	1.33e-25	NA	nucleic acid binding (GO:0003676); zinc ion binding (GO:0008270);	NA
lw_isotig00542	PTA00082932	glycine dehydrogenase (decarboxylating), mitochondrial	PREDICTED: glycine dehydrogenase (decarboxylating), mitochondrial-like [ <i>Musa acuminata subsp. malaccensis</i> ], XP_009401914.1	0	PLN02414 (cl31876)	0	glycine decarboxylation via glycine cleavage system (GO:0019464);	glycine dehydrogenase (decarboxylating) activity (GO:0004375); glycine binding (GO:0016594); pyridoxal phosphate binding (GO:0030170);	mitochondrion (GO:0005739); glycine cleavage complex (GO:0005960); chloroplast envelope (GO:0009941); apoplast (GO:0048046);
lw_isotig01420	PITA_000023899	flavonoid 3'-monooxygenase	CYP782B7 [ <i>Taxus chinensis</i> ], ATG29936.1	9.14e-136	cytochrome_P450 super family (cl41757)	5.06e-94	paclitaxel biosynthetic process (GO:0042617);	monooxygenase activity (GO:0004497); iron ion binding (GO:0005506); oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (GO:0016705); heme binding (GO:0020037);	integral component of membrane (GO:0016021);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
lw_isotig02138	PITA_000020633	E3 ubiquitin-protein ligase At1g12760-like	E3 ubiquitin-protein ligase At1g63170 isoformX1 [ <i>Selaginella moellendorffii</i> ], XP_024531855.1	2.37e-144	RING_Ubox Superfamily (c117238)	1.24e-16	NA	NA	integral component of membrane (GO:0016021);
lw_isotig02347	NP	NA	-	-	NA	-	NA	NA	NA
lw_isotig02842	PITA_000016857	WD-40 repeat-containing protein MSII	unknown [ <i>Picea sitchensis</i> ], ADE76527.1	6.82e-49	CAF1C_H4-bd (pfam12265)	1.18e-31	NA	protein binding (GO:0005515);	nucleus (GO:0005634);
lw_isotig03088	PTA00015290	Poly A polymerase	uncharacterized protein LOC18434369 [ <i>Amborella trichopoda</i> ], XP_006844503.1	1.62e-104	PcnB (COG0617)	1.19e-60	RNA processing (GO:0006396);	RNA binding (GO:0003723); nucleotidyltransferase activity (GO:0016779);	NA
lw_isotig04195	PITA_000001889	Late embryogenesis abundant protein	unknown [ <i>Picea sitchensis</i> ], ABK21301.1	1.06e-78	LEA_2 (pfam03168)	3.39e-08	NA	NA	intrinsic component of membrane (GO:0031224)
	PITA_000034486	Late embryogenesis abundant protein	unknown [ <i>Picea sitchensis</i> ], ABK21301.1	4.15e-84	LEA_2 (pfam03168)	3.00e-07	NA	NA	integral component of membrane (GO:0016021);
lw_isotig04204	PITA_000008362	Peptidoglycan-binding lysin domain	unknown [ <i>Picea sitchensis</i> ], ABK21059.1, ABR17490.1	6.28e-68	LysM (pfam01476)	1.53e-13	NA	NA	integral component of membrane (GO:0016021);
lw_isotig04306	PTA00014456	protein FATTY ACID EXPORT 2, chloroplastic	unknown [ <i>Picea sitchensis</i> ], ABR17562.1	1.97e-61	Tmemb_14 (pfam03647)	1.34e-20	NA	NA	integral component of membrane (GO:0016021);
lw_isotig04600	PTA00083895	trihelix transcription factor ASIL2	hypothetical protein AMTR_s00019p00163540 [ <i>Amborella trichopoda</i> ], ERN07191.1	1.32e-101	Myb_DNA-bind_4 (pfam13837)	1.08e-15	NA	transcription regulatory region sequence-specific DNA binding (GO:0000976);	nucleus (GO:0005634);
lw_isotig04931	PITA_000025726	dihydropyrimidinase	unknown [ <i>Picea sitchensis</i> ], ABR16534.1	0	metallo-dependent_hydrolases super family (c100281)	0	uracil catabolic process (GO:0006212); cellular response to nitrogen levels (GO:0043562);	dihydropyrimidinase activity (GO:0004157);	endoplasmic reticulum (GO:0005783); Golgi apparatus (GO:0005794);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
lw_isotig05123	PTA00083918	DUF674 domain-containing protein	unknown [ <i>Picea sitchensis</i> ], ABK24184.1	5.32e-133	DUF674 super family (cl04913)	1.83e-40	NA	NA	NA
lw_isotig06215	PTA00008461	Transcription factor MYB44	R2R3-MYB transcription factor MYB6 [ <i>Picea glauca</i> ], ABQ51222.1	0	PLN03091 super family (cl33633)	2.66e-26	regulation of transcription, DNA-templated (GO:0006355); regulation of transcription by RNA polymerase II (GO:0006357);	RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978); DNA-binding transcription factor activity, RNA polymerase II-specific (GO:0000981);	nucleus (GO:0005634);
lw_isotig06440	PITA_000080433	zinc finger protein 6	zinc finger protein 6 [ <i>Amborella trichopoda</i> ], XP_020521633.1	2.31e-11	NA	-	regulation of cellular process (GO:0050794); trichome morphogenesis (GO:0010090);	NA	NA
lw_isotig07383	PITA_000017622	protein RCC2 homolog isoform X1	hypothetical protein ZOSMA_1G02490 [ <i>Zostera marina</i> ], KMZ70272.1	3.69e-10	NA	-	NA	DNA binding (GO:0003677);	NA
lw_isotig10603	PITA_000023436	proline-rich receptor-like protein kinase PERK1	unknown [ <i>Picea sitchensis</i> ], ACN40429.1	0	PKc_like Superfamily (cl21453)	2.29e-106	protein phosphorylation (GO:0006468);	protein serine/threonine kinase activity (GO:0004674); ATP binding (GO:0005524);	integral component of membrane (GO:0016021);
lw_isotig11166	PITA_000059576	basic blue protein-like	APRN [ <i>Pinus tabuliformis</i> ], AJP06237.1	5.16e-39	Plantacyanin (cd11013)	2.62e-40	reproductive process (GO:0022414); multicellular organismal process (GO:0032501);	electron transfer activity (GO:0009055);	cell periphery (GO:0071944);
lw_isotig12667	PTA00083219	adenylyl-sulfate kinase 3 isoform X1	unknown [ <i>Picea sitchensis</i> ], ABR16196.1	0	CysC (COG0529)	1.03e-103	sulfate assimilation (GO:0000103); phosphorylation (GO:0016310); hydrogen sulfide biosynthetic process (GO:0070814);	adenylylsulfate kinase activity (GO:0004020); ATP binding (GO:0005524);	cytosol (GO:0005829); chloroplast (GO:0009507);

Locus	Pinus taeda gene ID	Potential function (Blast2GO)	Blast top hit description [organism], accession №	E-value (Blast2GO)	Specific hits (CDD NCBI)	E-value (CDD NCBI)	Biological Process	Molecular Function	Cellular Component
lw_isotig17679	PITA_000014378	protein NETWORKED 1D	PREDICTED: protein NETWORKED 1D-like [ <i>Nelumbo nucifera</i> ], XP_010256141.1	2.33e-120	KIP1	1.64e-39	endoplasmic reticulum organization (GO:0007029); endoplasmic reticulum membrane fusion (GO:0016320);	actin binding (GO:0003779); GTPase activity (GO:0003924); actin filament binding (GO:0051015);	cytoplasm (GO:0005737); endoplasmic reticulum (GO:0005783); plasma membrane (GO:0005886);
lw_isotig20215	PTA00083968	indole-3-acetaldehyde oxidase-like	indole-3-acetaldehyde oxidase [ <i>Amborella trichopoda</i> ], XP_006841111.1	0	PLN00192 super family (cl33422)	0	NA	oxidoreductase activity (GO:0016491); metal ion binding (GO:0046872); flavin adenine dinucleotide binding (GO:0050660); iron-sulfur cluster binding (GO:0051536);	NA
lw_isotig21953	PTA00006979	protein trichome birefringence-like 6	unknown [ <i>Picea sitchensis</i> ], ADE76095.1	1.79e-152	PC-Esterase (pfam13839) PMR5N (pfam14416)	5.95e-112 1.44e-23	NA	NA	cellular anatomical entity (GO:0110165);
lw_isotig26230	PITA_000038492	protein RSI-1	gasa5 like protein [ <i>Pinus pinaster</i> ], CBL95259.1	5.18e-45	GASA (pfam02704)	3.56e-22	gibberellic acid mediated signaling pathway (GO:0009740);	NA	extracellular region (GO:0005576); integral component of membrane (GO:0016021);
lw_isotig27940	PITA_000016972	auxin-responsive protein IAA9 isoform X3	auxin-induced protein 2 [ <i>Pinus taeda</i> ], AAP44405.1	0	AUX_IAA (pfam02309)	5.75e-103	regulation of transcription, DNA-templated (GO:0006355); auxin-activated signaling pathway (GO:0009734);	protein binding (GO:0005515);	nucleus (GO:0005634);

## REFERENCES

1. Pravdin L.F. *Scots Pine: Variation, Intraspecific Taxonomy and Selection*. London: Ann Arbor Humphrey Science Publishers Ltd, 1969. 208 p.
2. Mason W., Alía R. Current and future status of Scots pine (*Pinus sylvestris* L.) forests in Europe. *Forest Systems*. 2000. V. 9. P. 317–335.
3. Krakau U.-K., Liesebach M., Aronen T., Lelu-Walter M.-A. Scots Pine (*Pinus sylvestris* L.). In: *Forest Tree Breeding in Europe*. Ed. Pâques L.E. Dordrecht: Springer Netherlands, 2013. P. 267–323.
4. Cañellas I., Martínez García F., Montero G. Silviculture and dynamics of *Pinus sylvestris* L. stands in Spain. *Forest Systems*. 2000. V. 9. P. 233–253
5. Nilsson U., Luoranen J., Kolström T., Örlander G., Puttonen P. Reforestation with planting in northern Europe. *Scandinavian Journal of Forest Research*. 2010. V. 25. № 4. P. 283–294. doi: [10.1080/02827581.2010.498384](https://doi.org/10.1080/02827581.2010.498384)
6. Sukhbaatar G., Ganbaatar B., Jamsran T., Purevragchaа B., Nachin B., Gradel A. Assessment of early survival and growth of planted Scots pine (*Pinus sylvestris*) seedlings under extreme continental climate conditions of northern Mongolia. *Journal of Forestry Research*. 2020. V. 31. № 1. P. 13–26. doi: [10.1007/s11676-019-00935-8](https://doi.org/10.1007/s11676-019-00935-8)
7. Boratyński A. Range of natural distribution. In: *Genetics of Scots Pine*. Amsterdam: Elsevier, 1991. P. 19–30
8. Eriksson G. Evolutionary forces influencing variation among populations of *Pinus sylvestris*. *Silva Fennica*. 1998. V. 32. P. 173–184. doi: [10.14214/sf.694](https://doi.org/10.14214/sf.694)
9. Matías L., Jump A.S. Interactions between growth, demography and biotic interactions in determining species range limits in a warming world: The case of *Pinus Sylvestris*. *Forest Ecology and Management*. 2012. V. 282. P. 10–22. doi: [10.1016/j.foreco.2012.06.053](https://doi.org/10.1016/j.foreco.2012.06.053)
10. Mullin T.J., Andersson B., Bastien J.-C., Beaulieu J., Burdon R.D., Dvorak W.S., King J.N., Kondo T., Krakowski J., Lee S.J., et al. Economic importance, breeding objectives and achievements. In: *Genetics, Genomics and Breeding of Conifers*. Boca Raton: CRC Press, 2011. P. 40–127.
11. Ruotsalainen S., Persson T. Scots pine – *Pinus sylvestris* L. In: *Best Practice for Tree Breeding in Europe*. Uppsala: Skogforsk, 2013. P. 49–64.
12. Sannikov S.N., Petrova I.V., Sannikova N.S., Afonin A.N., Chernodubov A.I., Egorov E.V. Genetic-climatologic-geographical principles of seed zoning of pine forests in Russia. *Siberian Forest Journal*. 2017. № 2. P. 19–30. doi: [10.15372/SJFS20170203](https://doi.org/10.15372/SJFS20170203)
13. Wójkiewicz B., Cavers S., Wachowiak W. Current Approaches and Perspectives in Population Genetics of Scots Pine (*Pinus sylvestris* L.). *Forest Science*. 2016. V. 62. № 3. P. 343–354. doi: [10.5849/forsci.15-040](https://doi.org/10.5849/forsci.15-040)
14. Plomion C., Chagné D., Pot D., Kumar S., Wilcox P.L., Burdon R.D., Prat D., Peterson D.G., Paiva J., Chaumeil P., et al. Pines. In: *Forest Trees*. Berlin: Springer, 2007. P. 29–92.
15. Parchman T.L., Geist K.S., Grahnen J.A., Benkman C.W., Buerkle C.A. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics*. 2010. V. 11. № 1. P. 1–16. doi: [10.1186/1471-2164-11-180](https://doi.org/10.1186/1471-2164-11-180)
16. Mackay J., Dean J.F.D., Plomion C., Peterson D.G., Cánovas F.M., Pavy N., Ingvarsson P.K., Savolainen O., Guevara M.Á., Fluch S., et al. Towards decoding the conifer gigagenome. *Plant Molecular Biology*. 2012. V. 80. № 6. P. 555–569. doi: [10.1007/s11103-012-9961-7](https://doi.org/10.1007/s11103-012-9961-7)
17. Pellicer J., Leitch I.J. The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist*. 2020. № 226. P. 301–305. doi: [10.1111/nph.16261](https://doi.org/10.1111/nph.16261)

18. Pyhäjärvi T., Kujala S.T., Savolainen O. 275 years of forestry meets genomics in *Pinus sylvestris*. *Evolutionary Applications*. 2020. V. 13. № 1. P. 11–30. doi: [10.1111/eva.12809](https://doi.org/10.1111/eva.12809)
19. Lu M.Z., Szmidt A.E., Wang X.R. Inheritance of RAPD fragments in haploid and diploid tissues of *Pinus sylvestris* (L.). *Heredity*. 1995. V. 74. № 6. P. 582–589. doi: [10.1038/hdy.1995.82](https://doi.org/10.1038/hdy.1995.82)
20. Hurme P., Sillanpää M.J., Arjas E., Repo T., Savolainen O. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics*. 2000. V. 156. № 3. P. 1309–1322. doi: [10.1093/genetics/156.3.1309](https://doi.org/10.1093/genetics/156.3.1309)
21. Lerceteau E., Plomion C., Andersson B. AFLP mapping and detection of quantitative trait loci (QTLs) for economically important traits in *Pinus sylvestris*: a preliminary study. *Molecular Breeding*. 2000. V. 6. № 5. P. 451–458. doi: [10.1023/A:1026548716320](https://doi.org/10.1023/A:1026548716320)
22. Lerceteau E., Szmidt A.E., Andersson B. Detection of quantitative trait loci in *Pinus sylvestris* L. across years. *Euphytica*. 2001. V. 121. № 2. P. 117–122. doi: [10.1023/A:1012076825293](https://doi.org/10.1023/A:1012076825293)
23. Yazdani R., Nilsson J.E., Plomion C., Mathur G. Marker trait association for autumn cold acclimation and growth rhythm in *Pinus Sylvestris*. *Scandinavian Journal of Forest Research*. 2003. V. 18. № 1. P. 29–38. doi: [10.1080/0891060310002318](https://doi.org/10.1080/0891060310002318)
24. Yin T.M., Wang X.R., Andersson B., Lerceteau-Köhler E. Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. *Theoretical and Applied Genetics*. 2003. V. 106. № 6. P. 1075–1083. doi: [10.1007/s00122-003-1194-3](https://doi.org/10.1007/s00122-003-1194-3)
25. Abraham B., Miklossy I., Kovacs E., Tamas E., Meszaros I., Szilveszter S., Brezeanu A., Lanyi S. Genetic analysis of *Pinus sylvestris* L. and *Pinus sylvestris forma turfosa* L. using RAPD markers. *Notulae Scientia Biologicae*. 2010. V. 2. № 1. P. 129–132 doi: [10.15835/nsb213611](https://doi.org/10.15835/nsb213611)
26. Nowicka A, Ukalska J, Simińska J, Szyp-Borowska I. Characterization and mapping of QTL used in breeding of Scots pine (*Pinus sylvestris* L.). *Folia Forestalia Polonica. Series A. Forestry*. 2013. V. 55. № 4. P.168–173. doi: [10.2478/ffp-2013-0018](https://doi.org/10.2478/ffp-2013-0018)
27. Androsiuk P., Ciagło-Androsiuk S., Urbaniak L. Genetic diversity and differentiation of *Pinus sylvestris* L. from the IUFRO 1982 provenance trial revealed by AFLP analysis. *Archives of Biological Sciences*. 2015. V. 67. № 4. P. 1237–1249.
28. Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., van de Wiel C., Bredemeijer G., Vosman B., Matthes M., Daly A. et al. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*. 1997. V. 3. № 5. P. 381–390. doi: [10.1023/A:1009612517139](https://doi.org/10.1023/A:1009612517139)
29. Meudt H.M., Clarke A.C. Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*. 2007. V. 12. № 3. P. 106–117. doi: [10.1016/j.tplants.2007.02.001](https://doi.org/10.1016/j.tplants.2007.02.001)
30. Agarwal M., Shrivastava N., Padh H. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Reports*. 2008. V. 27. № 4. P. 617–631 doi: [10.1007/s00299-008-0507-z](https://doi.org/10.1007/s00299-008-0507-z)
31. Hamblin M.T., Warburton M.L., Buckler E.S. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PloS One*. 2007. V. 2. № 12. P. e136 doi: [10.1371/journal.pone.0001367](https://doi.org/10.1371/journal.pone.0001367)
32. Kalia R. K., Rai M. K., Kalia S., Singh R., Dhawan A. K. Microsatellite markers: an overview of the recent progress in plants. *Euphytica*. 2011. V. 177. № 3. P. 309–334. doi: [10.1007/s10681-010-0286-9](https://doi.org/10.1007/s10681-010-0286-9)

33. Brumfield R.T., Beerli P., Nickerson D.A., Edwards S.V. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution*. 2003. V. 18. № 5. P. 249–256. doi: [10.1016/S0169-5347\(03\)00018-1](https://doi.org/10.1016/S0169-5347(03)00018-1)
34. Choudhary S., Sethy N. K., Shokeen B., Bhatia S. Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theoretical and Applied Genetics*. 2009. V. 118. № 3. P. 591–608.
35. Nicolai M., Pisani C., Bouchet J.P., Vuylsteke M., Palloix A. Discovery of a large set of SNP and SSR genetic markers by high-throughput sequencing of pepper (*Capsicum annuum*). *Genetics and Molecular Research*. 2012. V. 11. № 3. P. 2295–2300. doi: [10.4238/2012.august.13.3](https://doi.org/10.4238/2012.august.13.3)
36. Hodel R.G., Segovia-Salcedo M.C., Landis J.B., Crowl A.A. Sun M., Liu X., Gitzendanner M.A., Douglas N.A., Germain-Aubrey C.C., Chen S., et al. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*. 2016. V. 4. № 6. P. 1600025. doi: [10.3732/apps.1600025](https://doi.org/10.3732/apps.1600025)
37. Flanagan S.P., Jones A.G. The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*. 2019. V. 28. № 3. P. 544–567. doi: [10.1111/mec.14988](https://doi.org/10.1111/mec.14988)
38. Li Y.-C., Korol A.B., Fahima T., Nevo E. Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution*. 2004. V. 21. № 6. P. 991–1007. doi: [10.1093/molbev/msh073](https://doi.org/10.1093/molbev/msh073)
39. Varshney R.K., Graner A., Sorrells M.E. Genic microsatellite markers in plants: features and applications. *TRENDS in Biotechnology*. 2005. V. 23. № 1. P. 48–55. doi: [10.1016/j.tibtech.2004.11.005](https://doi.org/10.1016/j.tibtech.2004.11.005)
40. Zuo L., Zhang S., Zhang J., Liu Y., Yu X., Yang M., Wang J. Primer development and functional classification of EST-SSR markers in *Ulmus* species. *Tree Genetics & Genomes*. 2020. V. 16. № 5. P. 1–11. doi: [10.1007/s11295-020-01468-6](https://doi.org/10.1007/s11295-020-01468-6)
41. Chagné D., Chaumeil P., Ramboer A., Collada C., Guevara A., Cervera M.T., Vendramin G.G., Garcia V., Frigerio J.-M., Echt C., et al. Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theoretical and Applied Genetics*. 2004. V. 109. № 6. P. 1204–1214. doi: [10.1007/s00122-004-1683-z](https://doi.org/10.1007/s00122-004-1683-z)
42. Lesser M.R., Parchman T.L., Buerkle C.A. Cross-species transferability of SSR loci developed from transcriptome sequencing in lodgepole pine. *Molecular Ecology Resources*. 2012. V. 12. № 3. P. 448–455. doi: [10.1111/j.1755-0998.2011.03102.x](https://doi.org/10.1111/j.1755-0998.2011.03102.x)
43. Telfer E., Graham N., Macdonald L., Sturrock S., Wilcox P., Stanbra L. Approaches to variant discovery for conifer transcriptome sequencing. *PLoS One*. 2018. V. 13. № 11. P. e0205835. doi: [10.1371/journal.pone.0205835](https://doi.org/10.1371/journal.pone.0205835)
44. Sebastiani F., Pinzauti F., Kujala S.T., González-Martínez S.C., Vendramin G.G. Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L. *Conservation Genetics Resources*. 2012. V. 4. № 2. P. 231–234. doi: [10.1007/s12686-011-9513-5](https://doi.org/10.1007/s12686-011-9513-5)
45. Liewlaksaneeyanawin C., Ritland C.E., El-Kassaby Y.A., Ritland K. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical and Applied Genetics*. 2004. V. 109. № 2. P. 361–369. doi: [10.1007/s00122-004-1635-7](https://doi.org/10.1007/s00122-004-1635-7)
46. Fang P., Niu S., Yuan H., Li Z., Zhang Y., Yuan L., Li W. Development and Characterization of 25 EST-SSR markers in *Pinus sylvestris* var. *mongolica* (Pinaceae). *Applications in Plant Sciences*. 2014. V. 2. № 1. P. 1300057. doi: [10.3732/apps.1300057](https://doi.org/10.3732/apps.1300057)
47. Zimin A., Stevens K.A., Crepeau M.W., Holtz-Morris A., Koriabine M., Marçais G., Puiu D., Roberts M., Wegrzyn J.L., de Jong P. J., et al. Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics*. 2014. V. 196. № 3. P. 875–890. doi: [10.1534/genetics.113.159715](https://doi.org/10.1534/genetics.113.159715)

48. Kumar S., Stecher G., Li M., Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*. 2018. V. 35. № 6. P. 1547–1549. doi: [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096)
49. Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talón M., Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005. V. 21. № 18. P. 3674–3676. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610)
50. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000. V. 28. № 1. P. 27–30. [10.1093/nar/28.1.27](https://doi.org/10.1093/nar/28.1.27)
51. Marchler-Bauer A., Bryant S.H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Research*. 2004. V. 32. № suppl\_2. P. W327–W331. doi: [10.1093/nar/gkh454](https://doi.org/10.1093/nar/gkh454)
52. Echt C.S., Saha S., Deemer D.L., Nelson C.D. Microsatellite DNA in genomic survey sequences and UniGenes of loblolly pine. *Tree Genetics & Genomes*. 2011a. V. 7. № 4. P. 773–780. doi: [10.1007/s11295-011-0373-7](https://doi.org/10.1007/s11295-011-0373-7)
53. Echt C.S., Saha S., Krutovsky K.V., Wimalanathan K., Erpelding J. E., Liang C., Nelson, C.D. An annotated genetic map of loblolly pine based on microsatellite and cDNA markers. *BMC Genetics*. 2011b. V. 12. № 1. P. 1–16. doi: [10.1186/1471-2156-12-17](https://doi.org/10.1186/1471-2156-12-17)
54. Li X., Liu X., Wei J., Li Y., Tigabu M., Zhao X. Development and Transferability of EST-SSR Markers for *Pinus koraiensis* from Cold-Stressed Transcriptome through Illumina Sequencing. *Genes*. 2020. V. 11. № 5. P. 500. doi: [10.3390/genes11050500](https://doi.org/10.3390/genes11050500)
55. Doolittle R.F. *Of URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences*. Sausalito, Calif: University Science Books, 1986. 112 p.
56. Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S., Rafalski A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*. 1996. V. 2. № 3. P. 225–238. doi: [10.1007/bf00564200](https://doi.org/10.1007/bf00564200)
57. Scalfi M., Troggio M., Piovani P., Leonardi S., Magnaschi G., Vendramin G.G., Menozzi P. A RAPD, AFLP and SSR linkage map, and QTL analysis in European beech (*Fagus sylvatica* L.). *Theoretical and Applied Genetics*. 2004. V. 108. № 3. P. 433–441. doi: [10.1007/s00122-003-1583-7](https://doi.org/10.1007/s00122-003-1583-7)
58. Song Q.J., Shi J.R., Singh S., Fickus E.W., Costa J.M., Lewis J., Gill B.S., Ward R., Cregan P.B. Development and mapping of microsatellite (SSR) markers in wheat. *Theoretical and Applied Genetics*. 2005. V. 110. № 3. P. 550–560. doi: [10.1007/s00122-004-1871-x](https://doi.org/10.1007/s00122-004-1871-x)
59. Chen L., Ma Q., Chen Y., Wang B., Pei D. Identification of major walnut cultivars grown in China based on nut phenotypes and SSR markers. *Scientia Horticulturae*. 2014. V. 168. P. 240–248. doi: [10.1016/j.scienta.2014.02.004](https://doi.org/10.1016/j.scienta.2014.02.004)
60. Jeennor S., Volkaert H. Mapping of quantitative trait loci (QTLs) for oil yield using SSRs and gene-based markers in African oil palm (*Elaeis guineensis* Jacq.). *Tree Genetics & Genomes*. 2014. V. 10. № 1. P. 1–14. doi: [10.1007/s11295-013-0655-3](https://doi.org/10.1007/s11295-013-0655-3)
61. Lin H., Li Y., Xue Z., Chen M., Zhu H., Wen Q. Analysis of SSR loci in transcriptome and development of molecular markers in *Brassica oleraceavar. botrytis* L. *Journal of Northwest A & F University-Natural Science Edition*. 2019. V. 47. № 3. P. 85–93.
62. Jiang Y., Xu S., Wang R., Zhou J., Dou J., Yin Q., Wang R. Characterization, validation, and cross-species transferability of EST-SSR markers developed from *Lycoris aurea* and their application in genetic evaluation of *Lycoris* species. *BMC Plant Biology*. 2020. V. 20. № 1. P. 1–14. doi: [10.1186/s12870-020-02727-3](https://doi.org/10.1186/s12870-020-02727-3)
63. Pavy N., Laroche J., Bousquet J., Mackay J. Large-scale statistical analysis of secondary xylem ESTs in pine. *Plant Molecular Biology*. 2005. V. 57. № 2. P. 203–224. doi: [10.1007/s11103-004-6969-7](https://doi.org/10.1007/s11103-004-6969-7)
64. Cairney J., Zheng L., Cowels A., Hsiao J., Zismann V., Liu J., Ouyang S., Thibaud-Nissen F., Hamilton J., Childs K., et al. Expressed Sequence Tags from loblolly pine



- embryos reveal similarities with angiosperm embryogenesis. *Plant Molecular Biology*. 2006. V. 62. № 4. P. 485–501.
65. Nystedt B., Street N.R., Wetterbom A., Zuccolo A., Lin Y.-C., Scofield D.G., Vezzi F., Delhomme N., Giacomello S., Alexeyenko A. et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013. V. 497. № 7451. P. 579–584. doi: [10.1038/nature12211](https://doi.org/10.1038/nature12211)
  66. De La Torre A.R., Piot A., Liu B., Wilhite B., Weiss M., Porth I. Functional and morphological evolution in gymnosperms: A portrait of implicated gene families. *Evolutionary Applications*. 2020. V. 13. № 1. P. 210–227. doi: [10.1111/eva.12839](https://doi.org/10.1111/eva.12839)
  67. Nikolic B., Kovacevic D., Mladenovic-Drinic S., Nikolić A., Mitić Z., Bojović S., Marin P.D. Relationships among some pines from subgenera *Pinus* and *Strobus* revealed by nuclear EST-microsatellites. *Genetika*. 2018. V. 50. № 1. P. 69–84. doi: [10.2298/GENSR1801069N](https://doi.org/10.2298/GENSR1801069N)
  68. Navascués M. *Genetic Diversity of The Endemic Canary Island Pine Tree, Pinus canariensis*: PhD Theses. Norwich: University of East Anglia, 2005. 230 p.
  69. Sarac Z., Aleksic J., Dodos T., Rajcevic N., Bojović S., Marin P.D. Cross-species amplification of nuclear EST-microsatellites developed for other *Pinus* species in *Pinus nigra*. *Genetika*. 2015. V. 47. № 1. P. 205–217. doi: [10.2298/GENSR1501205S](https://doi.org/10.2298/GENSR1501205S)
  70. Kalko G.V., Kotova T.M. The microsatellite markers for estimation of genetic diversity of Scots pine. *Proceedings of the Saint Petersburg Forestry Research Institute*. 2018. P. 17–30. doi: [10.21178/2079-6080.2018.3-4.17](https://doi.org/10.21178/2079-6080.2018.3-4.17)
  71. Kormutak A., Galgoci M., Manka P., Koubova M., Jopcik M., Sukenikova D., Bolecek P., Gómóry D. Field-based artificial crossings indicate partial compatibility of reciprocal crosses between *Pinus sylvestris* and *Pinus mugo* and unexpected chloroplast DNA inheritance. *Tree Genetics & Genomes*. 2017. V. 13. № 3. P. 1–13. doi: [10.1007/s11295-017-1152-x](https://doi.org/10.1007/s11295-017-1152-x)
  72. Kalko G.V. The testing of nuclear microsatellite markers of Scots pine. *Proceedings of the Saint Petersburg Forestry Research Institute*. 2017. P. 23–34. doi: [10.21178/2079-6080.2017.1.23](https://doi.org/10.21178/2079-6080.2017.1.23)
  73. Máchová P., Cvrčková H., Poláková L., Trčková O. Genetic variability of selected populations of Scots pine in the Czech Republic. *Reports of forestry research-Zpravylesnickenovyzkumu*. 2016. V. 61. № 3. P. 223–229.
  74. Tóth E.G., Vendramin G.G., Bagnoli F., Cseke K., Höhn M. High genetic diversity and distinct origin of recently fragmented Scots pine (*Pinus sylvestris* L.) populations along the Carpathians and the Pannonian Basin. *Tree Genetics & Genomes*. 2017. V. 13. № 2. P. 1–12. doi: [10.1007/s11295-017-1137-9](https://doi.org/10.1007/s11295-017-1137-9)
  75. Tóth E.G., Bede-Fazekas Á., Vendramin G.G., Bagnoli F., Höhn M. Mid-Pleistocene and Holocene demographic fluctuation of Scots pine (*Pinus sylvestris* L.) in the Carpathian Mountains and the Pannonian Basin: Signs of historical expansions and contractions. *Quaternary International*. 2019. V. 504. P. 202–213. doi: [10.1016/j.quaint.2017.11.024](https://doi.org/10.1016/j.quaint.2017.11.024)
  76. González-Díaz P., Jump A.S., Perry A., Wachowiak W., Lapshina E., Cavers S. Ecology and management history drive spatial genetic structure in Scots pine. *Forest Ecology and Management*. 2017. V. 400. P. 68–76. doi: [10.1016/j.foreco.2017.05.035](https://doi.org/10.1016/j.foreco.2017.05.035)
  77. Volkov V.A. Applying microsatellite DNA markers for establish fact of illegal logging of basic coniferous forest-forming species in North-West Russian area. *Sinergiâ Nauk*. 2018. V. 30. P. 2211–2219.
  78. Șofletea N., Mihai G., Ciocîrlan E., Curtu A.L. Genetic diversity and spatial genetic structure in isolated Scots Pine (*Pinus sylvestris* L.) populations native to Eastern and Southern Carpathians. *Forests*. 2020. V. 11. № 10. P. 1047. [10.3390/f11101047](https://doi.org/10.3390/f11101047)
  79. Żukowska W.B., Wójkiewicz B., Litkowiec M., Wachowiak W. Cross-amplification and multiplexing of cpSSRs and nSSRs in two closely related pine species (*Pinus*

- sylvestris* L. and *P. mugo* Turra). *Dendrobiology*. 2017. V. 77. P. 59–64. doi: [10.12657/denbio.077.005](https://doi.org/10.12657/denbio.077.005)
80. Ellis J.R., Burke J.M. EST-SSRs as a resource for population genetic analyses. *Heredity*. 2007. V. 99. № 2. P. 125–132. doi: [10.1038/sj.hdy.6801001](https://doi.org/10.1038/sj.hdy.6801001)
  81. Duran C., Singhania R., Raman H., Batley J., Edwards D. Predicting polymorphic EST - SSR s in silico. *Molecular Ecology Resources*. 2013. V. 13. № 3. P. 538–545. doi: [10.1111/1755-0998.12078](https://doi.org/10.1111/1755-0998.12078)
  82. Luro F.L., Costantino G., Terol J., Argout X., Allario T., Wincker P., Talon M., Ollitrault P., Morillon R. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC Genomics*. 2008. V. 9. № 1. P. 1–13. doi: [10.1186/1471-2164-9-287](https://doi.org/10.1186/1471-2164-9-287)
  83. Dutta S., Kumawat G., Singh B.P., Gupta D.K., Singh S., Dogra V., Gaikwad K., Sharma T.R., Raje R.S., Bandhopadhyaya T.K. et al. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biology*. 2011. V. 11. № 1. P. 1–13. doi: [10.1186/1471-2229-11-17](https://doi.org/10.1186/1471-2229-11-17)
  84. Wu J., Cai C., Cheng F., Cui H., Zhou H. Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Molecular Breeding*. 2014. V. 34. № 4. P. 1853–1866. doi: [10.1007/s11032-014-0144-x](https://doi.org/10.1007/s11032-014-0144-x)
  85. Kuchma O., Vornam B., Finkeldey R. Mutation rates in Scots pine (*Pinus sylvestris* L.) from the Chernobyl exclusion zone evaluated with amplified fragment-length polymorphisms (AFLPs) and microsatellite markers. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*. 2011. V. 725. № 1–2. C. 29–35. doi: [10.1016/j.mrgentox.2011.07.003](https://doi.org/10.1016/j.mrgentox.2011.07.003)
  86. Danusevičius D., Kerpauskaitė V., Kavaliauskas D., Fussi B., Konnert M., Baliuckas V. The effect of tending and commercial thinning on the genetic diversity of Scots pine stands. *European Journal of Forest Research*. 2016. V. 135. № 6. P. 1159–1174. doi: [10.1007/s10342-016-1002-7](https://doi.org/10.1007/s10342-016-1002-7)
  87. Shilkina E.A., Ibe A.A., Sheller M.A., Sukhikh T.V. Using methods of DNA-analysis in the examination of the illegal timber trade. *Siberian Forest Journal*. 2019. V. 3. P. 64–70. doi: [10.15372/SJFS20190308](https://doi.org/10.15372/SJFS20190308)
  88. Tian A.G., Wang J., Cui P., Han Y.-J., Xu H., Cong L.-J., Huang X.-G., Wang X.-L., Jiao Y.-Z., Wang B.-J., et al. Characterization of soybean genomic features by analysis of its expressed sequence tags. *Theoretical and Applied Genetics*. 2004. V. 108. № 5. P. 903–913. doi: [10.1007/s00122-003-1499-2](https://doi.org/10.1007/s00122-003-1499-2)
  89. Gong Y.M., Xu S.C., Mao W.H., Hu Q.Z., Zhang G.W., Ding J., Li Y.D. Developing new SSR markers from ESTs of pea (*Pisum sativum* L.). *Journal of Zhejiang University Science B*. 2010. V. 11. № 9. P. 702–707. doi: [10.1631/jzus.b1000004](https://doi.org/10.1631/jzus.b1000004)
  90. Cloutier S., Miranda E., Ward K., Radovanovic N., Reimer E., Walichnowski A., Datla R., Rowland G., Duguid S., Ragupathy R. Simple sequence repeat marker development from bacterial artificial chromosome end sequences and expressed sequence tags of flax (*Linum usitatissimum* L.). *Theoretical and Applied Genetics*. 2012. V. 125. № 4. P. 685–694. doi: [10.1007/s00122-012-1860-4](https://doi.org/10.1007/s00122-012-1860-4)
  91. Tóth G., Gáspári Z., Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research*. 2000. V. 10. № 7. P. 967–981. doi: [10.1101/gr.10.7.967](https://doi.org/10.1101/gr.10.7.967)
  92. Metzgar D., Bytof J., Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome research*. 2000. V. 10. № 1. P. 72–80.
  93. Lawson M.J., Zhang L. Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene*. 2008. V. 407. № 1–2. P. 54–62. doi: [10.1016/j.gene.2007.09.017](https://doi.org/10.1016/j.gene.2007.09.017)

94. Vinces M.D., Legendre M., Caldara M., Hagihara M., Verstrepen K.J. Unstable tandem repeats in promoters confer transcriptional evolvability. *Science*. 2009. V. 324. № 5931. P. 1213–1216. doi: [10.1126/science.1170097](https://doi.org/10.1126/science.1170097)
95. Ince A.G., Karaca M., Naci Onus A. Differential expression patterns of genes containing microsatellites in *Capsicum annuum* L. *Molecular Breeding*. 2010. V. 25. № 4. P. 645–658. doi: [10.1007/s11032-009-9362-z](https://doi.org/10.1007/s11032-009-9362-z)
96. Joshi-Saha A., Reddy K.S. Repeat length variation in the 5'UTR of myo -inositol monophosphatase gene is related to phytic acid content and contributes to drought tolerance in chickpea (*Cicer arietinum* L.). *Journal of Experimental Botany*. 2015. V. 66. № 19. P. 5683–5690. doi: [10.1093/jxb/erv156](https://doi.org/10.1093/jxb/erv156)
97. Kumar S., Bhatia S. A polymorphic (GA/CT) n-SSR influences promoter activity of Tryptophan decarboxylase gene in *Catharanthus roseus* L. *Scientific Reports*. 2016. V. 6. № 1. P. 1–12. doi: [10.1038/srep33280](https://doi.org/10.1038/srep33280)
98. Abedini D., Monfared S.R., Abbasi A. The effects of promoter variations of the N-Methylcanadine 1-Hydroxylase (CYP82Y1) gene on the noscapine production in opium poppy. *Scientific Reports*. 2018. V. 8. № 1. P. 1–11. doi: [10.1038/s41598-018-23351-0](https://doi.org/10.1038/s41598-018-23351-0)
99. Sangwan I., O'Brian M.R. Identification of a Soybean Protein That Interacts with GAGA Element Dinucleotide Repeat DNA. *Plant Physiology*. 2002. V. 129. № 4. P. 1788–1794. doi: [10.1104/pp.002618](https://doi.org/10.1104/pp.002618)
100. Zhang L., Zuo K., Zhang F., Cao Y., Wang J., Zhang Y., Sun X., Tang K. Conservation of noncoding microsatellites in plants: implication for gene regulation. *Bmc Genomics*. 2006. V. 7. № 1. P. 1–14. doi: [10.1186/1471-2164-7-323](https://doi.org/10.1186/1471-2164-7-323)
101. Mignone F., Gissi C., Liuni S., Pesole G. Untranslated regions of mRNAs. *Genome Biology*. 2002. V. 3. № 3. P. 1–10.
102. Heller G., Adomas A.G., Li G., Osborne J., van Zyl L., Sederoff R., Finlay R.D., Stenlid J., Asiegbu F.O. Transcriptional analysis of *Pinus sylvestris* roots challenged with the ectomycorrhizal fungus *Laccaria bicolor*. *BMC Plant Biology*. 2008. V. 8. № 1. P. 1–13. doi: [10.1186/1471-2229-8-19](https://doi.org/10.1186/1471-2229-8-19)
103. Fernández-Pozo N., Canale J., Guerrero-Fernández D., Villalobos D.P., Díaz-Moreno S.M., Bautista R., Flores-Monterroso A., Guevara M.Á., Perdiguero P., Collada C. et al. EuroPineDB: a high-coverage web database for maritime pine transcriptome. *BMC Genomics*. 2011. V. 12. № 1. P. 1–11. doi: [10.1186/1471-2164-12-366](https://doi.org/10.1186/1471-2164-12-366)
104. Niu S.H., Li Z.X., Yuan H.W., Chen X.Y., Li Y., Li W. Transcriptome characterisation of *Pinus tabuliformis* and evolution of genes in the *Pinus* phylogeny. *Bmc Genomics*. 2013. V. 14. № 1. P. 1–12. doi: [10.1186/1471-2164-14-263](https://doi.org/10.1186/1471-2164-14-263)
105. Du J., Zhang Z., Zhang H., Junhong T. EST–SSR marker development and transcriptome sequencing analysis of different tissues of Korean pine (*Pinus koraiensis* Sieb. et Zucc.). *Biotechnology & Biotechnological Equipment*. 2017. V. 31. № 4. P. 679–689. doi: [10.1080/13102818.2017.1331755](https://doi.org/10.1080/13102818.2017.1331755)
106. Chano V., López de Heredia U., Collada C., Soto Á. Transcriptomic analysis of juvenile wood formation during the growing season in *Pinus canariensis*. *Holzforschung*. 2017. V. 71. № 12. P. 919–937. doi: [10.1515/hf-2017-0014](https://doi.org/10.1515/hf-2017-0014)
107. Brown G.R., Bassoni D.L., Gill G.P., Fontana J.R., Wheeler N.C., Megraw R.A., Davis M.F., Sewell M.M., Tuskan G.A., Neale D.B. Identification of quantitative trait loci influencing wood property traits in loblolly pine (*Pinus taeda* L.). III. QTL verification and candidate gene mapping. *Genetics*. 2003. V. 164. № 4. P. 1537–1546. doi: [10.1093/genetics/164.4.1537](https://doi.org/10.1093/genetics/164.4.1537)
108. Dillon S.K., Nolan M.F., Wu H., Southerton S.G. Association genetics reveal candidate gene SNPs affecting wood properties in *Pinus radiata*. *Australian Forestry*. 2010. V. 73. № 3. P. 185–190. doi: [10.1080/00049158.2010.10676326](https://doi.org/10.1080/00049158.2010.10676326)

109. Lepoittevin C., Harvengt L., Plomion C., Garnier-Géré P. Association mapping for growth, straightness and wood chemistry traits in the *Pinus pinaster* Aquitaine breeding population. *Tree Genetics & Genomes*. 2012. V. 8. № 1. P. 113–126. doi: [10.1007/s11295-011-0426-y](https://doi.org/10.1007/s11295-011-0426-y)
110. Li Z., Hallingbäck H.R., Abrahamsson S., Fries A., Gull B.A., Sillanpää M.J., García-Gil M.R. Functional multi-locus QTL mapping of temporal trends in scots pine wood traits. *G3: Genes, Genomes, Genetics*. 2014. V. 4. № 12. P. 2365–2379. doi: [10.1534/g3.114.014068](https://doi.org/10.1534/g3.114.014068)
111. Calleja-Rodríguez A., Li Z., Hallingbäck H.R., Sillanpää M.J., Abrahamsson S., García-Gil M.R. Functional Quantitative Trait Loci (QTL) analysis for adaptive traits in a three-generation Scots pine pedigree. *bioRxiv*. 2018. P. 297986. doi: [10.1101/297986](https://doi.org/10.1101/297986)
112. Groover A., Devey M., Fiddler T., Lee J., Megraw R., Mitchel-Olds T., Sherman B., Vujcic S., Williams C., Nede D. Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics*. 1994. V. 138. № 4. P. 1293–1300. doi: [10.1093/genetics/138.4.1293](https://doi.org/10.1093/genetics/138.4.1293)
113. Chagné D., Brown G., Lalanne C., Madur D., Pot D., Neale D., Plomion C. Comparative genome and QTL mapping between maritime and loblolly pines. *Molecular Breeding*. 2003. V. 12. № 3. P. 185–195.
114. González-Martínez S.C., Wheeler N.C., Ersoz E., Nelson C.D., Neale D.B. Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics*. 2007. V. 175. № 1. P. 399–409. [10.1534/genetics.106.061127](https://doi.org/10.1534/genetics.106.061127)
115. Lorenz W.W., Dean J.F. SAGE profiling and demonstration of differential gene expression along the axial developmental gradient of lignifying xylem in loblolly pine (*Pinus taeda*). *Tree Physiology*. 2002. V. 22. № 5. P. 301–310. doi: [10.1093/treephys/22.5.301](https://doi.org/10.1093/treephys/22.5.301)
116. Kirst M., Johnson A.F., Baucom C., Erin U., Hubbard K., Staggs R., Paule C., Retzel E., Whetten R., Sederoff R. Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*. 2003. V. 100. № 12. P. 7383–7388. doi: [10.1073/pnas.1132171100](https://doi.org/10.1073/pnas.1132171100)
117. Yang S.H., van Zyl L., No E.G., Loopstra C.A. Microarray analysis of genes preferentially expressed in differentiating xylem of loblolly pine (*Pinus taeda*). *Plant Science*. 2004. V. 166. № 5. P. 1185–1195. doi: [10.1016/j.plantsci.2003.12.030](https://doi.org/10.1016/j.plantsci.2003.12.030)
118. Li X., Wu H.X., Dillon S.K., Southerton S.G. Generation and analysis of expressed sequence tags from six developing xylem libraries in *Pinus radiata* D. Don. *BMC Genomics*. 2009. V. 10. № 1. P. 1–18. doi: [10.1186/1471-2164-10-41](https://doi.org/10.1186/1471-2164-10-41)
119. Li X., Wu H.X., Southerton S.G. Seasonal reorganization of the xylem transcriptome at different tree ages reveals novel insights into wood formation in *Pinus radiata*. *New Phytologist*. 2010. V. 187. № 3. P. 764–776. doi: [10.1111/j.1469-8137.2010.03333.x](https://doi.org/10.1111/j.1469-8137.2010.03333.x)
120. Li X., Wu H.X., Southerton S.G. Identification of putative candidate genes for juvenile wood density in *Pinus radiata*. *Tree Physiology*. 2012. V. 32. № 8. P. 1046–1057. doi: [10.1093/treephys/tps060](https://doi.org/10.1093/treephys/tps060)
121. Zhang D., Hrmova M., Wan C.-H., Wu C., Balzen J., Cai W., Wang J., Densmore L.D., Fincher G.B., Zhang H., Haigler C.H. Members of a New Group of Chitinase-Like Genes are Expressed Preferentially in Cotton Cells with Secondary Walls. *Plant Molecular Biology*. 2004. V. 54. № 3. P. 353–372. doi: [10.1023/B:PLAN.0000036369.55253.dd](https://doi.org/10.1023/B:PLAN.0000036369.55253.dd)
122. Aspeborg H., Schrader J., Coutinho P.M., Stam M., Kallas A., Djerbi S., Nilsson P., Denman S., Amini B., Sterky F., et al. Carbohydrate-active enzymes involved in the secondary cell wall biogenesis in hybrid aspen. *Plant Physiology*. 2005. V. 137. № 3. P. 983–997. doi: [10.1104/pp.104.055087](https://doi.org/10.1104/pp.104.055087)

123. de Carvalho M.C.D.C.G., Caldas D.G.G., Carneiro R.T., Moon D.H., Salvatierra G.R., Franceschini L.M., de Andrade A., Celedon P.A.F., Oda S., Labate C.A. SAGE transcript profiling of the juvenile cambial region of *Eucalyptus grandis*. *Tree Physiology*. 2008. V. 28. № 6. P. 905–919. doi: [10.1093/treephys/28.6.905](https://doi.org/10.1093/treephys/28.6.905)
124. Sánchez-Rodríguez C., Bauer S., Hématy K., Saxe F., Ibáñez A. B., Vodermaier V., Konlechner C., Sampathkumar A., Rüggeberg M., Aichinger E., et al. CHITINASE-LIKE1/POM-POM1 and Its Homolog CTL2 Are Glucan-Interacting Proteins Important for Cellulose Biosynthesis in Arabidopsis. *The Plant Cell*. 2012. V. 24. № 2. P. 589–607. doi: [10.1105/tpc.111.094672](https://doi.org/10.1105/tpc.111.094672)
125. Pirttilä A., Laukkanen H., Hohtola A. Chitinase production in pine callus (*Pinus sylvestris* L.): a defense reaction against endophytes? *Planta*. 2002. V. 214. № 6. P. 848–852. doi: [10.1007/s00425-001-0709-x](https://doi.org/10.1007/s00425-001-0709-x)
126. Bischoff V., Nita S., Neumetzler L., Schindelasch D., Urbain A., Eshed R., Persson S., Delmer D., Scheible W.-R. TRICHOME BIREFRINGENCE and Its Homolog AT5G01360 Encode Plant-Specific DUF231 Proteins Required for Cellulose Biosynthesis in Arabidopsis. *Plant Physiology*. 2010. V. 153. № 2. P. 590–602. doi: [10.1104/pp.110.153320](https://doi.org/10.1104/pp.110.153320)
127. Pauly M., Ramírez V. New insights into wall polysaccharide O-acetylation. *Frontiers in Plant Science*. 2018. P. 1210. doi: [10.3389/fpls.2018.01210](https://doi.org/10.3389/fpls.2018.01210)
128. Haigler C.H., Singh B., Wang G., Zhang D. Genomics of cotton fiber secondary wall deposition and cellulose biogenesis. In: *Genetics and Genomics of Cotton*. Ed. Paterson A.H. New York: Springer. 2009. P. 385–417. (Plant Genetics and Genomics: Crops and Models, V. 3). doi: [10.1007/978-0-387-70810-2\\_16](https://doi.org/10.1007/978-0-387-70810-2_16)
129. Ubeda-Tomas S., Edvardsson E., Eland C., Singh S.K., Zadik D., Aspeborg H., Gorzsás A., Teeri T.T., Sundberg B., Persson P., et al. Genomic-assisted identification of genes involved in secondary growth in Arabidopsis utilising transcript profiling of poplar wood-forming tissues. *Physiologia Plantarum*. 2007. V. 129. № 2. P. 415–428. doi: [10.1111/j.1399-3054.2006.00817.x](https://doi.org/10.1111/j.1399-3054.2006.00817.x)
130. Hodgson-Kratky K., Perlo V., Furtado A., Choudhary H., Gladden J.M., Simmons B.A., Botha F. Henry R.J. Association of gene expression with syringyl to guaiacyl ratio in sugarcane lignin. *Plant Molecular Biology*. 2021. V. 106. № 1. P. 173–192. doi: [10.1007/s11103-021-01136-w](https://doi.org/10.1007/s11103-021-01136-w)
131. Gui J., Luo L., Zhong Y., Sun J., Umezawa T., Li L. Phosphorylation of LTF1, an MYB transcription factor in *Populus*, acts as a sensory switch regulating lignin biosynthesis in wood cells. *Molecular Plant*. 2019. T. 12. № 10. P. 1325–1337. doi: [10.1016/j.molp.2019.05.008](https://doi.org/10.1016/j.molp.2019.05.008)
132. Patzlaff A., McInnis S., Courtenay A., Surman C., Newman L.J., Smith C., Bevan M.W., Mansfield S., Whetten R.W., Sederoff R.R., Campbell M.M. Characterisation of a pine MYB that regulates lignification. *The Plant Journal*. 2003. V. 36. № 6. P. 743–754. doi: [10.1046/j.1365-313X.2003.01916.x](https://doi.org/10.1046/j.1365-313X.2003.01916.x)
133. Bedon F., Grima-Pettenati J., Mackay J. Conifer R2R3-MYB transcription factors: sequence analyses and gene expression in wood-forming tissues of white spruce (*Picea glauca*). *BMC Plant Biology*. 2007. V. 7. № 1. P. 1–17. doi: [10.1186/1471-2229-7-17](https://doi.org/10.1186/1471-2229-7-17)
134. Bomal C., Bedon F., Caron S., Mansfield S.D., Levasseur C., Cooke J.E.K., Blais S., Tremblay L., Morency M.-J., Pavy N., et al. Involvement of *Pinus taeda* MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. *Journal of Experimental Botany*. 2008. V. 59. № 14. P. 3925–3939. doi: [10.1093/jxb/ern234](https://doi.org/10.1093/jxb/ern234)
135. Ni Z., Han X., Yang Z., Xu M., Feng Y., Chen Y., Xu L. A. Integrative analysis of wood biomass and developing xylem transcriptome provide insights into mechanisms of lignin biosynthesis in wood formation of *Pinus massoniana*. *International Journal of*

- Biological Macromolecules*. 2020. V. 163. P. 1926–1937. doi: [10.1016/j.ijbiomac.2020.08.253](https://doi.org/10.1016/j.ijbiomac.2020.08.253)
136. Kim M.-H., Tran T.N.A., Cho J.-S., Park E.J., Lee H., Kim D.G., Hwang S., Ko J.H. Wood transcriptome analysis of *Pinus densiflora* identifies genes critical for secondary cell wall formation and NAC transcription factors involved in tracheid formation. *Tree Physiology*. 2021. V. 41. № 7. P. 1289–1305. doi: [10.1093/treephys/tpab001](https://doi.org/10.1093/treephys/tpab001)
  137. Ko J., Prassinos C., Han K. Developmental and seasonal expression of PtaHB1, a *Populus* gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). *New Phytologist*. 2006. V. 169. № 3. P. 469–478. doi: [10.1111/j.1469-8137.2005.01623.x](https://doi.org/10.1111/j.1469-8137.2005.01623.x)
  138. Côté C.L., Boileau F., Roy V., Ouellet M., Levasseur C., Morency M.J., Cooke J.E.K., Séguin A., MacKay J.J. Gene family structure, expression and functional analysis of HD-Zip III genes in angiosperm and gymnosperm forest trees. *BMC Plant Biology*. 2010. V. 10. № 1. P. 1–17. doi: [10.1186/1471-2229-10-273](https://doi.org/10.1186/1471-2229-10-273)
  139. Gao J., Lan T. Functional characterization of the late embryogenesis abundant (LEA) protein gene family from *Pinus tabulaeformis* (Pinaceae) in *Escherichia coli*. *Scientific Reports*. 2016. V. 6. № 1. P. 1–10. doi: [10.1038/srep19467](https://doi.org/10.1038/srep19467)
  140. Veluthakkal R., Dasgupta M.G. Pathogenesis-related genes and proteins in forest tree species. *Trees*. 2010. V. 24. № 6. P. 993–1006. doi: [10.1007/s00468-010-0489-7](https://doi.org/10.1007/s00468-010-0489-7)
  141. Mageroy M.H., Parent G., Germanos G., Giguère I., Delvas N., Maaroufi H., Bauce E., Bohlmann J., Mackay J.J. Expression of the  $\beta$ -glucosidase gene *Pg $\beta$ glu-1* underpins natural resistance of white spruce against spruce budworm. *The Plant Journal*. 2015. V. 81. № 1. P. 68–80. doi: [10.1111/tpj.12699](https://doi.org/10.1111/tpj.12699)
  142. Porth I., White R., Jaquish B., Ritland K. Partial correlation analysis of transcriptomes helps detangle the growth and defense network in spruce. *New Phytologist*. 2018. V. 218. № 4. P. 1349–1359. doi: [10.1111/nph.15075](https://doi.org/10.1111/nph.15075)
  143. De La Torre A.R., Puiu D., Crepeau M.W., Stevens K., Salzberg S.L., Langley C.H., Neale D.B. Genomic architecture of complex traits in loblolly pine. *New Phytologist*. 2019. V. 221. № 4. P. 1789–1801. doi: [10.1111/nph.15535](https://doi.org/10.1111/nph.15535)
  144. Elsik C.G., Minihan V.T., Hall S.E., Scarpa A.M., Williams C.G. Low-copy microsatellite markers for *Pinus taeda* L. *Genome*. 2000. V. 43. № 3. P. 550–555. doi: [10.1139/g00-002](https://doi.org/10.1139/g00-002)
  145. Komulainen P., Brown G.R., Mikkonen M., Karhu A., Garcia-Gil M. R., O'malley D., Lee B., Neale D.B., Savolainen O. Comparing EST-based genetic maps between *Pinus sylvestris* and *Pinus taeda*. *Theoretical and Applied Genetics*. 2003. V. 107. № 4. P. 667–678. doi: [10.1007/s00122-003-1312-2](https://doi.org/10.1007/s00122-003-1312-2)
  146. Waldmann P., García-Gil M.R., Sillanpää M.J. Comparing Bayesian estimates of genetic differentiation of molecular markers and quantitative traits: an application to *Pinus sylvestris*. *Heredity*. 2005. V. 94. № 6. P. 623–629. doi: [10.1038/sj.hdy.6800672](https://doi.org/10.1038/sj.hdy.6800672)
  147. Varis S., Santanen A., Pakkanen A., Pulkkinen P. The importance of being the first pollen in the strobili of Scots pine. *Canadian Journal of Forest Research*. 2008. V. 38. № 12. P. 2976–2980. doi: [10.1139/X08-138](https://doi.org/10.1139/X08-138)
  148. García-Gil M.R., Olivier F., Kamruzzahan S., Waldmann P. Joint analysis of spatial genetic structure and inbreeding in a managed population of Scots pine. *Heredity*. 2009. V. 103. № 1. P. 90–96. doi: [10.1038/hdy.2009.33](https://doi.org/10.1038/hdy.2009.33)
  149. Torimaru T., Wang X.-R., Fries A., Andersson B., Lindgren D. Evaluation of pollen contamination in an advanced Scots pine seed orchard. *Silvae Genetica*. 2009. V. 58. № 1–6. P. 262–269.
  150. Torimaru T., Wennström U., Lindgren D., Wang X.-R. Effects of male fecundity, interindividual distance and anisotropic pollen dispersal on mating success in a Scots pine (*Pinus sylvestris*) seed orchard. *Heredity*. 2012. V. 108. № 3. P. 312–321. doi: [10.1038/hdy.2011.76](https://doi.org/10.1038/hdy.2011.76)

151. Torimaru T., Wennström U., Andersson B., Almqvist C., Wang X.-R. Reduction of pollen contamination in Scots pine seed orchard crop by tent isolation. *Scandinavian Journal of Forest Research*. 2013. V. 28. № 8. P. 715–723. doi: [10.1080/02827581.2013.838298](https://doi.org/10.1080/02827581.2013.838298)
152. Wang X.-R., Torimaru T., Lindgren D., Fries A. Marker-based parentage analysis facilitates low input ‘breeding without breeding’ strategies for forest trees. *Tree Genetics & Genomes*. 2010. V. 6. № 2. P. 227–235. doi: [10.1007/s11295-009-0243-8](https://doi.org/10.1007/s11295-009-0243-8)
153. Baumanis I., Veinberga I., Ļubinskis L., Almqvist C., Wang X.R. Seed quality and genetic polymorphism in Scots pine seed orchards. *Mežzinātne*. 2012. V. 26. P. 74–87.
154. Abrahamsson S., Ahlinder J., Waldmann P., García-Gil M.R. Maternal heterozygosity and progeny fitness association in an inbred Scots pine population. *Genetica*. 2013. V. 141. № 1. P. 41–50. doi: [10.1007/s10709-013-9704-y](https://doi.org/10.1007/s10709-013-9704-y)
155. Korecký J., Klápště J., Lstibůrek M., Kobliha J., Nelson C.D., El-Kassaby Y.A. Comparison of genetic parameters from marker-based relationship, sibship, and combined models in Scots pine multi-site open-pollinated tests. *Tree Genetics & Genomes*. 2013. V. 9. № 5. P. 1227–1235. doi: [10.1007/s11295-013-0630-z](https://doi.org/10.1007/s11295-013-0630-z)
156. Demkovich A.E., Korshikov I.I., Politov D.V., Mudrik A., Los S.A. Genetic polymorphism of *Pinus sylvestris* L. plus trees and their progenies by SSR loci. *Plant Physiology and Genetics*. 2014. № 46. P. 395–405.
157. Sheykina O.V., Unzhenina O.V., Gladkov Yu.F. Selection of SSR-markers for identification of pine genotypes. In: *Reproduction of Forest Plants in Vitro Culture as a Basis for Plantation Forestry*. Yoshkar-Ola: Volga State University of Technology, 2014. P. 35–39.
158. Ilinov A.A., Raevsky B.V. Analysis of the *Pinus sylvestris* L. plus tree gene pool in Karelia using microsatellite loci. *Transactions of Karelian Research Centre of the Russian Academy of Sciences*. 2018. № 6. P. 124–134. doi: [10.17076/eb840](https://doi.org/10.17076/eb840)
159. Ilinov A.A., Raevsky B.V. Comparative evaluation of the genetic diversity of natural populations and clonal seed orchards of *Pinus sylvestris* L. and *Picea × fennica* (Regel) Kom. in Karelia. *Russian Journal of Genetics: Applied Research*. 2015. № 7. P. 607–616. doi: [10.1134/S2079059717060065](https://doi.org/10.1134/S2079059717060065)
160. Ilinov A.A., Raevsky B.V. The current state of *Pinus sylvestris* L. gene pool in Karelia. *Siberian Journal of Forest Science*. 2016. № 5. P. 45–54. doi: [10.15372/SJFS20160504](https://doi.org/10.15372/SJFS20160504)
161. Ilinov A.A., Raevsky B.V., Chirva O.V. The state of gene pool of the basic forest-forming species of the White Sea watershed (on the example of a *Picea × fennica* (Regel) Kom. and *Pinus sylvestris* L.). *Ecological Genetics*. №18. P. 185–202. doi: [10.17816/ecogen19006](https://doi.org/10.17816/ecogen19006)
162. Funda T., Wennström U., Almqvist C., Torimaru T., Gull B.A., Wang X.R. Low rates of pollen contamination in a Scots pine seed orchard in Sweden: the exception or the norm? *Scandinavian Journal of Forest Research*. 2015. V. 30. № 7. P. 573–586. doi: [10.1080/02827581.2015.1036306](https://doi.org/10.1080/02827581.2015.1036306)
163. Funda T., Wennström U., Almqvist C., Andersson Gull B., Wang X.R. Mating dynamics of Scots pine in isolation tents. *Tree Genetics & Genomes*. 2016. V. 12. № 6. P. 1–17. doi: [10.1007/s11295-016-1074-z](https://doi.org/10.1007/s11295-016-1074-z)
164. Wójkiewicz B., Litkowiec M., Wachowiak W. Contrasting patterns of genetic variation in core and peripheral populations of highly outcrossing and wind pollinated forest tree species. *AoB Plants*. 2016. V. 8. P. plw054. doi: [10.1093/aobpla/plw054](https://doi.org/10.1093/aobpla/plw054)
165. Čepl J., Holá D., Stejskal J., Korecký J., Kočová M., Lhotáková Z., Tomášková I., Palovská M., Rothová O., Whetten R.W., et al. Genetic variability and heritability of chlorophyll a fluorescence parameters in Scots pine (*Pinus sylvestris* L.). *Tree physiology*. 2016. V. 36. № 7. P. 883–895. doi: [10.1093/treephys/tpw028](https://doi.org/10.1093/treephys/tpw028)
166. Sheykina O., Gladkov Y. Simulation of the indices of genetic diversity depending on the number of plus trees of Scots pine. *Vestnik of Volga State University of Technology*.

- Les. Ekologiya. Prirodopolzovanie*. 2018. V. 1. № 37. P. 33–44. doi: [10.15350/2306-2827.2018.1.33](https://doi.org/10.15350/2306-2827.2018.1.33) (In Russ.)
167. González-Díaz P., Cavers S., Jason G., Booth A., Russell J., Jump A. S. Weak isolation by distance and geographic diversity gradients persist in Scottish relict pine forest. *iForest-Biogeosciences and Forestry*. 2018. V. 11. P. 449–458. doi: [10.3832/ifor2454-011](https://doi.org/10.3832/ifor2454-011)
168. Gladkov Y., Sheykina O. Genetic polymorphism of Scots pine trees from adjacent bawl and dry cenopopulations in nuclear microsatellite loci. *Vestnik of Volga State University of Technology. Les. Ekologiya. Prirodopolzovanie*. 2019. V. 4. № 44. P. 70–79. doi: [10.25686/2306-2827.2019.4.70](https://doi.org/10.25686/2306-2827.2019.4.70) (In Russ.)
169. Shuvaev D.N., Ibe A.A., Shcherba Yu.E., Sukhikh T.V., Shilkina E.A., Usova E.A., Lisotova E.V., Repyah M.V., Stupakova O.M. A panel of nuclear microsatellite markers for the identification of Scots pine illegal logs on the Krasnoyarsk territory. *Conifers Boreal Area*. 2020. № 38. P. 297–304.
170. Ganea S., Garcia Gil M.R. Multiplex nuclear SSR amplification in Scots pine (*Pinus sylvestris* L.). *Bulletin UASVM Horticulture*. 2011. V. 68. № 1. P. 47–53.
171. Ranade S.S., Ganea L.-S., Razzak A.M., García Gil M.R. Fungal infection increases the rate of somatic mutation in Scots pine (*Pinus sylvestris* L.). *Journal of Heredity*. 2015. V. 106. № 4. P. 386–394. doi: [10.1093/jhered/esv017](https://doi.org/10.1093/jhered/esv017)
172. García Gil M.R., Floran V., Östlund L., Mullin T.J., Andersson Gull B. Genetic diversity and inbreeding in natural and managed populations of Scots pine. *Tree Genetics & Genomes*. 2015. V. 11. № 2. P. 1–12. doi: [10.1007/s11295-015-0850-5](https://doi.org/10.1007/s11295-015-0850-5)
173. Ganea S., Ranade S.S., Hall D., Abrahamsson S., García-Gil M.R. Development and transferability of two multiplexes nSSR in Scots pine (*Pinus sylvestris* L.). *Journal of Forestry Research*. 2015. V. 26. № 2. P. 361–368. doi: [10.1007/s11676-015-0042-z](https://doi.org/10.1007/s11676-015-0042-z)
174. Bernhardsson C., Floran V., Ganea S.L., García-Gil M.R. Present genetic structure is congruent with the common origin of distant Scots pine populations in its Romanian distribution. *Forest Ecology and Management*. 2016. V. 361. P. 131–143. doi: [10.1016/j.foreco.2015.10.047](https://doi.org/10.1016/j.foreco.2015.10.047)
175. Zimmer K., Sønstebø J.H. A preliminary study on the genetic structure of Northern European *Pinus sylvestris* L. by means of neutral nuclear microsatellite markers. *Scandinavian Journal of Forest Research*. 2018. V. 33. № 1. P. 6–13. doi: [10.1080/02827581.2017.1337919](https://doi.org/10.1080/02827581.2017.1337919)
176. Nowakowska J.A. Application of DNA markers against illegal logging as a new tool for the Forest Guard Service. *Folia Forestalia Polonica. Series A. Forestry*. 2011. V. 53. № 2. P. 142–149.
177. Nowakowska J.A., Oszako T., Tereba A., Konecka A. Forest tree species traced with a DNA-based proof for illegal logging case in Poland. In: *Evolutionary Biology: Biodiversification from Genotype to Phenotype*. Cham: Springer International Publishing, 2015. P. 373–388.
178. Lewandowski A., Kowalczyk J., Litkowiec M., Urbaniak L., Rzońca M. Selection of elite plus trees of Scots pine and European larch for the establishment of 1.5 generation seed orchards. *Sylwan*. 2017. V. 161. № 11. P. 917–926.
179. Konecka A., Tereba A., Bieniek J., Nowakowska J.A. Comparison of the genetic variability of Scots pine trees and their progeny from nursery production based on DNA analyses. *Sylwan*. 2018. V. 162. P. 32–40.
180. Konecka A., Tereba A., Studnicki M., Nowakowska J.A. Rare and private alleles as a measure of gene pool richness in Scots pine planting material. *Sylwan*. 2019. V. 163. P. 948–956. doi: [10.26202/SYLWAN.2019068](https://doi.org/10.26202/SYLWAN.2019068)
181. Wójkiewicz B., Żukowska W.B., Urbaniak L., Kowalczyk J., Litkowiec M., Lewandowski A. Determination of the origin of the rychtal Scots pine (*Pinus sylvestris*



- L.) seed tree stands using microsatellite markers. *Sylwan*. 2019. № 163. P. 637–644. doi: [10.26202/SYLWAN.2019059](https://doi.org/10.26202/SYLWAN.2019059)
182. Kalko G., Kotova T. Real-Time PCR and analysis of amplicon melting curves to assess the suitability of SSR loci of Scots pine for multiplexing. *IOP Conference Series: Earth and Environmental Science*. 2019. V. 392. № 1. P. 012015. [10.1088/1755-1315/392/1/012015](https://doi.org/10.1088/1755-1315/392/1/012015)
183. Tereba A., Konecka A. Comparison of microsatellites and SNP markers in genetic diversity level of two Scots pine stands. *Environmental Sciences Proceedings*. 2021. V. 3. № 1. P. 1–3. doi: [10.3390/IECF2020-07776](https://doi.org/10.3390/IECF2020-07776)
184. Przybylski P., Mohytych V., Rutkowski P., Tereba A., Tyburski Ł., Fyalkowska K. Relationships between some biodiversity indicators and crown damage of *Pinus sylvestris* L. in natural old growth pine forests. *Sustainability*. 2021. V. 13. № 3. P. 1239. doi: [10.3390/su13031239](https://doi.org/10.3390/su13031239)
185. Saari S.K., Campbell C.D., Russell J., Alexander I.J., Anderson I.C. Pine microsatellite markers allow roots and ectomycorrhizas to be linked to individual trees. *New Phytologist*. 2005. № 165. P. 295–304. doi: [10.1111/j.1469-8137.2004.01213.x](https://doi.org/10.1111/j.1469-8137.2004.01213.x)
186. Kerpauskaite V., Danusevicius D., Kavaliauskas D., Fussi B., Konnert M., Baliuckas V., Augustaitis A. A methodological approach for assessment of the spatial genetic structure within Scots pine stands based on DNA markers. In: *Rural Development*. Kaunas: Aleksandras Stulginskis University, 2013. P. 324–331.
187. Kavaliauskas D. *Genetic Structure and Genetic Diversity of Scots Pine (Pinus sylvestris L.) Populations in Lithuania*: PhD Thesis. Kaunas: Aleksandras Stulginskis University, 2015. 146 p.
188. Danusevičius D., Kavaliauskas D., Fussi B. Optimum sample size for SSR-based estimation of representative allele frequencies and genetic diversity in Scots pine populations. *Baltic Forestry*. 2016a. V. 22. P. 194–202.
189. Danusevičius D., Buchovska J., Žulkus V., Daugnora L., Girininkas A. DNA markers reveal genetic associations among 11,000-year-old Scots pine (*Pinus sylvestris* L.) found in the Baltic Sea with the present-day gene pools in Lithuania. *Forests*. 2021. V. 12. № 3. P. 317. doi: [10.3390/f12030317](https://doi.org/10.3390/f12030317)
190. Łabiszak B., Zaborowska J., Wójkiewicz B., Wachowiak W. Molecular and paleoclimatic data uncover the impact of an ancient bottleneck on the demographic history and contemporary genetic structure of endangered *Pinus uliginosa*. *Journal of Systematics and Evolution*. 2021. V. 59. № 3. P. 596–610. doi: [10.1111/jse.12573\(2020\)](https://doi.org/10.1111/jse.12573(2020))
191. Sobierajska K., Wachowiak W., Zaborowska J., Łabiszak B., Wójkiewicz B., Sękiewicz M., Jasinska A.K., Sekiewicz K., Boratyńska K., Marcysiak K., Boratyński A. Genetic consequences of hybridization in relict isolated trees *Pinus sylvestris* and the *Pinus mugo* complex. *Forests*. 2020. V. 11. № 10. P. 1086. doi: [10.3390/f11101086](https://doi.org/10.3390/f11101086)
192. Verbylaitė R., Pliūra A., Lygis V., Suchockas V., Jankauskienė J., Labokas J. Genetic diversity and its spatial distribution in self-regenerating Norway spruce and Scots pine stands. *Forests*. 2017. V. 8. № 12. P. 470. doi: [10.3390/f8120470](https://doi.org/10.3390/f8120470)
193. Padutov A.V. Study of the genetic half-sib progeny structure of scots pine clones on forest-seed order II orchards. *Molekulyarnaya I Prikladnaya Genetika*. 2018. V. 25. P. 92–98. (In Russ.)
194. Torbik D., Bedrickaya T., Vlasova M., Sinelnikov I. Genetic diversity of natural populations of *Pinus sylvestris*. In: *Science - to the Forestry of the North*. Arkhangelsk: SevNIILH, 2019. P. 92–99. (In Russ.)
195. Hebda A., Wójkiewicz B., Wachowiak W. Genetic characteristics of Scots pine in Poland and reference populations based on nuclear and chloroplast microsatellite markers. *Silva Fennica*. 2017. V. 51. № 2. P. 1721. doi: [10.14214/sf.1721](https://doi.org/10.14214/sf.1721)

196. Pyataev A.S., Ibe A.A., Shilkina E.A. Genetic markers combination calculation in wood samples identification. In: *Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes* (SDM-2019). *CEUR Workshop Proceedings*, 2534. Berdsk: CEUR, 2019. P. 501–506.
197. Miao Y., Zhu X., Li Z., Jia F., Li W. Genetic evaluation of breeding resources of *Pinus sylvestris* var. *mongolica* from different improved generations. *Journal of Beijing Forestry University*. 2017. V. 39. № 12. P. 71–78. doi: [10.13332/j.1000-1522.20170194](https://doi.org/10.13332/j.1000-1522.20170194)

Received 12.05.2022.

Revised 14.06.2022.

Published 27.06.2022.

---



---

БИОИНФОРМАТИКА

---



---

## Функциональная характеристика EST-SSR маркеров сосны обыкновенной

Гуляева Е.Н.<sup>1</sup>, Тарелкина Т.В.<sup>2</sup>, Галибина Н.А.<sup>2</sup>

<sup>1</sup>*Отдел комплексных научных исследований, Карельский научный центр РАН, Петрозаводск, Россия*

<sup>2</sup>*Институт леса, Карельский научный центр РАН, Петрозаводск, Россия*

**Аннотация.** Полученные из библиотек EST локусы простых повторяющихся последовательностей (EST-SSR) являются важными инструментами для изучения генетического разнообразия, филогении, эволюции, сравнительной геномики, анализа QTL и ассоциативного картирования. Мы провели поиск в литературе известных EST-SSR, используемых для сосны обыкновенной (*Pinus sylvestris* L.) – одной из основных лесных пород в мире. 91 из 102 известных EST-SSR локусов для сосны обыкновенной, был вручную сопоставлен с эталонным геномом *Pinus taeda* L., а также с доступными генами *P. sylvestris*. Для 83 EST-SSR было определено местоположение в геноме и предполагаемая функция связанных с ними генов с помощью анализа консервативных доменов (CDD), функционального анализа известных гомологов Gene Ontology и анализа пути KEGG. Многие из маркеров располагались в нетранслируемых областях (преимущественно в 3'UTR), а также в кодирующих последовательностях генов сосны обыкновенной и сосны ладанной. Для восьми маркеров не удалось обнаружить гены ни у одного из видов. Из них семь маркеров были локализованы на участках скаффолдов *P. taeda*, не содержащих гены в текущем варианте сборки генома (v.1.0). Полученные результаты могут быть использованы в дальнейшем для популяционно-генетических исследований, а также изучения адаптивных признаков и картирования QTL *P. sylvestris* и других видов сосны.

**Ключевые слова:** EST-SSR, сосна обыкновенная, маркер-ген ассоциация, локализация маркера в гене, функциональная аннотация.