STUDY OF GENETIC DIVERSITY OF *Artemisia annua* FOR ARTEMISININ USING SSR MOLECULAR MARKERS

BY

ASEEM CHAUDHRY (061572)

MAY2010

Submitted in partial fulfillment of the Degree of Bachelor of Technology

DEPARTMENT OF
BIOTECHNOLOGY & BIOINFORMATICS
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN, HP, INDIA

Certificate

This is to certify that the project report entitled **"Study of Genetic Diversity of *Artemisia annua* for Artemisinin using SSR Molecular Markers"**, submitted by Aseem Chaudhry in partial fulfillment for the award of the degree of Bachelor of Technology in Biotechnology to Jaypee University of Information Technology, Waknaghat, Solan has been carried out under my supervision.

Date: 21/05/10

Dr. Harvinder Singh

Department of Biotechnology and Bioinformatics

JUIT, Waknaghat

Certified that this work has not been submitted partially or fully to any other University or Institute for the award of this or any other degree or diploma.

Aseem Chaudhry

Aseem Chaudhry

061572

Acknowledgement

Apart from my effort, my work could never have been complete without the constant help and support of a number of people, towards whom I would like to extend my sincerest gratitude.

First and foremost, it is my proud privilege to have the opportunity to thank my Project Guide, **Dr. Harvinder Singh**. It was an honor to carry out my project work under his expert guidance and supervision, without which my work would never have been possible.

I would also like to express my gratitude to the Head of the Department **Dr. R. S. Chauhan** and **Dr. Pradeep Naik** for their critical analysis of my work, which was of immense help to me.

I would also like to thank **Dr. Y. Medury,** the Vice Chancellor of the university, and **Brig. (Retd.) Balbir Singh,** the Registrar of the university for allowing us to work in the university's laboratories and providing us with all necessary facilities, equipment and chemicals.

I would indeed be failing in my duties if I did not acknowledge the constant support and immense patience of the staff of the biotechnology lab of the university. Lastly, I would like to thank my parents whose constant encouragement gave me the courage and enthusiasm to complete this work.

Date: 21/05/10

Aseem Chaudhry
Aseem Chaudhry

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER 3

RESULTS AND DISCUSSION

LIST OF ABBREVIATIONS

1. SSRs	: Simple Sequence Repeats
2. RFLP	: Restriction Fragment Length Polymorphism
3. ISSR	: Inter simple sequence repeats
4. RAPD	: Random amplification of polymorphic DNA
5. AFLP	: Amplified fragment length polymorphism
6. PCR	: Polymerase Chain Reaction
7. PIC	: Polymorphism Information Content
8. ACT	: Artemisinin Combination Therapies
9. PCA	: Principle Component Analysis
10. AMOVA	: Analysis of Molecular Variance
11. CTAB	: Cetyl Trimethyl Ammonium Bromide
12. EST	: Expressed Sequence Tag

ABSTRACT

The genetic diversity of *Artemisia annua* for artemisinin using SSR markers was studied. As of today there are ~ 85,282 ESTs developed from Artemisia *annua.* On the assumption that genes involved in biosynthesis of artemisinin will be orthologues and conserved, both in sequence, structure and function, between Artemisia genotypes localized at different altitudes, the available EST sequences of *Artemisia annua* were utilized in amplifying *Artemisia* genomic DNA and for identifying molecular markers. The genes of key enzymes involved in the biosynthesis of artemisinin and the genes of enzymes relevant to the biosynthesis of artemisinin have been sequenced and are available in Genbank.

On the basis of identified genes, primers were designed using BatchPrimer3. PCR was performed. 16 SSR markers generated a total of 38 alleles with PIC value in the range of 0.2-0.8. It was observed that there was considerable genetic diversity amongst sample genotypes collected from the two valleys viz. Leh and Nubra. The Nubra valley showed greater diversity as compared to Leh valley. Molecular variance between the two valleys is 23% whereas within the valleys, it is 77%. This variation may be helpful in making strategies for germplasm collection and evaluation. The best genotypes,yielding high artemisinin content, are of great potential for further research.

CHAPTER 1

INTRODUCTION

Artemisia annua L. (*Asteraceae* formerly *Compositae*), the natural source of Artemisinin, is an annual herb native to Asia but now also found in regions of Europe, North America, Central and South America. It is one of the most important medicinal plants of cold arid regions of India especially in Ladakh region. Traditionally, the plant is used for crafting of aromatic wreaths, flavouring of spirits, and its essential oil is distilled for perfumery and industrial use (Ferreira et al., 1996). The plant thrives well in dry cold region of Ladakh with sandy soils and marginally rock area. It is well adapted to survive at high altitudes of 2,500-5,000 m above mean sea level (MSL) and the temperature, nutrient and environmental stress that they are subjected to under the cold arid conditions.

Artemisinin, a sesquiterpene endoperoxide has been found to be potent antimalarial drug against strains of *Plasmodium falciparum, parasite* species, which accounts for maximum number of malarial cases. Strains of this specie have developed resistant against a number of previously used antimalarial drugs some of which are Amodiaquine, quinine, chloroquine and primaquine. Requirement of new drug led to study of more than thirty species of genus Artemisia, only *Artemisia annua* L. and *Artemisia Apiaca* Hance proved to be effective against Plasmodium falciparum and Plasmodium vivax.

But problem with individual Artemisinin use is that parasite is developing resistance against it. Parasite resistance to artemisinin has recently been confirmed in western Cambodia. Artemisinin Combination Therapies (ACTs) are being recommended and monotherapies using Artemisinin discouraged [so as to protect Artemisinin being developed resistance against by the Plasmodium parasite.

Using ACTs has been strongly supported by the global health communities with both funding and demand. ACTs use is expected to increase massively in coming future. The growing concern today about the availability of ACTs at high demands because production of high-quality artemisinin in large quantities is very difficult. This low availability of artemisinin accounts to the fact that plant based production of Artemisinin is very difficult because A. annua remains relatively undeveloped as a crop. The artemisinin content of plants from different origins varies considerably and is highly heritable, and being from the family of Asteraceae, this plant favours outcrossing over selfing. Chemical synthesis of the artemisinin, although available, is commercially non-viable, and efforts to produce it in cultured cells have, so far, not been very fruitful. Thus, the cultivated plant is the sole source of the drug (Sangwan et al., 1993).

Artemisinin action on Plasmodium species is noteworthy, both in vivo and in vitro. Artemisinin and its derivatives affect the erythrocytic stage of plasmodia, but is unable to affect the exoerythrocytic stage. There has been no conclusive evidence over a particular mechanism of action by artemisinin over parasite, but two mechanisms have been considered to be important in this process. A two-step mechanism proposed by (Meshnick et al.,1991) suggests the heme-catalyzed cleavage of the endoperoxide bridge forms a free radical followed by specific and selective alkylation of some malarial proteins. Hemoglobin is utilized as a major source of amino acids by the malarial parasite during its intraerythrocytic development and proliferation. Through a sequentially ordered process that involves cysteine as well as aspartic acid proteases, constant degradation of hemoglobin inside the parasite food vacuole occurs, and toxic free heme which is generated due to digestion of hemoglobin, is simultaneously detoxified by the malarial parasite through a specific mechanism of heme polymerization. The polymerized heme is commonly referred to as "hemozoin" or "malaria pigment" gets accumulated in the form of a crystalline, insoluble, black-brown pigment, and once the parasite life cycle is complete, this pigment is sequestered to various tissues of the host.

This pathway, heme polymerization, offers a potential biochemical target for designing antimalarials and is specific to the malarial parasite. Studies have shown that artemisinin taken up by the malarial parasite growing *in vitro* was selectively concentrated in the parasite food vacuole and was associated with hemozoin.Artemisinin also interacts with heme, forming covalent adducts.Studies provided by (Pandey et al ,1999*)* provide evidence that the antimalarial effect of artemisinin may at least partly be due to inhibition of malarial hemoglobin degradation pathway and heme detoxification system.

Genetic uniqueness is brought about by two factors, inheritance and mutations and since all genetic differences between individuals are laid down in the primary sequence of their genomic DNA, the most straight forward method is identifying an individual sequence for genomes under comparison (Krawczak and Schmidtke, 1994). Genome sequence composition is portrayed by DNA markers which enable us to detect the differences in the genetic information which is carried by the different individuals. Microsatellites or SSRs (*simple sequence repeats*) are ubiquitous short tandem duplications occurring in eukaryotic organisms. These sequences are among the best marker technologies applied in plant genetics and breeding (Luciano Carlos daMaia et al, 2008). SSRs can be found in different regions of genes, coding sequences, untranslated sequences (5_-UTR and 3_-UTR), and introns, where the expansions and/or contractions can lead to gene gain or loss of function. Also, there are evidences that genomic distribution of SSRs is related to chromatin organization, recombination, and DNA repair. SSRs are found throughout the genome, in both protein-coding and non-coding regions.

Artemisinin (extracted from *Artemisia annua* L) is currently at the heart of WHO's global fight against malaria, and a new generation of drugs, based on artemisinin, should reach the market in a few years time. Remarkably, there has still not been one documented clinical case of resistance to artemisinin or any of its derivatives. Indeed, when treating multi-drug resistant and non-drug resistant *falciparum* malaria, it is now recommended by the WHO that artemisinin be taken in combination with another drug (this is a strategy designed to slow the development of resistance, since during treatment with two-or more- drugs, the chance of a

mutant emerging which resistant to both is drastically reduced). In the continuing absence of an effective malaria vaccine, the development of new anti-malarial drugs- most likely derived from, or inspired by, artemisinin – will continue to be our primary in the fight against malaria. However, artemisinin is in short supply and unaffordable to most malaria suffers.

Despite much research, artemisinin remains the only known natural product to contain a 1,2,4-trioxane ring , and *Artemisia annua* continues to be the only known natural source. Several total syntheses of artemisinin are reported but, because of the complexity of the target compound, none is commercially competitive with extraction from natural source. In addition, although micropropagation (tissue culture) of *Artemisia annua* can be easily accomplished, it is equally unlikely that *in vitro* production of artemisinin will become a commercially viable proposition. It seems that the production of artemisinin from wild or field cultivated plants will be the preferred option for the foreseeable future. Thus development of *Artemisia* species with high artemisinin yield is the need of the hour. For improvement of *Artemisia* knowledge of magnitude of variation in available germplasm and association of characters with artemisinin content are a prerequisite and meager information is available on these aspects. Thus the development or selection of genetically superior genotypes of *Artemisia* is a high priority of various Departments and funding agencies of Govt. of India. However, lack of essential genome resources such as availability of a genetic map, identification of high artemisinin content genotypes, non-existance of molecular markers are major stumbling blocks in realizing the full potential of molecular approaches in genetic improvement of *Artemisia*.

Objectives of this study also were to compare the effectiveness of both the PCR-based molecular approaches to determine the genetic relationships among several genotypes of *Artemisia annua* from two valleys i.e. Leh and Nubra of Ladakh region in India. These valleys are separated from each other through natural mountain barrier. To the best of our knowledge, this is the first report of molecular genetic diversity studies in Artemisia annua from this region of the world.

1.1 Classification of *Artemisia annua*

- Kingdom: Plantae – Plants
- Subkingdom: Tracheobionta – Vascular plants
- Superdivision: Spermatophyta – Seed plants
- Division: Magnoliophyta – Flowering plants
- Class: Magnoliopsida – Dicotyledons
- Subclass: Asteridae
- Order: Asterales
- Family: Asteraceae – Aster family
- Genus: Artemisia L. – sagebrush
- Species: Artemisia annua L. – sweet sagewort

1.2 Physical Characteristics

- Annual growing, 3m by 1m at a fast rate.
- It is in flower from August to September, and the seeds ripen from September to October. The flowers are hermaphrodite and are pollinated by Insects.
- The plant prefers light (sandy) and medium (loamy) soils, requires well-drained soil and can grow in nutritionally poor soil. The plant prefers acid, neutral and basic (alkaline) soils. It can grow in semi-shade (light woodland) or no shade. It requires dry or moist soil.

1.3 Distribution of *Artemisia annua* in India

ARTEMISIA is majorly found in:

- Jammu and Kashmir (Leh–Ladakh)
- Himachal Pradesh

1.4 Medicinal Uses

- Qing Haosu, better known as sweet wormwood, is a traditional Chinese herbal medicine. An aromatic anti-bacterial plant, recent research has shown that it destroys malarial parasites, lowers fevers and checks bleeding. It is often used in the Tropics as an affordable and effective anti-malarial.
- The leaves are antiperiodic, antiseptic, digestive, febrifuge. An infusion of the leaves is used internally to treat fevers, colds, diarrhoea etc. Externally, the leaves are poulticed onto nose bleeds, boils and abscesses.
- The plant contains artemisinin, this substance has proved to be a dramatically effective anti-malarial against multi-drug resistant Plasmodium spp. Clinical trials have shown it to be 90% effective and more successful than standard drugs.

1.5 Edible Uses: (As Condiment).

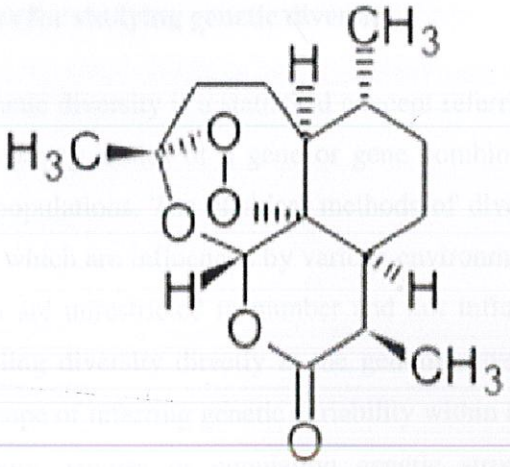
- An essential oil in the leaves is used as a flavouring in spirits such as vermouth. The plant yields 0.3% essential oil.

Other Uses:

- The plant is used in China as a medium for growing *Aspergillus* which is used in brewing wine.
- This has an agreeable, refreshing and slightly balsamic odour and has been used in perfumery.

Artemisia annua

1.6 CHEMICAL STRUCTURE OF ARTEMISININ



CHEMICAL NAME : Octahydro-3,6,9-trimethyl-3,12-epoxy-12*H*-pyrano[4,3]-1,2-enzodioxepin-10(3*H*)-one.

CHEMICAL FORMULA : C15H22O5

MOLECULAR MASS : 282.332 g/mol

As artemisinin itself has physical properties such as poor bioavailability that limit its effectiveness, semi-synthetic derivatives of artemisinin have been developed.

1.7 Chemically modified analogues

- Artesunate (water-soluble: for oral, rectal, intramuscular, or intravenous use)
- Artemether (lipid-soluble: for oral, rectal or intramuscular use)
- Dihydroartemisinin
- Artelinic acid
- Artenimol
- Artemotil

1.8 MOLECULAR MARKERS

Use of molecular markers for studying genetic diversity

In simple terms, genetic diversity is a statistical concept referring to the variations within the individual gene loci among alleles of a gene or gene combinations, between individual plants or between plant populations. The classical methods of diversity studies are based on morphological characters which are influenced by various environmental factors. However, the molecular markers which are unrestricted in number and not influenced by the environment have the ability of sampling diversity directly at the genome level. They provide increased accuracy and expanded scope of inferring genetic variability within and between populations of plant species. Traditionally, studies of population genetic structure have used proteins, isozymes, allozyme etc.These markers provide informative genetic markers, detecting useful levels of genetic variations within populations using straightforward laboratory procedures that

are relatively rapid and inexpensive, but there are some widely recognised limitations of their use. Only a little portion of the genome was covered when protein and isozymes were analysed and most part of the genome remained unanalysed. So biochemical characterization does not reflect the entire genome. Furthermore, all genetic changes that occur at the DNA level are not detected at the protein level. The detection of genetic variation is limited to protein coding loci which may not be the representative of the entire genome, hence may lead to underestimation of genetic diversity.Enzymes are tissue and species specific and change during development and differentiation. It is likely to get more variability in DNA than in proteins since much of the DNA does not code for active genes and is unexpressed. Analysis of such variations can be done in the form of DNA profiling.

The drawbacks of biochemical markers led many workers to shift to nuclear DNA markers such as RFLPs, RAPDs, ISSRs, AFLPs, etc. DNA-based molecular-marker techniques have been proved powerful in genetic diversity estimations.Molecular markers, unlike morphological markers are stable and have been found useful in population studies and phylogeny. Different types of marker systems have been used for biodiversity and phylogenetic analyses. These include restriction fragment length polymorphism (RFLP), simple sequence repeats (SSR), inter simple sequence repeats (ISSR), random amplification of polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP). RFLPs are well suited for the construction of linkage maps because of their high specificity and are also less polymorphic, more expensive and laborious compared to RAPD and ISSR. Currently AFLP is the method of choice for analysis of germplasm, genetic diversity and phylogeny, gene tagging and molecular map construction.

1.9 Various types of DNA markers

There are various types of DNA markers available presently to evaluate DNA polymorphism in sample genomes. Selection of a correct marker system depends upon the type of study to be undertaken and whether that marker system would fulfill at least a few of the mentioned characteristics such as easy availability, highly polymorphic nature, Mendelian inheritance, frequent occurrence in genome, selective neutral behavior, easy and fast assay, high reproducibility, free of epistasis and pleiotropy etc, (Welsh and Clelland, 1990). These markers are generally classified as hybridization based markers and polymerase chain reaction (PCR) based markers. In the hybridization-based markers, the DNA profiles are visualized by hybridizing the restriction enzyme digested DNA to a labeled probe which is a DNA fragment of known / unknown sequence. In case of PCR based markers, the primers of known sequence and length are used to amplify genomic and cDNA sequences which are visualised by gel electrophoresis technique. The invention of PCR which is a very versatile and extremely sensitive technique, (Saiki and Scharf et al., 1985) uses a thermostable DNA polymerase and has changed the total scenario of molecular biology and has also brought about a multitude of new possibilities in molecular marker research

1.10 SSRs

- Simple sequence repeats (microsatellites) are stretches of DNA consisting of tandemly repeating nucleotides of 2-8 bases, which are dispersed throughout eukaryotic genomes.
- Microsatellite polymorphism is revealed by PCR amplification of total genomic DNA using two unique primers (18-25 bp) which flank the microsatellite locus. Amplification products from different individuals can then be resolved on gels.

- The value of microsatellite markers is due to-

1. Multiallelic nature
2. Co-dominance
3. Ease of detection by PCR
4. Relative abundance
5. Requires only a small amount of DNA

1.11 Rationale of the Study

The level of gene conservation i.e. synteny in structure, function and chromosomal location is an important criterion in determining the extent to which comparative genomics can be applied across species. Extensive gene conservation, both in structure and function, has been reported in grass genomes(Bennetzen et al., 1998), Crucifers (Me et al., 2006) and solanaceous plants (Causse et al., 2005). The overall purpose and goal of understanding gene conservation across different genomes is to make use of model genome resource for practical applications. Off late, the candidate genes involved in the biosynthetic pathways are considered to be potential targets for identifying gene-specific molecular markers.Candidate gene approach has been successfully used in association mapping of silage corn digestibility and maize kernel composition and starch production in maize.These reports and the results obtained therein imply that there is possibility of identifying candidate genes involved in the biosynthetic pathways of a trait and then to explore the possibilities of developing molecular markers from the candidate genes. This is expected to speed up the process of gene discovery and their subsequent utilization.

SSRs are a class of molecular markers that have been identified from genic as well non-genic regions of genomes and the former have been suggested to be of great practical importance (Varshney et al., 2005). SSRs are versatile candidate markers and have been useful for integrating the genetic, physical and sequence maps in plant species, and simultaneously have provided researchers with an efficient tool to link phenotypic and genotypic variation. Off late, the development of molecular markers has been speeded up both in model genomes and other

genomes related to the model genome. The availability of a plethora of bioinformatics tools and low cost high throughput genotyping technologies has opened up avenues in exploring the model genomes for utilizing in resource-poor genomes.

1.12 Current status of research and development in the subject

National status

In India, studies made so far on *Artemisia annua* pertains to RAPD profile based genetic characterization of chemotypic variants of *Artemisia annua L.* (Sangwan et al., 1999). Work on transformation of artemisinic acid to artemisinin B and artemisinin both in vivo and in a cell free system from *Artemisia annua* has been carried out with successful results (Sangwan et al., 1993). Extraction of essential and terpenoids from *Artemisia annua* has been carried out. The yield of artemisinin and essential oils from *Artemisia annua* is very much influenced by the planting time. Cultivation, utilization and chemical studies of *Artemisia annua* was reported by (Sharma et al., 1991). The high yielding variety, Jeevanraksha of *Artemisia annua* containing up to 1.2% artemisinin in the leaves and flowers has been developed by CIMAP. The gene 1467bp for cyp71AV1, which catalyses the three step oxidation of amorpha-4,11-diene to artemisinic acid in the artemisinin biosynthetic pathway of *Artemisia annua* has been isolated, cloned and sequence-characterized by CIMAP.

International Status

Very limited information is available on various biochemical and molecular aspects of *Artemisia annua.* Nonenzymatic, photochemical conversion of dihydroartemisinic acid into artemisinin in vivo has been carried out by (Wallaart et al. 1999 a,b). Bertea et al. 2005 identified intermediates and enzymes involved in the biosynthesis of artemisinin in *Artemisia annua.* Several key enzymes involved in the biosynthesis of artemisinin have been discovered (Liu et al., 2006), such as 3-hydoxy-3-methlyglutaryl coenzyme A reductase (HMGR), farnesyl diphosphate synthase (FPPS), and sesquiterpene synthase (SES). An alternative

pathway of synthesis of artemisinin from farnesyl diphosphate by the activity of sesquiterpene cyclase and amorpha-4, 11-diene sythetase enzymes has been worked out. It is attractive to produce transgenic plants of *Artemisia annua,* which ensures a constant high production of artemisinin by overexpressing the enzymes in the terpene biosynthetic pathway or by inhibiting an enzyme of another pathway competing for artemisinin precursors. The genes of the key enzymes involved in artemisinin biosynthesis, such as farnesyl diphosphate synthase (FPS), amorpha-4, 11-diene sythetase (AMS), and the genes of the enzymes relevant to the biosynthesis of artemisinin including squalene synthase (SQS),have been cloned from *Artemisia annua*. The enzyme involved in the biochemical transformation of artemisinin B to artemisinin was purified. By genetic engineering, the key enzymes involved in biosynthesis of artemisinin, such as FPS and AMS, can be overexpressed and artemisinin production was significantly enhanced in the transgenic high yield *Artemisia annua.*Although the approach is targeted to establishment of alternative pathways for the conversion of precursors to artemisinin and development of transgenic plant to increase the yield, the methods developed in this study can be applied in many areas of plant biotechnology and functional genomics research particularly aimed at genetic enhancement of artemisinin content.

1.13 OBJECTIVE :

1. Insilico analysis and primer designing for target genes involved in the biosynthesis of artemisinin.
2. To study genetic diversity for artemisinin, in *Artemisia annua* using SSR markers.

CHAPTER 2

MATERIALS AND METHODS

2.1 Collection of Germplasms and In silico analysis of *Artemisia annua*.

Collection of different *Artemisia annua* was made from different altitude levels from Leh and Nubra valleys (J & K), and documented. Each sample was considered as one genotype and coded according to the site of collection. 10 samples each were collected from one site. Data on morphological characters and environmental factors like temperature, rain fall, altitude distance etc. was collected with respect to the site of collection and proper documentation of the data was done.

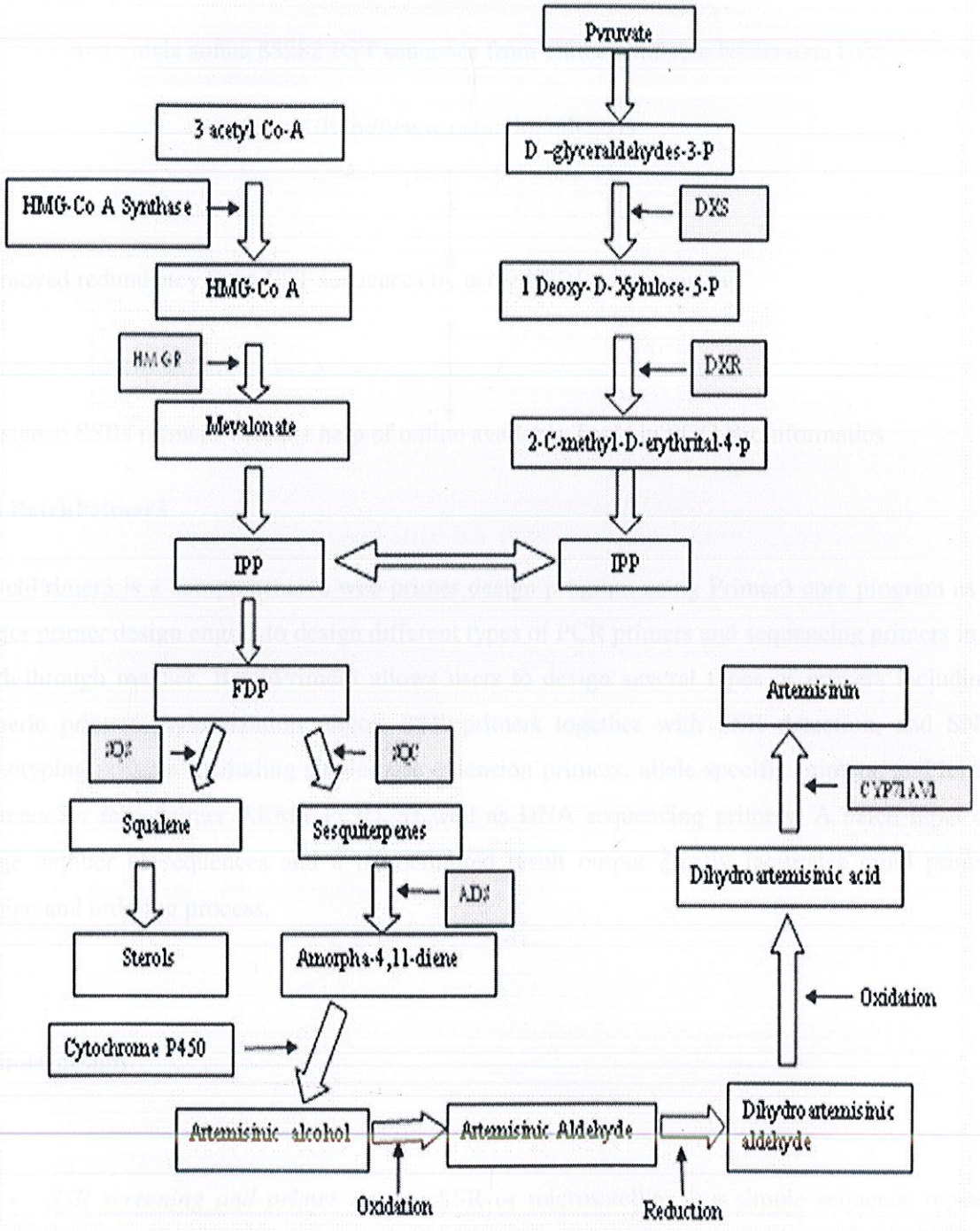
2.2 InSilico Analysis

BlastClust was performed on dataset of 85,282 EST sequences to remove the highly similar sequences using 60% sequence similarity and 100% sequence coverage and and the result was 68974 EST sequences.

Genes involved in artemisinin biosynthesis and considered for designing primers.

Enzyme	Gene	Function
Deoxyxylulose Synthase	dxs	1-deoxy-D-xylose-5-phosphate synthase activity
Deoxyxylulose reductase	dxr	Isomerase and oxidoreductase activity
3-hydroxy-3-methyl coenzyme A reductase	HMGR	HMG-Co A synthase
Franesyl diphosphate synthase	FPS	Synthesis of FDP
Epicedrol Synthase	EPS	8-epicedrol
Amorpha diene synthase	ADS	Amorpha 2, 11, diens
Squalene synthase	aasqs	Farnesyl-diphosphate famesyltransferase activity.
Squalenesynthase fragment	Sqs1	Transferase activity
CYP71AV1	CYP71AV1	Catalyzes 3 steps post ADS
Peroxidase 1	Pod1	Favored the bioconversion of artemisinine acid
	Idi1	Function unknown

PATHWAY INVOLVED IN THE BIOSYNTHESIS OF ARTEMISININ



Flow chart of Primer Designing

Aretemisia annua 85282 EST sequence from online available NCBI data base

(http://www.ncbi.nlm.nih.gov)

Removed redundancy from EST sequences by using PERL programming

Designed SSRs primers with the help of online available Tools in PGG Bioinformatics

2.3 BatchPrimer3

BatchPrimer3 is a comprehensive web primer design program using Primer3 core program as a major primer design engine to design different types of PCR primers and sequencing primers in a high-through manner. BatchPrimer3 allows users to design several types of primers including generic primers, hybridization oligos, SSR primers together with SSR detection, and SNP genotyping primers (including single-base extension primers, allele-specific primers, and tetra-primers for tetra-primer ARMS PCR), as well as DNA sequencing primers. A batch input of large number of sequences and a tab-delimited result output greatly facilitates rapid primer design and ordering process.

Primer design

- *SSR screening and primer design*: SSR or microsatellite is a simple sequence repeat, which is a useful genetic marker. SSR primers are picked from the SSR-flanking regions.

SSR motif screening varies in criteria of SSR definition, i.e., motif length and number of motif repeats or SSR length. Typically dinucleotide motifs to hexanucleotide motifs are detected with at least 12 nucleotides in length of SSRs. Since the different criteria of SSR screening will result in large difference in screening results, BatchPrimer3 provides flexible options to allow users to set the screening criteria. A regular expression was used to design an algorithm to detect the repeat motifs and number of repeats. The detected SSRs are masked as targets. Then Primer3 program picks the best pairs of primers flanking the targets.

Sequence Input

Two ways are available to input sequences:

1. Sequences can be copied and then pasted to a sequence text box. This approach has a maximum 256 kb size limit.
2. A FASTA file can be uploaded to the server and the sequence size limitation only depends on Internet speed and server machine memory.

A FASTA format is acceptable in the BatchPrimer3 program.

Result output

The BatchPrimer3 program produces four parts of outputs:

1. a main HTML page containing the primer design summary of all input sequences;
2. an HTML table page listing all designed primers and primer properties;
3. a tab-delimited text file with the same contents in the HTML table page,
4. and a detailed primer view page for each sequence with successfully designed primers. A simple click on the links on the main HTML page or HTML table page will display the primer view. The primer list can be directly saved as a text file or an Excel file for further editing or primer ordering.

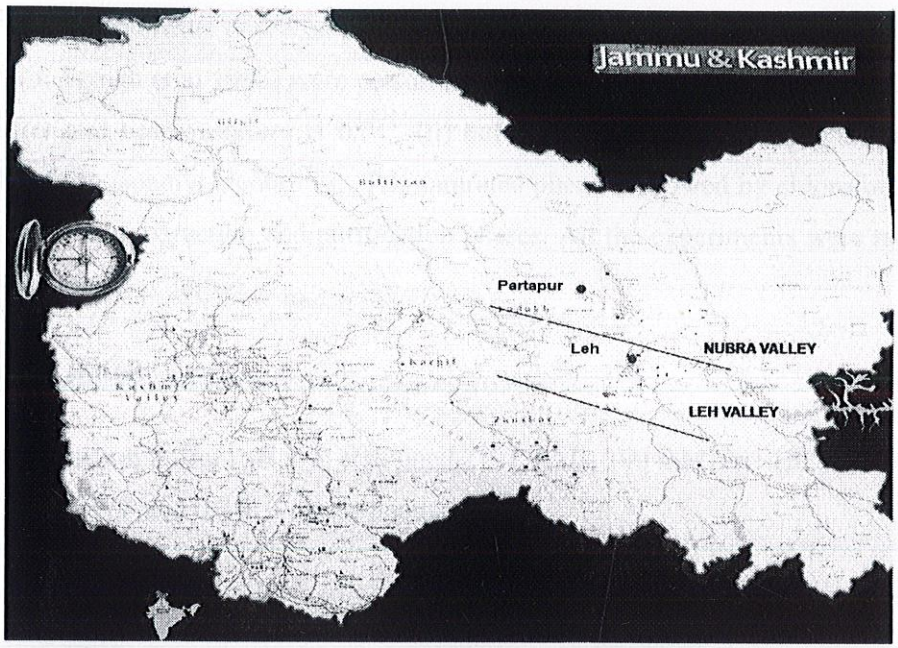


Fig. 1. Map of leh and Nubra valley

2.4 Plant materials

Twenty genotypes of Artemisia annua were collected from two valleys (Leh and Nubra) with altitude ranging from 9,600 m (Nubra) to 11,500 m (Leh) from the cold arid desert of Trans-Himalayas (Ladakh) (Fig. 1). Ten genotypes from each valley were obtained from the nursery of Defence Institute of High Altitude Research (DIHAR). The leaves were stored in laboratory at –20°C until further analysis. The interval between samples was 100-200m, the pair wise distance between valley divisions was 50– 250 Km.

2.5 DNA extraction and PCR amplification

Several experiments based on the available protocols (Doyle and Doyle 1990; Stewart and Via 1993; Stange et al 1998) were performed using fresh plant materials for (i) incubation time of buffer and tissue mixture at 65^0C, (ii) buffer to tissue ratio and (iii) extraction with phenol:chloroform:isoamyl alcohol vs. Tris saturated phenol followed by chloroform: isoamyl alcohol extraction in extraction and purification phases. All the experiments were repeated 3-4 times to check reproducibility.

2.6 Extraction Buffer

The extraction buffer (pH 8.0) contained 2% CTAB, 100 mM Tris-HCl, 20 mM EDTA, 1.4 M NaCl, 3% PVP and 0.2% β-mercaptoethanol.

2.7 Reagent and Solution

Tris saturated phenol, phenol:chloroform:isoamyl alcohol (25:24:1), chloroform:isoamyl alcohol (24:1), 70% and 80% ethanol, 4M NaCl, 3M sodium acetate (pH 5.2) and TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Solutions and buffers were autoclaved at 1210C at 15 psi pressure. The stock solution 10 mg/ml of RNase A was prepared as per the user's manual (Sigma USA).

2.8 DNA Extraction

Fresh young plant leaves plucked from the nursery were rinsed with distilled water and blotted gently with soft tissue paper, 0.1 g of this tissue, precooled using liquid nitrogen, was ground to a fine powder with a mortar and pestle along with 10 mg (2% of extraction buffer) of PVP (Sigma). The powdered tissue was scraped into a 2.0ml microcentrifuge tubes containing preheated (65^0C) extraction buffer in a 1:5 ratio (0.5 ml). β-Mercaptoethanol was then added to

the final concentration of 0.2 M and mixed well. The mixture was incubated in water bath at 65^0C for 90 min and cooled for 5 min. An equal volume of chloroform:isoamyl alcohol mixture (24:1) was added to the extract and mixed by gentle inversion for 5 to 10 min to form an uniform emulsion. The mixture was centrifuged at 8000 rpm for 8 min at room temperature. Chloroform: isoamyl alcohol extraction was repeated again. The aqueous phase was pipetted out gently, avoiding the interface. To the above solution, 5 M NaCl (to final concentration 2M) and 0.6 volume of isopropanol of the total solution was added and incubated at room temperature for 1 h. To the above solution, two volumes of 80% ethanol was added and incubated again for 10 min at room temperature for DNA precipitation. After incubation, the mixture was centrifuged at 10,000 rpm for 15 min. The white/translucent pellet was washed with 70% ethanol, dried and resuspended in 200 µL of TE buffer.

2.9 Purification phase

The sample was then treated with 20 µL of 10 mg/ml of RNase and incubated at 37^0C for 60 min. After incubation with RNase, equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added and mixed gently by inverting the microcentrifuge tube followed by centrifuged at 10,000 rpm for 5 min at room temperature. The supernatant was pipetted out into a fresh tube. The sample was then extracted with equal volumes of chloroform:isoamyl alcohol (24:1) twice. The DNA was precipitated with 0.6 volumes of isopropanol, 2.0 M NaCl (final concentration) and incubated for 10 min. To the above, 20 µL of sodium acetate and 1 volume of 80% ethanol were added, incubated for 30 min and centrifuged at 5,000 rpm for 3 min to pellet the DNA. The pellet was then washed with 70% ethanol twice; air-dried and finally suspended in 40-50 µL of TE buffer.

2.10 Evaluation of resultant DNA

The yield of the extracted DNA and purity was checked by running the samples on 0.8% agarose gels along with standard (non restriction enzyme digested), lamda DNA marker (Biogene, USA) (Fig.2)

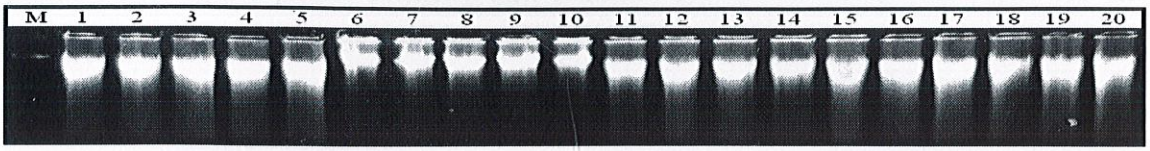


Fig. 2. Agrose gel electrophoresis showing purified high molecular weight *Artemisia annua* genomic DNA (The genomic DNA used for this study was extracted and purified with modified Doyle and Doyle method).

2.11 PCR

The **polymerase chain reaction (PCR)** is a technique in molecular biology to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase (after which the method is named) are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations.

2.12 AMPLIFICATION

Sixteen pairs of gene specific designed primers were used for amplification. Amplification reaction were performed in volumes of 25 µl containing 10 mM Tris- HCl (pH 9.0), 1.5 mM MgCl2, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 20 ng template DNA and 0.5unit of *Taq* polymerase (Sigma-Aldrich, USA). The first cycle consisted of denaturation of template DNA at 94°C for 4 min, primer annealing at 37 °C for 1 min, and primer extension at 72 °C for

2 min. In the next 40 cycles the period of denaturation was reduced to 1 min at 92°C, while the primer annealing and primer extension time remained the same as in the first cycle. The last cycle consisted of only primer extension (72 °C) for 5 min.

2.13 SSR Analysis

Sixteen SSR primers were obtained from 'Applied Biosciences, India' and PCR amplification was performed. Amplification reaction were performed in volumes of 25 µl containing 10 mM Tris- HCl (pH 9.0), 1.5 mM MgCl2, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 20 ng template DNA and 0.5unit of *Taq* polymerase (Sigma-Aldrich, USA). Initial denaturation for 4 min at 94 °C was followed by 40 cycles of 45 second at 94 °C, 30 second at specific annealing temperature (± 5 °C of Tm), 2 min at 72 °C and a 5 min final extension step at 72 °C (Table 2). PCR products were stored at 4 °C before analysis.

The amplification for each primer was performed twice independently with same procedure in order to ensure the fidelity of SSR markers. Amplification products were electrophoresed on 1.5 % at constant voltage (70 V) in 1X TAE for approximately 2 h, visualized by staining with ethidium bromide (0.5 µg ml-1) and a total of 2.5 µl loading buffer (6X) was added to each reaction before electrophoresis. After electrophoresis, the gels were documented on a gel documentation system (Alpha Innotech, Alphaimager, USA). Molecular size of amplicons was estimated using a 100 bp and 1 Kb DNA ladders ('Bangalore Genei, India).

2.14 Data collection and analysis

The banding patterns obtained from SSR were scored as present (1) or absent (0), each of which was treated as an independent character. Jaccard's similarity coefficient (J) was used to calculate similarity between pairs of accessions. The similarity matrix was subjected to cluster analysis by unweighted pair group method with arithmetic means (UPGMA) and a dendrogram was generated using the program NTSYSpc (Rohlf, 1992). POPGENE software

was used to calculate Nei's unbiased genetic distance among different genotypes with all markers. Data for observed number of alleles (Na), effective number of alleles (Ne), Nei's genetic diversity (H), Shannon's information index (I), number of polymorphic loci (NPL) and percentage polymorphic loci (PPL) were also analyzed (Zhao et al., 2006). Within species diversity (Hs) and total genetic diversity (Ht) (Nei, 1978) were calculated within the species and within two major groups (as per their collection site) using POPGENE software. The SSR data were subjected to a hierarchical analysis of molecular variance (AMOVA) (Excoffier et al., 1992), using two hierarchical levels; among valleys and among genotypes within each valley. GenAlEx software was used to calculate a principal coordinates analysis (PCA) that plots the relationship between distance matrix elements based on their first two principal coordinates (Peakall and Smouse, 2001). Polymorphism information content (PIC) value was calculated using the following formula:

$$\mathbf{PIC = 1 - \sum pi^2}$$

where p_i is the frequency of the *i*th (presence of band) allele.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 SSR analysis

Table 1. List of primers used for SSR amplification, GC content, total number of loci, the level of polymorphism and resolving power.

Primer	Primer Sequence (5' – 3')	GC (%)	Tm (^{0}C)	Allele number	PIC
1 (A) F	ATAATACGCATGAGCTGGTTAG	**40.9**	56.5	2	0.245
R	CCACTACCAATCACAATAACAG	40.9	56.5		
2 (B) F	GAATTGAGATTGTGGTCCTTAG	40.9	56.5	2	0.444
R	GGTTGCTAAGAATGTCGATTG	42.9	55.9		
4 (B) F	ATCGTATTACCTTGGTCACATC	40.9	56.5	2	0.245
R	TGTCATACTGACTTACACAGGG	45.5	58.4		
5 F	TAAGCCAAAGGCTCAAGTAAAC	40.9	56.5	3	0.198
R	GGATTGCTCATCTAGTGCTTAT	40.9	56.5		
6 (A) F	GCATGCATTTATGTTGGATCAC	40.9	56.5	2	0.565
R	CAGCAGCAACAACAACAACAG	47.6	57.9		
7 (B) F	GGAACAGATGATCTATATGCCT	40.9	56.5	2	0.340
R	GCATACTATGTGCAAGGTCTAGT	43.5	58.9		
8 F	TGGTAGAACTCCACCTACTAACT	43.5	58.9	2	0.426
R	TATAATAGTTGGGTGGTTCCTC	40.9	56.5		
9 F	GAGAAAGAGAAAGCCAAACAC	42.9	55.9	1	0
R	TAGCTCCATAGATCTCAAACCT	40.9	56.5		
10 F	GGATCATTAAGTTACGCTCCT	42.9	55.9	2	0.495
R	CCATGCTTTATGTTGTAGAGTG	40.9	56.5		
13 F	GTAAGTTATACCTGGTTTCCAGC	43.5	58.9	4	0.595
R	ACCACTACACCTTGCATTCTA	42.9	55.9		
14 F	CTCTCTTCCTCTTTGTGTGTCT	45.5	58.4	4	0.503

R	CAAGATGGTACGAATACTGTTG	40.9	56.5		
15 F	CGAGCAATCGGAGAGTTAGC	55.0	59.4	7	0.802
R	ATGCATCTCGCGAATCTTCT	45.0	55.3		
16 F	GTGTGAGGCCTCTGCTCTG	63.2	61	5	0.716
R	ACCGCCATGTCTTCTCCATA	50.0	57.3		
Total				38	

A total of 16 SSR loci generated from 68974 Artemisia EST sequences were tested on twenty Artemisia genotypes, of these 13 primer pairs generated single consistent amplification products (Fig. 3). The data scored for these 13 markers was analysed to establish the level of polymorphism. The 13 loci revealed 38 alleles with a range of 1 to 7 alleles and an average of 2.92 alleles per locus. 12 markers (75%) were polymorphic (Table.1) whereas a single marker generated single monomorphic product. The maximum number of alleles were detected with five markers generating alleles in the range of 3-7 alleles followed by seven markers with 2 alleles and one marker with 2 alleles per locus .

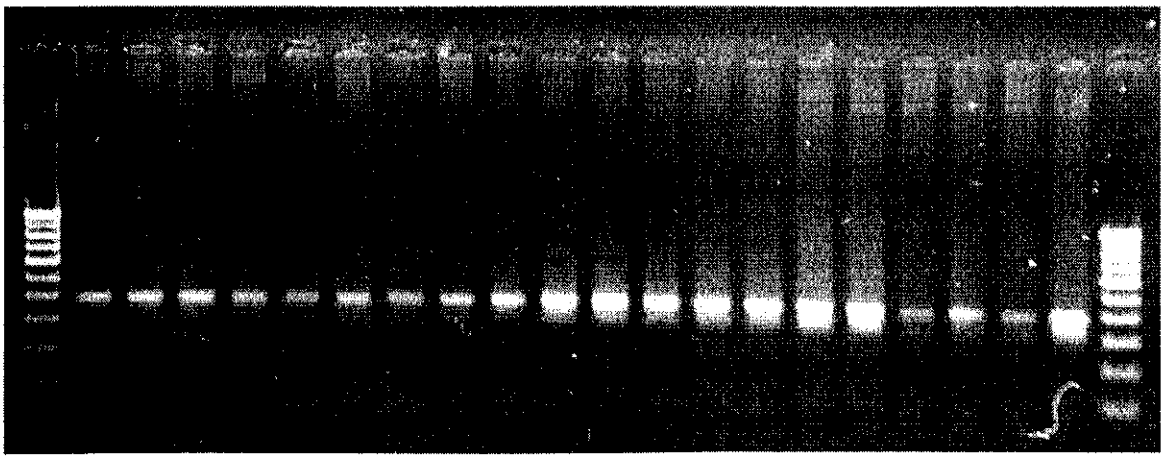


Fig. 3. Variability for SSR primer no.6 in 20 population of *Artemisia annua.*

Polymorphism information content (PIC) refers to the value of a marker for detecting polymorphism within a population or set of genotypes by taking into account not only the number of alleles that are expressed but also the relative frequencies of alleles per locus. PIC was calculated for the markers which generated 2 or more alleles, SSR marker no. 15 showed the highest level of polymorphism with PIV value of 0.802 followed by SSR marker no. 16 (0.716), SSR marker no.13 (0.595), SSR marker no.6 (0.565) and the PIC values for rest of the SSR markers were in the range of 0.198-0.495. In case of non amplifying SSR primers, different annealing temperatures (± 5 °C of Tm) in combination with different PCR reactions were tried but no amplification was observed.

3.2 SSR derived dendrogram analysis

The complete data was based on a total of 38 alleles and Jaccard's similarity coefficient ranged from 0.24 to 0.87. The genotypes were clustered into two clusters (I and II) where, cluster I, represents all the genotypes from partapur valley while, cluster II contains all the genotypes from leh valley (Fig. 4) . The results of PCA analysis were comparable to the cluster analysis (Fig. 5).

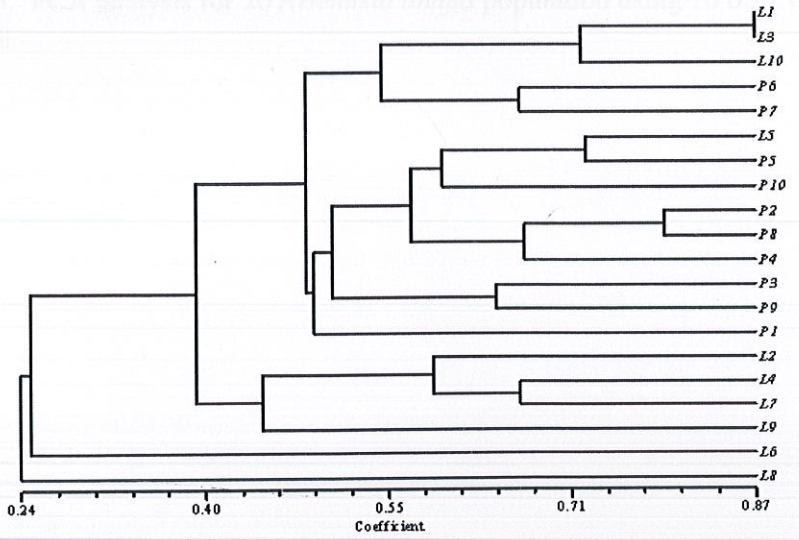


Fig. 4. Dendrograms generated using unweighted pair of group method with arithmetic average analysis, showing relationships between 20 *Artemisia annua* genotypes, using SSR data.

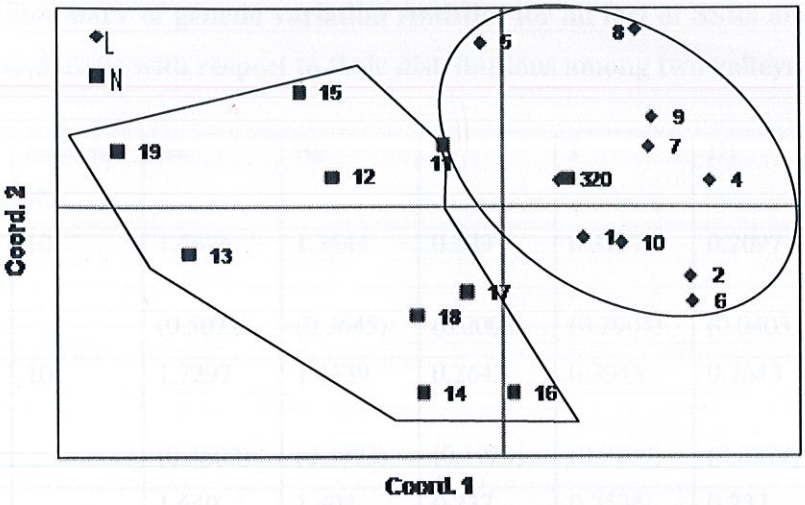


Fig. 5. PCA analysis for 20 *Artemisia annua* population using 16 SSR markers

3.3 Genetic Diversity Analysis

Table 2. Summary of genetic variation statistics for all loci of SSRs among the *Artemisia annua* populations with respect to their distributions among two valleys.

Valley	Sample size	Na	Ne	H	I	Ht	NPL	PPL
Leh	10	1.5676 (0.5022	1.3544 (0.3645)	0.2097 (0.2008)	0.3127 (0.2904)	0.2097 (0.0403)	21	56.76
Nubra	10	1.7297 (0.4502)	1.4539 (0.3773)	0.2643 (0.1929)	0.3945 (0.2707)	0.2643 (0.0372)	27	72.97
Mean		1.649	1.404	0.237	0.3536	0.237	-	-

Na = Observed number of alleles; Ne = Effective number of alleles; H = Nei's gene diversity; I = Shannon's Information index; Ht = Total genetic diversity; Hs = Genetic diversity in population; Gst = Genetic diversity between population; NPL = Number of Polymorphic Loci; PPL = percentage of Polymorphic Loci.

The respective values of Na, Ne, H, I, Ht, Hs, NPL and PPL were found higher for Nubra valley genotypes indicating that there is more variability in Nubra valley than in Leh valley genotypes (Table 2). Gst value 0.1808 indicated that 81.92 % of the genetic diversity resided within the population (Table 3). The details of overall genetic variability across 20 genotypes were given in Table 3. AMOVA helps in partitioning of the overall gene specific variations among groups and populations within the group. Molecular variance among valley is 23% and among population within valleys is 77% (Table 4) revealed higher variation.

Table 3. Overall genetic variability across all the 20 genotypes of Artemisia annua based on SSR analysis.

Na	Ne	H	I	Ht	Hs	Gst	NPL	PPL	Nm
1.9730 (0.1644)	1.4646 (0.3026)	0.2893 (0.1409)	0.4499 (0.1409)	0.2893 (0.0198)	0.2370 (0.0169)	0.1808	36	97.30	1.1325

Nm= Estimate of gene flow from Gst; Nm = 0.25 (1-Gst)/Gst; DI = Diversity index; EMR = Effective multiplex ratio; MI = Marker Index.

All the components of molecular variation were significant (P<0.001). This is helpful in making strategy for germplasm collection and evaluation. The estimated gene flow was 1.1325. In population genetics, a value of a gene flow (Nm) <1.0 (less than one migrant per generation into a population) or, equivalently, a value of gene differentiation (Gst) >0.25 is generally regarded as the threshold quantity beyond which significant population differentiation occurs (Slatin, 1987).

Table 4. Summary of nested analysis of molecular variance (AMOVA) based on SSR, among the populations of Artemisia annua. Levels of significance are based on 1000 iteration steps.

Source of variation	Df	S.S.D.	Variance component	Percentage	P-value
Among valley	1	19.350	1.448	23	<0.001
Among genotypes/valley	18	87.70	4.872	77	<0.001

Where d.f. = degree of freedom; S.S.D = sum of square deviation; P-value = probability of null distribution.

CONCLUSION

Artemisia annua is a very important medicinal plant. The important secondary metabolite of *Artemisia annua* is artemisinin which is commercially being used as an antimalarial drug. Due to the huge demand of artemisinin in commercial market, so it is very important to focus on proper conservation and management of *Artemisia* sp. growing in wild. In present study a very high level of polymorphism was attained and the high Gst value indicated that there was significant population differentiation with 81.92 % of the genetic diversity resided within the population. Molecular variance between the population of two valley was observed to be less as compared to within valleys, which revealed higher variation.

BIBLIOGRAPHY

1. B. Li, Q. Xia, C. Lu, Z. Zhou, and Z. Xiang, "Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes," *Genomics Proteomics & Bioinformatics*, vol. 2, no. 1, pp. 24–31, 2004.
2. Bohle, D. S., Dinnebier, R. E., Madsen, S. K., and Stephens, P. W. (1997) *J. Biol. Chem.* **272,** 713–716
3. Darbinyan V, Kteyan A, Panossian A, Gabrielian E, Wikman G and Wagner H (2000) *Rhodiola rosea* in stress induced fatigue: A double blind cross-over study of a standardized extract SHR-5 with a repeated low-dose regimen on the mental performance of healthy physicians during night duty. Phytomedicine 7:365-71
4. Dondorp A. M. (2009*) N. Engl. J. Med.* **361**, 455 .
5. Fitch, C. D., and Kanjananggulpan, P. (1987) *J. Biol. Chem.* **262,** 15552–15555
6. Foley, M., and Tilley, L. (1998) *Pharmacol. Ther.* **79,** 55–87
7. Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerboker, A., Muller, R., Bryant, M. L., Sherman, D. R., Russel, D. G., and Goldberg, D. E. (1994*) EMBO J.* **13,** 306–313
8. Francis, S. E., Sullivan, D. J., Jr., and Goldberg, D. E. (1997) *Annu. Rev. Microbiol.* **51,** 97–123
9. Gluzman, I. Y., Francis, S. E., Oksman, A., Smith, C. E., Duffin, K. L., and Goldberg, D. E. (1994) *J. Clin. Invest.* **93,** 1602–1608
10. Goldberg, D. E., Slater, A. F. G., Beavis, R., Chait, B., Cerami, A., andHenderson, G. B. (1991) *J. Exp. Med.* **173,** 961–969
11. Goldberg, D. E., Slater, A. F. G., Cerami, A., and Henderson, G. B. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2931–2935
12. Hong, Y. L., Yang, Y. Z., and Meshnick, S. R. (1994) *Mol. Biochem. Parasitol.* **63,** 121–128

13. M. Morgante, M. Hanafey, and W. Powell, "Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes," *Nature Genetics*, vol. 30, no. 2, pp. 194–200, 2002.

14. Meshnick, S. R., Taylor, T. E., and Kamchonwongpaisan, P. (1996) *Microbiol. Rev.* **60,** 301–315

15. Meshnick, S. R., Thomas, A., Ranz, A., Xu, C. M., and Pan, H. Z. (1991) *Mol. Biochem. Parasitol.* **49,** 181–189

16. Ohba H (1987) Biogeography of genus *Rhodiola* (Crassulaceae). Acta Phytotax Geobot 38:211-223.

17. Pandey, A. V., and Chauhan, V. S. (1998) *Curr. Sci.* **75,** 911–918

18. Pandey, A. V., Tekwani, B. L., and Pandey, V. C. (1995) *Biomed. Res.* **16,** 115–120

19. Pandey, A.V., Tekwani B. L., Singh R. L., and Chauhan V.S. (1999) J. Biol. Chem. **274**, 383–388

20. S. Subramanian, R. K. Mishra, and L. Singh, "Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions," *Genome Biology*, vol. 4, no. 2, p. R13, 2003.

21. Salas, F., Fichmann, J., Lee, G. K., Scott, M. D., and Rosenthal, P. J. (1995) *Infect. Immun.* **63,** 2120–2125

22. Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance," National Academy of Sciences 2004, www.nap.edu/catalog/11017.html. Global Malaria Action Plan, Report of the 2008 Artemisinin Conference, 8 to 10 October, York, UK (www.york.ac.uk/org/cnap/artemisiaproject/pdfs/AEconference-report-web.pdf).

23. Schwarzer, E., Turrini, F., Ulliers, D., Giribaldi, G., Ginsburg, H., and Arese, P. (1992) *J. Exp. Med.* **176,** 1033–1041

24. Sherry, B. A., Alava, G., Tracey, K. J., Martiney, J., Cerami, A., and Slater, A. F. G. (1995) *J. Inflamm.* **45,** 85–96

25. Slater, A. F., Swiggard, W. J., Orton, B. R., Flitter, W. D., Goldberg, D. E., Cerami, A., and Henderson, G. B. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 325–329

26. Slatin M (1987). Gene flow and geographic structure of natural populations. Science,236:787-792

27. Vander Jagt, D. L., Hunsaker, L. A., and Campose, N. M. (1986) *Mol. Biochem. Parasitol.* **18,** 389–400

28. Vander Jagt, D. L., Hunsaker, L. A., Campos, N. M., and Scaletti, J. V. (1992) *Biochim. Biophys. Acta* **1122,** 256–264

29. Yang, Y. Z., Asawamahasakda, W., and Meshnick, S. R. (1993) *Biochem.*