Conservation Genet Resour (2012) 4:231–234
DOI 10.1007/s12686-011-9513-5

TECHNICAL NOTE

Novel polymorphic nuclear microsatellite markers for *Pinus sylvestris* L.

F. Sebastiani · F. Pinzauti · S. T. Kujala · S. C. González-Martínez · G. G. Vendramin

Received: 18 August 2011 / Accepted: 28 August 2011 / Published online: 9 September 2011
© Springer Science+Business Media B.V. 2011

Abstract Scots pine (*Pinus sylvestris* L.) is one of the most widespread forest trees in the world, ranging from southern Mediterranean mountains to eastern Siberia. Ten polymorphic microsatellite loci were isolated from Scots pine cDNA sequences and were screened for variability in three natural populations. High levels of genetic variability were observed with effective number of alleles per locus ranging from 1.0 to 4.6 and average expected heterozygosity per population of 0.79. With only two exceptions, Hardy–Weinberg expectations were confirmed. All loci were in linkage equilibrium and there was little evidence for confounding null alleles. These new markers will be used to resolve population structure and gene flow patterns in this major Eurasian forest tree.

Keywords Scots pine · Microsatellites · Genetic diversity · Population genetics · Demography

Scots pine (*Pinus sylvestris* L.) is a wind-pollinated, wind-dispersed, predominantly outcrossing conifer (Kärkkäinen et al. 1996) and the most widely spread among pines, extending from south of Spain (38°N) to eastern Siberia (140°E). The taxon is best thought of as a complex of different evolutionary units (Moritz 1994) and several subspecies or varieties have been recognized (Farjon 1998). Neutral genetic markers provide powerful tools for identification of differentiated gene pools within widespread species. Many forest trees with large geographical distributions, such as Scots pine, show genetic signatures resulting from recent patterns of colonization from genetically differentiated glacial refugia (Hewitt 2004; see Naydenov et al. 2007; Pyhäjärvi et al. 2008 for Scots pine). When considering neutral nuclear markers, the genetic variation of Scots pine is high and accumulated mainly within populations (e.g., Dvornyk 2001). Soto et al. (2010), in a comparative study across six Iberian native pines, suggested that contemporary high levels of genetic diversity in Scots pine result from higher past effective population sizes boasted by its remarkable cold-tolerance.

Nuclear microsatellites (simple sequence repeats, nuSSRs) have proved to be useful to study phylogeographic and gene flow patterns in conifers (e.g., Bagnoli et al. 2009; González-Martínez et al. 2010) and are increasingly being used to infer demographic history in tree species (e.g., Daïnou et al. 2010). Apart from their interest per se, demographic inferences can also be used to obtain null hypotheses to test for signatures of selection in functional genetic markers (e.g., candidate genes, SNPs), which in turn provide insights on the molecular basis of adaptive evolution in forest trees. Unfortunately, only few nuSSRs are available for Scots pine (Kostia et al. 1995; Soranzo et al. 1998; González-Martínez et al. 2004; Liewlaksaneeyanawin et al. 2004) mainly because reliable nuSSRs are difficult to develop for conifer species due to their large genome size (estimated in 22,474 Mbp for Scots pine;

F. Sebastiani · F. Pinzauti · G. G. Vendramin (✉)
National Research Council, Plant Genetic Institute,
Via Madonna del Piano 10, 50019 Sesto Fiorentino, FI, Italy
e-mail: giovanni.vendramin@igv.cnr.it

S. T. Kujala
Department of Biology, University of Oulu, 90014 Oulu, Finland

S. T. Kujala
Biocenter Oulu, University of Oulu, 90014 Oulu, Finland

S. C. González-Martínez
Department of Forest Ecology and Genetics, Center of Forest Research, CIFOR-INIA, Carretera de A Coruña Km 7.5,
28040 Madrid, Spain

Plant DNA C-values Database, release 5.0, December 2010) and the extensive repetitive nature of their DNA (Scotti et al. 2002).

In order to provide a set of easily scorable nuSSRs, di-, tri- and tetra-nucleotide repeats were isolated from EST libraries and characterized for their level of polymorphism in three natural populations. A total of 6,641 *P. sylvestris* EST sequences were obtained from cDNA libraries developed within the EU-EVOLTREE project (libraries WZ0APSBA, WZ0APSBB, WZ0APSBC; http://www.evoltree.eu/index.php/elab-start/elab-wizard), which were assembled in 6,107 unique sequences using CodonCode software (CodonCode Corporation, USA). The sequences were screened for the presence of di-, tri- and tetra-nucleotide repeats using Msatfinder v.2.0 software (http://www.genomics.ceh.ac.uk/cgi-bin/msatfinder/msatfinder.cgi). A total of 55 primer pairs were selected for testing. Twelve primer pairs generated easily scorable amplification products of the expected size while the others showed no amplification, multi-banding patterns, or too pronounced stutters. To confirm marker usability and characterize the selected twelve SSR markers for variation and presence of null alleles, a total of 44 individuals from two Russian and one Finnish populations were analysed (Table 1). Ten out of the twelve selected nuSSRs displayed consistent and polymorphic patterns.

Amplification reactions were carried out following the PCR method by Schuelke (2000) in a final volume of 10 µl containing 30 ng of template DNA, 1× PCR buffer, 0.2 mM of each dNTP, 1 U of Go*Taq* polymerase (Promega, Madison, WI), 1.5 mM of $MgCl_2$, 0.2 µM of the reverse and the M13 universal primer (the latter labeled with FAM, NED, VIC or PET to the 5′ end), and 0.07 µM of the modified forward primer with the M13 primer sequence (18 bp) added at its 5′ end. The PCR profile was: denaturation at 94°C for 4 min followed by 35 cycles at 94°C (30 s), 55°C (30 s), 72°C (40 s), and a final step at 72°C for 8 min. Amplification reactions were carried out in a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA) and amplified products were run in an ABI 3130xl automatic sequencer (Applied Biosystems, Foster City, CA). Electropherograms were analyzed using GeneMapper v4.0 (Applied Biosystems, Foster City, CA). Primer sequences, repeat motifs, and GenBank accession numbers are shown in Table 1.

Standard genetic diversity parameters and departure from Hardy–Weinberg equilibrium (HWE) were estimated using GENALEX, v.6 (Peakall and Smouse 2006). Null allele frequencies were estimated using FreeNA (Chapuis and Estoup 2007), and linkage disequilibrium between pairs of loci, applying Bonferroni corrections, using FSTAT 2.9.3.2 (Goudet 2002).

Considering the three populations separately, the overall number of alleles per locus ranged from 1 to 6 (Table 1). Expected heterozygosity ranged from 0.095 to 0.785 in

Table 1 Primer sequences and characteristics of the 10 selected polymorphic nuclear microsatellite markers in *P. sylvestris*

Locus name	Primers sequence 5′–3′	Repeated motif	Allele size range (bp)	Total no. alleles	GenBank accession no.
psyl2	F: TTGCTTTTGCAGAACATTCG R: GTCCTGCAGGCAATCAAAAT	$(gct)_5$	199–211	4	HQ113935
psyl16	F: GCTCTGCCCATGCTATCACT R: TGATGCTACCCAATGAGGTG	$(at)_7$	202–210	5	HQ113936
psyl17	F: TGGTCTGCAAATCAATCGAA R: GGGTAGGAATGCAAGTTAGGC	$(ta)_7$	219–251	6	HQ113937
psyl18	F: ACTACCTGGCATTCGTCCTG R: GGATCTGGTCCATTTCGTGT	$(gca)_7$	297–306	3	HQ113938
psyl19	F: GGCTGTAATTGGCACAGGTT R: CGAGGTGGTACACAGCAACA	$(gct)_7$	315–324	3	HQ113939
psyl25	F: CAGCACGCGTTCTTTGTATC R: ACCGTTGCTCGTTGTCTTCT	$(gca)_5$	214–244	5	HQ113940
psyl36	F: TATCATCGAGAGCCCCAAAA R: GAAAGGCGAAAGCAAAAGTG	$(gtc)_7$	245–257	5	HQ113941
psyl42	F: CAACTTCAGCCTTGCAACAA R: CGACTTCATTTGGAACACCA	$(tc)_9$	171–179	4	HQ113942
psyl44	F: TCCAAGTTCGGTTCCTTGTC R: GACACGATGGATTCCCTGAT	$(cgg)_5$	166–175	4	HQ113943
psyl57	F: CCCCACATCTCTACAGTCCAA R: TGCTCTTGGATTTGTTGCTG	$(acc)_7$	187–202	6	HQ113944

KAR population, from 0.000 to 0.698 in LAD population, and from 0.000 to 0.772 in PUN population (Table 2). *Psyl17* and *psyl16* loci showed significant HWE departures in KAR and PUN populations, respectively, probably due to a moderate presence of null alleles. In fact, the estimated frequency of null alleles was very low (<5%, but generally <1%) with the only exceptions of *psyl16* (19.7%) in PUN population and *psyl17* (17.6%) in KAR population (Table 2). No significant linkage disequilibrium among loci was detected ($P < 0.05$).

Because of the high polymorphism, almost no deviation from HWE due to low null alleles frequency, and amenability to score in multiplex reactions, these markers are likely to be valuable tools for population genetic studies, in particular to elucidate fine-scale population structure and demographic history, and for parentage assignment in *P. sylvestris*.

Table 2 Genetic diversity parameters, deviation from HW equilibrium and frequency of null alleles for the three natural populations of *P. sylvestris*

Population	Locus	Sample size	N_a	N_e	H_o	H_e	F	Deviation from HW equilibrium	Null allele frequency
KAR	psyl2	10	2	1.220	0.200	0.180	−0.111	ns	0.000
	psyl16	10	4	3.509	0.400	0.521	0.232	ns	0.049
	psyl17	10	6	4.651	0.500	0.785	0.363	*	0.164
	psyl18	10	2	1.105	0.100	0.095	−0.053	ns	0.000
	psyl19	10	2	1.105	0.100	0.095	−0.053	ns	0.000
	psyl25	10	3	1.361	0.300	0.265	−0.132	ns	0.000
	psyl36	10	4	2.041	0.500	0.510	0.020	ns	0.000
	psyl42	10	4	3.077	0.700	0.675	−0.037	ns	0.006
	psyl44	10	2	1.105	0.100	0.095	−0.053	ns	0.000
	psyl57	10	5	2.174	0.500	0.540	0.074	ns	0.000
LAD	psyl2	10	4	1.527	0.400	0.345	−0.159	ns	0.000
	psyl16	9	4	2.348	0.556	0.574	0.032	ns	0.000
	psyl17	10	5	2.985	0.700	0.665	−0.053	ns	0.000
	psyl18	9	2	1.246	0.222	0.198	−0.125	ns	0.000
	psyl19	10	2	1.220	0.200	0.180	−0.111	ns	0.000
	psyl25	10	3	1.227	0.200	0.185	−0.081	ns	0.000
	psyl36	10	1	1.000	0.000	0.000	ND	ND	ND
	psyl42	9	4	3.306	1.000	0.698	−0.434	ns	0.000
	psyl44	9	1	1.000	0.000	0.000	ND	ND	ND
	psyl57	9	2	1.670	0.556	0.401	−0.385	ns	0.000
PUN	psyl2	24	4	1.538	0.333	0.350	0.047	ns	0.019
	psyl16	23	5	3.492	0.391	0.714	0.452	***	0.197
	psyl17	23	6	4.390	0.739	0.772	0.043	ns	0.023
	psyl18	24	1	1.000	0.000	0.000	ND	ND	ND
	psyl19	24	1	1.000	0.000	0.000	ND	ND	ND
	psyl25	24	2	1.043	0.042	0.041	−0.021	ns	0.000
	psyl36	24	4	1.186	0.167	0.157	−0.061	ns	0.000
	psyl42	24	4	3.545	0.625	0.718	0.129	ns	0.041
	psyl44	24	4	1.136	0.125	0.120	−0.043	ns	0.000
	psyl57	23	4	2.031	0.478	0.508	0.058	ns	0.000

KAR Kartashevskaja (Russia, 59°24′N, 30°04′E), *LAD* Northern Ladoga (Russia, 61°07′N, 29°59′E), *PUN* Punkaharju (Finland, 61°48′N, 29°19′E)

ns Not significant, *ND* not determined

N_a number of different alleles, N_e effective number of alleles = 1/(sum p_i^2), H_o and H_e observed and expected (1 − sum p_i^2) heterozygosities, *F* fixation index = $(H_e - H_o)/H_e = 1 - (H_o/H_e)$, where p_i is the frequency of the ith allele and sum p_i^2 is the sum of the squared allele frequencies

* $P < 0.05$, *** $P < 0.001$

Acknowledgments This research was funded by European Union—NOVELTREE (Novel tree breeding strategies, contract no. 211868) and EVOLTREE (Evolution of trees as drivers of terrestrial biodiversity, contract no. 016322) projects. We are grateful to Katri Kärkkäinen (Finnish Forest Research Institute, Finland) and Olga Lisitsyna (Department of Geology, University of Oulu, Finland) for providing needles of the three populations.

References

Bagnoli F, Vendramin GG, Buonamici A, Doulis AG, González-Martínez SC, La Porta N, Magri D, Sebastiani F, Raddi P, Fineschi S (2009) Is *Cupressus sempervirens* native in Italy? An answer from genetic and palaeobotanical data. Mol Ecol 18:2276–2286

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24:621–631

Daïnou K, Bizoux JP, Doucet JL, Mahy G, Hardy OJ, Heuertz M (2010) Forest refugia revisited: SSRs and cpDNA sequence support historical isolation in a wide-spread African tree with high colonization capacity, *Milicia excelsa* (*Moraceae*). Mol Ecol 19:4462–4477

Dvornyk V (2001) Genetic variability and differentiation of geographically marginal scots pine populations from Ukraine. Silvae Genet 50:64–69

Farjon A (1998) World checklist and bibliography of conifers. Royal Botanical Gardens at Kew, Richmond

González-Martínez SC, Robledo-Arnuncio JJ, Collada C, Díaz A, Williams CG, Alía R, Cervera MT (2004) Cross-amplification and sequence variation of microsatellite loci in Eurasian hard pines. Theor Appl Genet 109:103–111

González-Martínez SC, Dubreuil M, Riba M, Vendramin GG, Sebastiani F, Mayol M (2010) Spatial genetic structure of *Taxus baccata* L. in the western Mediterranean Basin: past and present limits to gene movement over a broad geographic scale. Mol Phyl Evol 55:805–815

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). URL: http//www2.unil.ch/popgen/softwares/fstat.htm

Hewitt GM (2004) Genetic consequences of climatic changes in the quaternary. Philos Trans R Soc Lond B Biol Sci 359:183–195

Kärkkäinen K, Koski V, Savolainen O (1996) Geographical variation in the inbreeding depression of Scots pine. Evolution 50:111–119

Kostia S, Varvio SL, Vakkari P, Pulkkinen P (1995) Microsatellite sequences in a conifer, *Pinus sylvestris*. Genome 38:1244–1248

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Moritz C (1994) Defining evolutionarily significant units for conservation. TREE 9:373–375

Naydenov K, Senneville S, Beaulieu J, Tremblay F, Bousquet J (2007) Glacial vicariance in Eurasia: mitochondrial DNA evidence from Scots pine for a complex heritage involving genetically distinct refugia at mid-northern latitudes and in Asia Minor. BMC Evol Biol 7:233. doi:10.1186/1471-2148-7-233

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Pyhäjärvi T, Salmela MJ, Savolainen O (2008) Colonization routes of *Pinus sylvestris* inferred from distribution of mitochondrial DNA variation. TGG 4:247–254

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234

Scotti I, Magni F, Paglia GP, Morgante M (2002) Trinucleotide microsatellites in Norway spruce (*Picea abies*): their features and the development of molecular markers. Theor Appl Genet 106:40–50

Soranzo N, Provan J, Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Mol Ecol 7:1260–1261

Soto A, Robledo-Arnuncio JJ, González-Martínez SC, Smouse PE, Alía R (2010) Climatic niche and neutral genetic diversity of the six Iberian pine species: a retrospective and prospective view. Mol Ecol 19:1396–1409