CHARACTERIZATION OF G-SSR MARKERS IN *Rhododendron arboreum*

Dissertation submitted to the Central University of Punjab

For the award of
Masters
In
Biosciences

BY

Harmeen Kaur

Supervisor
Dr. Pankaj Bhardwaj

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

July, 2014

CERTIFICATE

I declare that the dissertation entitled "CHARACTERIZATION OF G-SSR 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.

Harmeen Kaur

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

Date:

CERTIFICATE

I certify that Harmeen Kaur has prepared her dissertation entitled "CHARACTERIZATION OF G-SSR MARKERS IN *Rhododendron arboreum*", for the award of M.Sc. 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:

ABSTRACT

Characterization of G-SSR markers in *Rhododendron arboreum*

Name of student: Harmeen Kaur
Registration number: CUPB/M.Sc./SBAS/BIO/2012-13/07
Degree for which submitted: Master of Science
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, Gene Flow, SSR characterization, Population structure

Himalayan region is one of the important hotspot, having *Rhododendron arboreum* as major species. Anthropogenic activities and unsustainable use of the plant for its fruits, flower and wood by local people has raised questions about the sustainability of the species. In order to answer these questions, it is important to first know the genetic structure of the population. This study is focused on the characterization of the G-SSR markers for Rhododendrons and to study the population genetic structure. A set of 38 primers out of 41 are characterized on three population of *Rhododendron arboreum,* a total of 30 individuals. Allele number at different loci came to be in range of 2 to 14 with an average of 5.2 alleles per locus. Level of heterozygosity observed was high which varies from 0.422 to 0.917 (H_e) and 0.167 to 0.933 (H_o) with a polymorphic information content 0.104 to 0.811. Out of 38, about 19 loci deviated from the Hardy Weinberg Equilibrium. Genetic diversity was high in these populations with an average gene flow of 5.436. Based on the information from the loci, there was no observed linkage disequilibrium and bottleneck effect. Software DARwin and STRUCTURE, irrespective of geographical location, revels the presence of two original populations with admixed genetic stock. These studied loci will help in the further genetic mapping of the genes and genotyping studies in *Rhododendron arboreum.*

Signature of Student   Signature of Supervisor

ACKNOWLEDGEMENTS

With utmost modesty, I bow before God 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 and lab mates especially, Gaganpreet Kaur, Shruti Choudhary, Sapna Thakur and Richa Mehra. I am also thankful to technical staff for their help and support.

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

Harmeen Kaur

TABLE OF CONTENTS

S. No.	Contents	Page No.
1.	Introduction	1-3
2.	Review of Literature	4-10
3.	Material and Methods	11-15
4.	Results and Discussion	16-29
5.	Summary	30-32
6.	References	33-39
7.	Appendix	40-43

LIST OF TABLES

Table No.	Table Description	Page No.
1.	List of studies conducted on the genetic structure studies using SSR markers	9-10
2.	Geographical locations of 30 sampled individuals of R. arboretum	12
3.	Characteristics of 38 G-SSR loci developed from *R. arboretum*	20-21
4.	Diversity parameters for three populations of *R. arboreum*	22
5.	*R. arboreum* populations with percentage of polymorphic loci and their genetic variability parameters detected by SSR analysis	23
6.	Nei's Original Measures of Genetic Identity (upper diagonal) and Genetic distance (lower diagonal)	23
7.	Population with maximum gene flow	24
8.	Nei's gene diversity analysis	25
9.	Allele frequency based mutation- drift equilibrium of SSR loci	28

LIST OF FIGURES

Fig. No.	Description of Figures	Page No.
1.	Location of sampling sites on map of H.P., India	13
2.	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™).	17
3	Amplification profile generated with primer RF29. Lanes 1- 30 represent sampled individuals of *R. arboreum* as presented in Table 3; Ld: 100 bp DNA ladder as size standard (Bangalore Genei™).	18
4.	Dendrogram based on Nei's genetic distance for 3 populations of *R. arboreum* based on 38 SSR loci	24
5.	Unrooted tree constructed by DARwin software. Each colored branch represents single individual collected from corresponding inferred population (pop1: Dalhousie; pop2: Ghatasani; pop3: Multhan)	26
6.	Most appropriate value of K for a range of three populations using the second order statistics (ΔK) given by Evanno *et al.* (2005)	27
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.	28
8.	Graphic representation of proportion of alleles and their distribution in populations of *R. arboreum*	29

LIST OF APPENDICES

Appendix Serial	Description of Appendices	Page No.
A.	Composition of chemicals	39-40
B.	DNA isolation and purification protocol	41

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 Sequence Repeats	SSR
31.	Thymine	T
32.	Un-rooted Pair Group Method Analysis	UPGMA

CHAPTER 1

Introduction

Rhododendron, belongs to family Ericaceae, has around 1000 species. Rhododendrons are widely distributed in the Himalayan regions of India and China and is a key species, constituting a larger part of the flora. In India, Rhododendrons are widely present in Jammu and Kashmir, Uttaranchal, Nepal, Himachal Pradesh, Tamil Nadu, Manipur, Meghalaya, Mizoram and Sikkim (Swaroop *et al.*, 2005). *Rhododendron arboreum* is the most abundant species among the Indian Rhododendrons, which plays a crucial role in the conservation of soil and water. Local people use plant parts for many purposes like fuel, timber, jellies, squashes and antioxidant properties. Along with that, Rhododendrons also have anti-diabetic, anti-inflammatory and anti-nociceptive properties.

With the increasing global warming, the mean annual temperature is also raising which will have a negative impact on the flora and fauna of the region. These variations will affect the growth of the plant, its reproduction and the behavior of the pollinators. Also, there is a rapid decrease in the forest cover due to anthropogenic activities, which result in the diminishing of Rhododendron population in the entire Himalayan region. This decrease in the native species will create a huge imbalance in the region leading to a loss of many other ecologically important species (Bruni et al, 2012).

Occurrence of the species in an area and its diversity is determined by many factors such as gene flow, migration, selection, genetic drift and mutation. Any environmental change in the region may lead to change in the occurrence of the species in that area. For the conservation of species, better breeding programs and further exploration of species, not only the ecological knowledge is required but also the molecular level studies are necessary (Bhardwaj *et al.*, 2013). These studies allow us to have a better insight of the plant and also reveal the evolutionary relationships.

In the recent years, molecular marker techniques are used to interpret the genetic variation and gene flow in the ecologically important plant species for their long term preservation. Various types of marker based studies have been conducted in the past which have used different genetic markers like randomly amplified fragment length polymorphism (RAPD), Amplified fragment length polymorphism (AFLP), restricted fragment length polymorphism (RFLP) in different

species of Ericaceae. These markers, used in the studies are dominant in nature and have limited utility. There is a need for the development of co-dominant markers which can provide us with more information about the species and one such type of marker is simple sequence repeats (SSR) (Miah *et al.*, 2013).

SSRs or microsatellites are being widely used because they are readily amplified by the PCR, highly polymorphic and informative, species specific, co-dominant, multi-allelic, reproducible, abundant in the genome and are helpful in the relevance of specific traits (Hasan *et al.*, 2006). SSRs frequency in plants is high which comes around one SSRs region in every 6-7 kilo-base of DNA (Cardle *et al.*, 2000). SSRs, due to these characteristics, are widely used for the strain improvement, marker assisted selection (MSA), cultivar identification, gene mapping and development of the conservation strategies (Miah *et al.*, 2013).

. Considering the significance of the work, following objectives were undertaken to provide a genetic background for the species:

- Sample collection and isolation of DNA
- Characterization of the previously developed G-SSRs marker in the selected germplasm.

CHAPTER 2

Review of Literature

Himalayas are one of the major biodiversity hotspots, having a great deal of ecological value. *Rhododendron arboreum* is an important forest tree of the region having household and medicinal uses. It is highly adapted to the region and play important role in the maintenance of the ecosystem.

2.1 Classification/ Taxonomic Rank:

Rhododendron arboreum belongs to a family Ericaceae, which is a largest family in the order Ericales. Ericaceae contain 126 genera and nearly 4000 species. These are generally found in the higher altitude regions with lower temperature. Classification of Rhododendron is as below (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser):

Kingdom: Plantae

Phylum: Magnolliophyta

Class: Angiospermae

Order: Ericales

Family: Ericaceae

Genus: *Rhododendron*

Species: *arboreum*

2.2 Distribution and Occurrence:

Rhododendrons cover a vast area in upper temperate location in Asia. Of all the natural occurring population of the Rhododendrons, about 90% are found in the vast section of south-eastern Asia. About 50 Indian species, consisting of 95%, of the Rhododendrons are found in the Himalayan region of the India (Singh *et al.*, 2003). Occurrence of Rhododendrons is in alpine and sub alpine regions and reproduction is largely dependent on the snowmelt gradient in the region which is dependent on the local environment conditions. The higher fruit yield was reported with earlier snowmelt whereas the latent flowering due to unfavorable conditions, results in the abortion (Kudo, 1993; Xu *et al.*, 2009).

Rhododendrons are extensively variable in the size, flower color, and leaf characteristics. It can grow up to the altitude of 4500 to 10,500 ft and can have a height up to 100ft (Rai and Rai, 1994). Rhododendrons are largely used by the local residents for its wood, fruits and flower. Along with this, Rhododendrons have anti-inflammatory, anti-nociceptive, anti-diabetic, anti-diarrheal, anti-oxidant and hepatoprotective activity, due to the presence of the phenolic, flavinoids and other secondary metabolites (Srivastava, 2012; Swaroop *et al.*, 2005). Rhododendrons are threatened species due to high overexploitation of it in its natural ecologies.

2.3 Factors Shaping the Populations:

The genetic diversity play an important role in ecosystem functioning, its survival, predicting the vulnerability of a species towards extinction and the ability of a population to adapt in changed conditions. High genetic diversity increases the chances of survival of plant in long term as it is able to impart the resistance to plant towards various pathogens (Ehlers *et al.*, 2008; Johnson *et al.*, 2006). Genetic diversity of a plant is attributed by the various factors *viz.*, genetic diversity, gene flow, genetic drift, breeding system and natural selection. The different ecological locations and the different seasons may change the rate of these processes, even for the same species present in different ecologies. Demography and spatial distribution of the populations, in turn, govern these processes (Eckert *et al.*, 2008).

Geographical isolation or barrier in-between the populations leads to reduced gene flow, thus giving birth to the uniquely adapted populations with specific allelic combination for particular climatic conditions. Gene flow is also determined by the dispersal of the pollens or the seeds. Due to the presence of the barrier in the population and limited pollen and seed dispersal, the populations tend to exhibit the bottleneck effect (Eckert *et al.*, 2008). On the other hand, hampered gene flow reduces the genetic variation within the population, but increases the diversity between the populations (Zawko *et al.*, 2001).

Naturally, populations exhibit a dynamic equilibrium between genetic drift and gene-flow. The isolated small population experiences inbreeding depression

which results in the reduced adaptability to the environmental changes and reduced diversity (Bruni *et al.*, 2012b). Population size also determines the rate at which the traits are lost due to genetic drift; smaller populations enhance the rate at which the traits are lost and vice versa (Eckert *et al.*, 2008). Selection pressure also plays an important role in the gene flow. Low selection pressure favors seed dispersal hence reducing the genetic variances in between the populations. And high selection pressure will favors the dispersal of pollens and inhibiting the seed dispersal. Population divergence by seed dispersal is dependent on the fitness and dominance in case of presence of selection pressure (Loveless and Hamrick, 1984).

For the study of the population diversity or heterozygosity in the samples, most commonly used parameter is Nei's unbiased genetic diversity (F). Average heterozygosity can be estimated from the small number of individual if loci used in the study are in large number having low heterozygosity (Nei, 1978).

2.4 Molecular Markers and their Implications:

The shape of the population is determined by the underlying genetic structure of the individuals which in turn is determined by the factors like gene flow, genetic diversity, inbreeding and genetic drift. Information about the genetic makeup of a species reveals current scenario as well as past and future of the species (Fenster *et al.*, 2003). The presence of factors like gene flow, random mating, and large population size favors the growth of population and on other hand, factors like inbreeding and genetic drift have a negative impact on the population (Charlesworth and Charlesworth, 1987). Molecular genetics approaches can produce noteworthy information about the factors that can further help conservation program for the particular species.

DNA markers are extensively used for population study and are one of the major tools for getting an insight into the population structure. Molecular markers help to get information about the levels of heterozygosity on loci. Markers can be dominant or co-dominant based on the type of information it depict. Various types of markers are restriction fragment length polymorphism (RFLP), random

amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites or simple sequence repeats (SSRs) markers and single nucleotide polymorphism (SNP). While selecting a marker for study, it is very important to consider the following factors (Miah *et al.*, 2013):

- Marker inheritance
- Quantity and quality of DNA and size of population
- Cost associated with the technique
- Main objective of the study for which markers to be used and their transferability

RAPD, RFLP, and AFLP, being dominant markers, are not to a great extent informative because they are unable to depict heterozygosity. While on the other hand, the co-dominant markers such as SSRs are much more informative, more polymorphic, reproducible, and are abundantly present in the genome (Park *et al.*, 2009).

Dominant marker based studies have been carried out on the various species of Rhododendrons; RAPD performed on 18 species of Rhododendrons for population structure study (Lanying *et al.*, 2008), AFLP on *Rhododendron delavayi* and *Rhododendron decorum* for study of hybrids between two species (ZHA *et al.*, 2008) and RFLP for the determination of origin of *Rhododendron ponticum* which is naturalized in the British Isles (Milne and Abbott, 2000). In the light of the drawback of dominant marker used, it is very important to perform studies to generate the co-dominant marker and their implications. The reason behind the huge success of the SSRs is its ability to reveal co-dominantly inherited multi-allelic loci that can be mapped (Cordeiro *et al.*, 2000).

2.5 SSRs: The Marker of Choice

SSRs (term coined by Litt and Luty) are the simple repeat motifs of 1-6 nucleotide repeated as many as 100 times at each locus, present abundantly in the genome. Microsatellite can have a repeat region of the di, tri or tetra nucleotide but the occurrence of di-nucleotide is most abundant in the plant genome, especially AT/TA followed by AG/CT repeats (Cardle *et al.*, 2000).

Tautz was first to report polymorphism at the SSRs locus. SSRs are generated due to slippage of enzymes during DNA replication and are hyper-variable in length (Tautz, 1989). Variations in the SSRs are independent of the natural selection because of their major presence in the non-coding region of DNA (Abdul-Muneer, 2014).

Polymorphism can be studied by the separation of PCR amplified alleles present on SSRs loci on a polyacrylamide gel in the form of bands. Due to high levels of variability, and capability to isolate large number of loci, SSR marker system is able to differentiate even between two very closely related species. Differentiation is on the basis of differences in the size of two alleles, even of 1 - 2 nucleotides, which is detectable on polyacrylamide gel electrophoresis (PAGE) (Abdul-Muneer, 2014).

SSRs are widely used for depicting true structure of the populations. Various studies have been conducted and are enlisted in table below (Table 1).

Table 1: List of studies conducted on the genetic structure studies using SSR markers

S.No	Title	Result	Reference
1.	Development and characterization of genomic SSR markers in *Cynodon transvaalensis* Burtt-Davy	From a [CA]n, [GA]n, [AAG]n, and [AAT]n enriched library, 1,426 markers were designed. About 93% of the loci were polymorphic and about 917 markers were found to be associated with heritable alleles.	(Tan *et al.*, 2014)
2.	SSR marker based DNA fingerprinting and diversity assessment in superior tea germplasm cultivated in western Himalaya	21 markers were used to evaluate 15 tea accessions which showed high levels of heterozygosity. Genetic similarity (GS-0.28) was used to cluster population in two major groups with exclusive china and china hybrids.	(Bhardwaj *et al.*, 2013)
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. Dendrogram grouped the populations on the basis of high bootstrap.	(Hirai *et al.,* 2010)
4.	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)

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.	Characterization of microsatellite markers from sugarcane (*Saccharum* sp.), a highly polyploidy species	Primer set of 100 out of 124 synthesised were used to test five cultivars of sugarcane in which 91% of polymorphism was detected in primers high PIC value was observed making them suitable for genetic identification.	(Cordeiro *et al.*, 2000)
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.	Analysis of genetic diversity in the *Brassica napus* L. gene pool using SSR markers	30 SSR primers were used to estimate the genetic diversity of 96 individuals. UPGMA clustering identified 4 major groups with high genetic variance observed in spring kale group	(Hasan *et al.*, 2006)
9.	Application of simple sequence repeat (SSR) markers in apricot breeding: molecular characterization, protection, and genetic relationships	SSR analysis reveals the presence of seven groups used in the study. High degree of homology was observed between apricot and peach, thus depicting low levels of breeding barriers i.e. high gene flow. SSRS of apricot showed transferability.	(Sánchez-Pérez *et al.*, 2005)
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.*, 2012a)

CHAPTER 3

Materials and Methods

3.1 Sample collection:

Three locations in the Himachal Pradesh viz. Dalhousie, Multhan, and Ghatasani were selected for the collection of plant samples. These three geographical locations were marked in the map (Figure1) and details are given (Table 2). From these three locations, three populations consisting of a total of 30 individuals were collected.

Table 2: 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

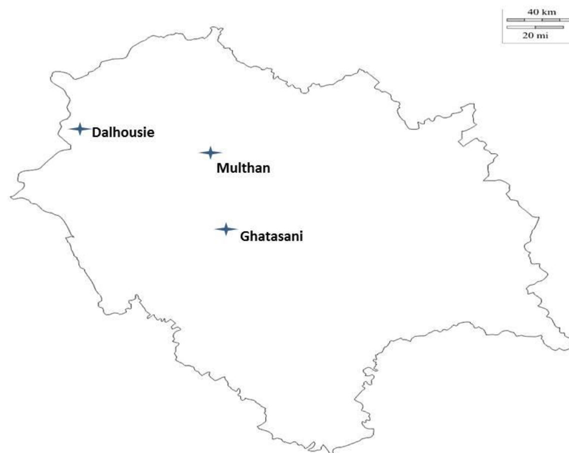

Figure 1: Location of sampling sites on map of H.P., India

3.2 DNA Isolation:

DNA was isolated from these thirty samples by using the CTAB protocol (Doyle, 1987) with some modifications. Isolated genomic DNA samples were then quantified on Nanodrop 2000 Spectrophotometer and quality was accessed on the 0.8% agarose gel.

3.3 G-SSR Marker Characterization:

38 previously designed G-SSR markers which were designed from unique, sufficiently long SSRs flanking regions were used for characterization over three Rhododendron populations for further genetic diversity analysis. Firstly, amplification of the genomic DNA was done in a PCR in the total reaction 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 MgCl2), 2.5 mM of each dNTPs, 5-10 ng each of forward and reverse primer in T100 Thermal Cycler (BIORAD). PCR Cycle consists of first initial denaturation cycle of 3 min at 94°C. This initial denaturation step is then followed by the 35 cycles of denaturation, annealing, and elongation and an addition step of final elongation at 72°C for 8 min. In the cycle, denaturation is done at 94°C for 1 min, annealing at required T_m for 1 min and elongation at 72°C for 1 min.

Annealing temperature for each primer is estimated on the basis of melting temperature of the primer, which is ±5°C.

After that reaction product is denatured by mixing with equal volume of denaturing dye and subsequently heating it at 94°C for 5 min. Above denatured product was then quantified on 6% urea polyacrylamide gel. Before gel preparation, the plates were cleaned very carefully by wiping the plates with tissue soaked in ethyl alcohol and iso-propenol. While pouring of the gel, special attention must be given so that no bubble is formed in the gel as it will affect the migration of the sample.

Samples were then loaded on polyacrylamide gel and electrophoresed at 60W for desired time depending on size of the fragment. Amplified DNA fragments were then visualized by the silver staining. The plate with the gel was kept in the fixer for an hour or overnight followed by the water washing of the plate for about 45 min. Plate is then treated with the sliver stain again for 45 min and then developer is used for the development of bands which then can be scored.

3.4 Data Analysis

Profiles for each of the Microsatellite primer pair were scored as a single band signifying a single allele. Various parameters were estimated for the determination of the genetic diversity which include the effective number of alleles (Ne), observed and expected heterozygosity (Ho, He), and Shannon's information index (I). POPGENE was used to calculate the heterozygosity, effective alleles, deviations from Hardy Weinberg Equilibrium (HWE), Nei's genetic diversity and linkage disequilibrium both locus wise and population wise. UPGMA dendrogram was also construction using POPGENE software v1.32 which was constructed on the basis of the genetic distance. Polymorphic information index (PIC) was also estimated for the determination of discrimination ability of a locus.

Software DARwin v5.0.158 and STRUCTURE v2.3.4 were used which perform the Bayesian analysis to depict the phylogenetic relationship among populations. Other parameters, like population wise genetic distance was predicted by the GeneALEx v6.5 which also gave us information about fixation index, in turn, the information about the differentiation in the population. The level

of heterozygosity on a locus, either deficiency or the excess, was determined by BOTTLENECK v1.2.02.

CHAPTER 4

Results and Discussion

4.1 SSRs Primer Evaluation:

Out of the total of 41 primers designed previously in the lab, only 38 primers, about 93% were showing polymorphism in all the three populations and were selected for the evaluation the genetic structure. These sequences were submitted to the NCBI database and accession numbers were obtained (given in table 3).

4.2 Characterization of SSR markers:

Each SSR loci was amplified, a PCR reaction with the single primer pair was carried out. PCR product was then resolved in a polyacrylamide gel after denaturing. Gel was silver stained and then bands were scored. Each band present corresponds to the single allele (Figure 2 & 3). Every lane can have a maximum of two bands in expected range. Number of alleles present on a locus and its size was firstly determined by careful examination of the position of the each band with respect to the DNA 100 bp ladder. For every visible band in a lane, alleles were scored and no amplification in a lane corresponds to null allele.

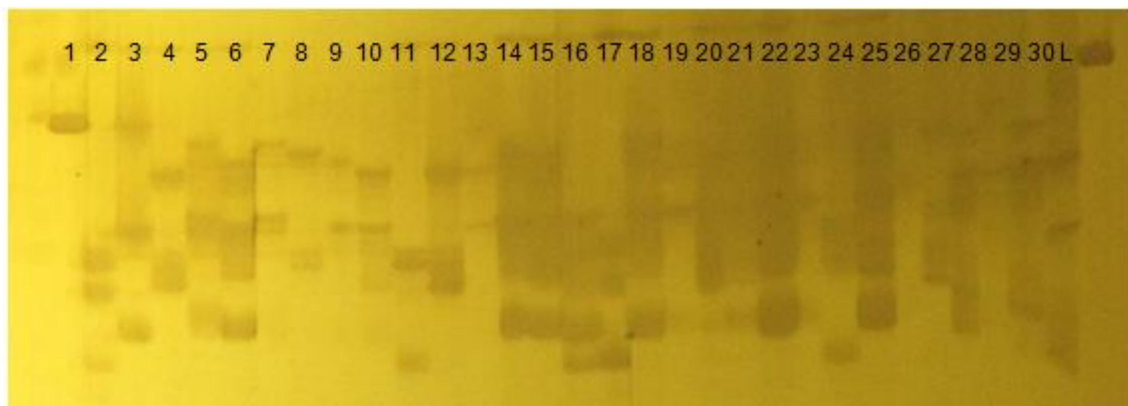

Figure 2: Amplification profile generated with primer RA430. Lanes 1- 30 represent sampled individuals of *R. arboreum* as presented in Table 3; Ld: 100 bp DNA ladder as size standard (Bangalore Genei^{TM}).

Figure 3: Amplification profile generated with primer R25. Lanes 1- 30 represent sampled individuals of *R. arboreum* as presented in Table 3; Ld: 100 bp DNA ladder as size standard (Bangalore Genei^{TM}).

Among the thirty eight primers used to study three populations of the Rhododendrons, fifteen primers were found to be highly informative and showed considerable variations. Locus amplification observed was higher which comes out to be 93%. This success rate was higher than the previous studies carried on other species of Rhododendron *viz. Rhododendron aureum* (Li *et al.*, 2011) and *Rhododendron simsii* (Tan *et al.*, 2009).

Total of 224 polymorphic bands were generated within the range from 2 – 14 alleles per locus and an average of 5.2 alleles per locus. The highest no of polymorphism was observed with RA25 (14), R422 (12), and RF98 (10). The fragment length of sequences was expected to be 83-228 bp but in turn it comes out to be in range of 100-420 bp. This larger length of fragments corresponds to the higher polymorphic potential of repeats (Herrera *et al.*, 2014). Expected heterozygosity (He) varied within the range of 0.422 to 0.917 (average 0.723) and the observed hetrozygosity (H_o) comes out to be 0.167 to 0.933 (average 0.523). The estimated PIC value ranged in between 0.104 to 0.811 (average 0.464). The most informative primers were RF304 (0.811), RF98 (0.720) and RF71 (0.710).

This data revealed high genetic diversity and gene flow in the studied populations which also indicate that the populations are outcrossing. According to the studies, low levels of the genetic diversity within population is in case of genetically restricted population but high levels of genetic diversity within population was observed in case of out crossing species (Torres *et al.*, 2003; Wagner *et al.*, 2011).

HWE was also studied, in which about 19 loci were found to in equilibrium but the other half, 19 loci were deviating from the HWE. Possible explanation for this deviation can be due to natural selection, genetic drift, gene flow and mutations (Alvarez, 2008; Van Ooijen, 1999). As a whole, no significant linkage disequilibrium was detected at population level ($P<0.05$). All details of the 38 loci viz. sequences of primer, repeat motif, annealing temperature of primer, number of observed alleles (N_a) (except null alleles), approximate size of the fragment, observed and expected heterozygosity, PIC value and accession number of the sequences in the table below (Table 3).

Also the population wise diversity parameters were estimated using GenAlEx software which give us estimate about the number of alleles, observed and expected heterozygosity (Table 4). The highest number of alleles was observed in Ghatasani population with average of 5.184 followed by Multhan with average of 4.789. Effective number of alleles (N_e) was also observed highest in Ghatasani followed by Multhan and Dalhousie (Table 5). Higher allelic number in Ghatasani indicated the shared gene pool with surrounding populations. These results are also supported by the lesser values of fixation index or Nei's gene diversity.

Table 3: 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 4: 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

The percentage polymorphism of loci was estimated and population wise effective number of alleles, Nei's gene diversity and Shannon's Informative index were also calculated (Table 5). The percentage of polymorphism observed was comparable or higher than earlier reports on generic or co-generic species of *Rhododendron* (Li *et al.*, 2011; Wang *et al.*, 2014)

Table 5: *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

Nei's original measure of genetic distance (Nei, 1978) reveals the genetic identity along with genetic distance between the pairs of population (Table 6) Dalhousie population came out to be most differentiated of three which can be because of isolation of population from other naturally occurring populations. Also, the samples from Ghatasani and Multhan were more genetically similar to each other as shown in dendrogram also (Figure 4). Dendrogram based on UPGMA modified from neighbor procedure of PHYLIP version 3.5, detects the presence of two clusters, one consist of Dalhousie population and the other Ghatasani and Multhan.

Table 6: 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	---

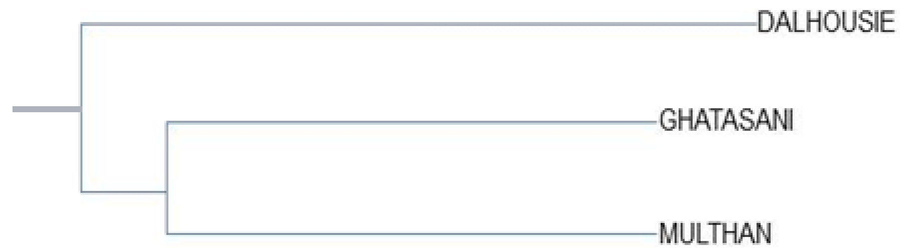

Figure 4: Dendrogram based on Nei's genetic distance for 3 populations of *R. arboreum* based on 38 SSR loci

Gene Flow in between the populations was also estimated on the basis of the genetic differentiation between subpopulations relative to the total genetic diversity which is denoted as F_{st} (Fixation index) (Table 7). Highest gene flow was observed between Ghatasani and Multhan confirming they are more alike as depicted by the dendrogram. Also, the higher value of genetic differentiation in the Dalhousie and Multhan depicts that the two populations are distantly related with low levels of gene flow.

Table 7: 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}

Total gene diversity within population was analyzed on the basis of Nie's F-statistics which uses fixation index (F_{is}) or inbreeding coefficient as a measure to evaluate diversity. Total gene diversity for all populations (F_{is}), within population diversity (F_{it}), genetic differentiation between subpopulations (F_{st}) and gene flow [$Nm= \{0.25\ (1-F_{st})\}/\ F_{st}$] were calculated at each locus (Table 8). The total gene diversity within population (F_{it}) was 0.320 and among (sub-population) (F_{is}) was 0.268.

If N_m is higher than 4, gene flow between populations is sufficient to create genetic differentiation. In our case, high values of N_m were there with an average of 5.036. Highest observed gene flow was at locus RA337 with a value of 23.625

and least fixation index of 0.010 whereas least gene flow (N_m = 1.070) was at locus R356 having F_{st} =0.189. The higher values of Nm in the study indicate the absence geographical barrier and are geographically closer. Or high gene flow can be due to human interference in dispersion of seeds and pollens (Ayres and Ryan, 1999).

Table 8: 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

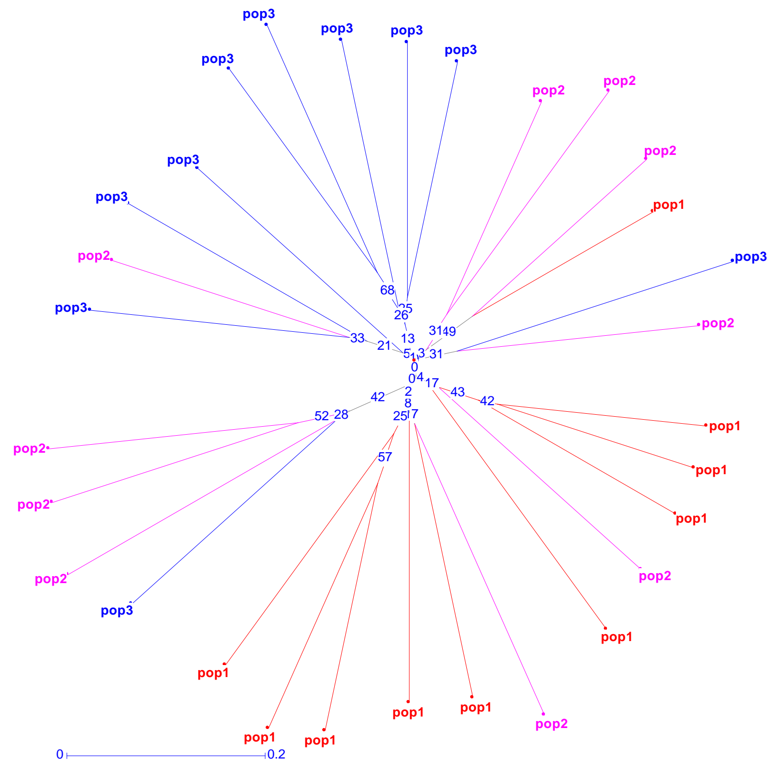

Figure 5: Unrooted tree constructed by DARwin software. Each coloured branch represents single individual collected from corresponding inferred population (pop1- Dalhousie; pop2- Ghatasani; pop3- Multhan)

On the basis of the evolutionary dissimilarities, DARwin is used to develop diversity and phylogenetic analysis. It uses different algorithms to perform factorial analyses and to construct trees and to assess their reliability or to compare several trees on the same data set. The un-rooted tree constructed by the software on the basis of genetic distance showed two major clusters with further sub clustering (Figure 5). Not a single cluster is either comprised of a single population hence concluding the presence of gene flow among populations.

To know the actual structure of the population of *R. arboreum*, a Bayesian model clustering program, STRUCTURE is used. Clustering or grouping of the population's relative distinctiveness was done on the basis of genetically homogeneous 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.

. Here also on the basis of the given K values, ΔK values were calculated. The set of values are then used to plot a graph, in which the K value corresponding to the highest value of ΔK, gives the actual number of populations

(Figure 6). Highest value of ΔK is at K=2, number of real populations is two which further supports UPGMA dendrogram.

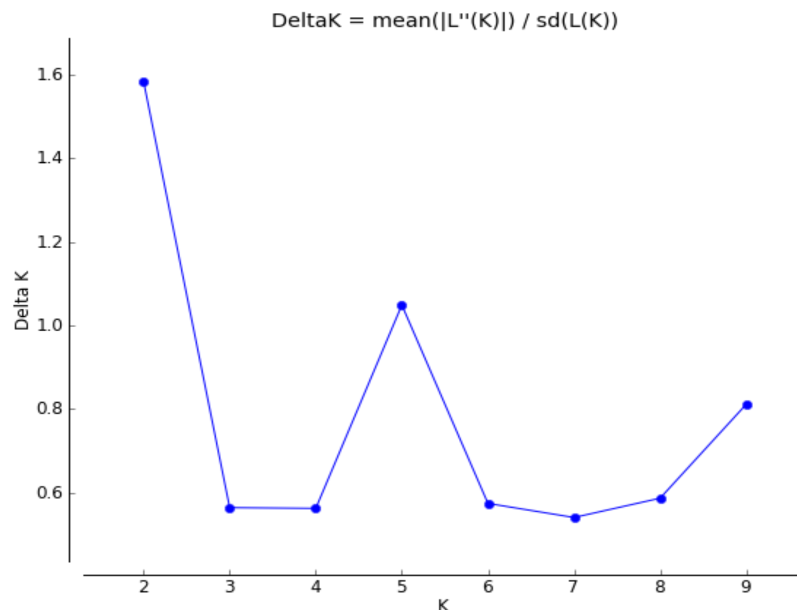

Figure 6: Most appropriate value of K for a range of 3 populations using the second order statistics (ΔK) given by Evanno *et al.*, (2005)

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 Ghatasani), 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.

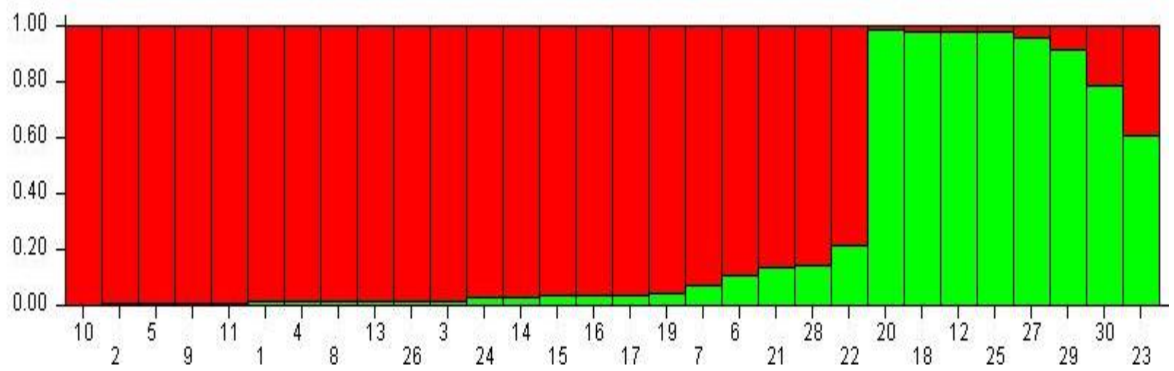

Figure 7: Structure plot with membership coefficient (Q) of individuals on y- axis. Two clusters inferred by Structure are combined from clustering of all individuals. Each color represents different genetic stock and each vertical line broken by colors represent each individual

Also the populations were to be studied for any detectable reduction of the number of alleles or population size, which can ultimately lead to the phenomenon called bottleneck. The BOTTLENECK software assumes that all the loci are in HWE and mutation- drift equilibrium is verified under two mutation models: Infinite allele model (IAM) and one-step stepwise mutation model (SMM). The BOTTLENECK program computes the distribution of the heterozygosity (H_e) for each population sample and each locus which were expected from the observed number of alleles and sample size and ultimately detects the heterozygosity excess or deficit for each locus.

Table 9: Allele frequency based mutation- drift equilibrium of SSR loci

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

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

In IAM, at 31 loci, heterozygosity excess was observed which is significant at 5% significance level in the three, Sign test, Standardized difference and

Wilcoxon test (Table 9). But on the other hand, heterozygosity excess for 23 loci was not significant as observed under SMM.

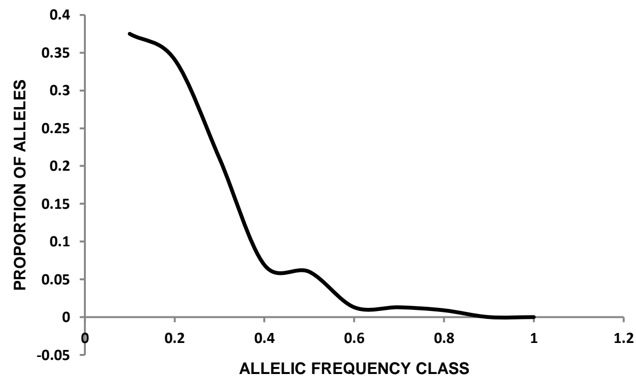

Figure 8: Graphic representation of proportion of alleles and their distribution in populations of *R. arboreum*

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.

Summary

Rhododendron arboretum is key species in Himalayan region and is ecologically significant. Due to anthropogenic activity, over exploitation by the local people and changes in the environmental conditions, there is need for organization of conservation strategies so that species can be maintained in the future. Markers are useful to get an insight about the genetic structure of the plant. For the revelation of the genetic diversity and population structure of the species, co-dominant microsatellite markers are a competent tool. 38 novel primers were characterized on the three populations of the *Rhododendron arboretum* from Dalhousie, Ghatasani and Multhan on 6% denaturing PAGE. Length of amplified fragment came out to be 100-420 bp which was expected to be 83- 228 bp. Number of observed alleles was in range of 2-14 with average of 5.2 alleles. In the diversity analysis, expected heterozygosity ranged from 0.422 to 0.917 (average 0.723) and observed heterozygosity varied from 0.167 to 0.933 (average 0.523). Polymorphic information content (PIC) 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. There was no bottleneck observed in the population and linkage disequilibrium, also, came out be zero at $P<0.05$. High levels of polymorphism for the loci and gene flow (average $N_m= 5.036$) was observed which indicate the high genetic diversity within the populations. Maximum fixation index was observed in the Dalhousie population which tells us that the population is most genetically differentiated from the other two. High effective number of alleles was observed in Ghatasani population. Multhan and Ghatasani populations were closely related due to the gene flow also shown in UPGMA based dendrogram and Dalhousie was most distinct of all. Results of UPGMA were further supported by the DARwin and STRUCTURE. Both these software grouped the populations into two major clusters. No cluster is comprised of a single population indicating that the genetic stock was inter-mixed due to the presence of the germplasm migration among the populations.

The polymorphism of the loci is highly informative data which will be helpful in managing conservation policies for *R. arboretum* and helpful in the genotyping. This information will also help in the determination of large population's gene flow and effects of other evolutionary forces disturbing HWE. On the basis of the

benefits and large application range of SSRs, we can say that the SSRs are a very important tool for genetics.

References

Abdul-Muneer, P. (2014). Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies. *Genetics research international* 2014.

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

Ayres, D.R. and Ryan, F.J. (1999). Genetic diversity and structure of the narrow endemic Wyethia reticulata and its congener *W. bolanderi* (Asteraceae) using RAPD and allozyme techniques. *American Journal of Botany* **86**: 344-353.

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**: 748-763.

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

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

Cardle, L., Ramsay, L., Milbourne, D., Macaulay, M., Marshall, D. and Waugh, R. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* **156**: 847-854.

Charlesworth, D. and Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual review of ecology and systematics*: 237-268.

Cordeiro, G.M., Taylor, G. and Henry, R.J. (2000). Characterisation of microsatellite markers from sugarcane (< i> *Saccharum*</i> *sp.*), a highly polyploid species. *Plant Science* **155**: 161-168.

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

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

Ehlers, A., Worm, B. and Reusch, T.B. (2008). Importance of genetic diversity in eelgrass *Zostera marina* for its resilience to global warming. *Marine ecology progress series* **355**: 1-7.

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**: 995-1007.

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**: 227-236.

Hasan, M., Seyis, F., Badani, A., Pons-Kühnemann, J., Friedt, W., Lühs, W. and Snowdon, R. (2006). Analysis of genetic diversity in the *Brassica napus L.* gene pool using SSR markers. *Genetic Resources and Crop Evolution* **53**: 793-802.

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**: 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**: 75-84.

Johnson, M.T., Lajeunesse, M.J. and Agrawal, A.A. (2006). Additive and interactive effects of plant genotypic diversity on arthropod communities and plant fitness. *Ecology Letters* **9**: 24-34.

Kudo, G. (1993). Relationships between flowering time and fruit set of the entomophilous alpine shrub, *Rhododendron aureum* (Ericaceae), inhabiting snow patches. *American Journal of Botany*: 1300-1304.

Lanying, Z., Yongqing, W. and Li, Z. (2008). Genetic diversity and relationship of Rhododendron species based on RAPD analysis. *Am. Eur. J. Agric. Environ. Sci* **3**: 626-631.

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**: e250-e252.

Loveless, M.D. and Hamrick, J.L. (1984). Ecological determinants of genetic structure in plant populations. *Annual review of ecology and systematics*: 65-95.

Miah, G., Rafii, M.Y., Ismail, M.R., Puteh, A.B., Rahim, H.A., Islam, K.N. and Latif, M.A. (2013). A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. *International journal of molecular sciences* **14**: 22499-22528.

Milne, R.I. and Abbott, R.J. (2000). Origin and evolution of invasive naturalized material of *Rhododendron ponticum L.* in the British Isles. *Molecular Ecology* **9**: 541-556.

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**: 405-412.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

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**: 4546-4569.

Rai, T. and Rai, L. (1994). *Trees of the Sikkim Himalaya*: Indus Publishing.

Sánchez-Pérez, R., Ruiz, D., Dicenta, F., Egea, J. and Martínez-Gómez, P. (2005). Application of simple sequence repeat (SSR) markers in apricot breeding: molecular characterization, protection, and genetic relationships. *Scientia horticulturae* **103**: 305-315.

Singh, K., Kumar, S., Rai, L. and Krishna, A. (2003). Rhododendrons conservation in the Sikkim Himalaya. *current science-bangalore-* **85**: 602-606.

Srivastava, P. (2012). *Rhododendron arboreum*: An overview.

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**: 649-652.

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**: 1553-1555.

Tautz, D. (1989). Hypervariabflity of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research* **17**: 6463-6471.

Torres, E., Iriondo, J.M. and Pérez, C. (2003). Genetic structure of an endangered plant, *Antirrhinum microphyllum* (Scrophulariaceae): allozyme and RAPD analysis. *American Journal of Botany* **90**: 85-92.

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

Wagner, V., Durka, W. and Hensen, I. (2011). Increased genetic differentiation but no reduced genetic diversity in peripheral vs. central populations of a steppe grass. *American Journal of Botany* **98**: 1173-1179.

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.

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**: 520-530.

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

ZHA, H.G., Milne, R.I. and Sun, H. (2008). Morphological and molecular evidence of natural hybridization between two distantly related Rhododendron

species from the Sino-Himalaya. *Botanical Journal of the Linnean Society* **156**: 119-129.

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. 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.