Association mapping for resin yield in *Pinus roxburghii* Sarg. using microsatellite markers

By A. RAWAT^{*}, S. BARTH WAL and H. S. GINWAL

Division of Genetics and Tree Propagation, Forest Research Institute,
P.O.I.P.E. Kaulagarh Road, Dehradun-248195, Uttarakhand, India

(Received 27th May 2014)

Abstract

Association mapping is a method for detection of gene effects based on linkage disequilibrium (LD) that complements QTL analysis in the development of tools for molecular plant breeding. A total of 240 genotypes of *Pinus roxburghii* (Himalayan Chir Pine) from a natural population in Chakrata division (Tiunee range), Uttarakhand (India) were evaluated for resin yield. Based on the phenotypic data and stable resin production in consecutive years, 53 genotypes were selected after excluding the individuals with similar resin production. The selected 53 individuals were best representatives of the variation in resin yield in Chakrata population which varied between 0.25 and 8.0 kg/tree/year and were used for genotyping and association analysis using SSR markers. Out of 80 primers initially screened, a total of 19 polymorphic SSRs (11 cpSSR and 8 nSSR) were used in the study. Model based clustering using 19 polymorphic SSR markers identified five subpopulations among these genotypes. LD was evaluated using the entire population. The squared allele frequency correlation, r^2 was estimated for each pair of SSR loci. The comparison wise significance (p -values) of SSR marker pairs was determined by performing 100,000 permutations. The genetic divergence ranged from 50 to 100%. The UPGMA based hierarchical clustering grouped the genotypes in accordance with their resin yield. Model based clustering suggested the existence of five subpopulations in the sample. However, the distribution of *P. roxburghii* genotypes into five subpopulations had no correlation with their resin yield thus ruling out the possibility of any ancestral relationship among the genotypes with similar resin yield. AMOVA suggested that the variation among *P. roxburghii* genotypes at the molecular level was related with the variation in resin yield and not their site of collection thus highlighting the genetic basis of the trait. LD based association analysis revealed two chloroplast SSRs Pt71936 and Pt87268 and one nuclear SSR pm09a to be in significant association with resin yield. The two associated chloroplast SSRs showed significant LD ($p < 0.01$). One of the chloroplast SSR Pt87268 showing association with resin yield was also found to be in significant LD with the nuclear SSR pm07, further showing the probability of this marker also to be associated with resin yield.

Key words: *Pinus roxburghii*, SSR, association mapping, Linkage disequilibrium (LD).

Introduction

Pine forests are of great economic importance as a source of wood, paper, resins, charcoal, food and ornamentals (LE MAITRE, 1998). Resin is a commercially important product, having huge export potential. *Pinus roxburghii* (Sarg.) commonly called as long leaf pine or 'Chir pine' yields the highest amount of oleoresin in India (Coppen and Hone, 1995). It is found in the lower Himalayan region between latitudes 26°N and 36°N and longitudes 71°E and 93°E (GHILDYAL *et al.*, 2009). In India, it covers approximately 6, 77,813 ha area in the states of Himachal Pradesh, Jammu and Kashmir and Uttarakhand out of which Uttarakhand alone contains a major portion of 4,12,000 ha of chir pine forests (SINGH and KUMAR, 2004). Oleoresins from pines are composed of two components, volatile turpentine oil and the remaining solid transparent material called as Rosin. Turpentine oil is mainly used as a solvent in industries and has medicinal qualities as well. Rosin is used in paper manufacturing, paper sizing, chemicals and pharmaceuticals, synthesis of ester gums, synthetic resins, paint, varnishes, printing inks, soap, rubber, surface coatings, floor coverings, adhesives, plastics, etc. India stood at sixth position among the top ten-resin production countries across the world (COPPEN and HONE, 1995). As per FAO reports (COPPEN and HONE, 1995), crude resin production in India has fallen steadily since 1975–76. As a result of the loss of substantial indigenous production of crude resin and the demands of Indian industry for naval stores products, India became a net importer of both rosin and turpentine. Large scale exploitation using old and outdated methods of resin tapping have caused severe damage to the pine trees. There is a need to identify pine trees with high resin yield to avoid the damage to naturally occurring pine forests.

Diverse habitat of chir pine in different geographical regions of Himalayas and Shivalik range supports the existence of natural variation. Forest Research Institute, Dehradun conducted studies on resin yield in nine different provenances during 1926–1927 and found considerable variation among them (SHARMA *et al.*, 2006). This was subsequently confirmed by KEDHARNATH (1971) who reported significant variation in resin production among nine provenances of chir pine. The resin yield varied between 4 and 7 kg per tree. Significant genetic variation in resin production among the *P. roxburghii* genotypes was reported by SHARMA *et al.* (2001), which

^{*}) Author for correspondence: ANITA RAWAT. Ph: +91-9719408989
Fax: +91-135-2756865. E-Mail: anitasrawat@gmail.com

indicate the possibility of exploiting this variation for development of high resin yielding genotypes.

Resin yield is a phenotypic trait and it can be scored only in mature trees which are 15–20 years or have attained a diameter greater than 20–25 cm (COPPEN and HONE, 1995). Therefore, identification of plus trees for high resin yield through quantitative estimation of the resin yield is highly time consuming. Pines have long gestation periods with vegetative phase extending over hundred years and because of which multiple generations are not readily obtained and traditional approaches of tree improvement involving the identification of mature trees with desirable phenotypes, followed by their incorporation into breeding programs are rather slow processes. However, if it is possible to identify the high resin yielding genotypes at the nursery stage, then plantations can be raised solely for the purpose of resin production. This will reduce the harm to the naturally occurring forests of chir pine as well as the time period and cost required for the quantitative detection.

The identification of trait specific molecular markers has been successfully attempted in many agricultural crops through linkage mapping (XUELIAN *et al.*, 2014; YANG *et al.*, 2013; REN *et al.*, 2009). However, identification of such trait specific markers is difficult and tedious in tree species due to lack of experimental populations attributed to longer gestation periods. An alternate strategy to identify trait specific marker is through association mapping, which is based on the concept of linkage disequilibrium (LD) (ZONDERVAN and CARDON, 2004). Originally developed for human genetics, this approach exploits the candidate gene sequence variation and relies on the existence of LD (non-random association between alleles at the linked loci) between detectable sequence polymorphism. The advantage of this approach over anonymous markers is that once a major effect gene is identified and validated, Marker Assisted Selection can be practiced directly on the gene. LD mapping can be applied to wild, unstructured and un-pedigreed (RISCH, 2000) populations.

Population based-association study is advantageous over traditional QTL-mapping in bi-parental crosses due to availability of broader genetic variations with wider background for marker-trait correlations (ABDURAKHMONOV and ABDUKARIMOV, 2008). In association mapping, unaccounted subdivisions in the sample, referred to as population structure (PRITCHARD *et al.*, 2000a) may result in false positives. The presence of related subgroups in the sample could create covariances among individuals that, if not included explicitly in the model, generate bias in the estimates of allele effects (KENNEDY *et al.*, 1992). Understanding the population structure and linkage disequilibrium in an association panel can effectively avoid spurious associations and improve the accuracy in association mapping (ZHAO *et al.*, 2014). A Bayesian approach for inference of population structure based on unlinked markers was implemented in the software Structure (PRITCHARD *et al.*, 2000a). This program assigns individuals to subpopulations, and that assignment is considered in testing associations of markers with dichotomous traits (PRITCHARD *et al.*, 2000b).

Work on identification of trait specific molecular markers is essential for future tree improvement and conservation of Pines. Development of markers for resin production is needed for better utilization and conservation of an important commercial conifer species. Keeping in view, the above facts, the study was initiated with a broad objective of identifying DNA based markers associated with the resin production by studying polymorphism in phenotypically varying population of Himalayn Chir Pine (*P. roxburghii*).

Materials and Methods

Plant material and field experiment

P. roxburghii trees were evaluated for resin yield at three sites viz. Chakrata division (Uttarakhand), Nahan division (Himachal Pradesh) and Udhampur division (Jammu and Kashmir) in a study conducted by Forest Research Institute, Dehradun, India (NEGI and MALIK, 2009). The data revealed maximum variation in resin yield in Chakrata division (Uttarakhand). As per the study, correlation of tree diameter, altitude and site quality with resin yield was found to be not significant in Chakrata. Based on these two observations, Chakrata site (Uttarakhand) was selected for carrying out molecular characterization of pine genotypes for the identification of markers associated with resin yield. A total of 240 genotypes of chir pine from Chakrata division (Tiunee range), Uttarakhand were evaluated for resin yield. The experiment was laid in the natural forest of chir pine at an altitude ranging from 1000 to 1500 m above the mean sea level covering southern aspect (A1) and northern aspect (A2), each with two sites having different site qualities (S1 and S2). For each site quality, three plots (0.25 ha each) were selected at random, comprising total area of 0.75 ha. The plots were considered as replications. For each replication, entire area of 0.25 ha was surveyed for the collection of data. Since, the genotypes showing maximum variation in the trait are highly recommended for conducting association studies (ZHAO *et al.*, 2014) so the individuals with similar resin yield were excluded. Fifty-three genotypes that were best representatives of the variation in resin yield were selected for genotyping and association mapping. Geographical details along with the morphological data of the selected trees in terms of diameter, height and annual resin yield is tabulated in *Table 1*. Young needles or sapwood (in case needles were not available due to extreme height of trees) samples were collected from the site and stored at -80°C .

Estimation of resin yield

Rill method of resin tapping was used keeping the blaze area uniform (45×20) cm^2 for all the trees. Month wise resin yield was recorded from the month of June till November and finally the annual resin yield was determined for all the trees (*Table 1*). The resin yield ranged between 0.25 and 8.0 kg/tree/year with an average yield of 3 kg/tree/year. The individuals with resin yield less than 3 kg/year were grouped as low resin yielders while those with resin yield more than 3 kg/year were grouped as high resin yielders.

Table 1. – Geographical details and resin yield data of various samples of *P. roxburghii*.

S. No.	Sample	Location	Longitude (°E)	Latitude (°N)	Altitude (m amsl)*	Diameter (m)	Height (m)	Resin Yield (kg/year)
1	A-01	Southern Aspect, Site Quality 01; Chatra	77°56' 16.9"E	30°58' 02.3"N	1301	1.3	36.0	2.20
2	A-10	Southern Aspect, Site Quality 01; Chatra	77°56' 18.4"E	30°58' 02.2"N	1298	1.6	32.0	3.30
3	A-12	Southern Aspect, Site Quality 01; Chatra	77°56' 42.9"E	30°56' 54.9"N	1437	2.5	37.0	8.00
4	A-13	Southern Aspect, Site Quality 01; Chatra	77°56' 58.0"E	30°51' 26.2"N	1203	1.8	24.8	3.35
5	A-19	Southern Aspect, Site Quality 01; Chatra	77°56' 17.4"E	30°58' 02.7"N	1307	1.5	35.0	5.20
6	A-02	Southern Aspect, Site Quality 01; Chatra	77°56' 42.9"E	30°56' 53.3"N	1423	2.3	37.0	6.50
7	A-24	Southern Aspect, Site Quality 01; Chatra	77°56' 20.3"E	30°58' 01.0"N	1292	1.1	23.0	2.80
8	A-25	Southern Aspect, Site Quality 01; Chatra	77°56' 20.3"E	30°58' 01.6"N	1292	1.5	24.5	2.90
9	A-28	Southern Aspect, Site Quality 01; Chatra	77°56' 43.3"E	30°56' 52.9"N	1421	1.1	29.0	6.20
10	A-03	Southern Aspect, Site Quality 01; Chatra	77°56' 43.0"E	30°56' 53.3"N	1423	1.7	36.8	0.90
11	A-06	Southern Aspect, Site Quality 01; Chatra	77°56' 43.2"E	30°56' 53.6"N	1425	1.4	34.0	1.40
12	A-07	Southern Aspect, Site Quality 01; Chatra	77°56' 18.0"E	30°68' 03.3"N	1293	2.1	28.9	4.10
13	A-09	Southern Aspect, Site Quality 01; Chatra	77°56' 42.5"E	30°56' 54.4"N	1432	2.8	33.4	5.80
14	B-10	Southern Aspect, Site Quality 02; Chatra	77°56' 47.4"E	30°57' 48.2"N	1454	2.6	27.0	2.00
15	B-12	Southern Aspect, Site Quality 02; Chatra	77°56' 48.4"E	30°56' 49.3"N	1451	1.1	28.4	4.60
16	B-13	Southern Aspect, Site Quality 02; Chatra	77°56' 46.3"E	30°56' 42.4"N	1443	1.8	28.0	3.60
17	B-14	Southern Aspect, Site Quality 02; Chatra	77°56' 47.6"E	30°56' 49.3"N	1442	2.5	28.0	0.25
18	B-18	Southern Aspect, Site Quality 02; Chatra	77°56' 28.4"E	30°57' 53.7"N	1302	2.1	45.0	3.35
19	B-19	Southern Aspect, Site Quality 02; Chatra	77°56' 22.1"E	30°57' 57.6"N	1315	1.4	37.0	2.70
20	B-02	Southern Aspect, Site Quality 02; Chatra	77°56' 46.4"E	30°57' 47.5"N	1450	2.5	29.2	2.50
21	B-24	Southern Aspect, Site Quality 02; Chatra	77°56' 46.6"E	30°56' 47.1"N	1438	2.5	28.0	5.70
22	B-25	Southern Aspect, Site Quality 02; Chatra	77°56' 46.8"E	30°56' 42.2"N	1437	1.4	34.0	6.40
23	B-26	Southern Aspect, Site Quality 02; Chatra	77°56' 21.7"E	30°57' 58.3"N	1306	1.3	28.0	2.80
24	B-03	Southern Aspect, Site Quality 02; Chatra	77°56' 46.7"E	30°56' 47.4"N	1450	2.6	29.0	4.90
25	B-04	Southern Aspect, Site Quality 02; Chatra	77°56' 22.7"E	30°57' 58.0"N	1316	2.3	30.0	0.80
26	B-06	Southern Aspect, Site Quality 02; Chatra	77°56' 47.0"E	30°56' 47.3"N	1452	1.4	31.0	4.50
27	B-07	Southern Aspect, Site Quality 02; Chatra	77°56' 21.6"E	30°57' 58.0"N	1324	2.1	30.0	4.30
28	B-08	Southern Aspect, Site Quality 02; Chatra	77°56' 47.3"E	30°57' 48.5"N	1462	1.5	27.0	2.25
29	B-09	Southern Aspect, Site Quality 02; Chatra	77°56' 47.0"E	30°56' 48.3"N	1454	2.2	27.0	4.70
30	C-01	Northern Aspect, Site Quality 01; Chatra	77°56' 36.5"E	30°57' 09.8"N	1379	1.5	35.0	4.20
31	C-10	Northern Aspect, Site Quality 01; Chatra	77°56' 48.6"E	30°56' 47.9"N	1406	1.5	34.0	2.25
32	C-12	Northern Aspect, Site Quality 01; Chatra	77°56' 37.2"E	30°57' 10.1"N	1364	2.9	36.0	2.25
33	C-15	Northern Aspect, Site Quality 01; Chatra	77°56' 47.7"E	30°56' 49.3"N	1404	1.8	35.0	2.90
34	C-20	Northern Aspect, Site Quality 01; Chatra	77°56' 46.6"E	30°56' 50.3"N	1401	1.5	32.0	2.90
35	C-03	Northern Aspect, Site Quality 01; Chatra	77°56' 47.8"E	30°56' 46.1"N	1404	2.5	40.0	5.00
36	C-04	Northern Aspect, Site Quality 01; Chatra	77°56' 37.2"E	30°57' 09.8"N	1379	1.9	42.8	2.70
37	C-07	Northern Aspect, Site Quality 01; Chatra	77°56' 48.4"E	30°56' 47.7"N	1404	2.5	38.0	4.50
38	C-08	Northern Aspect, Site Quality 01; Chatra	77°56' 38.8"E	30°57' 11.2"N	1360	1.2	36.0	4.00
39	C-09	Northern Aspect, Site Quality 01; Chatra	77°56' 49.5"E	30°56' 47.5"N	1405	2.0	39.0	5.60
40	D-11	Northern Aspect, Site Quality 02; Chatra	77°56' 30.4"E	30°58' 04.9"N	1171	1.6	29.5	2.60
41	D-14	Northern Aspect, Site Quality 02; Chatra	77°56' 31.0"E	30°58' 05.1"N	1171	2.5	36.6	2.30
42	D-24	Northern Aspect, Site Quality 02; Chatra	77°56' 37.0"E	30°58' 06.3"N	1160	1.4	29.2	1.75
43	D-26	Northern Aspect, Site Quality 02; Chatra	77°56' 35.5"E	30°57' 03.2"N	1352	2.2	34.2	1.80
44	D-27	Northern Aspect, Site Quality 02; Chatra	77°56' 38.2"E	30°57' 04.4"N	1360	1.3	36.0	0.90
45	D-29	Northern Aspect, Site Quality 02; Chatra	77°56' 38.0"E	30°57' 06.2"N	1353	1.3	27.9	2.10
46	D-30	Northern Aspect, Site Quality 02; Chatra	77°56' 38.3"E	30°57' 06.5"N	1355	2.2	40.2	1.60
47	D-31	Northern Aspect, Site Quality 02; Chatra	77°56' 43.7"E	30°56' 53.9"N	1396	1.5	35.0	4.20
48	D-33	Northern Aspect, Site Quality 02; Chatra	77°56' 42.4"E	30°56' 53.5"N	1393	2.5	40.0	5.00
49	D-37	Northern Aspect, Site Quality 02; Chatra	77°56' 42.4"E	30°56' 54.1"N	1432	2.5	38.0	4.50
50	D-38	Northern Aspect, Site Quality 02; Chatra	77°56' 43.2"E	30°56' 53.4"N	1423	1.2	36.0	4.00
51	D-39	Northern Aspect, Site Quality 02; Chatra	77°56' 42.2"E	30°56' 54.3"N	1421	2.0	39.0	5.60
52	D-07	Northern Aspect, Site Quality 02; Chatra	77°56' 38.8"E	30°57' 04.5"N	1370	1.6	28.3	1.30
53	D-08	Northern Aspect, Site Quality 02; Chatra	77°56' 41.8"E	30°57' 07.2"N	1365	1.8	29.2	1.10

* amsl-above mean sea level. Blaze area: 45 × 20 cm².

DNA extraction and quantification

DNA was extracted from young needles using a combination of the methods described by STANGE *et al.* (1998) and DOYLE and DOYLE (1990) and from the sapwood following a combination of the protocols given by ASIF and CANNON (2005) and DOYLE and DOYLE (1990). The quality of DNA was tested on 0.8% agarose gel and the DNA concentration was quantified using BioPhotometer

(Eppendorf 6131, Germany). DNA samples were diluted to the required concentration (15 ng/μl) for polymerase chain reaction (PCR) amplification.

SSR analysis

A total of 80 SSR markers (47 nuclear SSRs and 33 chloroplast SSRs) from different species of pines (nuclear SSRs were from *P. resinosa*, *P. taeda*,

Table 2. – SSRs from different *Pinus* species used to amplify *P. roxburghii* microsatellite regions.

S. No.	Locus		Primer Sequence (5'-3')	Ta (°C)	Size range (bp)	Repeat motif	Reference
1	pdms 011	Forward	TGCTCAACTATATTGCTACCAAC	55	141-147	(GT) ₁₁	Watanabe et al., 2006
		Reverse	CGTCAATGATTCAAAATCC				
2	pdms 221	Forward	GAGAGTTGTATGACGGAATAC	55	151-201	(GA) ₅ G ₂ (GA) ₅	Watanabe et al., 2006
		Reverse	CCCACACAAAAGTGACTTC				
3	pm 05	Forward	GAGTCTAATTGCAACCCCA	60	98-120	(TG) ₅	Nurtjahjarningsih et al., 2005
		Reverse	TGGAGATCTACCACTTTTC				
4	pm 07	Forward	GAATCTAAGCATATGAAATGAG	55	190-230	(AC) ₃ (AT) ₄	Nurtjahjarningsih et al., 2005
		Reverse	CTTGTAAAGTACAGTAAAG				
5	PtTX 3025	Forward	CACGCTGTATAATAACAATCTA	55	273-290	(CAA) ₁₀	Elsik et al., 2000
		Reverse	TTCTATATTCGCTTTTAGTTTC				
6	RPTst6	Forward	AGGATTCCAACAGCATCACC	60	73-77	TGC	Chagne et al., 2004
		Reverse	CTGACATGAAAGCGCAGTGT				
7	RPTst9	Forward	CCAGACAACCAAAAGGAGG	60	314-357	AGC	Chagne et al., 2004
		Reverse	GCCTGCTATCGAATCCAGAA				
8	Pt1254	Forward	CAATTGGAATGAGAAGATAGG	57.8	146-161	(T) ₁₇	Vendramin et al., 1996
		Reverse	TGCGTTGCACTTCGTTATAG				
9	Pt71936	Forward	TTCATTGGAAATACACTAGCCC	55	141-143	(T) ₆	Vendramin et al., 1996
		Reverse	AAAACCGTACATGAGATTC				
10	Pt87268	Forward	GCCAGGGAAATCGTAGG	55	138-182	(T) ₁₄	Vendramin et al., 1996
		Reverse	AGACGATTAGACATCCAACCC				
11	pm 09a	Forward	CCCTCTCATTCGATATGCAC	58.5	108-115	(AT) ₅ (GT) ₁₃ (AT) ₂	Nurtjahjarningsih et al., 2005
		Reverse	ATTAAGGTTATATGGGGCT				
12	PCP26106	Forward	AATCCGACAAAAAGATTCGG	58.5	144-155	(A) ₁₄	Provan et al., 1998
		Reverse	GCTCCATTTCACGTGGTTG				
13	PCP30277	Forward	TGTTGATGCTAGCGGAAG	58.5	124-182	(A) ₁₄ (G) ₁₀	Provan et al., 1998
		Reverse	ATGAATGAATCACTTCCCCC				
14	Pt30204	Forward	TCATAGCGGAAGATCTCTTT	56	143-145	(A) ₁₂ (G) ₁₀	Vendramin et al., 1996
		Reverse	CGGATTGATCTCAACATACC				
15	Pt45002	Forward	AAGTTGGATTTTACCCAGGTG	58	154-165	(T) ₁₅	Vendramin et al., 1996
		Reverse	GAACAAGAGGATTTTCTCATACA				
16	Pt79951	Forward	CTTTTGTTTTCAACAATTGCA	58.5	124-126	(T) ₁₇	Vendramin et al., 1996
		Reverse	ACATCTATCTCCATATCGGC				
17	PCP9434	Forward	AAACTGACGTAGATGCCATGG	58.5	114-117	(A) ₁₁	Provan et al., 1998
		Reverse	GCGGTATGAGGGGAAGAGC				
18	Pt36480	Forward	TTTTGGCTTACAAAATAAAGAGG	52	137-145	(T) ₁₇	Vendramin et al., 1996
		Reverse	AAATTCCTAAAGAAGAGAGCA				
19	PCP41131	Forward	AAAGCATTTCCAGTTGGGG	58.5	137-142	(T) ₁₇	Provan et al., 1998
		Reverse	GGTCAGGATTCATGTTCTTCC				

P. merkussi and *P. densiflora* whereas, chloroplast SSRs were from *P. thunbergii* and *P. sylvestris*) were screened for amplification in *P. roxburghii*. Forty eight SSRs showed successful amplification but only 19 SSR markers (8 nuclear SSRs and 11 chloroplast SSRs) were found to be polymorphic (Table 2). The polymorphic SSRs were screened on 275 adult trees of *P. roxburghii* from a single large population in its natural range of distribution. PCR was performed in a 15 µl reaction volume (VENDRAMIN *et al.*, 1996) containing 15 ng of template DNA, 1X *Taq* buffer, 3.0 mM MgCl₂, 0.2 mM dNTPs, 0.2 µM of each primer and 0.06U of *Taq* DNA

polymerase (Bangalore Genei Pvt. Ltd., India). PCR amplification was carried out at 5 min. at 95°C followed by 30 cycles of 1 min. at 94°C, 1 min. at 55°C to 60°C (as per the annealing temperature of the primer) and 1 min. at 72°C and a final extension of 8 min. at 72°C. Amplified products were electrophoresed on 3% (w/v) metaphor agarose: agarose (3:1) gel with 1X TBE buffer and stained with ethidium bromide (0.5 µg/ml) (Fig. 1). DNA fragments were visualized under UV light and documented with the gel documentation imaging system (GelDoc-It System, UVP Ltd.). The primers which were not resolved on metaphor-agarose gel were then sepa-

Lane M: 100 bp DNA ladder; Lanes 1-53: *P. roxburghii* genotypesFigure 1. – Gel image showing amplification of SSR primer pm05 in different genotypes of *P. roxburghii* on 3% metaphor-agarose gel.

rated on 8% (w/v) polyacrylamide gel casted in 'MEGA-GEL High Throughput Vertical Unit' (model C-DASG-400-50) marketed by C.B.S Scientific Co. (Del Mar, CA, USA) with 1X TBE buffer.

Scoring of data

The molecular size of the different fragments of the DNA ladders (GeneRuler™ 100 bp ladder and Fermentas O' GeneRuler™ Ultra low range DNA ladder) were plotted (scatter plot) against the distance travelled by each fragment of the ladder and with the help of MS Excel 'Chart Wizard', a trend line was applied to the scatter plot. Polynomial curves with powers from two to four were used to produce the closest fit to the marker curve. Polynomial coefficients for calculating the formula were derived from the regression equation of the trend line displayed on the same chart (LORENZ *et al.*, 1997). The distance travelled by each amplified DNA fragment was used to calculate their molecular weight by extrapolating the graph using the regression equation of the trend line. For accuracy, the distance migrated by each fragment of the DNA ladder was used to back calculate their molecular weight. Since chloroplast genome does not genetically recombine, or exist in heterozygous state, so the first homozygous allele (heaviest fragment) was scored as AA, second homozygous allele was scored as BB and so on and so forth. For nuclear SSRs, the gels were scored in a specified data format. Presence of single homozygous allele (heaviest fragment) was scored as AA; second homozygous allele was scored as BB and so on and so forth. Presence of heterozygous alleles (allele A and B or B and C or A and C)

were scored as AB, BC and AC respectively. Following this method of scoring, all the 53 genotypes were scored.

Statistical analysis

Polymorphism and primer informativeness

The genetic diversity parameters like per cent polymorphism, total number of bands amplified per primer and number of polymorphic bands were calculated. Genotypic data obtained for different markers was used for assessing the discriminatory power of primers and determining the utility of each marker system by evaluating the parameters: Polymorphism Information Content (PIC) (ROLDAN-RUIZ *et al.*, 2000), Marker index (MI) (POWELL *et al.*, 1996) and Resolving Power (RP) (Prevost and Wilkinson).

Cluster analysis

Genetic dissimilarity was calculated based on Jaccard's dissimilarity index using the software DARwin ver 5.0.158 (PERRIER and JACQUEMOUD-COLLET, 2006), where "0" and "1" were standardized as the least and maximum dissimilarity respectively. The dissimilarity matrix was used for tree construction following hierarchical clustering method using UPGMA algorithm implemented in the software DARwin ver 5.0.158. Confidence limits of different clades were tested by bootstrapping 1000 times to assess the repetitiveness of genotype clustering (FELSENSTEIN, 1985).

Population structure analysis

For the analysis of population structure, a model-based (Bayesian) cluster analysis was performed. This

Table 3. – SSR marker polymorphism in *P. roxburghii* genotypes.

SSR Marker	Total no. of alleles	Polymorphic alleles	% polymorphism	Polymorphism Information Content (PIC)	Marker Index (MI)	Resolving Power (RP)
pdms011	2	1	50	0.153	0.077	0.378
pdms221	2	1	50	0.115	0.057	0.264
pm05	2	2	100	0.376	0.752	1.094
pm07	2	1	50	0.211	0.106	0.604
PtTX3025	2	2	100	0.497	0.994	1.850
RPtest 6	2	1	50	0.164	0.082	0.416
RPtest 9	2	1	50	0.100	0.050	0.226
Pt1254	2	2	100	0.414	0.827	1.170
Pt71936	2	2	100	0.499	0.998	1.926
Pt87268	3	3	100	0.324	0.971	1.548
pm09a	3	3	100	0.339	1.018	1.622
PCP26106	3	3	100	0.348	1.045	1.434
PCP30277	2	2	100	0.391	0.782	0.340
Pt30204	4	4	100	0.292	1.168	1.850
Pt45002	5	4	80	0.262	0.838	1.888
Pt 79951	2	2	100	0.459	0.917	1.510
PCP41131	2	2	100	0.486	0.971	1.660
Pt36480	2	2	100	0.303	0.607	0.792
PCP9434	2	2	100	0.473	0.946	1.546
Total	46	40	-	-	-	-
Minimum	2	1	50	0.100	0.050	0.226
Maximum	5	4	100	0.499	1.168	1.926
Average	2.42	2.10	85.79	0.327	0.695	1.164

analysis was implemented in the software STRUCTURE ver 2.2 (PRITCHARD *et al.*, 2000a and 2000b) which identify subgroups of accessions with distinct allele frequencies within the germplasm. STRUCTURE computes a Q matrix defined as an $n \times p$ population structure incidence matrix where n is the number of individuals assayed and p is the number of sub-populations assumed; Q is inferred from Pritchard's STRUCTURE estimates with p (Pritchard's K) sub-populations. The model based cluster analysis was used to test the hypothesis of one to ten sub-populations ($K=1$ to $K=10$) assuming admixture and correlated allele frequencies in

different subpopulations. 100,000 iterations and a burn-in period of 100,000 were carried out for each run. Ten independent STRUCTURE runs were performed separately for each K. The value of K was detected by an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK) (EVANNO *et al.*, 2005)

$$\Delta K = m (|L(K+1) - 2L(K) + L(K-1)|) / s[L(K)]$$

Where, $L(K)$ is $\ln P(D)$, the posterior probability of the data for a given K, $\Pr(X|K)$ in STRUCTURE output, $s[L(K)]$ is the standard deviation of $L(K)$, and m is mean

Figure 2. – Dendrogram showing genetic relationship among *P. roxburghii* genotypes varying in resin yield using SSR markers.

in the parenthesis. ΔK shows a clear peak at the true value of K.

Analysis of Molecular Variance (AMOVA)

The population genetic structure was inferred by an analysis of molecular variance framework (AMOVA) according to EXCOFFIER *et al.* (1992) using the Arlequin software version 3.11 (EXCOFFIER *et al.*, 2005). The type of hierarchical AMOVA implemented here was with genotypic data, one group of populations and number within individual level. This technique treats genetic distances as deviations from a group mean position, and uses squared deviations as variances. The total sum of squares of genetic distances were partitioned into components that represent the within population and among population mean squares

Linkage disequilibrium

The software TASSEL ver 2.1 (BRADBURY *et al.*, 2007) was used to measure the extent of LD as squared allele frequency correlations estimates (r^2 , WEIR, 1996) and to measure significance of r^2 for each pair of loci. For multiple alleles, a weighted average of r^2 between each locus pair was calculated (FARNIR *et al.*, 2000). Only alleles with frequencies equal or greater than 0.05 were considered for LD calculations (THORNSBERRY *et al.*, 2001). Significance of LD for SSR pairs was determined by 100,000 permutations for each pair (WEIR, 1996). The number of marker pairs with LD probability values less than threshold values of 0.01 and 0.001 were counted.

Linkage disequilibrium based association analysis

The general linear model (GLM) implemented in the software TASSEL ver 2.1 (Trait analysis by Association, Evolution and Linkage) was used for association analysis taking into consideration the Q matrix produced by

the software STRUCTURE ver 2.2 for five sub-populations existing in the sample. To select appropriate significance thresholds for association analysis, probability values of association between single alleles from nineteen SSR marker loci and resin yield based on the GLM-Q association model were permuted 1,000 times (CHURCHILL and DOERGE, 1994). A polymorphic site was deemed to have a significant association if the p-value was below the 5% empirically derived value. Only alleles with frequency more than 5% were considered in the analysis. Single alleles were tested for association.

Results

Information content of SSR markers

A total of nineteen out of eighty initially screened SSR markers were selected on the basis of polymorphism (Table 2). PCR amplification of the *P. roxburghii* genotypes using nineteen SSR markers produced a total of 46 bands, out of which 40 were polymorphic. The total number of polymorphic bands amplified per marker varied from 1 to 4 with an average of 2.10 per marker (Table 3). Genetic divergence in terms of per cent polymorphism ranged from 50 to 100% with an average of 85.79% per marker. The PIC ranged from 0.100 (RPtest 9) to 0.499 (Pt 71936) with an average of 0.327 per marker. There was a strong correlation between polymorphism and PIC ($r^2=0.864$). The MI ranged from a minimum of 0.050 (RPtest9) to a maximum of 1.168 (Pt30204) with an average of 0.695 per marker. Direct correlations were observed between the number of polymorphic bands and MI. The SSR primer Pt 30204 showing maximum number of polymorphic bands had highest value for MI (1.168) and the primer with lowest MI (0.050) produced least number of polymorphic bands. There was a strong correlation between polymorphism

Posterior probabilities, LnP(D) as function of K, where K = 1-10

Fig. 3a. – Bayesian posterior probability of data [LnP(D)] with increasing K for SSR markers.

An ad hoc quantity based on the second order rate of change of the likelihood function with respect to K (ΔK)

Fig. 3b. – Magnitude of ΔK as a function of K for SSR markers.

As per the results revealed by STRUCTURE, the posterior probability of data, LnP(D), steadily improved until K=5, and then continued to increase slightly until K = 10 (Fig. 3a). Based on the four steps for the graphical method allowing detection of the true number of groups K suggested by EVANNO *et al.* (2005), true value for K was detected. The height of modal value of the distribution of ΔK (the second order rate of change of the likelihood function with respect to K) located at K indicated the strength of the signal detected by STRUCTURE. With respect to K, (ΔK) showed a clear peak at the true value of K. The real structure showing a clear peak of the genotypes was set at K=5 (Fig. 3b).

Figure 4. – Estimated genotypic structure plot for all the genotypes of *P. roxburghii* using SSR markers.

The graph is based on Structure run of real set K estimated for SSR data. Each genotype is represented by a bar, partitioned into different segments corresponding to its membership coefficient in inferred clusters. Each colour represents a different cluster, and black segments separate the different genotypes. Left-to-right colour grouping represented in plot is in accordance with the estimated cluster ID.

Table 4a. – Analysis of molecular variance (AMOVA) of *P. roxburghii* genotypes based on collection site using SSR markers.

Source of variation	df	Sum of squares	Mean Sum of squares	Variance components	Percentage of variation	F-statistic
Among populations	3	15.51	5.17	0.018Va	0.38	$F_{ST} = 0.003$
Within populations	49	241.31	4.92	4.92 Vb	99.62	
Total	52	256.83		4.94		

Table 4b. – Analysis of molecular variance (AMOVA) of *P. roxburghii* genotypes based on resin yield using SSR markers.

Source of variation	df	Sum of squares	Mean Sum of squares	Variance components	Percentage of variation	F-statistic
Among populations	2	27.77	13.88	0.57 Va	10.96	$F_{ST} = 0.109^{***}$
Within populations	50	231.66	4.63	4.63 Vb	89.04	
Total	52	259.43		5.20		

*** denote significance at 0.1% probability level.

Figure 5. – Linkage disequilibrium measure (r^2 , above diagonal line) and probability value (p, below diagonal line) for 17 SSR markers in 53 genotypes of *P. roxburghii*.

This is a disequilibrium matrix for polymorphic SSRs where polymorphic sites are plotted on both the X- axis and Y- axis. Pair-wise calculations of LD (r^2) are displayed above the diagonal with the corresponding p-values for Fisher's exact test displayed below the diagonal.

and MI ($r^2=0.942$), and MI was found to be positively correlated with PIC ($r^2=0.822$). RP ranged from 0.226 (RPtest9) to 1.926 (Pt71936) with an average of 1.164. There was a strong correlation between polymorphism and RP ($r^2=0.725$). RP showed positive correlation with PIC ($r^2=0.712$).

Cluster analysis

The hierarchical clustering using UPGMA implemented in the software DARwin ver 5.0.158 revealed the existence of two distinct major clusters: Cluster-I and Cluster-II (Fig. 2). Cluster-I with a bootstrap value of 102, grouped twenty-eight genotypes together, out of which twenty-two were low resin yielders (< 3 kg/year) and only six genotypes had high resin yield (> 3 kg/year). Cluster-II with a bootstrap value of 101, grouped twenty-five genotypes, out of which twenty-two had high resin yield and only three genotypes were low resin yielders.

The genetic dissimilarity index revealed high genetic diversity among the fifty-three genotypes of *P. roxburghii* used in the study. The dissimilarity coefficients ranged from 0.04 to 0.72. The genotypes D-26 (2.1 kg/year) and B-13 (1.7 kg/year) both having low resin yield were found to be most similar, whereas A-24 (2.8 kg/year) and B-18 (3.35 kg/year) were found to be the most dissimilar genotypes.

Genetic structure analysis

As per the STRUCTURE results, the log likelihood steadily improved until $K=5$, and then continued to increase slightly until $K=10$ (Fig. 3a). The results showed that the peak value of Evanno's ΔK was at $K=5$, suggesting five genetic clusters (Fig. 3b). With five as the optimum population structure, inferred ancestries (Q matrix) of individuals were determined. Each individual is represented by a vertical line broken into K colored segments, with lengths proportional to each of the K inferred clusters.

Beyond $K=5$, the probability of the data did not peak and hence it was considered that five clusters captured the entire divisions of the sample (Fig. 4a). In total, twenty-three genotypes (43.40% out of 53 genotypes) were clearly assigned to each single population, where 80% of their inferred ancestry was derived from one of the model populations. On the other hand, thirty genotypes (56.60% out of 53 genotypes) in the sample were categorized as having admixed ancestry. Each cluster had ten individuals on an average, the highest in cluster 4 and the lowest in cluster 1. The levels of differentiation between subgroups were variable with F_{ST} ranging from 0.54 to 0.61. The distribution of *P. roxburghii* genotypes into these five sub-populations had no correlation with their resin yield (Fig. 4b).

Table 5. – Fixed effect Linear model solution for phenotypes and genotypes of *P. roxburghii*.

Locus	p Marker	R ² Marker	p adj Marker	R ² Model
pdms011	0.6267	0.0051	1	0.1576
pdms221	0.6585	0.0035	1	0.1755
pm05	0.4392	0.0116	0.999	0.2805
pm07	0.1054	0.0563	0.5445	0.1573
RPtest6	0.9213	0.0001	1	0.1722
RPtest9	0.0818	0.0549	0.3457	0.22
Pt1254	0.3021	0.0191	0.98	0.193
Pt71936	0.0183	0.0957*	0.0509	0.2641*
Pt87268	0.0034	0.1501**	0.007	0.3296**
pm09a	0.0001	0.2433***	0.00099	0.3983***
PCP26106	0.047	0.0765*	0.1738	0.2907
PCP30277	0.1506	0.0439	0.7313	0.2249
Pt30204	0.0916	0.0581	0.4346	0.222
Pt79951	0.2766	0.0221	0.977	0.4031
PCP9434	0.4777	0.0095	1	0.371
Pt36480	0.2386	0.0286	0.9291	0.242
PCP41131	0.1747	0.0395	0.8232	0.2157

p marker – p-value on account of marker alone; p adj marker –p-value adjusted on account of full model.

R² model – portion of total variation explained by the full model;

R² marker – portion of total variation explained by the marker.

*, **, *** denote significant at 5%, 1% and 0.1% level of probability.

Partitioning of variance using SSR markers

AMOVA analysis revealed that 99.62% of the total variation in studied populations of *P. roxburghii* was structured within populations and only 0.38% was among populations (Table 4a). Similarly, it revealed that 89.04% of the variation with respect to the resin yield lies within populations and rest 10.96% variation was among populations. There was negligible population genetic differentiation ($F_{ST}=0.003$) between the studied populations for the molecular variation (Table 4a). However the population genetic differentiation was moderate for resin yield ($F_{ST}=0.10$) (Table 4b) in *P. roxburghii*.

Linkage disequilibrium

Of the total forty-six SSR alleles amplified, thirty-nine alleles were used for estimating LD between all pairs of SSR alleles. The reduction in the total number of alleles was due to the entire data set being filtered to eliminate alleles with a frequency less than 5%. The r^2 and the P value representing LD were assessed for seventeen SSR markers. The markers PCP30277 and PCP26106 were detected to be in significant LD at $P<0.01$ (Fig. 5). The markers Pt71936 and Pt87268 were found to be associated with resin yield and showed significant linkage disequilibrium. Further, one of the trait associated marker (Pt87268) was also found to be in significant linkage disequilibrium with the marker pm07 showing the probability of this marker also to be associated with resin yield.

Marker-trait association analysis using SSR markers

Association analysis with a total of forty polymorphic SSR loci revealed two cpSSR markers (Pt71936 and Pt

87268) and one nSSR (pm09a) to be in significant association with resin yield (Table 5). Pt71936 was able to explain 9.57% of the total variation in resin yield when association was tested on account of the marker alone. Whereas, when full model was used (including population structure), it could explain 26.41% of the total variation in resin yield ($P<0.05$). The SSR marker Pt87268 which could explain 15.01% of the total variation in absence of population structure was able to explain 32.96% of the total variation when population stratification was taken into account ($P<0.01$). Similarly, the marker pm09a which explained 24.33% of the variation in absence of population structure was able to explain 39.83% of the total variation after considering the population stratification ($P<0.001$). All these markers showed association with resin yield in the absence as well as presence of population structure and the percentage of variation explained by these markers was increased after taking the population stratification into account. The SSR marker PCP26106 which showed association with resin yield in the absence of population structure ($P<0.05$), lost its significance in the presence of population structure.

Discussion

The amplification of *P. roxburghii* genotypes using nineteen SSR primers resulted in a total of 46 bands with 40 bands showing polymorphism (86.9%). The total number of polymorphic bands amplified per marker varied from as low as 1 to a maximum of 4 (Pt30204 and Pt45002), with an average of 2.105 per marker. This is in agreement with the results earlier reported by

CHAUHAN, 2011 where Pt30204 (localized 524 bp upstream of the *clpP* gene of *P. thunbergii*) was found to be the most variable locus in *P. roxburghii* which detected maximum number of alleles. The locus Pt 30204 was reported to be a “mutational hotspot” (VENDRAMIN *et al.*, 1999) in *Abies alba* (Pinaceae) where a total of fifteen size variants were detected in seventy individuals. Genetic divergence in terms of per cent polymorphism ranged from 50 to 100% which was in agreement with the results reported earlier in *P. roxburghii* by CHAUHAN, 2011. Among all the SSR primer pairs tested in *P. roxburghii* genotypes, Pt71936 was the most informative with high PIC, MI and RP values.

The hierarchical clustering using UPGMA implemented in the software DARwin ver 5.0.158 clustered the genotypes distinctly on the basis of their resin yield. The dissimilarity coefficients ranged from 0.04 to 0.72 suggesting high genetic variability among the genotypes. The grouping of *P. roxburghii* genotypes was on the basis of resin yield and not their site of collection suggesting the genetic basis of the trait.

Model based clustering of the *P. roxburghii* genotypes using SSR markers revealed the occurrence of five subpopulations in the sample. The distribution of genotypes into different subpopulations had no correlation with their resin yield suggesting that resin yield of the genotypes was not attributed to their ancestry but it was because of their genetic constitution.

The AMOVA analysis showed that most of the variation in *P. roxburghii* lies within populations, a result compatible with woody perennial, out breeding plant species, especially conifers (HAMRICK *et al.*, 1992). Population genetic differentiation is negligible ($F_{ST}=0.003$) between the studied populations and indicate that there is no hindrance in the gene flow among the selected populations resulting in homogeneous genetic structures. For the interpretation of F_{ST} , it has been suggested that a value lying in the range 0–0.05 indicates little genetic differentiation; a value between 0.05 and 0.15, moderate differentiation; a value between 0.15 and 0.25, great differentiation; and values above 0.25, very great genetic differentiation (WRIGHT, 1978; HARTL and CLARK, 1997). The AMOVA analysis also showed that most of the variations (89.04%) with respect to the resin yield lie within populations than among populations (10.96%) in *P. roxburghii*. The F_{ST} value indicated that there was moderate genetic differentiation among the groups when the genotypes were grouped based on their resin yield.

Linkage disequilibrium based association mapping have been a research objective in plants beginning with the model organism as *Arabidopsis*, and now extended to crops as maize, barley, durum wheat, spring wheat, rice, sorghum, sugarcane, sugar beet, soybean and grape as well as in forest tree species and forage grasses (ABDURAKHMONOV and ABDUKARMINOV, 2008).

Linkage disequilibrium is the non random association of alleles at different loci which play an integral role in association mapping, and determines the resolution of an association study. The mating system of the species

(selfing versus outcrossing), and phenomena such as population structure and recombination hotspots, can strongly influence patterns of LD. Generally, LD decays more rapidly in outcrossing species as compared to selfing species (NORDBORG, 2000). This is because recombination is less effective in selfing species, where individuals are more likely to be homozygous, than in outcrossing species. Admixture results in the introduction of chromosomes of different ancestry and allele frequencies. Often, the resulting LD extends to unlinked sites, even on different chromosomes, but breaks down rapidly with random mating (PRITCHARD and ROSENBERG, 1999). Genome wide LD has been quantified for many forest tree species that extended upto 16–34 kb in *Populus trichocarpa* (YIN *et al.*, 2004); <500 bp in *Populus termula* (INGVARSSON, 2005); 2000 bp in *Pinus taeda* (BROWN *et al.*, 2004); 1000 bp in *Pseudotsuga menziensis* (KRUTOVSKY and NEALE, 2005) and 100–200 bp in *Picea abies* (RAFALSKI and MORGANTE, 2004).

Since *P. roxburghii* is a highly outcrossing species so a rapid decay of LD is expected leading to a fine resolution mapping which makes this species appropriate for association studies. Since SSR markers specific for *P. roxburghii* are not available so SSRs from different *Pinus* species including *P. taeda*, *P. thunbergii*, *P. densiflora*, *P. merkussii*, *P. sylvestris* and *P. resinosa* were tested for cross amplification in *P. roxburghii*. It was found that the transferability of nuclear SSRs was much less (CHAUHAN, 2011) as compared to chloroplast SSRs and since the nuclear genome of *P. roxburghii* is not yet sequenced but its partial chloroplast genome sequence is known, so a combination of nuclear and chloroplast SSRs were used in the present investigation. Although candidate gene based approach has been employed in many association studies, the success of this approach depends upon the correct choice of which genes/pathways to study. Therefore, *a priori* hypothesis about biological function is required, which is exposed to the risk of arbitrariness. A more comprehensive and unbiased approach is to employ markers encompassing the entire genome (EBERLE *et al.*, 2007). Several genome-wide association studies (GWAS) for complex diseases have been completed, as reviewed in MANOLIO *et al.* (2008). Genome wide association studies hold the promise to relatively complete genetic effects (additive and non additive) and pleiotropy in an unbiased way (STRANGER *et al.*, 2011).

There are a number of reports where dominantly coded (present versus absent) marker data of SSRs were successfully used in genome wide LD analyses and LD-based association mapping in plants (KRAAKMAN *et al.*, 2004; KRAAKMAN *et al.*, 2006; HANSEN *et al.*, 2001; TOMMASSINI *et al.*, 2007; IWATA *et al.*, 2007; MALOSETTI *et al.*, 2007 and GEBHARDT *et al.*, 2004), demonstrating the feasibility of dominantly coded molecular data in revealing of haplotypic associations. So, the SSR data was dominantly coded as present vs. absence in the present study. Similar strategy was reported by ZHAO *et al.*, 2014 where some SSRs were considered as codominant and others as dominant while conducting LD based association studies in (*Gossypium hirsutum* L.) germplasm.

Association analysis revealed that two chloroplast SSRs Pt71936 and Pt87268 and one nuclear SSR pm09a to be in significant association with the resin yield. The cpSSR markers showing association with resin yield, were present at a distance 15kb apart from one another in the chloroplast genome and thus are supposed to be linked. In a previous report in *P. densiflora* (KIM *et al.*, 2009), the chloroplast was found to be involved in resin biosynthesis pathway. They have discussed that the ABS gene product catalyzes the cyclization of geranylgeranyl diphosphate to abietadiene as the first committed step of resin biosynthesis. Abietadiene is the precursor of the major pine resin acids, abietic, neoabietic and dehydroabietic acids (JOYE and LAWRENCE, 1967). Therefore, the enzyme could serve as a site marker for resin biosynthesis. ABS is known to occur in the chloroplast (Ro and BOHLMANN, 2006), which means that the initial cyclization of GGPP in resin acid biosynthesis and biosynthesis of the building blocks take place in the same organelle. In an earlier report in *Triticum aestivum* (LIU *et al.*, 2010), a total of 10 linked SSR markers were identified to be associated with the six traits (plant height, spike length, spikelets per spike, spikelets density, grains per spike and thousand kernel weight) at the 0.01 probability level, and each QTL explained 4.85% to 20.59% of the phenotypic variation. In the present study, the two cpSSR markers Pt71936 and Pt 87268 which were found to be associated with resin yield showed significant linkage disequilibrium ($P < 0.01$) with r^2 values higher than 0.05. Similar results were reported earlier in Barley (ELEUCH *et al.*, 2008) where the markers Bmag 749 and HVHOTRI located on chromosome 2H were found to be salinity and showed linkage disequilibrium with r^2 values higher than 0.05. In *Gossypium hirsutum* L., the average r^2 of global marker pairs was reported to be 0.0132 (ZHAO *et al.*, 2014).

Further in this study, one of the trait associated cpSSR (Pt87268) was also found to be in significant linkage disequilibrium ($p < 0.01$) with the nuclear SSR pm07 showing the probability of this marker also to be associated with resin yield. This suggests the presence of cytonuclear disequilibrium in *P. roxburghii*. Cytonuclear disequilibrium is the nonrandom association of alleles or genotypes at a nuclear locus with haplotypes at cytoplasmically inherited organeller DNA (ASMUSSEN *et al.*, 1987; SCHNABEL and ASMUSSEN, 1989). In a previous study, in a natural population of *Ponderosa pine* (LATTA *et al.*, 2001), cytonuclear disequilibrium was measured between eleven nuclear allozymes loci and both mitochondrial and chloroplast DNA haplotypes. Three allozymes loci (*Fe*, *Got* and *Udp*) showed significant associations ($P < 0.05$) with mitochondrial DNA variation, while two other loci (*Per* and *Sdh*) showed significant association with cpDNA. The overall magnitude (normalized disequilibrium) of associations was greater for maternally inherited mtDNA than for paternally inherited cpDNA, though this difference was neither large nor significant ($p > 0.1$). The nonrandom association of nuclear alleles or genotypes with organeller haplotypes can arise from a number of evolutionary forces that fall into three categories (ASMUSSEN

et al., 1987): (i) nonrandom mating, including patterns of admixture, migration and hybridization; (ii) interactive fitness across genomes; and (iii) the historical sampling of gametes in finite populations (drift). Given the many factors that may give rise to cytonuclear disequilibrium, SCRIBNER *et al.* (1999) cautioned that in the absence of independent information it can be difficult to ascribe a particular evolutionary process to observed patterns.

Limited data exist for cytonuclear disequilibria involving a paternally inherited organelle. The data of L1 (1995) for jack pine (*P. banksiana*) was analyzed by BASTEN and ASMUSSEN (1997) but did not reveal significant associations. The existence of significant associations between nuclear and cpSSRs in *P. roxburghii* suggests that founding events occurred through the paternal lines in much the same way as through the maternal lineages as in the case of *Ponderosa pine*.

References

- ABDURAKHMONOV, I. Y. and A. ABDUKARIMONOV (2008): Application of association mapping to understanding the genetic diversity of plant germplasm resources. *International Journal of Plant Genomics*, 2008: 574927.
- ASIF, M. J. and C. H. CANNON (2005): DNA Extraction from Processed Wood: A case study for the identification of an endangered timber species (*Gonystylus bancanus*). *Plant Molecular Biology Reporter*, **23**: 185–192.
- ASMUSSEN, M. A., J. ARNOLD and J. C. AVISE (1987): Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics*, **115**: 755–768.
- BASTEN, C. J. and M. A. ASMUSSEN (1997): The exact test for cytonuclear disequilibria. *Genetics*, **146**: 1165–1171.
- BRADBURY, P. J., Z. ZHANG, D. E. KROON, T. M. CASSTEVEN, Y. RAMDOSS and E. S. BUCKLER (2007): TASSEL: software for association mapping of complex traits in diverse samples. *Genetics and Population Analysis*, **23**(19): 2633–2635.
- BROWN, G. R., G. P. GILL, R. J. KUNTZ, J. A. BEAL, D. NELSON, N. C. WHEELER, B. PENTTILA, J. ROERS and D. B. Neale (2004): Associations of candidate gene single nucleotide polymorphisms with wood property phenotypes in loblolly pine. In: *Plant and Animal Genomes XII Conference*, San Diego, CA, USA. Abstract: W98.
- CHAUHAN, P. (2011): Molecular genetic analysis of Chir pine (*Pinus roxburghii* Sarg.) through microsatellite markers. Ph.D thesis submitted to Forest Research Institute University.
- CHURCHILL, G. A. and R. W. DOERGE (1994): Empirical threshold values for quantitative trait mapping. *Genetics*, **138**: 963–971.
- COPPEN, J. J. W. and G. A. HONE (1995): Gum naval stores: turpentine and rosin from pine resin. *FAO, Non-wood Forest Products*, **2**: 62 p.
- DOYLE, J. J. and J. L. DOYLE (1990): A rapid total DNA preparation procedure for fresh plant tissue. *Focus*, **12**: 13–15.
- EBERLE, M. A., K. KUHN, P. C. NG, L. ZHOU, D. A. PEIFFER, L. GALVER, K. A. VIAUD-MARTINEZ, C. T. LAWLEY, K. L. GUNDERSON and R. SHEN *et al.* (2007): Power to detect risk alleles using genome-wide tag SNP panels. *PLoS Genet*, **3**(10): 1827–1837.

- ELSIK, C. G., V. T. MINIHAN, S. E. HALL, A. M. SCARPA and C. G. WILLIAMS (2000): Low-copy microsatellite markers for *Pinus taeda* L. *Genome*, **43**: 550–555.
- EVANNO, G., S. REGNAUT and J. GOUDET (2005): Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**: 2611–2620.
- EXCOFFIER, L., P. E. SMOUSE and J. M. QUATTRO (1992): Analysis of molecular variance inferred from metric distances among DNA haplotypes; Application to human mitochondrial DNA restriction data. *Genetics*, **131**: 479–491.
- EXCOFFIER, L., G. LAVAL and S. SCHNEIDER (2005): Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics online*, **1**: 47–50.
- FARNIR, F., W. COPPIETERS, J. J. ARRANZ, P. BERZI and N. CAMBISANO (2000): Extensive genome wide linkage disequilibrium in cattle. *Genome Research*, **10**: 220–227.
- FELSENSTEIN, J. (1985): Confidence limits in phylogenies: an approach using bootstrap. *Evolution*, **39**: 783–791.
- GEBHARDT, C., A. BALLVORA, B. WALKEMEIER, P. OBERHAGEMANN and K. SCHULER (2004): Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Molecular Breeding*, **13**: 93–102.
- GHILDYAL, S. K., C. M. SHARMA and S. GAIROLA (2009): Environmental variation in seed and seedling characteristics of *Pinus roxburghii* Sarg. from Uttarakhand, India. *Applied Ecology and Environmental research*, **7**(2): 121–129.
- HANSEN, M., T. KRAFT, S. GANESTAM, T. SALL and N. O. NILSSON (2001): Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genetic Research*, **77**(1): 61–66.
- HAMRICK, J. L., M. J. W. GODT and S. L. SHERMAN-BROYLES (1992): Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**: 95–124.
- HARTL, D. L. and A. G. CLARK (1997): *Principles of Population Genetics*, 3rd edn. Sinauer Associates, Inc, Sunderland, MA.
- INGVARSSON, P. K. (2005): Nucleotide polymorphism and linkage disequilibrium within and among natural populations of European Aspen (*Populus tremula* L. Salicaceae). *Genetics*, **169**: 945–953.
- IWATA, H., Y. UGA, Y. YOSHIOKA, K. EBANA and T. HAYASHI (2007): “Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms,” *Theoretical and Applied Genetics*, **114**(8): 1437–1449.
- JOYE, N. M. JR. and R. V. LAWRENCE (1967): Resin acid composition of pine oleoresins. *Journal of Chemical and Engineering Data*, **12**: 279–282.
- KEDHARNATH, S. (1971): Evolving high oleoresin yielding strains of Chir pine (*Pinus roxburghii*) through breeding. In: *Proceedings of seminar on the role of pine resin in economic and industrial development of India*. Sympine. New Delhi, India. 13–14 April, D-1-D-5.
- KENNEDY, B. W., M. QUINTON and J. A. M. VANARENDONK (1992): Estimation of effects of single genes on quantitative traits. *J. Anim. Sci.* **70**: 2000–2012.
- KIM, Y. B., S. M. KIM, M. K. KANG, T. KUZUYAMA, J. K. LEE, S. C. PARK, S. C. SHIN and S. U. KIM (2009): Regulation of resin acid synthesis in *Pinus densiflora* by differential transcription of genes encoding multiple 1-deoxy-D-xylulose 5-phosphate synthase and 1-hydroxy-2-methyl-2(E)-butenyl 4-diphosphate. *Tree Physiology*, **29**: 737–749.
- KRAAKMAN, A. T. W., R. E. NIKS, P. M. M. VAN DEN BERG, P. STAM and F. A. VAN EEUWIJK (2004): Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, **168**: 435–446.
- KRAAKMAN, A. T. W., F. MARTINEZ, B. MUSSIRALIEV, F. A. VAN EEUWIJK and R. E. NIKS (2006): Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars. *Molecular Breeding*, **17**: 41–58.
- KRUTOVSKY, K. V. and D. B. NEALE (2005): “Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality related candidate genes in douglas fir.” *Genetics*, **171**(4): 2029–2041.
- LATTA, R. G., Y. B. LINHART and J. B. MITTON (2001): Cytonuclear disequilibrium and genetic drift in a natural population of Ponderosa pine. *Genetics*, **158**: 843–850.
- LE MAITRE, D. C. (1998): Pines in cultivation: a global view. In: D. M. RICHARDSON (Eds.), *Ecology and biogeography of Pinus*, Cambridge University Press, New York, USA. pp. 407–431.
- LIU, L., L. WANG, J. YAO, Y. ZHENG and C. ZHAO (2010): Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). *Molecular Plant Breeding*, **1**(5): 1–10.
- LORENZ, E., S. LEETON and R. J. OWEN (1997): A simple method for sizing large fragments of bacterial DNA separated by PAGE. *Cabios*, **13**: 485–486.
- MALOSETTI, M., C. G. VAN DER LINDEN, B. VOSMAN and F. A. VAN EEUWIJK (2007): “A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato,” *Genetics*, **175**(2): 879–889.
- MANOLIO, T. A., L. D. BROOKS and F. S. COLLINS (2008): A HapMap harvest of insights into the genetics of common disease. *J. Clin. Invest.* **118** (5) (2008), pp. 1590–1605.
- NAUTIYAL, S., M. TYAGI, S. S. NEGI, R. K. SRIVASTAVA and S. NAUTIYAL (2006): Improvement of Chir pine (*Pinus roxburghii* Sarg.). In: (Eds.) *Studies on Himalayan Pines*. Pub. FRI, Dehradun, India. pp. 50–63.
- NEGI, S. S. and S. MALIK (2009): Development of prediction models for resin production of *Pinus roxburghii*. FRI-66(N) funded by ICFRE.
- NORDBORG, M. (2000): Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self fertilization. *Genetics*, **154**: 923–29.
- PERRIER, X. and J. P. JACQUEMOUD-COLLET (2006): DARwin software. Genetic improvement of vegetatively propagated crops. <http://darwin.cirad.fr/Home.php>
- POWELL, W., G. MACHRAY and J. PROVAN (1996): Polymorphism revealed by simple sequence repeats. *Trends in Plant Sciences*, **1**: 215–222.
- PREVOST, A. and M. WILKINSON (1999): A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theoretical and Applied Genetics*, **98**: 107–112.
- PRITCHARD, J. K. and N. A. ROSENBERG (1999): Use of unlinked genetic markers to detect population stratification in association studies. *American Journal of Human Genetics*, **65**: 220–228.

- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY (2000a): Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945–959.
- PRITCHARD, J. K., M. STEPHENS, N. A. ROSENBERG and P. DONNELLY (2000b): Association mapping in structured population. *American Journal of Human Genetics*, **67**: 170–181.
- RAFALSKI, D. L. and M. MORGANTE (2004): Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics*, **20**: 103–111.
- REN, Y., Z. ZHANG, J. LIU, J. E. STAUB, Y. HAN and Z. CHENG *et al.* (2009): An Integrated Genetic and Cytogenetic Map of the Cucumber Genome. *PLoS ONE* **4**(6): e5795.
- RISCH, N. J. (2000): Searching for genetic determinants in the new millennium. *Nature* **405**: 847–856.
- RO, D. K. and J. BOHLMANN (2006): Diterpene resin acid biosynthesis in loblolly pine (*Pinus taeda*): functional characterization of abietadiene/levopimaradiene synthase (PtTPS-LAS) cDNA and subcellular targeting of PtTPS-LAS and abietadienol/abietadienal oxidase (PtAO, CYP720B1). *Phytochemistry*, **67**: 1572–1578.
- ROLDAN-RUIZ DENDAUW, J., E. VAN BOCKSTAELE, A. DEPICKER and M. DE LOOSE (2000): AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Molecular Breeding*, **6**: 125–134.
- SCHNABEL, A. and M. A. ASMUSSEN (1989): Definition and properties of disequilibria within nuclear-mitochondrial-chloroplast and other nuclear-dicytoplasmic systems. *Genetics*, **123**: 199–215.
- SCRIBNER, K. T., S. DATTA, J. ARNOLD and J. C. AVISE (1999): Empirical evaluation of cytonuclear models incorporating genetic drift, and tests for neutrality of mtDNA variants: data from experimental *Gambusia* hybrid zones. *Genetica*, **105**: 101–108.
- SHARMA, C. M., S. K. GHILDYAL and D. P. NAUTIYAL (2001): Plus tree selection and their seed germination in *Pinus roxburghii* from Garhwal Himalaya. *Indian Journal of Forestry*, **24**(1): 48–52.
- SHARMA, R., S. KUMAR and K. S. THAKUR (2006): Genetic improvement of Chir pine (*Pinus roxburghii* Sargent.) in India – A review. *The Indian Forester*, **132**(3): 314–328.
- SINGH, V. and S. KUMAR (2004): Seed quality as affected by mid cone diameter in *Pinus roxburghii* Sarg. *Indian Forester*, **130**(7): 757–761.
- STANGE, C., D. PREHN and P. A. JOHNSON (1998): Isolation of *Pinus radiata* genomic DNA suitable for RAPD analysis. *Plant Molecular Biology Reporter*, **16**: 1–8.
- STRANGER, B. E., E. A. STAHL and T. RAJ (2011): Progress and Promise of Genome-Wide Association Studies for Human Complex Trait Genetics. *Genetics*, **187**(2): 367–383.
- THORNSBERRY, J. M., M. M. GOODMAN, J. DOEBLEY, S. KRESOVICH, D. NIELSEN and E. S. BUCKLER (2001): Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics*, **28**: 286–289.
- VENDRAMIN, G. G., P. LELLIL ROSSI and M. MORGANTE (1996): A set of primers for the amplification of 20 chloroplast microsatellites in Pinaceae. *Molecular Ecology*, **5**: 595–598.
- VENDRAMIN, G. G., B. DEGEN, R. J. PETIT, M. ANZIDIE, A. MADAGHIELE and B. ZIEGENHAGEN (1999): High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology*, **8**: 1117–1126.
- WEIR, B. S. (1996): Genetic data analysis II. Methods for discrete population genetic data. Sinauer Associates, Sunderland. University of Washington, 376 p.
- WRIGHT, S. (1978): Evolution and the Genetics of Population, Variability Within and Among Natural Populations. The University of Chicago Press, Chicago.
- XUELIAN, S., Y. TAO, H. JUNJIE, Z. XIAOYAN, F. REBECCA, J. JUNYE, A. FANG, G. JIANPING and Z. XUXIAO (2014): SSR genetic linkage map construction of pea (*Pisum sativum* L.) based on Chinese native varieties, *The Crop Journal* **2**, 170–174.
- YANG, H., Y. TAO, Z. ZHENG, Q. ZHANG, G. ZHOU, M. W. SWEETINGHAM, J. G. HOWIESON and C. LI (2013): Draft Genome Sequence, and a Sequence-Defined Genetic Linkage Map of the Legume Crop Species *Lupinus angustifolius* L. *PLoS One*. **8**(5): e64799.
- YIN, T. M., S. P. DIFAZIO, L. E. GUNTER, S. S. JAWDY, W. BOERJAN and G. A. TUSKAN (2004): “Genetic and physical mapping of Melampsora rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence,” *New Phytologist*, **164**(1): 95–105.
- ZHAO, Y., H. WANG, W. CHEN and Y. LI (2014): Genetic Structure, Linkage Disequilibrium and Association Mapping of Verticillium Wilt Resistance in Elite Cotton (*Gossypium hirsutum* L.) Germplasm Population. *PLoS ONE* **9**(1): e86308. doi:10.1371/journal.pone.0086308
- ZONDERVAN, K. T. and L. R. CARDON (2004): The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* **5**: 89–100.