DEVELOPMENT AND CHARACTERIZATION OF GENOMIC MICROSATELLITE MARKERS IN *Rhododendron arboreum*

Dissertation submitted to the Central University of Punjab

For the award of
Master of Philosophy
In
Biosciences

BY

Shruti Choudhary

Supervisor
Dr. Pankaj Bhardwaj

Centre for Biosciences
School of Basic and Applied Sciences
Central University of Punjab, Bathinda

August, 2014

CERTIFICATE

I declare that the dissertation entitled "DEVELOPMENT AND CHARACTERIZATION OF GENOMIC MICROSATELLITE MARKERS IN *Rhododendron arboreum*" has been prepared by me under the guidance of Dr. Pankaj Bhardwaj, Assistant Professor, Centre for Biosciences, School of Basic and Applied Sciences, Central University of Punjab. No part of this dissertation has formed the basis for the award of any degree or fellowship previously.

Shruti Choudhary

Centre for Biosciences,
School of Basic and Applied Sciences,
Central University of Punjab,
Bathinda-151001

Date: August , 2014

CERTIFICATE

I certify that Shruti Choudhary has prepared her dissertation entitled "DEVELOPMENT AND CHARACTERIZATION OF GENOMIC MICROSATELLITE MARKERS IN *Rhododendron arboreum*", for the award of M.Phil. degree of the Central University of Punjab, under my guidance. She has carried out this work at the Centre for Biosciences, School of Basic and Applied Sciences, Central University of Punjab.

Dr. Pankaj Bhardwaj

Centre for Biosciences,
School of Basic and Applied Sciences,
Central University of Punjab,
Bathinda – 151001

Date: August , 2014

ABSTRACT

Development and characterization of genomic microsatellite markers in *Rhododendron arboreum*

Name of student: Shruti Choudhary

Registration number: CUPB/MPh-PhD/SBAS/BIO/2012-13/02

Degree for which submitted: Master of Philosophy

Name of supervisor: Dr. Pankaj Bhardwaj

Name of centre: Centre for Biosciences

Name of school: School of Basic and Applied Sciences

Keywords: *Rhododendron arboreum,* Microsatellites, Selective hybridization, SSR characterization, Population structure

Rhododendron arboreum, a key species inhabiting Indian Himalayas, has ecological and economic importance. Global climate change and anthropogenic activities can pose a threat to the biodiversity richness. Population genetics characteristics are the fundamentals of conservation and management practices. However, no genetic background for the species is presently available. To provide an insight into the population structure of the species, an effective marker resource was developed. A genomic library enriched for microsatellites was constructed using biotin- streptavidin hybridization technique, followed by sequencing of positive clones. With 54% enrichment rate, 41 primers were designed from SSR clones, consisting of perfect or interrupted repeats of AG/CT, AC/GT motifs. 38 loci showed successful amplification on genotypes of three populations with number of alleles ranging from 2 to 14, observed and expected heterozygosity of 0.167 to 0.933 and 0.422 to 0.917, respectively and average polymorphic information content of 0.104 to 0.811. A high gene flow (N_m= 5.436) demonstrated high genetic diversity within the populations. 19 loci displayed significant deviations from Hardy Weinberg equilibrium. Overall no linkage disequilibrium and bottleneck was detected among sampled populations. DARwin and STRUCTURE analysis grouped the populations into two original clusters and admixed genetic stock irrespective of their geographical locations. These novel loci will support further genotyping studies in *R. arboreum.*

(Shruti Choudhary)
Signature of Student

(Dr. Pankaj Bhardwaj)
Signature of Supervisor

ACKNOWLEDGEMENTS

With utmost modesty, I bow before God, the Almighty, for showering me with blessings and strength.

I express the deepest gratitude to my supervisor, **Dr. Pankaj Bhardwaj**, Assistant Professor, Centre for Biosciences, Central University of Punjab for encouraging me throughout the course of this study and helping me grow as a researcher.

I am grateful to **Prof. R. G. Saini**, Centre Coordinator, Biosciences, for his generous advice. My appreciations to **Prof. P. Ramarao**, Acting Vice- Chancellor and **Prof. Jai Rup Singh**, former Vice chancellor, Central University of Punjab, for providing infrastructure and other technical facilities. The cooperation of all the faculty members of Centre for Biosciences is highly acknowledged.

My special thanks to seniors, lab mates and friends especially, Richa Mehra, Sapna Thakur, Shishir Upadhyay, Harmeen and Gaganpreet Kaur. I am also thankful to technical staff for their help and support.

Warmest regards to my family for their blessings and valuable suggestions.

At last, I thank Indian Council of Medical Research (ICMR), New Delhi for awarding me fellowship and Department of Science and Technology (DST), Govt. of India for funding the research work.

(Shruti Choudhary)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

S. No.	Full Form	Abbreviation
1.	Adenine	A
2.	Amplified Fragment Length Polymorphism	AFLP
3.	Base pair	bp
4.	Cetyl trimethylammonium bromide	CTAB
5.	Cytosine	C
6.	Degree Celsius	°C
7.	Deoxyribonucleic acid	DNA
8.	Deoxyribonucleotide triphosphates	dNTPs
9.	Genomic SSRs	G-SSRs
10.	Gram	g
11.	Guanine	G
12.	Hardy- Weinberg Equilibrium	HWE
13.	Hour(s)	hr(s)
14.	Magnesium Chloride	$MgCl_2$
15.	Melting temperature	T_m
16.	Microgram	µg
17.	Microliter	µl
18.	Milligram	mg
19.	Milliliter	ml
20.	Millimolar	mM
21.	Minute	min
22.	Nanogram	ng
23.	Polyacrylamide gel electrophoresis	PAGE
24.	Polymerase Chain Reaction	PCR
25.	Polymorphic Information Content	PIC
26.	Potassium Chloride	KCl
27.	Randomly Amplified Polymorphic DNA	RAPD
28.	Restriction Fragment Length Polymorphism	RFLP
29.	Second	sec
30.	Simple Seqeunce Repeats	SSR
31.	Sodium- Saline Citrate	SSC
32.	SSR Identification Tool	SSRIT
33.	Thymine	T
34.	Unrooted Pair Group Method Analysis	UPGMA

CHAPTER 1

Introduction

India has a rich source of diversity in terms of its biotic components supported by a regime of climatic conditions. The Himalayas, a hotspot for biodiversity, harbor variety of flora and fauna. One of the endemic plant species native to the region is *Rhododendron arboreum*, an Ericaceae, the most abundant among Indian Rhododendrons. The plant parts are utilized for general purpose (fuel, timber, making incense, jellies, and squashes) as well as for ornamental, landscaping and medicinal use (attributed to their constituting antioxidants, flavinoids, saponins, phenolics and other secondary metabolites). In addition, the forest cover of *R. arboreum* supports a range of various ecosystem components dwelling in sub-alpine to alpine transition zone. With a projected 1°C rise in mean annual temperature, a snowline shift has been predicted as a consequence of global warming. This overall temperature variation will affect plant reproduction, flowering and vegetative characteristics and behaviour of seasonal pollinators. *Rhododendron* species, for instance, are now reported to flower earlier than usual (Xu *et al.*, 2009). Being a key species, any threat to the diversity can pose a risk to the associated biota too. Anthropogenic activities including, habitat destruction and deforestation along with overexploitation and poor management practices exaggerate the scenario further hampering the existence of species and thus, necessitating the conservation of biodiversity as a whole.

Genetic information of a species is significant not only from conservation but from evolutionary point of view, which can be efficiently accounted from a measure of genetic variation within and among populations (Godt and Hamrick, 1998; Park *et al.*, 2009). Diversity in turn is sculptured by factors *viz.,* natural selection, mutation, migration, population size, gene flow and genetic drift. Also, the population structure and evolutionary statistics are vital to recognize the extinction and ecological risks posed by environmental changes (Bruni *et al.*, 2012).

The advent in the era of genomics can be projected to elucidate distribution pattern of genetic resources in order to interpret genetic variation and gene flow in threatened species enlightening towards planning long- term preservation strategies (Boyle *et al.*, 1990). Although numerous genomic markers *viz*., randomly amplified fragment polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and restricted fragment length polymorphism (RFLP), have

found applications among different members of the family Ericaceae (Albert *et al.*, 2003; De Riek *et al.*, 1999; Debnath, 2007; Zawko *et al.*, 2001); but being dominant, their utility is limited. To know more about evolutionary relationships among species, more co-dominant markers are needed to be developed.

SSRs (simple sequence repeats or microsatellites) are being designed and recommended for quantifying genetic diversity within populations; to relate differences in genetic structure and distribution patterns among the species and to assess their implications on plant conservation measures (Guerrant *et al.*, 2004). Defined as a tandem repeat of more than two base pairs and owing to highly polymorphic, genome- wide abundant, co- dominant, multiallelic, locus specific and analytically simple characteristics, SSRs are accredited amongst extensively employed molecular markers in a variety of plant breeding, population genetics, ecological and genomic studies *viz*., genetic diversity, cultivar identification, fingerprinting, marker-assisted selection and linkage mapping (Lioi and Galasso, 2013; McCOUCH *et al.*, 2002; Suwabe *et al.*, 2002). Among various approaches to isolate microsatellite, selective hybridization coupled with screening of genomic insert libraries with a labelled probe and sequencing the positive clones is commonly adopted strategy.

So far relatively meager genetic base is available that can be related to diversity, population structure and other characteristics of *R. arboreum*. Considering the significance of the work, following objectives were undertaken to provide a genetic background for the species:

- Development of enriched genomic library for SSR markers in *Rhododendron arboreum.*
- Characterization of the developed G-SSRs in the selected germplasm.

CHAPTER 2

Review of Literature

Indian Himalayan Region, one of the biodiversity hotspot is a home to thousands of plant species. *Rhododendron*, the largest genus of Ericaceae, is endemic to temperate subalpine and alpine Himalayan zones has ecological and economic significance. The evergreen tree cover adorns the hills with magnificent blooms during spring season. Of the all 1025 species dispersed throughout tropical south-east Asia to north-east Australia (Kron *et al.*, 2002), 87 species (with 12 subspecies and 8 varieties) are concentrated in Indian states of Jammu-Kashmir, Himachal Pradesh, Uttaranchal, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, West Bengal, Arunachal Pradesh, Sikkim and Tamil Nadu and in neighboring countries. A few rare or endangered species of the genera might be at the verge of extinction in near future. Among all the Indian Rhododendrons, *R. arboreum* is most abundant. In addition to aromatic and fuel wood values, the plant is equipped with anti-inflammatory, antioxidative and hepatoprotective properties and is recommended as a remedy for numerous health related disorders by local population (Swaroop *et al.*, 2005).

2.1 Formulation of population structure and genetic diversity

The distribution of populations over time and space is non- random, that also holds true for individuals, genotypes and alleles within a population. Gene flow, genetic drift, natural selection, life history, environment and other ecological constrains determine population structure (Knowles *et al.*, 1992). Gene flow, a phenomenon whereby parental genes reassociates in progeny, shapes this structure by influencing the distribution pattern of individuals and vice versa. Although the rates may vary for different ecologies and even with seasons yet, populations of same species can reflect distinct patterns. Establishment history of a population during different evolutionary periods may cause differences in genetic structure (Boyle *et al.*, 1990).

Population is defined as a dynamic equilibrium of genetic drift and gene flow. A higher gene drift reduces genetic variation within the population, but leads to genetic differentiation among populations (Wolf *et al.*, 2004). Spatial isolation, in turn, prevents gene flow among different sites and gives rise to different allele frequencies and genetic distinction. Although geographical barriers are known to restrict the number and variety of species however, higher endemism is seen in

mountainous areas probably due to favourable conditions, such as higher rainfall nourishing the flora (Eckert *et al.*, 2008). Habitat destruction and less growth densities can lead to genetic isolation among populations. If gene flow is limited due to limited pollen and seed dispersal, it would result in higher level of similarity among individuals forming sub-populations. Clonal propagation or self- pollination is occasionally reported in such cases. On the contrary, long distance dispersal enables germplasm mixing among diverse territories. (Cain *et al.*, 2000; Debout *et al.*, 2011).

2.2 Assessment of risk to the biodiversity

Impact of ecological constraints upon enhancing vulnerability is most pronounced in population of small- size or species located in narrow geographic areas due to natural or human interferences. For illustration, scarcity of suitable habitats promotes self- pollination, clonal propagation or inbreeding in small populations, limiting the level of heterozygosity and long distance gene exchange by hindering seed dispersal and germination (Dick *et al.*, 2008). In addition to gene flow, drift and allele fixation prevent genetic differentiation especially in perennial species (Besnard *et al.*, 2003). This means low survival and hence, low viability or adaptability of populations to changing environments and in extreme cases, extinction (Young *et al.*, 1996). Henceforth, one should recommend *in situ* conservation along with continuous management and monitoring of populations to preserve the diversity in case of such scattered and rare plant species.

From a past few years, *R. arboreum* is gradually becoming a victim of overexploitation due to unsustainable human use, poor management policies, habitat fragmentation and deforestation. An isothermic shift due to climate change aggravates the situation by being a factor responsible for progressively depleting populations of rhododendrons turning dense green forests into barren lands and consequently, threatening the associated biodiversity by devoiding the living world of their natural habitat and making the soil prone to erosion as well (Xu *et al.*, 2009). These risks need to be addressed immediately, requiring a study of genetic structure in depth. Such estimates can provide an insight towards maintaining the overall ecology.

2.3 Genetic diversity and molecular markers

Availability of genetic resources is a key to predict the past and future history of a population (Fenster *et al.*, 2003). In this context, understanding the pattern and level of gene flow is vital to ecological and evolutionary studies. Gene flow can be estimated by comparing genotypes within and among different populations. Population structure of a species provides the necessary statistics for diversity estimations which in turn will facilitate the optimization of management strategies (Escudero *et al.*, 2003). Therefore, a genetic base is prerequisite in planning a conservation programme. Molecular approach can generate significant data in terms of molecular genetics research, especially for a naïve plant species. Formerly adopted morphological characterization and identification is highly ambiguous and is known to be influenced by environment. DNA markers gradually proved their stronghold in elucidating genetic variation in threatened species and to know more about their phylogeny and adaptations that can aid the formulations of long-term conservation or restoration projects.

Various types of dominant markers *viz*., RAPD (Iqbal *et al.*, 1995), AFLP (De Riek *et al.*, 1999; Pornon *et al.*, 2000) and RFLP (Dunemann *et al.*, 1999) have been inferred for DNA fingerprinting and genetic diversity in *Rhododendron* genera as well as in other Ericaceae, such as *Arbutus* (Takrouni and Boussaid, 2010), *Vaccinium* spp. (Albert *et al.*, 2005; Albert *et al.*, 2003; Debnath, 2007), *Leucopogon obtectus* (Zawko *et al.*, 2001). On the other hand, codominant SSRs are highly informative in terms of their polymorphic nature (varying number of repeats) and provide extensive genome coverage (owing to their abundance). These characteristics have contributed to their widespread practice in fingerprinting, inbreeding, population and genetic structure studies (Park *et al.*, 2009). Table 1 enlists a few studies depicting development and characterization using SSRs in different plant species.

Table 1: List of studies concerning development and characterization using microsatellites

S.No	Title	Result	Reference
1.	Isolation and characterization of microsatellites of *Rhododendron metternichii* Sieb. et Zucc. var. *hondoense* Nakai.	SSR loci analysis revealed the mode of reproduction and high genetic polymorphisms in the populations.	(Naito *et al.*, 1999)
2.	Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.)	A highly efficient enrichment protocol used to design SSRs directed a high level of polymorphism in cultivated accessions	(He *et al.*, 2003)
3.	Microsatellite markers for *Vaccinium* from EST and genomic libraries	Microsatellites designed from genomic and EST libraries showed an average of 83% transferability on 12 species.	(Boches *et al.*, 2005)
4.	Eight microsatellite markers for sympatric alpine shrubs, *Phyllodoce aleutica* and *P. caerulea* (Ericaceae)	SSRs developed for the alpine shrubs found in northern Japan were polymorphic and amplified in other species as well. Except for few, majority of the loci showed significant linkage disequilibrium in both the species.	(Kameyama *et al.*, 2006)
5.	Characterization of microsatellite loci in the myco-heterotrophic plant *Monotropa hypopitys* (Ericaceae) and amplification in related taxa	A first report on development of SSR markers (Glenn and Schable, 2005) revealed polymorphism in the red and yellow coloured varieties of this non- photosynthetic herb and were also characterized in other generic and congeneric species.	(Klooster *et al.*, 2009)
6.	Development and characterization of eight polymorphic microsatellites for *Rhododendron simsii* Planch. (Ericaceae)	A high level of heterozygosity without any significant linkage disequilibrium was observed both at species and population level from the SSR analysis	(Tan *et al.*, 2009)
7.	Development and characterization of 15 microsatellite loci for *Rhododendron delavayi* Franch. (Ericaceae)	SSRs were quantified with high levels of heterozygosity and no linkage disequilibrium in *R. delavyi* and demonstrated successful cross-species amplification in *R. agastum* and *R. decorum*	(Wang *et al.*, 2010)
8.	Isolation and characterization of microsatellite loci for *Menziesia goyozanensis*, an endangered shrub species endemic to Mt. Goyo in Northern Japan	GSSRs designed were characterized for the species with high level of heterozygosity and no linkage disequilibrium. The markers also showed polymorphism in a close relative, *M. pentandra.*	(Abe *et al.*, 2011)
9.	Development of microsatellite markers in a Riparian shrub, *Spiraea thunbergii* (Rosaceae)	Successful application of SSR primers to study population structure and genetic diversity in the threatened shrub that also showed transferability on other congeneric species	(Ashizawa *et al.*, 2012)
10.	Genomic and EST microsatellites for *Rhododendron aureum* (Ericaceae) and cross-amplification in other congeneric species	In *R. aureum,* GSSRs developed using FIASCO protocol and ESTs using *Rhododendron* dbEST database of GenBank were characterized with effective amplification in *R. brachycarpum* and *R. dauricum*.	(Li *et al.*, 2011)

11.	Analysis of genetic relationship among *Arbutus unedo* L. genotypes using RAPD and SSR markers	Both RAPD and SSRs (designed for *Vaccinium* spp.) evidenced high genetic variability in *A. unedo* genotypes irrespective of any geographical correlation and found latter to be more polymorphic and informative.	(Gomes *et al.*, 2013)
12.	Development of genomic simple sequence repeat markers from an enriched genomic library of grass pea (*Lathyrus sativus* L.)	Affinity capture technique was employed to develop an SSR enriched library. Characterization revealed polymorphism in four different grass pea germplasm accessions and high transferability among *L. cicera*, *L. ochrus*, *L. tingitanus* and *Pisum sativum*	(Li *et al.*, 2011)
13.	Development and characterization of microsatellite markers for *Morus* spp. and assessment of their transferability to other closely related species	Genomic and genic SSRs reported high polymorphic and heterozygosity levels among mulberry species and genotypes. A majority of the primers exhibited transferability to other related species like *Ficus*, fig and jackfruit.	(Balachandran *et al.*, 2013)
14.	Mining new microsatellite markers for Siberian apricot (*Prunus sibirica* L.) from SSR- enriched genomic library	Natural variation in apricots assessed with polymorphic SSRs inferred high polymorphism information content (PIC) in apricot and related species.	(Wang *et al.*, 2014)

Amongst various enrichment cloning techniques available for development of SSRs, selective hybridization on nylon membrane or streptavidin- coated magnetic bead (Armour *et al.*, 1994; Karagyozov *et al.*, 1993; Kijas *et al.*, 1994; Nunome *et al.*, 2006; Prochazka, 1996; Sekar and Srivastava, 2010), FIASCO (Fast Isolation by AFLP of Sequences COntaining repeats) protocol based on the AFLP procedure (Vos *et al.*, 1995) are prevalent and have been optimized accordingly by various researchers (Zane *et al.*, 2002). Briefly, small fragments generated either by mechanical or enzymatic (one or a combination of different restriction endonucleases) disruptions of genomic DNA may either be size selected on agarose gel first or, ligated to adaptor, amplified first and then size selected. After this, there is crucial step of hybridizing SSR containing fragments with a labelled probe followed by removal of nonspecific binding. The eluted DNA, finally said to be enriched, is cloned into a suitable vector either by using a restriction site on the known flanking regions or by TA cloning. Transformed recombinant clones obtained can be directly sequenced or screened for the presence of repeats using Southern blotting or PCR strategies (probe as one of the three primers), termed as secondary enrichment depending on the efficiency of the whole procedure. Finally the positive clones can be sequenced to design the primers from the region flanking the repeat motif.

Compared to the dominant marker systems, codominant SSRs are advantageous in targeting true population genetic structure (Table 2). This includes an accurate phylogenetic classification of species.

Table 2: List of genetic structure studies using microsatellites

S.No	Title	Result	Reference
1.	Patterns and levels of gene flow in *Rhododendron* metternichii var. *hondoense* revealed by microsatellite analysis	Designed SSRs were highly variable and analysis proved a low level of gene flow among subpopulations but long-distance gene flow from outside the quadrat responded by vegetative, geographical, reproductive and genetic characteristics within subpopulations.	(Kameyama *et al.*, 2001)
2.	Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*	A set of primers developed with 63% efficiency displayed transferability between two species. Syntenic groups were defined based on their map position.	(Brondani *et al.*, 1998)
3.	Genetic diversity and phylogenetic relationships of the endangered species *Vaccinium sieboldii* and *Vaccinium ciliatum* (Ericaceae)	Inspite of large geographical distance, *V. sieboldii*, *V. ciliatum* and *V. oldhami* share close genetic diversity as observed from SSR data. Dendrograms grouped the populations based on high bootstrap.	(Hirai *et al.,* 2010)
4.	Development and characterization of polymorphic microsatellite markers for *Conopholis americana* (Orobanchaceae)	From the primer loci designed, low levels of heterozygosity and deviations from linkage equilibrium was reported, concluding that self-pollination dominated among the species	(Rodrigues *et al.*, 2012)
5.	Analysis of genetic relationship among *Arbutus unedo* L. genotypes using RAPD and SSR markers	Both RAPD and SSRs (designed for *Vaccinium* spp.) evidenced high genetic variability in *A. unedo* genotypes irrespective of any geographical correlation and found latter to be more polymorphic and informative.	(Gomes *et al.*, 2013)
6.	Development and application of 15 novel polymorphic microsatellite markers for sect. *Paeonia* (*Paeonia* L.)	Magnetic bead enrichment generated SSRs that enabled the distinction of species from cultivars. Cluster analysis was also helpful in distinguishing wild and cultivated Paeonies.	(Ji *et al.*, 2014)
7.	Population structures in *Rhododendron metternichii* var. *hondoense* assessed with microsatellites and their implication for conservation	SSR loci study in two populations growing in different habitats resolved polymorphisms within and among the mesic valley and mountain ridge populations advising sexual propagation with respect to its conservation.	(Naito *et al.*, 1999)
8.	Genetic diversity in the highbush blueberry evaluated with microsatellite markers	National Clonal Repository was managed through genotyping of the wild and cultivated germplasm distinguishing the accessions into separate groups.	(Boches *et al.*, 2005)

9.	Spatial genetic structure in the Laperrine's olive (*Olea europaea* subsp. *laperrinei*), a long-living tree from the central Saharan mountains	This locally threatened olive species was clustered into two distinct groups based on the genomic SSR data confirming the geographical barrier limiting the gene flow and also suggesting conservation measures	(Besnard *et al.*, 2003)
10.	Genetic variability of relict *Rhododendron ferrugineum* L. populations in the Northern Apennines with some inferences for a conservation strategy	Low gene flow was found between the Apennine and Alpine populations, with less heterozygosity in the former, and molecular data also aided clustering into separate groups. A combination of *ex* and *in situ* conservation measure was proposed in view of geographical and ecological conditions of these populations and impact of climate changes on their habitat.	(Bruni *et al.*, 2012)

CHAPTER 3

Materials and Methods

3.1 Plant sampling and DNA extraction

Three populations of *R. arboreum*, a total of 30 individuals were sampled from three geographical locations (Table 3) within Himachal Pradesh, India. The positions on map are presented in Figure 1.

Table 3: Geographical locations of 30 sampled individuals of *R. arboreum*

Sample No.	Location	Latitude	Longitude
1.	Sadar Bazar, Dalhousie	32°32´03.46´´N	75°57´57.42´´E
2.	Garam Sadak, Dalhousie	32°32´03.05´´N	75°58´46.94´´E
3.	Subash Chowk, Dalhousie	32°32´06.87´´N	75°58´03.29´´E
4.	Dalhousie bus stand	32°32´03.46´´N	75°59´00.62´´E
5.	Swas boli, Dalhousie	32°32´05.42´´N	75°59´01.57´´E
6.	Tikka lane, Dalhousie Cantt.	32°32´41.55´´N	75°57´19.74´´E
7.	Chandri Ghat, Dalhousie	32°32´00.47´´N	75°58´06.47´´E
8.	Subash Bowli, Dalhousie	32°32´08.45´´N	75°59´07.53´´E
9.	Panchpula, Dalhousie	32°32´03.05´´N	75°58´46.94´´E
10.	Bihar Regiment, Dalhousie Cantt.	32°32´40.55´´N	75°57´18.74´´E
11.	Near Ghatasni bus stand	32°28´37.24´´N	75°55´52.43´´E
12.	Ghatasni bus stand	32°28´38.25´´N	75°55´54.45´´E
13.	Ghatasni Road towards Padhar	32°28´40.28´´N	75°55´56.48´´E
14.	Jhitkari Village, Tehsil Padhar	32°28´40.41´´N	75°55´41.81´´E
15.	Tindunala,Tehsil Padhar	32°28´35.07´´N	75°55´52.06´´E
16.	Fenul glu, Tehsil Padhar	32°28´39.54´´N	75°55´56.01´´E
17.	Tikkin Village, Tehsil Padhar	32°28´38.70´´N	75°55´39.04´´E
18.	Bradhan bus stand, Tehsil Padhar	32°08´22.14´´N	75°55´14.16´´E
19.	Bradhan Village, Tehsil Padhar	32°08´22.18´´N	75°55´14.19´´E
20.	Bradhan Village, Tehsil Padhar	32°08´22.38´´N	75°55´14.29´´E
21.	Multhan, Tehsil Multhan, Distt. Kangra	32°04´28.01´´N	76°52´06.94´´E
22.	Multhan, Tehsil Multhan, Distt. Kangra	32°04´28.09´´N	76°52´06.14´´E
23.	Dayot, Tehsil Multhan, Distt. Kangra	32°28´35.07´´N	76°55´52.06´´E
24.	Dayot, Tehsil Multhan, Distt. Kangra	32°28´35.14´´N	76°55´52.19´´E
25.	Garla, Tehsil Multhan, Distt. Kangra	32°04´47.15´´N	76°52´27.48´´E
26.	Garla, Tehsil Multhan, Distt. Kangra	32°04´47.01´´N	76°52´27.11´´E
27.	Bada Gaon, Tehsil Multhan, Distt. Kangra	32°06´25.09´´N	76°55´40.45´´E
28.	Bada Gaon, Tehsil Multhan, Distt. Kangra	32°06´25.15´´N	76°55´40.15´´E
29.	Bada Gaon, Tehsil Multhan, Distt. Kangra	32°06´25.22´´N	76°55´40.02´´E
30.	Bada Gaon, Tehsil Multhan, Distt. Kangra	32°06´25.28´´N	76°55´40.12´´E

Genomic DNA was isolated from young leaves by modified CTAB protocol (Doyle and Doyle, 1987) given in Appendix B. Quantity and quality of DNA was assessed on 0.8% agarose gel and Nanodrop- 2000 spectrophotometer (Thermo Scientific, USA).

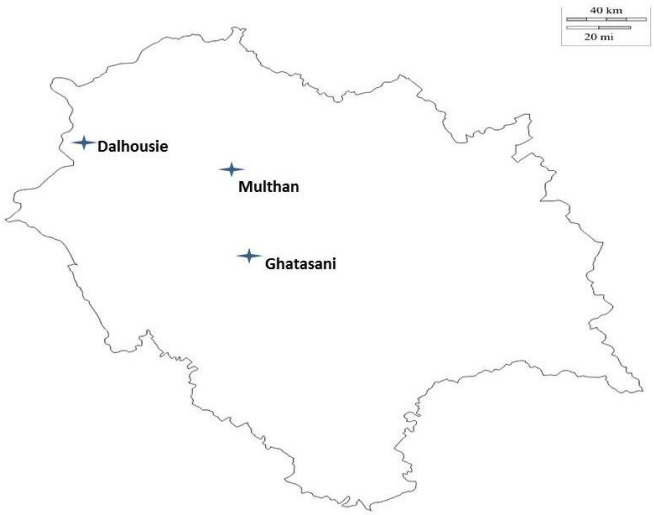

Figure 1: Location of sampling sites on the map of Himachal Pradesh, India

3.2 Construction of microsatellite enriched library

A single SSR library enriched for dinucleotide repeats (GA) was constructed from genomic DNA using biotin streptavidin capture protocol (Bhardwaj *et al.*, 2013) with some modifications. After digesting 5µg DNA with two restriction enzymes (MseI and EcoRI, New England Biolabs, USA) at 37°C, fragments ranging from 250- 1000 bp size were eluted from 1.2% agarose gel and purified by Promega gel extraction kit. The eluted fragments were ligated to the Eco and Mse double stranded adapters (Appendix C) at their respective restriction site by T4 DNA ligase (New England Biolabs, USA) and amplified using selective primers (Appendix C). The DNA was diluted, denatured at 95˚C then, allowed to hybridize to biotinylated probe (GA_{10}) at 65˚C for 4hrs followed by subsequent capturing of microsatellite containing fragments with streptavidin coated magnetic beads at room temperature for 20 min with continuous agitation. Highly stringent washing consecutively with 2X, 1X, 0.5 X SSC buffer (supplemented 0.1% SDS) at 65˚C and then at room temperature aided the removal of unbound fragments. Finally DNA enriched with microsatellites was separated from beads by denaturation at 95˚C, eluted in MilliQ water and amplified using the above stated primers.

Enriched amplified DNA was cloned into chemically competent *Escherichia coli* cells using TA cloning kit (Promega). Blue- white screening method was used to distinguish transformed host colonies. For secondary enrichment, the colonies

containing recombinant vector were identified by colony PCR using Eco, Mse and probe (GA_{10}) primers (Appendix C). The positive clones were selected by electrophoresis on 1.5% agarose gel (Figure 2). Plasmid was isolated from positive clones (displaying more than single band) using Mini Prep kit (Nucleopore) and sequenced with universal M13 sequence as primer on ABI 3730xl DNA Analyzer (Applied Biosystems).

3.3 Primer designing

Sequences containing SSR regions were identified with the help of online available SSR Identification Tool (SSRIT) (Temnykh *et al.*, 2001) with tetramer as maximum motif length and five as minimum number of repeats. Sequences were classified as perfect, compound or interrupted motifs of AG, AC, ACA repeats (or their complements). Raw sequences were cleared from plasmid and adapter sequences and checked for redundancy and duplicity by ClustalW (http://www.genome.jp/tools/clustalw). Primers were designed only from the unique, sufficiently long sequences flanking the SSR region using Primer3 software (Untergasser *et al.*, 2012) keeping in mind the primer size, product size, annealing temperature and GC content.

3.4 Characterization of SSR primers

Genetic diversity of three populations of *R. arboreum* was analyzed using SSRs. Amplification reactions were carried out in a total volume of 10µl, constituting 25 ng of template DNA, 0.3 U *Taq* DNA polymerase (Bangalore Genei^{TM}), 1X PCR buffer (1mM Tris pH 9.0, 50 mM KCl, 0.01% gelatin, 1.5 mM $MgCl_2$), 2.5 mM of each dNTPs, 5-10 ng each of forward and reverse primer in T100 Thermal Cycler (BIORAD). PCR profile comprised of an initial denaturation step at 94°C for 3 min followed by 35 cycles of denaturation at 94°C for 1 min, annealing at required T_m for 1 min and elongation at 72°C for 1 min and a final elongation at 72°C for 8 min. PCR products were quantified on 6% urea polyacrylamide gels. Reactions were mixed with equal volume o*f 10X sample loading buffer (Appendix B), denatured at 94°C for 3 min and electrophoresed at 60W for desired time depending on size of the fragment and visualized by silver staining (Figure 3a and 3b).

3.5 Data Analysis

Profiles for each of the SSR primer pair were scored as a single band signifying a single allele. Data analysis was done through Popgene version 1.32 (Yeh *et al.*, 1999). Overall diversity parameters were determined which included, total (N_a) and effective number (N_e) of alleles per locus, observed heterozygosity (H_o), expected heterozygosity (H_e), deviations from Hardy- Weinberg equilibrium (HWE) and linkage disequilibrium (LD), Nei's gene diversity analysis, Shannon's Informative Index (I) and dendrograms. Polymorphic information content (PIC) and percentage polymorphism was calculated, which gives an estimate of discrimination ability of a locus.

Population wise diversity parameters and genetic distance were predicted from GeneAlEx version 6.5 (Peakall and Smouse, 2012). Fixation index, a measure of population differentiation, was estimated based on variance of allele frequencies in populations and its significance among the population was tested using a hierarchical, nonparametric permutation approach. The heterozygosity excess or deficiency was determined using BOTTLENECK (Luikart *et al.*, 1998). The phylogenetic relationship among populations was depicted from DARwin version 5.0.158 and STRUCTURE version 2.3.4 (Pritchard *et al.*, 2000). Based on multi-locus Bayesian analysis, the hidden population clusters were distinguished and K values determined (Evanno *et al.*, 2005).

CHAPTER 4

Results

4.1 Assessment of Enriched Library

A library enriched for microsatellites, comprising of 1200 transformed clones, was developed for *R. arboreum.* The enriched library was screened for the presence of clones containing recombinant vector by three- primer amplification reactions on bacterial colony suspensions. The clones showing more than one distinct bands on 1.5% agarose gel were selected (Figure 2). Based on the length of the fragment amplified, 30% of the resulting 500 positive clones were considered unsuitable for sequencing. This reduced the overall cost of primer development by reducing the number of clones to be sequenced.

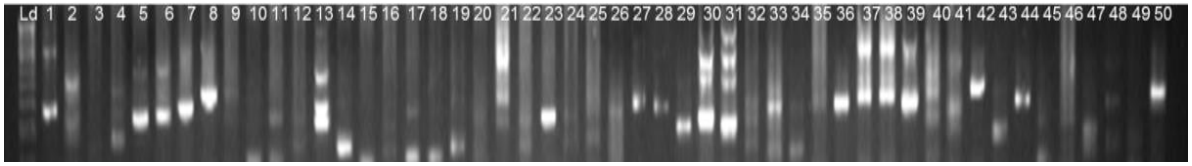

Figure 2: Amplification profile generated using three-primer PCR. Lanes 1- 50 depict bacterial clones and more than single band represent SSR- containing clones; Ld: 100bp DNA ladder as size standard (Bangalore Genei^{TM}).

A total of 351 clones were sequenced, out of which 191 sequences (54.4%) were identified with di- and tri- repeats of length varying from 5-23 using SSRIT. A statistic of genomic library constructed for *R. arboreum* is represented in Table 4.

Table 4: Statistic for library enriched for G-SSR development for *R. arboreum*

FACT	TOTAL
Total number of positive colonies obtained after blue- white screening	1200
Number of clones sequenced (percentage of the whole library)	351 (29.3%)
Clones containing SSR loci (percentage of clones sequenced)	191 (54.4%)
Non-suitable SSR clones (percentage of clones sequenced)	150 (78.5%)
Unique SSR clones (percentage of clones sequenced)	41 (11.7%)

A threshold of 10 bp was considered as repeat motif. Majority (93.7%) of the SSRs consisted of dinucleotide motifs of the type used for enrichment (AG/GA or CT/TC), followed by AC/CA (4.7%) and AT/TA (1.1%) repeats. Only a single trinucleotide motif (TGT) was sequenced. Frequency and type of motifs observed across the enriched library is presented in Table 5. The repeat number of motifs ranged from 5 to 23 and contained >50% compound or interrupted repeats.

Of the 191 SSR- containing clones, 78.5% were rendered unsuitable for amplification reactions either due to presence of SSR region close to either ends or sequence was too short to design the primer.

Table 5: Frequency and types of SSR motifs found across enriched library for *R. arboreum*

Type	Number	Percentage
Dinucleotide SSRs	190	
AG/GA (CT/TC)	179	94.2%
AC/CA (GT/TG)	8	4.2%
AT/TA (GC/CG)	3	1.5%
Trinucleotide SSRs	1	

4.2 Microsatellite evaluation

Only 41 unique sequences were of adequate size to design forward and reverse SSR primers using Primer3, with an expected PCR product size of 100-300 bp, GC content of 41- 55%, melting temperature (T_m) of 50- 60°C, and primer lengths of 18- 25 bp (Table 6). Of the 41 primer pairs, 40 showed successful amplifications in germplasm. 38 of all the amplified loci (95%) were polymorphic across all the three populations of *R. arboreum* and detected an overall allele number (including null alleles) ranging from 3-15 (average 4.8). Figure 3a and 3b depicts the amplification profile generated with primer RA267 and RF29. 16 (42%) of these 38 loci contained perfect repeats and rest were compound or interrupted repeats of AG/GA or CT/TC motif. The sequences were submitted to the NCBI database (accession numbers given in Table 6).

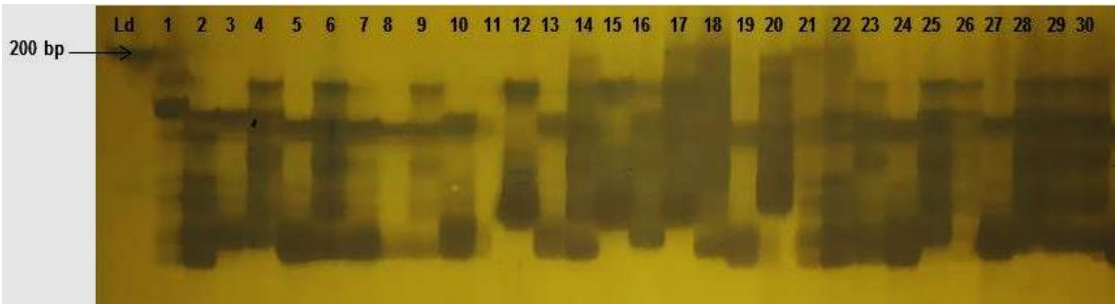

Figure 3a: Amplification profile generated with primer RA267. Lanes 1- 30 represent sampled individuals of *R. arboreum* as presented in Table 3; Ld: 100 bp DNA ladder as size standard (Bangalore Genei™).

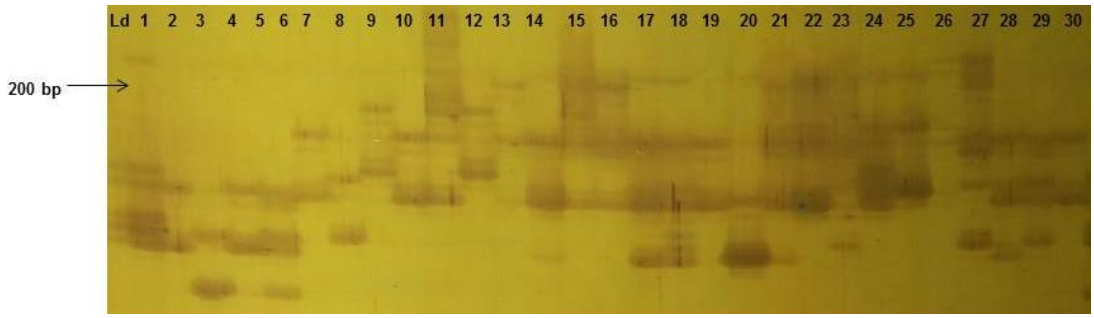

Figure 3b: Amplification profile generated with primer RF29. Lanes 1- 30 represent sampled individuals of *R. arboreum* as presented in Table 6; Ld: 100 bp DNA ladder as size standard (Bangalore Genei^{TM}).

4.3 SSR marke r characterization:

Several parameters were used to examine genetic diversity of *R. arboreum* populations. Total number of alleles (except null) ranged from 2- 14 (average 5.2) with observed fragment length of range 100- 420 bp. Estimated PIC value varied within the range of 0.104 to 0.911 (average 0.464). Observed heterozygosity (H_o) varied from 0.167 to 0.933 (average 0.523) and expected heterozygosity (H_e) ranged from 0.422 to 0.917 (average 0.723). 19 loci reported significant deviations from HWE (Table 6) which might be due to natural selection, genetic drift, gene flow and mutations. As a whole, no significant linkage disequilibrium was detected at population level ($P<0.05$).

Population wise diversity parameters were estimated using GenAlEx (Table 7). The highest values of N_a, N_e, H_o, H_e and I were observed for the population from Ghatasani. Nei's genetic diversity (F) was found highest for the population of Dalhousie (Table 8).

Table 8: *R. arboreum* populations with percentage of polymorphic loci and their genetic variability parameters detected by SSR analysis

Population	Sample Size	N_e	F	I	Percentage of polymorphic loci
Dalhousie	10	3.362	0.283	1.251	97.37%
Ghatasani	10	3.654	0.264	1.374	100.00%
Multhan	10	3.547	0.254	1.296	100.00%

N_e- Effective number of alleles; F- Nei's genetic diversity; I- Shannon's Informative Index

Table 6: Characteristics of 38 SSR loci developed from *Rhododendron arboreum*

S.No.	Locus	Primer Sequence	Repeat motif	T_a	N_a	Approx.Size Range (bp)	Heterozygosity		PIC	Accession No.
							H_o	H_e		
1.	R394	F: GGAAAGTGTGGGTGTTAGTGC R: TTGAGAGATGGCGAGAGAGAG	$(TC)_{16}$	59 ˚C	4	145-165	0.6667*	0.7181	0.104	KJ851157
2.	RE101	F: GACGGGAATGAGCAAGGTTG R: CTTCAATTCTGCAAGCCCGA	$(AG)_{16}$	55 ˚C	6	210-240	0.7333*	0.8418	0.480	KJ851180
3.	RF87A	F: TGGGTCATGTTCTGGAAGGT R: TGAACTAACCCTAGCCACACT	$(AG)_{10}(AG)_{10}(GA)_9$	55 ˚C	6	140-170	0.5677***	0.7689	0.436	KJ851183
4.	RA50	F: ACTCCCTCCTGTCGTTCCTT R: AATCGTGCATCCGTATCCTG	$(TC)_{16}$	58 ˚C	5	216-234	0.6667	0.7588	0.310	KJ851162
5.	RA324	F: GCGTACAACATGCCCAAATA R: CCCTGTTCTCATTGCTCACA	$(AG)_8(GA)_8(GA)_9$	55 ˚C	2	184-194	0.4667	0.7994	0.709	KJ851168
6.	RA351	F: TGTCGCTCTCTCACTGATCG R: TTTGTAGTTTTCCCGTGTCCTT	$(AG)_{12}(AG)_{11}$	59 ˚C	5	338-360	0.5677*	0.7525	0.374	KJ851170
7.	RA346	F: CGGAGCAAGCTCTCTTATCG R: CCTCTCCTGTGTAGCAAGTCG	$(TC)_9$	59 ˚C	8	100-116	0.5333*	0.7023	0.396	KJ851184
8.	RA321	F: AGAGATGGGTTTGTGTAAAGTCTG R: TATTTCGCTGCCACCCTAAC	$(GA)_9$	55 ˚C	4	264-280	0.4677*	0.7124	0.373	KJ851185
9.	RA470	F: AGGGACAAGAAGAAGCCACA R: TCGCGCTTATTACAGCTCTTC	$(GA)_{14}(AG)_{10}(AG)_{10}$	55 ˚C	3	150-154	0.3000**	0.5994	0.438	KJ851173
10.	RF304	F: TCCTAGGGTTTGTTCGCAAT R: TGCTAGCGATTTCCTAGGGT	$(AG)_{13}(GT)_9(AG)_5$ $(GT)_9$	55 ˚C	5	218-230	0.2333*	0.7339	0.811	KJ851178
11.	RA254	F: AGTAGCAACACCCACACACT R: GGAGGGGCTGTAGTCTGATT	$(CT)_{16}(CT)_{10}$	55 ˚C	7	150-164	0.8000*	0.8260	0.413	KJ851179
12.	RA337	F: GAGCGAGAGAGAGGTGTTGG R: ATTCACGGGAATCTTCACCA	$(AG)_6$	59 ˚C	2	206-208	0.2000	0.4316	0.266	KJ851169
13.	RA267	F: ACGGAGAAGCAGTGAGCATT R: TGCACAGGAACACCCAATAA	$(GA)_{11}(AG)_{10}$	59 ˚C	3	196-200	0.3000**	0.7102	0.510	KJ851163
14.	RA272A	F: GCCCCGGTGACTCATAAAAT R: TGGTACAAGTGGGACACGA	$(CT)_8(CT)_{11}$	59 ˚C	3	188-194	0.2667*	0.7525	0.683	KJ851164
15.	R460	F:CCCTACTTCTTTCATCACATACAA R: CAACTCCGGTCATTTTTGGT	$(GA)_{13}(AG)_{12}$	59 ˚C	5	188-196	0.4333	0.7345	0.621	KJ851158
16.	R97	F: AGCAGCAACAATGGTGTCC R: TCTAGAAGGCCTCCCATTCC	$(AG)_{10}(AG)_{13}$	59 ˚C	2	188-190	0.5000	0.6672	0.381	KJ851160
17.	RA430	F: GCGTAAATCGAGTTCGGAAG R: CTCTCTCTAATCGAATTCCCG	$(TC)_{10}(CT)_6$	59 ˚C	7	166-180	0.6333	0.8627	0.610	KJ851171
18.	R356	F: GAGCTAAGCACGCCGTATTC R: AAATTCGACGGCAAAGAGG	$(TC)_9$	59 ˚C	4	186-194	0.6667*	0.6774	0.104	KJ851166
19.	RA134	F: GGAGAGAGAGGCCGAGAGAG R: ACGTCGCTCTTGTCAAGCAT	$(GA)_6(GA)_6(AG)_8$	59 ˚C	3	220-228	0.3000	0.4220	0.140	KJ851167

20.	RF103	F: GATAGAGAGACAGGGGCAGC R: TGTACGCCAAGACTCCCATT	$(GA)_{14}$	59 ˚C	5	288-300	0.5333	0.8023	0.482	KJ851181
21.	RF29	F: ACAGACAGAAGCAGCGGAAC R: AAGGGGAGGAGATCGAGTTG	$(GA)_{10}(AG)_{13}$	59 ˚C	9	138-160	0.9333	0.8254	0.266	KJ851174
22.	RF43	F: AATTCGATGGGTTGGTGGTA R: GCCTTCTCTGTTCTCGGTTTT	$(AG)_{10}$	59 ˚C	4	190-200	0.3667*	0.5294	0.237	KJ851175
23.	RF245	F: GGGTTTTTGATCTTCATACGG R: AATCGGTTCAAGAGGGGTTT	$(AG)_{11}$	55 ˚C	8	198-220	0.5333	0.8576	0.693	KJ851177
24.	RF87B	F: GGGGAAAGGTCATTGGAGAT R: TTCTGAACTAACCCTAGCCACA	$(GA)_{10}(GA)_{10}(AG)_{10}$	55 ˚C	3	310-316	0.4000*	0.7249	0.536	KJ851183
25.	RF98	F: AATCCCATCCCCTAACTTGG R: CCGTGGCTTTACCTTTCACT	$(AG)_{22}$	59 ˚C	10	170-196	0.6667	0.9169	0.720	KJ851182
26.	RA7	F: GTCTACAATGCTTGCTTCCG R: CCTTATTTATTCTCTCTCT	$(AG)_{10}(AG)_{9}$	55 ˚C	5	110-120	0.7333*	0.7881	0.370	KJ851150
27.	RA19	F: AGCCAAAAACTTTCTTCTTTCC R: CTGTCGGCTGTCAGAGTTGA	$(GA)_{19}$	55 ˚C	4	124-138	0.5333	0.7073	0.381	KJ851151
28.	RF71	F: GCGTACAACATGCCCAAATA R: GTCGTTGCAGTTCAATCTCG	$(AG)_{12}$	55 ˚C	5	300-312	0.4667	0.8147	0.710	KJ851176
29.	RA443	F: CCATGCCTGAAGCAAACAC R: AGACTCCAAAGTCCTATCTGTGC	$(AG)_{12}$	59 ˚C	8	184-200	0.4000*	0.8282	0.649	KJ851162
30.	RA272B	F: ATGCAATGGAAATGGGAAAG R: GGAACGGGTAATTCGGATCT	$(CT)_{8}(CT)_{11}$	59 ˚C	3	184-194	0.3333	0.5085	0.194	KJ851164
31.	R372	F: GGTTGGGTGGATGGAGTAAC R: GCAATTTGCATAGCACTGTAAT	$(AG)_{15}(AG)_{15}$	55 ˚C	5	230-244	0.4667	0.8237	0.600	KJ851156
32.	R422	F: GCGGTACTGTTCCGATCAC R: TCCCAGCTCATCCACACATA	$(AG)_{12}(GA)_{13}$	55 ˚C	12	146-172	0.7333	0.8593	0.624	KJ851155
33.	R304	F: TCCTAGGGTTTGTTCGCAAT R: GCGTATTGTTGCACGAAAAA	$(AC)_{9}(CT)_{12}(TC)_{5}$ $(AC)_{9}(TC)_{13}$	55 ˚C	4	380-410	0.3000**	0.6582	0.482	KJ851161
34.	R79	F: AACGTGAAAACTGAGCAGAAC R: CCCGTGGGTAGGAAAATCAT	$(AG)_{10}(AG)_{5}(AG)_{10}$ $(AG)_{7}$	55 ˚C	4	236-250	0.2667	0.5316	0.358	KJ851159
35.	RA276	F: GCCAAAAAGCATCAAGTCGT R: TGTGATTTTTGTGTTGGATGG	$(TC)_{11}$	55 ˚C	6	390-420	0.4000*	0.6605	0.421	KJ851165
36.	R25	F: CCAACAACCCGAGAAAAAGA R: AGTGGGTTTCCGAGACAAAG	$(AG)_{12}(GA)_{11}$	55 ˚C	14	164-200	0.7667	0.9119	0.709	KJ851154
37.	RA30	F: TGTGATTTTTGTGTAGGATGGT R: AGCAACCCACTTCTCCTCTTC	$(AG)_{17}$	55 ˚C	2	150-160	0.1667*	0.5814	0.651	KJ851152
38.	RA54	F: CAAAAATGGCCAACAAGGAT R: ATTGCCTCCATACACAACCA	$(CT)_{5}(CT)_{6}$ $(CT)_{17}$	55 ˚C	2	190-200	0.5000	0.6605	0.429	KJ851153

T_a- Annealing temperature; N_a- Total number of alleles; H_o- Observed heterozygosity; H_e- Expected heterozygosity; PIC- Polymorphic information content; Significant deviations from Hardy- Weinberg equilibrium at *p<0.05, **p<0.01, ***p<0.001

Table 7: Diversity parameters for three populations of *R. arboreum*

LOCUS	DALHOUSIE (N=10)			GHATASANI (N=10)			MULTHAN (N=10)		
	N_a	H_o	H_e	N_a	H_o	H_e	N_a	H_o	H_e
R394	4	0.900	0.655	3	0.500**	0.645	4	0.800	0.690
RE101	4	0.700	0.615	6	0.700	0.755	5	0.800	0.780
RF87A	4	0.600	0.595	6	0.600**	0.750	6	0.500**	0.780
RA50	6	0.700*	0.815	5	0.500	0.685	5	0.800	0.630
RA443	4	0.200*	0.730	7	0.700	0.810	6	0.300*	0.760
RA351	5	0.500	0.630	5	0.600	0.665	5	0.600	0.725
RA346	5	0.500	0.600	3	0.500**	0.645	6	0.500	0.685
RA321	4	0.600	0.695	4	0.500*	0.615	3	0.300*	0.665
RA470	4	0.200	0.585	3	0.200*	0.460	3	0.500**	0.595
RF304	3	0.000***	0.340	4	0.300	0.625	5	0.400**	0.745
RA254	7	1.000*	0.780	7	0.800	0.750	6	0.600	0.685
RA337	3	0.100*	0.395	3	0.200*	0.460	3	0.300*	0.405
RA267	4	0.200*	0.595	4	0.200**	0.740	3	0.500	0.645
RA272A	3	0.000***	0.620	4	0.300	0.700	3	0.500	0.620
R460	4	0.500*	0.715	6	0.600	0.710	4	0.200***	0.660
R97	3	0.500**	0.655	3	0.600**	0.620	3	0.400**	0.620
RA430	8	0.600	0.810	7	0.700**	0.830	7	0.600	0.810
R356	3	0.700	0.560	4	0.400	0.565	2	0.900**	0.495
RA134	2	0.300	0.255	4	0.200	0.465	4	0.400	0.480
RF103	5	0.700	0.725	6	0.700	0.765	5	0.200***	0.760
RF29	7	0.900	0.790	8	1.000	0.770	8	0.900	0.825
RF43	1	0.000[M]	0.000	5	0.600*	0.740	4	0.500	0.535
RF245	9	0.700	0.815	8	0.400	0.805	7	0.500	0.835
RF87B	4	0.300	0.700	4	0.400	0.580	4	0.500	0.625
RF98	9	0.400*	0.880	10	0.700	0.865	10	0.900	0.855
RA7	4	0.600*	0.715	5	0.900	0.720	5	0.700*	0.745
RA19	5	0.500	0.720	5	0.700	0.680	4	0.400	0.580
RF71	5	0.200***	0.680	6	0.500	0.765	6	0.700	0.810
R443	3	0.100*	0.395	3	0.400*	0.560	3	0.500**	0.505
R372	5	0.600	0.775	6	0.500	0.740	5	0.300**	0.785
R422	9	0.600*	0.755	7	0.600	0.785	9	1.000	0.865
R304	3	0.500**	0.665	5	0.300*	0.485	3	0.100**	0.645
R79	5	0.400	0.630	5	0.300**	0.675	2	0.100	0.095
RA324	6	0.600*	0.815	6	0.600*	0.800	5	0.200*	0.665
RA276	5	0.500	0.540	3	0.300*	0.615	3	0.100*	0.485
R25	11	0.800	0.850	11	0.800	0.895	10	0.600	0.870
RA30	2	0.000**	0.480	3	0.300**	0.605	3	0.200*	0.460
RA54	3	0.700***	0.665	3	0.400**	0.620	3	0.400**	0.620
Mean	**4.763**	**0.471**	**0.638**	**5.184**	**0.513**	**0.683**	**4.789**	**0.492**	**0.659**

N_a- Total number of alleles; H_o- Observed heterozygosity; H_e- Expected heterozygosity; M- Monomorphic; Significant deviations from Hardy- Weinberg equilibrium at *p<0.05, **p<0.01, ***p<0.001

Dendrogram based on Nei's genetic distance method (UPGMA modified from Neighbour procedure of PHYLIP version 3.5) displayed two clusters, one consisting of Dalhousie population and the other of Ghatasani and Multhan (Figure 4) which is consistent with genetic distance and gene flow data for the populations (Table 9).

Table 9: Populations with maximum gene flow

POPULATION	F_{st}	N_m
Dalhousie- Ghatasani	0.052	0.723
Dalhousie- Multhan	0.065	0.512
Ghatasani- Multhan	0.046	0.921

F_{st}- Genetic differentiation between subpopulations relative to the total genetic diversity;
N_m- Estimate of gene flow from F_{st}

According to Nei's original measure of genetic distance (1978), Dalhousie population was the most differentiated of all. Genetic identity confirms the genetic distance findings that Multhan and Ghatasani populations are more similar to each other genetically (Table 10).

Table 10: Nei's Original Measures of Genetic Identity (upper diagonal) and Genetic distance (lower diagonal)

Pop Id	Dalhousie	Ghatasani	Multhan
Dalhousie	---	0.7925	0.7491
Ghatasani	0.2350	---	0.8018
Multhan	0.2889	0.2209	---

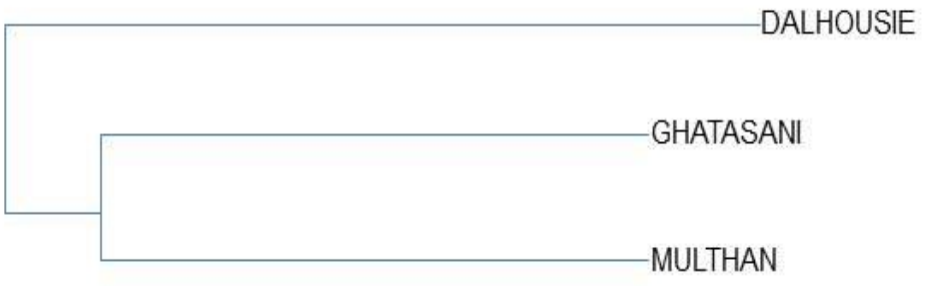

Figure 4: Dendrogram based on Nei's genetic distance for 3 populations of *R. arboreum* based on 38 SSR loci

Fixation index (F_{is}) or inbreeding coefficient, based on Nei's F- statistics, was taken a measure to evaluate variability among population. It gives an estimate

of heterozygosity at various levels of population structure. Total gene diversity for all populations (F_{is}), within population diversity (F_{it}), genetic differentiation between subpopulations (F_{st}) and gene flow [N_m= {0.25 (1-F_{st})}/ F_{st}] were calculated at each locus (Table 11). The total gene diversity within population (F_{it}) was 0.320 and among (sub-population) (F_{is}) was 0.268.

Table 11: Nei's genetic diversity analysis

Locus	Sample Size	F_{is}	F_{it}	F_{st}	N_m
R394	30	-0.003	0.056	0.058	4.044
RE101	30	-0.023	0.114	0.134	1.613
RF87A	30	0.200	0.251	0.063	3.706
RA50	30	0.061	0.106	0.048	4.915
RA443	30	0.478	0.509	0.059	4.012
RA351	30	0.158	0.234	0.090	2.525
RA346	30	0.210	0.265	0.070	3.338
RA321	30	0.291	0.334	0.060	3.898
RA470	30	0.451	0.491	0.073	3.195
RF304	30	0.591	0.677	0.210	0.940
RA254	30	-0.084	0.015	0.091	2.498
RA337	30	0.524	0.529	0.010	23.625
RA267	30	0.545	0.570	0.055	4.304
RA272A	30	0.588	0.640	0.126	1.732
R460	30	0.376	0.400	0.038	6.383
RF97	30	0.208	0.238	0.037	6.460
RA430	30	0.224	0.253	0.037	6.447
R356	30	-0.235	-0.001	0.189	1.070
RA134	30	0.250	0.277	0.036	6.667
RF103	30	0.289	0.324	0.049	4.821
RF29	30	-0.174	-0.150	0.021	11.925
RF43	30	0.137	0.296	0.184	1.112
RF245	30	0.348	0.368	0.030	8.183
RF87B	30	0.370	0.439	0.109	2.041
RF98	30	0.231	0.261	0.039	6.190
RA7	30	-0.009	0.054	0.062	3.759
RA19	30	0.192	0.233	0.051	4.641
RF71	30	0.379	0.417	0.062	3.801
R372	30	0.391	0.424	0.053	4.423
R422	30	0.085	0.132	0.051	4.625
R304	30	0.499	0.536	0.076	3.060
R79	30	0.429	0.490	0.107	2.079
RA324	30	0.386	0.406	0.033	7.277
RA276	30	0.451	0.498	0.086	2.674
R25	30	0.159	0.188	0.034	7.004
RA30	30	0.676	0.708	0.099	2.272
RA54	30	0.213	0.230	0.022	10.990
R443	30	0.315	0.333	0.027	9.125
Mean	30	0.268	0.320	0.071	5.036

F_{is}- Total gene diversity for all populations; F_{it}- Within population diversity; F_{st}- Genetic differentiation between subpopulations; N_m- Gene flow

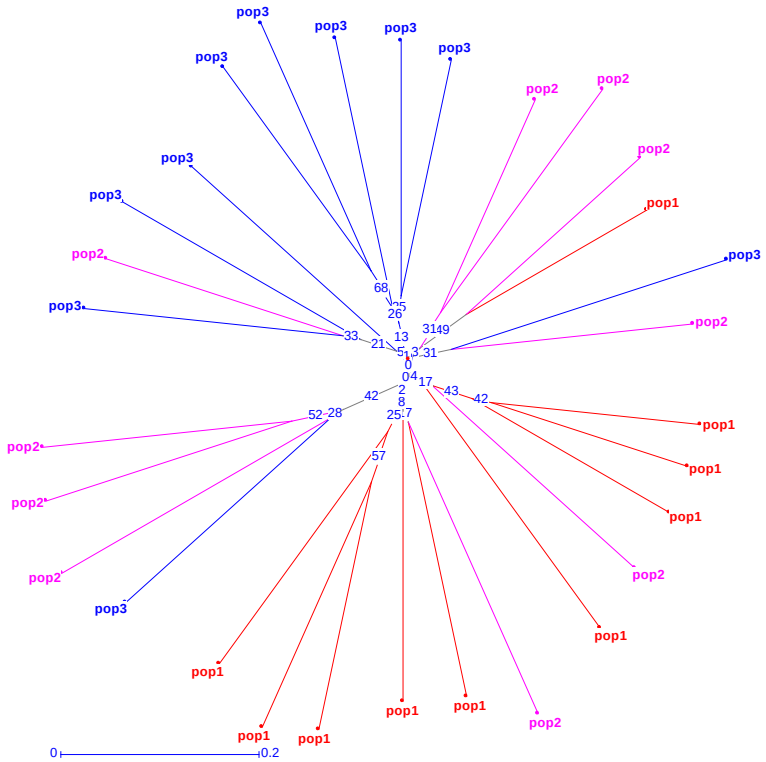

Figure 5: Unrooted tree constructed by DARwin software. Each coloured branch represents single individual collected from corresponding inferred population (pop1- Dalhousie; pop2- Ghatasani; pop3- Multhan)

Unrooted tree from DARwin was used to estimate phylogeny between individuals on the basis of genetic distance and showed two clusters of original genetic stock and no migration from the original clusters was found (Figure 5). The two clusters were further divided into three sub-clusters. No cluster comprised entirely of a single population concluding the presence of gene flow among populations.

To reveal the actual population structure among genotypes of different populations of *R. arboreum,* STRUCTURE based on Bayesian model clustering was approached. Clustering or grouping of the populations was done on the basis of genetic similarity among the groups irrespective of geographical locations of sampling. Two genetic stocks were observed among three populations based on the highest value of ΔK. The grouping was in accordance with the UPGMA dendrogram.

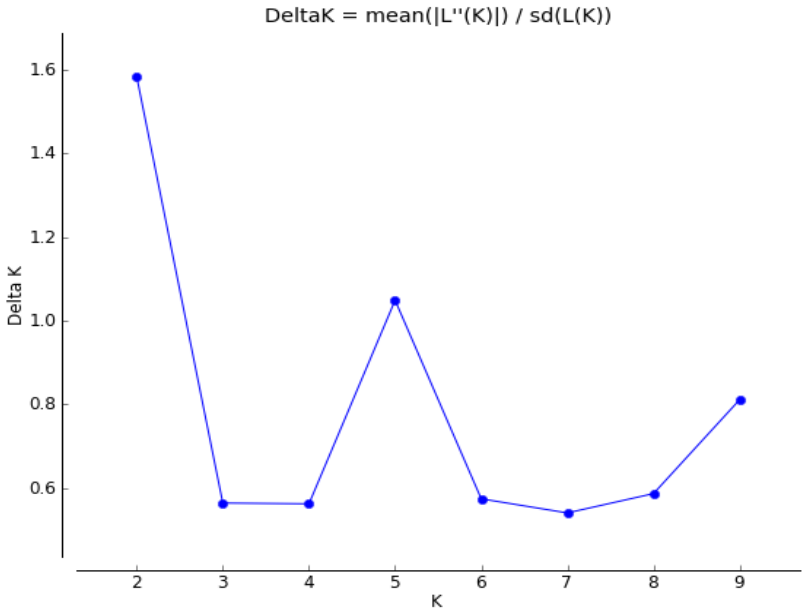

Figure 6: Most appropriate value of K for a range of 3 populations using the second order statistics (ΔK) given by Evanno *et al.*, (2005)

As established from ΔK distribution, the graphical method detected most of the genetic clusters at K=2 indicating true number of populations (K) (Figure 6). Out of 30 individuals, 5 were found in one original stock, 1 individual (Sample No. 20) from Ghatasani in the second original stock and the rest in the admixed stock (Figure 7). The larger genetic cluster constituted 16.7% (4 individuals from Dalhousie and 1 from Ghatsani), the smaller one comprised 3.3% (single individual) and the admixed stock formulated 80% comprising rest of the 24 individuals from the three populations. None of the cluster contained an entire population showing gene flow among populations.

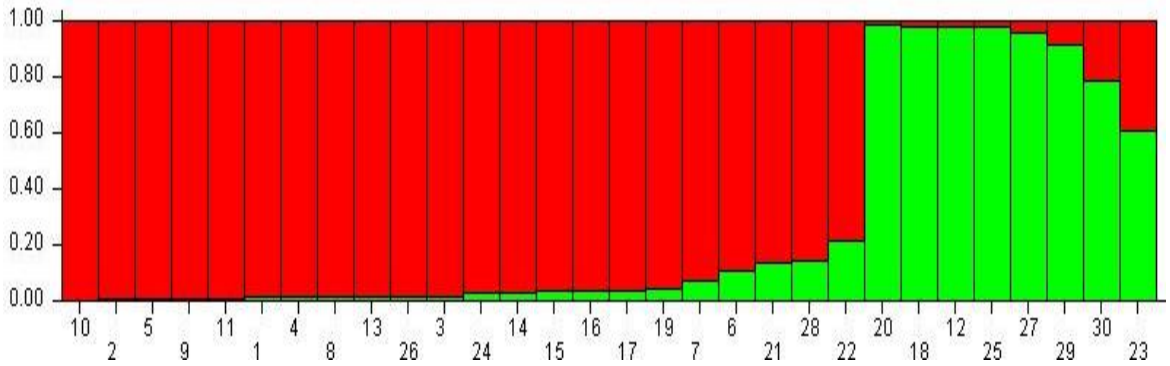

Figure 7: Structure plot with membership coefficient (Q) of individuals on y- axis. Two clusters inferred by Structure are combined from clustering of all individuals. Each color represents different genetic stock and each vertical line broken by colors represent each individual

Threatened plant species often suffers from an overall reduction in population size and reduced allele diversity, a phenomenon called bottleneck. It is testified in terms of heterozygosity deficiency or excess. The BOTTLENECK software assumes that all the loci are in HWE and mutation- drift equilibrium is verified under two mutation models.

Table 12: Allele frequency based mutation- drift equilibrium of SSR loci

Mutation Model	Sign Test	Standarization difference test	Wilcoxon test
IAM	H_{ee}= 22.20	T2= 4.475	P (one tail for H_d) 0.99999
	H_d= 7	P= 0.00000	P (one tail for H_e) 0.00001
	H_e= 31		P (two tails for H_e and H_d) 0.00002
	P= 0.00220		
SMM	H_{ee}= 22.71	T2= -0.678	P (one tail for H_d) 0.73617
	H_d= 15	P= 0.24893	P (one tail for H_e) 0.26856
	H_e= 23		P (two tails for H_e and H_e) 0.53711
	P= 0.5328		

IAM- Infinite allele model; SMM- Stepwise mutation model; H_{ee}- Expected heterozygosity excess; H_d- Heterozygosity deficiency; H_e- Heterozygosity excess

Heterozygosity excess as per the Sign test was discovered among 31 loci under IAM and ascertained to be significant at 5% significance level for both Standardized difference and Wilcoxon test (Table 12). On the other hand, heterozygosity excess for 23 loci was not significant as observed under SMM.

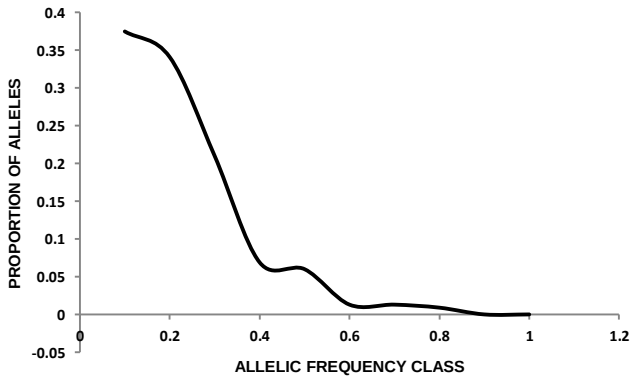

Figure 8: Graphic representation of proportion of alleles and their distribution in populations of *R. arboreum*

However, normal L- shaped distribution plot of allele proportion versus frequency distribution across the populations (Figure 8) confirmed that there is no recent bottleneck and no reduction in the effective population size.

CHAPTER 5

Discussion

Codominant microsatellites have emerged as a preferred marker system in various fields of molecular genetics since they exhibit a higher degree and resolution in allelic diversity (Buschiazzo and Gemmell, 2006; Tóth *et al.*, 2000). Enrichment hybridization followed by library screening has been recommended as a highly robust, reproducible and cost-effective approach for isolating large number of microsatellites in plants (Tan *et al.*, 2014; Zane *et al.*, 2002). In view of the high frequency of GA/AG repeats in plant genome (Gupta and Varshney, 2000), $(GA)_n$ oligonuleotide was employed for enrichment to develop novel microsatellites (Du *et al.*, 2013; Gao *et al.*, 2013; Liu *et al.*, 2013). AG/GA array, constituting 94% of the library, was the most abundant motif while proportions of other repeats were quite negligible. The selection and number of restriction enzyme(s) is critical to enrichment level, frequency and distribution of SSR-containing clones across the library (Hamilton and Fleischer, 1999; Vos *et al.*, 1995). 54% enrichment efficiency (proportion of SSR- clones to the total clones sequenced) was higher than peanut (Macedo *et al.*, 2012), *Monotrapa hypopytis* (Klooster *et al.,* 2009) and grasspea (Lioi and Galasso, 2013). 78.5% of the SSR-containing sequences were rendered unsuitable for primer designing due to insufficiently long flanking regions. Successful locus amplification rate of 93% was more than *Rhododendron simsii* (Tan *et al.,* 2009) and *Rhododendron aureum* (Li *et al.,* 2011). An inverse relation of motif length to mutation rate explains the high polymorphic potential of di- repeats (Herrera *et al.*, 2014). Percentage of polymorphic loci (99%) was comparable or higher than earlier reports on generic or co-generic species of *Rhododendron* (Li *et al.*, 2011; Wang *et al.*, 2014). 14 highly informative loci (PIC> 0.5) and all other loci (except 3) with moderate PIC (>0.25) were described as an important resource for genetic studies [according to the criteria of (Botstein *et al.*, 1980)].

Characterization based on microsatellites has been utilized to analyse genetic diversity and population structure (Aerts *et al.*, 2013; Besnard *et al.*, 2003; Rico *et al.*, 2014). Chi square test for multiple populations reported significant divergence from HWE. Deviations of observed from expected genotypic frequencies infers the presence of gene flow in the populations (Alvarez, 2008; Van Ooijen, 1999). Assuming HWE and independent segregation for all loci, BOTTLENECK estimates the level of heterozygosity, which accounts for high of

diversity level of the populations. The software follows two models (IMM and SAM) based on mutation drift equilibrium. Majority of the novel loci for *R. arboruem* displayed heterozygosity excess under IAM model and statistical significance was determined from Sign test, Standardized differences and Wilcoxon sign- rank test. Allele frequency distribution was L-shaped confirming no recent bottleneck in the population.

A higher genetic differentiation (F_{st}) among Dalhousie and Multhan populations agreed with dendrogram examination showing that the two are more distantly related and had lower gene flow (N_m) when compared to other population pairs. There is a correlation between genetic and geographical distance and is being influenced by forces (such as, natural selection, genetic drift) and, nowadays, human interferences (Vergara *et al.*, 2014). Average N_m of 5.036 revealed a high rate of gene flow which was in accordance with the low F_{st} value (0.071), both being inversely proportional to each other, typical of outbreeding species. According to the stepping stone model, $N_m>4$ is sufficient enough to balance genetic drift (Slatkin, 1993). Populations of Ghatasani and Multhan displayed more gene flow which might be due to their close physical distance and so were the results revealed by Nei's unbiased estimations on genetic identity and distance. STRUCTURE plot examination, however, revealed the true subpopulations. Two distinct genetic stocks and two admixed stocks, neither of which covered an entire population, suggested a high level of gene flow.

Not only individuals of different populations, but individuals within a population showed genetic dissimilarity as predicted from DARwin software. Henceforth, the loci designed in this study can be recommended as an effective marker system for distinguishing even closely related individuals. When associated to previous accounts with AFLP, RAPD and RFLP markers (Wolf *et al.*, 2004; Zawko *et al.*, 2001), an overall high genetic diversity was perceived with the novel microsatellites. *R. arboreum,* being a cross- pollinated species, dwells of its own in natural wild habitats of Himalayan ecology. This may be responsible for the origin of countless adaptations in the species, which when passed to the next generation leads to genetic variability at population and species level. Another possible explanation to higher heterozygosity levels might be the genome wide coverage and mutiallelic nature of microsatellites. The identified loci will deliver a better

understanding of population structure in *R. arboreum* and provide a genetic background for conducting genotyping, ecological and evolutionary studies.

Summary

Rhododendron arboreum (family Ericaceae), a key Himalayan species, has ecological and medicinal significance. Threatened by climate change, anthropogenic activities and poor management practices and without prior availability of genetic background, population dynamics becomes prerequisite for comprehensive organization of conservation strategies. Codominant microsatellites, an efficient molecular tool, were developed to elucidate genetic diversity and population structure of the species. Restricted, adapter ligated genomic DNA was enriched for SSRs using biotin-streptavidin hybridization technique. From a library of 1200 transformed clones, 351 were selected by secondary screening and subsequently sequenced. Of these, 191 were identified with perfect or interrupted motifs of di- and tri- repeats (AG, AT, AC, ACA) of lengths ranging from 5-23. The sequences were used to design 41 novel primer pairs of which, 38 were characterized among three populations of *R. arboreum* (each comprising of 10 individuals) from Dalhousie, Ghatasani, Multhan on 6% denaturing PAGE. Number of alleles ranged from 2-14 (average 5.2) with fragment length range of 100-420 bp. From overall diversity analysis, observed and expected heterozygosity varied from 0.167 to 0.933 (average 0.523) and 0.422 to 0.917 (average 0.723), respectively and PIC values ranged from 0.104 to 0.911 (average 0.464). 19 loci reported significant deviations from HWE which might be due to natural selection, gene flow or other factors. As a whole, no significant linkage disequilibrium and no bottleneck was observed at the population level ($P<0.05$). High levels of gene flow (average $N_m= 5.036$) were reported for the loci, evidence of high genetic diversity within the populations. Fixation index (F), a measure of genetic differentiation, was highest for Dalhousie population; while Ghatasani showed the highest value of Shannon's informative index or allelic diversity (I) and effective number of alleles (N_e). UPGMA based dendrogram displayed close genetic relationship between Multhan and Ghatasani populations, while Dalhousie population came out to be the most distinct of all. DARwin and STRUCTURE analysis grouped the populations into two original clusters and one admixed genetic stock irrespective of their geographical locations. Neither of the clusters comprised entirely of a single population demonstrating germplasm migration among the populations.

The polymorphic loci defined here are highly informative and will be useful for further genotyping or population genetics studies for managing conservation policies for *R. arboreum* and also, for investigating gene flow and effects of other evolutionary forces disturbing HWE. Although the direct costs involved in SSR development are relatively high, augmenting their application range can considerably enhance the cost- benefit ratio.

References

Abe, H. Maki, M. Horie, S. and Suyama, Y. (2011). Isolation and characterization of microsatellite loci for *Menziesia goyozanensis*, an endangered shrub species endemic to Mt. Goyo in northern Japan. *Conservation Genetics Resources* **3**(3): 569-571.

Aerts, R. Berecha, G. Gijbels, P. Hundera, K. Glabeke, S. Vandepitte, K. Muys, B. Roldán-Ruiz, I. and Honnay, O. (2013). Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian Montane rainforests. *Evolutionary Applications* **6**(2): 243-252.

Albert, T. Raspé, O. and Jacquemart, A. L. (2005). Diversity and spatial structure of clones in *Vaccinium uliginosum* populations. *Canadian Journal of Botany* **83**(2): 211-218.

Albert, T. Raspé, O. and Jacquemart, A. L. (2003). Clonal structure in *Vaccinium myrtillus* L. revealed by RAPD and AFLP markers. *International Journal of Plant Sciences* **164**(4): 649-655.

Alvarez, G. (2008). Deviations from Hardy- Weinberg proportions for multiple alleles under viability selection. *Genetics Research* **90**(02): 209-216.

Armour, J. A. Neumann, R. Gobert, S. and Jeffreys, A. J. (1994). Isolation of human simple repeat loci by hybridization selection. *Human Molecular Genetics* **3**(4): 599-605.

Ashizawa, K. Kimura, M. K. Takahashi, A. Lian, C. and Kuramoto, N. (2012). Development of microsatellite markers in a riparian shrub, *Spiraea thunbergii* (Rosaceae). *American Journal of Botany* **99**(7): e283-e285.

Balachandran, M. T. Kadam, N. N. Biradar, J. Reddy, S. H. Ankaiah, M. Narayanan, M. J. Makarla, U. Khurana, P. and Sreeman, S. M. (2013). Development and characterization of microsatellite markers for *Morus* spp. and assessment of their transferability to other closely related species. *BMC Plant Biology* **13**(1): 194.

Besnard, G. Achere, V. Faivre Rampant, P. Favre, J. and Jeandroz, S. (2003). A set of cross-species amplifying microsatellite markers developed from DNA sequence databanks in *Picea* (Pinaceae). *Molecular Ecology Notes* **3**(3): 380-383.

Bhardwaj, P. Kumar, R. Sharma, H. Tewari, R. Ahuja, P. S. and Sharma, R. K. (2013). Development and utilization of genomic and genic microsatellite markers in Assam tea (*Camellia assamica* ssp. *assamica*) and related *Camellia* species. *Plant Breeding* **132**(6): 748-763.

Boches, P. Bassil, N. and Rowland, L. (2005). Microsatellite markers for *Vaccinium* from EST and genomic libraries. *Molecular Ecology Notes* **5**(3): 657-660.

Botstein, D. White, R. L. Skolnick, M. and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**(3): 314.

Boyle, T. Liengsiri, C. and Piewluang, C. (1990). Genetic structure of black spruce on two contrasting sites. *Heredity* **65**(3): 393-399.

Brondani, R. P. V. Brondani, C. Tarchini, R. and Grattapaglia, D. (1998). Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theoretical and Applied Genetics* **97**(5-6): 816-827.

Bruni, I. De Mattia, F. Labra, M. Grassi, F. Fluch, S. Berenyi, M. and Ferrari, C. (2012). Genetic variability of relict *Rhododendron ferrugineum* L. populations in the Northern Apennines with some inferences for a conservation strategy. *Plant Biosystems* **146**(1): 24-32.

Buschiazzo, E. and Gemmell, N. J. (2006). The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* **28**(10): 1040-1050.

Cain, M. L. Milligan, B. G. and Strand, A. E. (2000). Long-distance seed dispersal in plant populations. *American Journal of Botany* **87**(9): 1217-1227.

De Riek, J. Dendauw, J. Mertens, M. De Loose, M. Heursel, J. and Van Bockstaele, E. (1999). Validation of criteria for the selection of AFLP markers to assess the genetic variation of a breeders' collection of evergreen azaleas. *Theoretical and Applied Genetics* **99**(7-8): 1155-1165.

Debnath, S. (2007). An assessment of the genetic diversity within a collection of wild cranberry (*Vaccinium macrocarpon* Ait.) clones with RAPD-PCR. *Genetic Resources and Crop Evolution* **54**(3): 509-517.

Debout, G. Doucet, J.-L. and Hardy, O. (2011). Population history and gene dispersal inferred from spatial genetic structure of a Central African timber tree, *Distemonanthus benthamianus* (Caesalpinioideae). *Heredity* **106**(1): 88-99.

Dick, C. W. Hardy, O. J. Jones, F. A. and Petit, R. J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology* **1**(1): 20-33.

Doyle, J. and Doyle, J. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19:** 11-15.

Du, F. K. Xu, F. Qu, H. Feng, S. Tang, J. and Wu, R. (2013). Exploiting the transcriptome of Euphrates Poplar, *Populus euphratica* (Salicaceae) to develop and characterize new EST-SSR markers and construct an EST-SSR database. *PLoS One* **8**(4): e61337.

Dunemann, F. Kahnau, R. and Stange, I. (1999). Analysis of complex leaf and flower characters in *Rhododendron* using a molecular linkage map. *Theoretical and Applied Genetics* **98**(6-7): 1146-1155.

Eckert, C. Samis, K. and Lougheed, S. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**(5): 1170-1188.

Escudero, A. Iriondo, J. M. and Torres, M. E. (2003). Spatial analysis of genetic diversity as a tool for plant conservation. *Biological Conservation* **113**(3): 351-365.

Evanno, G. Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**(8): 2611-2620.

Fenster, C. B. Vekemans, X. and Hardy, O. J. (2003). Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). *Evolution* **57**(5): 995-1007.

Gao, Z. Wu, J. Wang, L. Ren, H. and Shu, Q. (2013). Rapid microsatellite development for tree peony and its implications. *BMC Genomics* **14**(1): 886.

Glenn, T. C. and Schable, N. A. (2005). Isolating microsatellite DNA loci. *Methods in Enzymology* **395:** 202-222.

Godt, M. and Hamrick, J. (1998). Allozyme diversity in the endangered pitcher plant *Sarracenia rubra* ssp. *alabamensis* (Sarraceniaceae) and its close relative *S. rubra* ssp. *rubra*. *American Journal of Botany* **85**(6): 802-802.

Gomes, F. Costa, R. Ribeiro, M. M. Figueiredo, E. and Canhoto, J. M. (2013). Analysis of genetic relationship among *Arbutus unedo* L. genotypes using RAPD and SSR markers. *Journal of Forestry Research* **24**(2): 227-236.

Guerrant, E. O. Havens, K. and Maunder, M. (Eds.). (2004). *Ex situ plant conservation: supporting species survival in the wild* (Vol. 3). Island Press, Washington.

Gupta, P. K. and Varshney, R. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**(3): 163-185.

Hamilton, M. and Fleischer, R. (1999). Brief communication. Cloned microsatellite repeats differ between 4-base restriction endonucleases. *Journal of Heredity* **90**(5): 561-563.

He, G. Meng, R. Newman, M. Gao, G. Pittman, R. N. and Prakash, C. S. (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biology* **3**(1): 3.

Herrera, C. M. Medrano, M. and Bazaga, P. (2014). Variation in DNA methylation transmissibility, genetic heterogeneity and fecundity-related traits in natural

populations of the perennial herb *Helleborus foetidus*. *Molecular Ecology* **23**(5): 1085-1095.

Hirai, M. Yoshimura, S. Ohsako, T. and Kubo, N. (2010). Genetic diversity and phylogenetic relationships of the endangered species *Vaccinium sieboldii* and *Vaccinium ciliatum* (Ericaceae). *Plant Systematics and Evolution* **287**(1-2): 75-84.

Iqbal, M. J. Paden, D. and Rayburn, A. L. (1995). Assessment of genetic relationships among rhododendron species, varieties and hybrids by RAPD analysis. *Scientia Horticulturae* **63**(3): 215-223.

Ji, L. Teixeira da Silva, J. A. Zhang, J. Tang, Z. and Yu, X. (2014). Development and application of 15 novel polymorphic microsatellite markers for sect. Paeonia (*Paeonia* L.). *Biochemical Systematics and Ecology* **54:** 257-266.

Kameyama, Y. Isagi, Y. and Nakagoshi, N. (2001). Patterns and levels of gene flow in *Rhododendron metternichii* var*. hondoense* revealed by microsatellite analysis. *Molecular Ecology* **10**(1): 205-216.

Kameyama, Y. Kasagi, T. and Kudo, G. (2006). Eight microsatellite markers for sympatric alpine shrubs, *Phyllodoce aleutica* and *P. caerulea* (Ericaceae). *Molecular Ecology Notes* **6**(2): 402-404.

Karagyozov, L. Kalcheva, I. D. and Chapman, V. M. (1993). Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. *Nucleic Acids Research* **21**(16): 3911.

Kijas, J. Fowler, J. Garbett, C. and Thomas, M. (1994). Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. *Biotechniques* **16**(4): 656-660, 662.

Klooster, M. Hoenle, A. and Culley, T. (2009). Characterization of microsatellite loci in the myco-heterotrophic plant *Monotropa hypopitys* (Ericaceae) and amplification in related taxa. *Molecular Ecology Resources* **9**(1): 219-221.

Knowles, P. Perry, D. J. and Foster, H. A. (1992). Spatial genetic structure in two tamarack [*Larix laricina* (Du Roi) K. Koch] populations with differing establishment histories. *Evolution* 572-576.

Kron, K. A. Judd, W. Stevens, P. Crayn, D. Anderberg, A. Gadek, P. Quinn, C. and Luteyn, J. (2002). Phylogenetic classification of Ericaceae: molecular and morphological evidence. *Botanical Review* **68**(3): 335-423.

Li, L.-F. Yin, D.-X. Song, N. Tang, E.-H. and Xiao, H.-X. (2011). Genomic and EST microsatellites for *Rhododendron aureum* (Ericaceae) and cross-amplification in other congeneric species. *American Journal of Botany* **98**(9): e250-e252.

Lioi, L. and Galasso, I. (2013). Development of genomic simple sequence repeat markers from an enriched genomic library of grass pea (*Lathyrus sativus* L.). *Plant Breeding* **132**(6): 649-653.

Liu, S.-R. Li, W.-Y. Long, D. Hu, C.-G. and Zhang, J.-Z. (2013). Development and Characterization of Genomic and Expressed SSRs in Citrus by Genome-Wide Analysis. *PLoS One* **8**(10): e75149.

Luikart, G. Allendorf, F. Cornuet, J. and Sherwin, W. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**(3): 238-247.

Macedo, S. E. Moretzsohn, M. C. Leal-Bertioli, S. C. Alves, D. M. Gouvea, E. G. Azevedo, V. C. and Bertioli, D. J. (2012). Development and characterization of highly polymorphic long TC repeat microsatellite markers for genetic analysis of peanut. *BMC Research Notes* **5**(1): 86.

McCOUCH, S. R. Teytelman, L. Xu, Y. Lobos, K. B. Clare, K. Walton, M. Fu, B. Maghirang, R. Li, Z. and Xing, Y. (2002). Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Research* **9**(6): 199-207.

Naito, K. Isagi, Y. Kameyama, Y. and Nakagoshi, N. (1999). Population structures in *Rhododendron metternichii* var*. hondoense* assessed with microsatellites and their implication for conservation. *Journal of Plant Research* **112**(4): 405-412.

Nunome, T. Negoro, S. Miyatake, K. Yamaguchi, H. and Fukuoka, H. (2006). A protocol for the construction of microsatellite enriched genomic library. *Plant Molecular Biology Reporter* **24**(3-4): 305-312.

Park, Y. J. Lee, J. K. and Kim, N. S. (2009). Simple sequence repeat polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. *Molecules* **14**(11): 4546-4569.

Peakall, R. and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics* **28**(19): 2537-2539.

Pornon, A. Escaravage, N. Thomas, P. and Taberlet, P. (2000). Dynamics of genotypic structure in clonal *Rhododendron ferrugineum* (Ericaceae) populations. *Molecular Ecology* **9**(8): 1099-1111.

Pritchard, J. K. Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**(2): 945-959.

Prochazka, M. (1996). Microsatellite hybrid capture technique for simultaneous isolation of various STR markers. *Genome Research* **6**(7): 646-649.

Rico, Y. Holderegger, R. Boehmer, H. J. and Wagner, H. H. (2014). Directed dispersal by rotational shepherding supports landscape genetic connectivity in a calcareous grassland plant. *Molecular Ecology* **23**(4): 832-842.

Rodrigues, A. G. Colwell, A. E. and Stefanović, S. (2012). Development and characterization of polymorphic microsatellite markers for *Conopholis americana* (Orobanchaceae). *American Journal of Botany* **99**(1): e4-e6.

Sekar, K. C. and Srivastava, S. K. (2010). Rhododendrons in Indian Himalayan region: diversity and conservation. *American Journal of Plant Science* **1:** 131-137.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**(1): 264-279.

Suwabe, K. Iketani, H. Nunome, T. Kage, T. and Hirai, M. (2002). Isolation and characterization of microsatellites in *Brassica rapa* L. *Theoretical and Applied Genetics* **104**(6-7): 1092-1098.

Swaroop, A. Gupta, A. P. and Sinha, A. K. (2005). Simultaneous Determination of Quercetin, Rutin and Coumaric Acid in Flowers of *Rhododendron arboreum* by HPTLC1. *Chromatographia* **62**(11-12): 649-652.

Takrouni, M. M. and Boussaid, M. (2010). Genetic diversity and population's structure in Tunisian strawberry tree (*Arbutus unedo* L.). *Scientia Horticulturae* **126**(3): 330-337.

Tan, C. Wu, Y. Taliaferro, C. M. Bell, G. E. Martin, D. L. and Smith, M. W. (2014). Development and characterization of genomic SSR markers in *Cynodon transvaalensis* Burtt-Davy. *Molecular Genetics and Genomics* 1-9.

Tan, X.-X. Li, Y. and Ge, X.-J. (2009). Development and characterization of eight polymorphic microsatellites for *Rhododendron simsii* Planch. (Ericaceae). *Conservation Genetics* **10**(5): 1553-1555.

Temnykh, S. DeClerck, G. Lukashova, A. Lipovich, L. Cartinhour, S. and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* **11**(8): 1441-1452.

Tóth, G. Gáspári, Z. and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Research* **10**(7): 967-981.

Untergasser, A. Cutcutache, I. Koressaar, T. Ye, J. Faircloth, B. C. Remm, M. and Rozen, S. G. (2012). Primer3- new capabilities and interfaces. *Nucleic Acids Research* **40**(15): e115-e115.

Van Ooijen, J. W. (1999). LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* **83**(5): 613-624.

Vergara, R. Gitzendanner, M. A. Soltis, D. E. and Soltis, P. S. (2014). Population genetic structure, genetic diversity, and natural history of the South American species of *Nothofagus* subgenus *Lophozonia* (Nothofagaceae) inferred from nuclear microsatellite data. *Ecology and Evolution* (DOI: 10.1002/ece3.1108).

Vos, P. Hogers, R. Bleeker, M. Reijans, M. van De Lee, T. Hornes, M. Friters, A. Pot, J. Paleman, J. and Kuiper, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**(21): 4407-4414.

Wang, N. Qin, Z. C. Yang, J. B. and Zhang, J. l. (2010). Development and characterization of 15 microsatellite loci for *Rhododendron delavayi* Franch.(Ericaceae). *HortScience* **45**(3): 457-459.

Wang, Z. Liu, H. Liu, J. Li, Y. Wu, R. and Pang, X. (2014). Mining new microsatellite markers for Siberian apricot (*Prunus sibirica* L.) from SSR-enriched genomic library. *Scientia Horticulturae* **166:** 65-69.

Wolf, P. Doche, B. Gielly, L. and Taberlet, P. (2004). Genetic structure of *Rhododendron ferrugineum* at a wide range of spatial scales. *Journal of Heredity* **95**(4): 301-308.

Xu, J. Grumbine, R. E. Shrestha, A. Eriksson, M. Yang, X. Wang, Y. and Wilkes, A. (2009). The melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods. *Conservation Biology* **23**(3): 520-530.

Yeh, F. Yang, R. C. and Boyle, T. (1999). PopGene Version 131: Microsoft Window-based freeware for population genetic analysis. *University of Alberta and Centre for International Forestry Research* 11-23.

Young, A. Boyle, T. and Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution* **11**(10): 413-418.

Zane, L. Bargelloni, L. and Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**(1): 1-16.

Zawko, G. Krauss, S. Dixon, K. and Sivasithamparam, K. (2001). Conservation genetics of the rare and endangered *Leucopogon obtectus* (Ericaceae). *Molecular Ecology* **10**(10): 2389-2396.

Appendices

APPENDIX- A

COMPOSITION OF CHEMICALS

I. DNA EXTRACTION

DNA EXTRACTION BUFFER

1 M TrisCl (pH 8.0)	: 10.0 ml
0.5 M Na.EDTA (pH 8.0)	: 4.0 ml
5 M NaCl	: 35.0 ml
β- mercaptoethanol	: 0.2 ml (add just before use)
4% CTAB	: 15.0 ml
10% PVP	: 15.0 ml

Make final volume upto 100 ml by adding autoclaved distilled water

TAE (50X)

Tris	: 242 g
Glacial acetic acid	: 57.1 ml
0.5 M EDTA	: 100 ml

$T_{10}E_1$ Buffer

10mM Tris.Cl (pH 8.0)
1mM Na.EDTA (pH 8.0)

II. RESTRICTION

Genomic DNA	: 5 µg
EcoRI (NEB)	: 10 units/ µg of DNA
MseI (NEB)	: 10 units/ µg of DNA
Cut Smart Buffer (NEB)	: 1X
Total volume	: 40 µl (by adding MilliQ water)

Incubated at 37˚C for 10-15 min. Stopped the enzymatic activity by incubating at 65˚C for 20 min.

III. LIGATION

Restricted DNA	: 40 µl
EcoRI adapters	: 20 pmol
MseI adapters	: 20 pmol
Ligase	: 10 units/ µg of DNA
Buffer	: 1X
Total volume	: 50 µl (by adding MilliQ water)

Incubated at 16˚C for 12 hrs. Stopped the enzymatic activity by incubating at 65˚C for 20 min.

IV. DENATURING PAGE

1. 6% Gel Pouring Solution

Urea	: 45g
Acrylamide: bis-acrylamide (19:1)	: 30 ml
5XTBE	: 20 ml
APS (prepare fresh)	: 0.75 mg/ml
TEMED (v/v)	: 0.044% (add just before use)

Make final volume upto 100 ml by adding distilled water.

2. 5XTBE

Tris base	: 54g
Boric acid	: 27.5 g
EDTA (pH 8.0)	: 3.72 g

Make final volume upto 1000 ml by adding distilled water

3. Silver stain (prepare fresh)

Silver Nitrate	: 2 g
Formaldehyde	: 3 ml

Make final volume upto 2000 ml by adding distilled water

4. Developer Solution (prepare fresh)

Sodium carbonate	: 30 g
Formaldehyde	: 1.5 ml
Sodium thiosulfate	: 200 µl (add just before use)

Make final volume upto 1000 ml by adding distilled water

5. Fixer (prepare fresh)

Glacial acetic acid	: 200 ml
Distilled water	: 800 ml

6. 10X Sample loading buffer

Bromophenol blue	: 25 mg
Xylene cyanol	: 25 mg
0.5M EDTA (pH 8.0)	: 200 µl
99% formamide	: 9.8 ml

Mixed PCR product and denaturing dye (1:1) and denatured for 3 min at 95 °C, snap cooled and loaded

APPENDIX- B

DNA Isolation Protocol (Doyle and Doyle, 1990)

- 2- 5 g of young healthy leaves were homogenized to fine powder in liquid nitrogen using pestle mortar. Allowed liquid nitrogen to evaporate completely and added the powder to a centrifuge tube
- Added 5ml of pre-warmed DNA extraction buffer into the centrifuge tube, mixed and incubated at 65 ˚C for 1½ hrs with continuous gentle inversions
- Emulsified the mixture with equal volume of chloroform- isoamylalcohol (24:1) and inverted gently for 45 min. Centrifuged at 13,000 rpm for 30 min at 4˚C
- Took the upper aqueous layer into a fresh centrifuge tube and added 2/3 volume of ice- cold isopropanol. Centrifuged again at 13,000 rpm for 15 min
- Discarded the supernatant and washed the DNA pellet first with absolute alcohol and then with 70% ethanol
- Dried the pellet overnight. Added 1X TE buffer to dissolve the pellet

DNA Purification

The crude isolated DNA contains a lot of impurities *viz.,* RNA, proteins etc. which may inhibit further processing, making it essential to get rid of these contaminants.

- Incubated the dissolved DNA pellet with Ribonuclease A (10 mg/ml) at 37℃ for 1 hr
- Added equal volume of phenol: chloroform: isoamylalcohol (25:24:1), inverted to mix and centrifuged at 13,000 rpm for 5 min
- Took the upper aqueous layer to a fresh eppendorf and repeated the above step twice, but with equal volume of chloroform: isoamylalcohol
- Finally, precipitated the DNA with 2/3 volume of ice- chilled isopropanol and 1/10 volume of sodium acetate
- Washed the pellet with twice the volume of absolute alcohol and then with 70% ethanol

- Dried the pellet and finally dissolved in 1X TE buffer.

APPENDIX- C

SEQUENCES FOR OLIGONUCLEOTIDES

I. ADAPTERS

EcoRI-1 : CTCGTAGACTGCGTACC
EcoRI-2 : AATTGGTACGCAGTCTAC
MSEI-1 : GACGATGAGATCCTGAG
MSEI-2 : TACTCAGGACTCAT

II. PRIMERS

EcoRI : GACTGCGTACCAATTC
MseI : GATGAGTCCTGAGTAA
Probe : GAGAGAGAGAGAGAGAGAGA

Publication

Choudhary, S. Thakur, S. Saini, R. G. and Bhardwaj, P. (2014). Development and characterization of genomic microsatellite markers in *Rhododendron arboreum. Conservation Genetics Resources* **(DOI:10.1007/s12686-014-0246-0) (Accepted & Published)**.

Conservation Genet Resour
DOI 10.1007/s12686-014-0246-0

MICROSATELLITE LETTERS

Development and characterization of genomic microsatellite markers in *Rhododendron arboreum*

Shruti Choudhary · Sapna Thakur · Ram Gopal Saini · Pankaj Bhardwaj

Received: 22 May 2014 / Accepted: 2 June 2014

Abstract Population genetics characteristics are the fundamentals of conservation and management practices. *Rhododendron arboreum*, a key biodiversity component inhabiting Indian Himalayas, suffers from overexploitation and global warming. Using biotin–streptavidin hybridization technique, 41 microsatellite markers were designed from an enriched DNA library to provide a genetic background and an insight into the population structure of the species. With a range of 2–14 alleles amplified from 38 loci, the populations were reported with observed and expected heterozygosity of 0.167–0.933 and 0.422–0.917 respectively. Some of the loci showed significant deviations from Hardy–Weinberg equilibrium and overall no linkage disequilibrium was detected. These markers will support genetic diversity and further genotyping studies in *R. arboreum.*

Keywords *Rhododendron arboreum* · Genomic microsatellites · Selective hybridization

Rhododendron arboreum (a member of Ericaceae), a keystone species of the Indian Himalayan region, is an evergreen tree of ecological and medicinal importance. Inhabiting subalpine to alpine transition zone, the Rhododendron forest line supports a regime of biodiversity components. Snow cover favours the survival of biota at higher altitudes during harsh climate. An isothermic shift due to global warming; overexploitation and other anthropogenic activities can pose a threat to the species richness of Himalayan region in near future (Xu et al. 2009). Since no genetic base is available for this species, it becomes prerequisite to unravel the population dynamics essential for planning and adopting conservation measures. Co-dominant microsatellites have emerged as potent molecular markers for population and evolutionary studies. In view of the above context, we developed genomic SSRs to interpret genetic diversity and population structure of the species.

Genomic DNA was isolated from leaves according to the CTAB procedure (Doyle and Doyle 1987). A library enriched for dinucleotide repeats (GA) was constructed using two restriction enzymes (Mse1 and EcoR1) following the protocol standardized by Bhardwaj et al. (2013). The size selected adapter ligated fragments were enriched using biotin–streptavidin capture with $(GA)_{10}$ as probe, amplified and cloned into chemically competent *Escherichia coli* host using pGEM-T EASY vector. Of the 1,200 transformed colonies screened, 351 clones were subsequently sequenced by 3,730 × l DNA Analyzer (Applied Biosystems) with universal M13 as primer. 191 sequences were identified with di- and tri- repeats of lengths ranging from 5 to 23 using SSR Identification Tool and were classified as perfect, compound or interrupted motifs of AG, AC, ACA repeats (or their complements). The sequences were used to design 41 unique, non-redundant primer pairs from Primer 3.

Of the isolated markers, 38 loci were amplified in 10 µl reaction volume, constituting 25 ng template, 0.3 U *Taq* DNA polymerase (Bangalore Genei^{TM}), 1X *Taq* buffer (1 mM Tris pH 9.0, 50 mM KCl 0.01 % gelatin, 1.5 mM $MgCl_2$), 2.5 mM of each dNTPs, 5 ng each of forward and reverse primer. PCR thermal profile comprised of an initial denaturation step at 94 °C for 3 min followed by 35 cycles

Electronic supplementary material The online version of this article (doi:10.1007/s12686-014-0246-0) contains supplementary material, which is available to authorized users.

S. Choudhary · S. Thakur · R. G. Saini · P. Bhardwaj (✉)
Molecular Genetics Laboratory, Centre for Biosciences, Central University of Punjab, Bathinda 151001, India
e-mail: pankajihbt@gmail.com; pankajbhardwaj@cup.ac.in

Published online: 11 June 2014

Table 1 Characteristics of 38 SSR loci developed from *Rhododendron arboreum*

S. no.	Locus name	Primer sequences	Repeat motif	T_a (°C)	N_a	Approx. size range (bp)	Heterozygosity		PIC	Accession no.
							H_o	H_e		
1.	R394	F: GGAAAGTGTGGGTGTTAGTGC R: TTGAGAGATGGCGAGAGAGAG	$(TC)_{16}$	59	4	145–165	0.6667*	0.7181	0.104	KJ851157
2.	RE101	F: GACGGGAATGAGCAAGGTTG R: CTTCAATTCTGCAAGCCCGA	$(AG)_{16}$	55	6	210–240	0.7333*	0.8418	0.480	KJ851180
3.	RF87A	F: TGGGTCATGTTCTGGAAGGT R: TGAACTAACCCTAGCCACACT	$(AG)_{10}(AG)_{10}$ $(GA)_9$	55	6	140–170	0.5677***	0.7689	0.436	KJ851183
4.	RA50	F: ACTCCCTCCTGTCGTTCCTT R: AATCGTGCATCCGTATCCTG	$(TC)_{16}$	58	5	216–234	0.6667	0.7588	0.310	KJ851162
5.	RA324	F: GCGTACAACATGCCCAAATA R: CCCTGTTCTCATTGCTCACA	$(AG)_8(GA)_8$ $(GA)_g$	55	2	184–194	0.4667	0.7994	0.709	KJ851168
6.	RA351	F: TGTCGCTCTCTCACTGATCG R: TTTGTAGTTTTCCCGTGTCCTT	$(AG)_{12}(AG)_{11}$	59	5	338–360	0.5677*	0.7525	0.374	KJ851170
7.	RA346	F: CGGAGCAAGCTCTCTTATCG R: CCTCTCCTGTGTAGCAAGTCG	$(TC)_9$	59	8	100–116	0.5333*	0.7023	0.396	KJ851184
8.	RA321	F: AGAGATGGGTTTGTGTAAAGTCTG R: TATTTCGCTGCCACCCTAAC	$(GA)_9$	55	4	264–280	0.4677*	0.7124	0.373	KJ851185
9.	RA470	F: AGGGACAAGAAGAAGCCACA R: TCGCGCTTATTACAGCTCTTC	$(GA)_{14}(AG)_{10}$ $(AG)_{10}$	55	3	150–154	0.3000**	0.5994	0.438	KJ851173
10.	RF304	F: TCCTAGGGTTTGTTCGCAAT R: TGCTAGCGATTTCCTAGGGT	$(AG)_{13}(GT)_9$ $(AG)_5(GT)_9$	55	5	218–230	0.2333*	0.7339	0.811	KJ851178
11.	RA254	F: AGTAGCAACACCCACACACT R: GGAGGGGCTGTAGTCTGATT	$(CT)_{16}(CT)_{10}$	55	7	150–164	0.8000*	0.8260	0.413	KJ851179
12.	RA337	F: GAGCGAGAGAGAGGTGTTGG R: ATTCACGGGAATCTTCACCA	$(AG)_6$	59	2	206–208	0.2000	0.4316	0.266	KJ851169
13.	RA267	F: ACGGAGAAGCAGTGAGCATT R: TGCACAGGAACACCCAATAA	$(GA)_{11}(AG)_{10}$	59	3	196–200	0.3000**	0.7102	0.510	KJ851163
14.	RA272A	F: GCCCCGGTGACTCATAAAAT R: TGGTACAAGTGGGACACGA	$(CT)_8(CT)_{11}$	59	3	188–194	0.2667*	0.7525	0.683	KJ851164
15.	R460	F: CCCTACTTCTTTCATCACATACAA R: CAACTCCGGTCATTTTTGGT	$(GA)_{13}(AG)_{12}$	59	5	188–196	0.4333	0.7345	0.621	KJ851158
16.	R97	F: AGCAGCAACAATGGTGTCC R: TCTAGAAGGCCTCCCATTCC	$(AG)_{10}(AG)_{13}$	59	2	188–190	0.5000	0.6672	0.381	KJ851160
17.	RA430	F: GCGTAAATCGAGTTCGGAAG R: CTCTCTCTAATCGAATTCCCG	$(TC)_{10}(CT)_6$	59	7	166–180	0.6333	0.8627	0.610	KJ851171

Table 1 continued

S. no.	Locus name	Primer sequences	Repeat motif	T_a (°C)	N_a	Approx. size range (bp)	Heterozygosity		PIC	Accession no.
							H_o	H_e		
18.	R356	F: GAGCTAAGCACGCCGTATTC R: AAATTCGACGGCAAAGAGG	$(TC)_9$	59	4	186–194	0.6667*	0.6774	0.104	KJ851166
19.	RA134	F: GGAGAGAGAGGCCGAGAGAG R: ACGTCGCTCTTGTCAAGCAT	$(GA)_6(GA)_6$ $(AG)_8$	59	3	220–228	0.3000	0.4220	0.140	KJ851167
20.	RF103	F: GATAGAGAGACAGGGGCAGC R: TGTACGCCAAGACTCCCATT	$(GA)_{14}$	59	5	288–300	0.5333	0.8023	0.482	KJ851181
21.	RF29	F: ACAGACAGAAGCAGCGGAAC R: AAGGGGAGGAGATCGAGTTG	$(GA)_{10}(AG)_{13}$	59	9	138–160	0.9333	0.8254	0.266	KJ851174
22.	RF43	F: AATTCGATGGGTTGGTGGTA R: GCCTTCTCTGTTCTCGGTTTT	$(AG)_{10}$	59	4	190–200	0.3667*	0.5294	0.237	KJ851175
23.	RF245	F: GGGTTTTTGATCTTCATACGG R: AATCGGTTCAAGAGGGGTTT	$(AG)_{11}$	55	8	198–220	0.5333	0.8576	0.693	KJ851177
24.	RF87B	F: GGGGAAAGGTCATTGGAGAT R: TTCTGAACTAACCCTAGCCACA	$(GA)_{10}(GA)_{10}$ $(AG)_{10}$	55	3	310–316	0.4000*	0.7249	0.536	KJ851183
25.	RF98	F: AATCCCATCCCCTAACTTGG R: CCGTGGCTTTACCTTTCACT	$(AG)_{22}$	59	10	170–196	0.6667	0.9169	0.720	KJ851182
26.	RA7	F: GTCTACAATGCTTGCTTCCG R: CCTTATTTATTCTCTCTCT	$(AG)_{10}(AG)_9$	55	5	110–120	0.7333*	0.7881	0.370	KJ851150
27.	RA19	F: AGCCAAAAACTTTCTTCTTTCC R: CTGTCGGCTGTCAGAGTTGA	$(GA)_{19}$	55	4	124–138	0.5333	0.7073	0.381	KJ851151
28.	RF71	F: GCGTACAACATGCCCAAATA R: GTCGTTGCAGTTCAATCTCG	$(AG)_{12}$	55	5	300–312	0.4667	0.8147	0.710	KJ851176
29.	RA443	F: CCATGCCTGAAGCAAACAC R: AGACTCCAAAGTCCTATCTGTGC	$(AG)_{12}$	59	8	184–200	0.4000*	0.8282	0.649	KJ851162
30.	RA272B	F: ATGCAATGGAAATGGGAAAG R: GGAACGGGTAATTCGGATCT	$(CT)_8(CT)_{11}$	59	3	184–194	0.3333	0.5085	0.194	KJ851164
31.	R372	F: GGTTGGGTGGATGGAGTAAC R: GCAATTTGCATAGCACTGTAAT	$(AG)_{15}(AG)_{15}$	55	5	230–244	0.4667	0.8237	0.600	KJ851156
32.	R422	F: GCGGTACTGTTCCGATCAC R: TCCCAGCTCATCCACACATA	$(AG)_{12}(GA)_{13}$	55	12	146–172	0.7333	0.8593	0.624	KJ851155
33.	R304	F: TCCTAGGGTTTGTTCGCAAT R: GCGTATTGTTGCACGAAAAA	$(AC)_9(CT)_{12}(TC)_5$ $(AC)_9(TC)_{13}$	55	4	380–410	0.3000**	0.6582	0.482	KJ851161
34.	R79	F: AACGTGAAAACTGAGCAGAAC R: CCCGTGGGTAGGAAAATCAT	$(AG)_{10}(AG)_5(AG)_{10}(AG)_7$	55	4	236–250	0.2667	0.5316	0.358	KJ851159

Table 1 continued

S. no.	Locus name	Primer sequences	Repeat motif	T_a (°C)	N_a	Approx. size range (bp)	Heterozygosity		PIC	Accession no.
							H_o	H_e		
35.	RA276	F: GCCAAAAAGCATCAAGTCGT R: TGTGATTTTTGTGTTGGATGG	$(TC)_{11}$	55	6	390–420	0.4000*	0.6605	0.421	KJ851165
36.	R25	F: CCAACAACCCGAGAAAAAGA R: AGTGGGTTTCCGAGACAAAG	$(AG)_{12}(GA)_{11}$	55	14	164–200	0.7667	0.9119	0.709	KJ851154
37.	RA30	F: TGTGATTTTTGTGTAGGATGGT R: AGCAACCCACTTCTCCTCTTC	$(AG)_{17}$	55	2	150–160	0.1667*	0.5814	0.651	KJ851152
38.	RA54	F: CAAAAATGGCCAACAAGGAT R: ATTGCCTCCATACACAACCA	$(CT)_5(CT)_6(CT)_{17}$	55	2	190–200	0.5000	0.6605	0.429	KJ851153

T_a annealing temperature, N_a total number of alleles, H_o observed heterozygosity, H_e expected heterozygosity, *PIC* polymorphic information content

Significant deviations from Hardy–Weinberg equilibrium at * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

of denaturation at 94 °C for 1 min, annealing at specific temperature (standardized for each locus as depicted in Table 1) for 1 min and elongation at 72 °C for 1 min; and a final elongation step at 72 °C for 8 min. These loci were characterized among three populations of *R. arboreum* (each comprising of 10 individuals) from Dalhousie, Ghatasni, Multhan (India; geographical coordinates given in Supplementary Table 2) on 6 % denaturing polyacrylamide gel. Number of alleles ranged from 2 to 14 (an average of 5.2) with fragment length of range 100–420 bp. An overall population structure analysis was performed using PopGene version 1.31 (Yeh et al. 1999). Polymorphic information content and observed and expected heterozygosity varied within the range of 0.104–0.911 (with an average of 0.464), 0.167–0.933 (with a mean of 0.523) and 0.422–0.917 (with an average of 0.723) respectively. 19 loci reported significant deviations from Hardy–Weinberg equilibrium (Table 1) which might be due to natural selection or other factors. As a whole, no significant linkage disequilibrium was detected at the population level ($p < 0.05$). Also, the population-wise diversity parameters (Supplementary Table 1) were estimated using GenAlEx version 6.5 (Peakall and Smouse 2012). The polymorphic loci defined here are highly informative and will be used further in genotyping or population genetics studies for managing conservation policies for *R. arboreum* and also, for predicting gene flow and profound effects of other evolutionary forces disturbing HWE.

Acknowledgments The research work present in the manuscript is funded by Department of Science and Technology, Govt. of India. This is CUPB communication no P78/14.

References

Bhardwaj P, Kumar R, Sharma H, Tewari R, Ahuja PS, Sharma RK (2013) Development and utilization of genomic and genic microsatellite markers in Assam tea (*Camellia assamica* ssp. *assamica*) and related *Camellia* species. Plant Breed 132(6): 748–763

Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel: population genetic software for teaching and research-update. Bioinformatics 28(19):2537–2539

Xu J, Grumbine RE, Shrestha A, Eriksson M, Yang X, Wang Y, Wilkes A (2009) Melting Himalayas: cascading effects of climate change on water, biodiversity, and livelihoods. Conserv Biol 23(3):520–530

Yeh F, Yang RC, Boyle T (1999) PopGene Version 1.31: Microsoft window based freeware for population genetic analysis. University of Alberta and Centre for International Forestry Research 11–23