Study of Genetic Variations in *Artemisia annua*

Using ISSR and RAPD Molecular markers

By

AMAN AGGARWAL (061502)

MOHIT SAINI (061512)

MAY2010

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

Dr. Pradeep Kumar Naik
Assistant Professor
Biotechnology & Bioinformatics

Jaypee University of
Information Technology
Waknaghat-173 215
Solan, Himachal Pradesh
Phone No.: 91-1792-239227
Fax No.: 91-1792-245362

Certificate

This is to certify that the thesis entitled, "Study of Genetic Variations in *Artemisia annua* Using ISSR and RAPD Molecular markers in partial fulfillment for the award of Bachelor of Technology in Bioinformatics of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Place : Waknaghat
Date : May,2010

(Dr. Pradeep Kumar Naik)
Project Guide

Acknowledgements

At the outset, we acknowledge and express our deep sense of gratitude to Dr. P.K. Naik, project coordinator, Assistant Professor in Biotechnology and Bioinformatics Department, JUIT, Waknaghat, Solan (HP), for his valuable guidance , kind and constant encouragement immense patience and constructive criticism during the investigation and preparation of this manuscript and above all his cordial and fatherly treatment which goes much beyond his formal obligations.

We are indebted to all those who provided reviews and suggestions for improving the material and topics covered in this Project.

Place : Waknaghat, (HP)

Dated : May,2010

Aman Aggarwal

Mohit Saini

Contents

Abstract

The phylogenetic relationships of 20 locally grown *Artemisia annua* genotypes which are collected from two valleys *viz*. Nubra (9,600 ft) and Leh (11,500 ft) of trans-Himalayan region were analyzed using 31 PCR markers (20 RAPDs and 11 ISSRs). This is the first report of molecular genetic diversity studies in *Artemisia annua* genotypes from the Ladhak region. RAPD analysis yielded 139 fragments, of which 136 were polymorphic, with an average of 6.8 polymorphic fragments per primer. ISSR analysis produced 58 bands, of which 56 were polymorphic, with an average of 5.09 polymorphic fragments per primer. The primers based on (CT)n produced maximum number of bands (nine) while, (AT)n and many other motifs gave no amplification. RAPD markers were found more efficient with regards to polymorphism detection, as they detected 97.84 % as compared to 96.5 % for ISSR markers. Clustering of genotypes within groups was not similar when RAPD and ISSR derived dendrogram were compared, whereas the pattern of clustering of the genotypes remained more or less the same in RAPD and combined data of RAPD + ISSR. The results of PCA analysis were comparable to the cluster analysis. These analyses, allowed us to identify the groups corresponding to the two collection sites.

CHAPTER 1

Introduction

Artemisia annua L. (*Asteraceae* formerly *Compositae*) is an annual herb native to Asia now distributed throughout many countries such as 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 essential oil also possesses strong insecticidal potential (Charles et al., 1991). Besides the volatile terpenoids, the plant synthesizes and accumulates substantial quantities of many derivatives of a cadiene skeleton including artemisinin (an endoperoxide seco sesquiterpene lactone), which is currently the most effective agent against multidrug resistant strains of *Plasmodium* species, the malarial parasites (Phillipson and Wright, 1991; Klayman, 1993). Due to this discovery, *Artemisia annua* is now rated as one of the top ten industrial crops of the modern world. Artemisia thrives well in dry cold region of Ladakh having marginal rocky and sandy soils. It is well adapted to survive at high altitudes of 2,500-6,000 m above mean sea level (MSL) and the temperature, nutrient and environmental stress that they are subjected to under the cold arid conditions.

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). This has generated worldwide interest in studying the genetic diversity, of *Artemisia annua* populations, clonal variants, chemotypes, and ecotypes, and in the synthesis of pure-line cultivars. Over the year, the Artemisia populations in the Ladakh region have developed considerable variability. No reports are available so far regarding the genetic characterization of this plant from Ladakh region and detailed investigation is required. Moreover, the analysis of genetic

diversity and relatedness between or within different genotypes is a prerequisite towards effective utilization and protection of plant genetic resources (Weising et al., 1995). It will then be possible to achieve rational conservation, and the identification of diagnostic or agronomic traits linked to molecular markers. With DNA being the only basis of genetic differences between distinct organisms, DNA fingerprinting is presently the ultimate method of biological individualization. In principle, 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).

Unlike the morphological and biochemical markers which may be affected by environmental factors and growth practices (Xiao et al., 1996; Ovesna et al., 2002; Higgins 1984), DNA markers portray genome sequence composition, thus, enabling to detect differences in the genetic information carried by the different individuals. A wide variety of DNA-based markers have been developed in the past few years. Restriction fragment length polymorphism (RFLP) was the first molecular marker (Bostein et al., 1980), generated for genome analysis and mapping. However, the development of the polymerase chain reaction (PCR) technology has introduced a considerable number of useful molecular markers, e.g., random amplified polymorphic DNA (RAPDs) (Williams et al., 1990) and inter simple sequence repeats (ISSR) (Zeitkiewicz et al., 1994) which are independent of environmental factors and unaffected by developmental stages of the plant. These markers have been used both for DNA fingerprinting (Gupta et al., 2008) and population genetic studies (Alam et al., 2009). For these PCR based marker-assisted selection, suitably modified extraction method for genomic DNA in terms of quality and quantity is essential for a particular crop. Currently, there are several methods of DNA isolation available as variants of the few principal protocols (Dellaporta et al., 1984; Saghai-Maroof et al, 1984; Doyle and Doyle, 1987; Webb and Knapp, 1990). Experience has shown that plant protocols need to be tailored to each plant species (and sometimes even each tissue) due to the presence of secondary metabolites that vary in nature and/or composition. These phytochemicals may not only hinder the *su moto* application of the other methods, but they may also interfere

with subsequent amplification and/or restriction-digestion of the isolated DNA. Sometimes, the contaminants are not detectable but are sufficiently high to interfere in the analysis of polymorphisms. Particularly, medicinal and aromatic plants like *Artemisia annua* are rich in the myriad of natural products and they require additional efforts to find an appropriate procedure for reliable and consistent results.

Therefore, we describe a method for isolation of DNA from *Artemisia annua.* The DNA is clean and intact and is suitable for PCR amplification. The 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 were separated from each other through natural mountain barrier. To our knowledge this is the first reported protocol for isolation of DNA from the metabolite-rich species like *Artemisia annua* as well as genetic characterizations of Artemisia from the trans-Himalayan region with RAPD and ISSR molecular markers.

1.1 Importance of Artemisinin and it's derivatives

The leaves of Artemisia annua has many medicinal properties such as Antibacterial(kills bacteria),Antiperiodic(counteracts recurring illnesses such as malaria), Antiseptic(preventing sepsis, decay or putrefaction,it destroys or arrests the growth of micro organisms),Carminative(reduces flatulence and expels gas from the intestines), Digestive(aids digestion),Febrifuge(reduces fever).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. 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 leaves are harvested in the summer, before the plant comes into flower, and are dried for later use. The plant contains artemisinin, this substance has proved to be a dramatically effective anti-malarial against multi-drug resistant Plasmodium . Clinical trials have shown it to be 90% effective and more successful than standard drugs. In a trial of 2000 patients, all were cured of the disease. The seeds are used in the treatment of flatulence, indigestion and night sweats.

Artemisinin is used for the treatment of Malaria, Parasite, Cancer, HIV. The artemesinins are not used for malaria prophylaxis (prevention) because of the extremely short activity of the drug. To be effective, it would have to be administered multiple times each day.

The World Health Organization is pressuring manufacturers to stop making the pure drug, saying it would be a loss if the parasites would build up resistance for the only known drug the parasites have not developed resistance to *in vitro* experiments have been able to generate a resistant strain of the parasite and resistant strains have been found from field samples. Artemisinin is widely used in China and Southeast Asia for treatment of malaria.

The World Health Organisation has recommended that a switch to this should be made in all countries where the malaria parasite has developed resistance to chloroquine. Artemisinin and its derivatives are now standard components of malaria treatment in China, Vietnam, and some other countries in Asia and Africa, where they have proved to be safe and effective anti-malarial drugs. They have minimal adverse side effects. Not all plants of this species contain artemisinin. Apparently it is only produced when the plant is subjected to certain conditions, most likely biotic or abiotic stress. It can be synthesized from artemisinic acid. Artemisinin is a drug used to treat multi-drug resistant strains of falciparum malaria. Use of the drug by itself as a monotherapy is explicitly discouraged by the World Health Organization as there have been signs that malarial parasites are developing resistance to the drug. Combination therapies that include artemisinin are the preferred treatment for malaria and are both effective and well tolerated in patients. The drug is also being studied as a treatment for cancer. Consensus amongst researchers is that artemisinin is losing its potency in Cambodia and increased efforts are required to prevent drug-resistant malaria from spreading across the globe. Thus the drugs are used in combination therapies. The side effects from the artemisinins themselves are similar to the symptoms of malaria: nausea, vomiting, anorexia, and dizziness. The combination drugs may have additional side effects.

The plant *Artemisia annua* L. contains the well-established antimalarial compound artemisinin. In 1971, scientists demonstrated that the plant extracts had antimalarial activity in primate models, and in 1972 the active ingredient, artemisinin was isolated and its chemical structure described. Chemical formula of artemisin is $C_{15}H_{22}O_5$. Artemisinin is a sesquiterpene lactone with an endoperoxide bridge and has been produced semi-synthetically as an antimalarial drug. Most collections of artemisia derive from natural

stands with highlyvariable artemisinin content, some as low of 0.01%. Selections from Chinese origin vary from 0.05 to 0.21%. Swiss researcher N. Delabays reports a clonal selection derived from Chinese material which produces 1.1% artemisin but is very late flowering; proprietary hybrids have been obtained with somewhat lower content but flower earlier. Members of this genus are rarely if ever troubled by browsing deer.

Artemisinins can be used alone, but this leads to a high rate of recrudescence (return of parasites) and other drugs are required to clear the body of all parasites and prevent recurrence. The World Health Organization is pressuring manufacturers to stop making the uncompounded drug available to the medical community at large, saying it would be a significant loss if the malaria parasite developed resistance to Artemisinin. The World Health Organisation has recommended that a switch to artemisinin combination therapies (ACT) be made in all countries where the malaria parasite has developed resistance to chloroquine. Artemisinin and its derivatives are now standard components of malaria treatment in China, Vietnam, and some other countries in Asia and Africa, where it has been proven to be a safe and effective anti-malarial treatment. Fixed-dose combinations are preferred as this guarantees that the partner drug is present to eradicate the last parasites while the artemisinin component removes the majority at the start of the treatment. A large number of fixed-dose ACTs are now available containing an artemisinin component and a partner drug which has a long half-life, such as mefloquine (ASMQ), lumefantrine (Coartem), amodiaquine (ASAQ), piperaquine (Duo-Cotecxin) and antifolates (Ariplus). Most are made to GMP standard. A separate issue concerns the quality of some artemisinin-containing products being sold in Africa and South-East Asia. Artemisinins are not used for malaria prophylaxis (prevention) because of the extremely short activity of the drug. To be effective, it would have to be administered multiple times each day.

Artemisinin is undergoing early research and testing for the treatment of cancer, primarily by researchers at the University of Washington. Chinese scientists have shown artemisinin has significant anticancer effects against human hepatoma cells that Artemisinin has a peroxide lactone group in its structure. It is thought that when the peroxide comes into contact with high iron concentrations (common in cancerous cells), the molecule becomes

unstable and releases reactive oxygen species. It has been shown to reduce angiogenesis and the expression of vascular endothelial growth factor in some tissue cultures.

Serendipitous discovery was made in China while searching for novel anthelmintics for schistosomiasis. Artemisinin was effective against schistosomes, the human blood flukes, which are the second most prevalent parasitic infections, after malaria. Artemisinin and its derivatives are all potent anthelmintics. They are later found to possess a broad spectrum of activity against a wide range of trematodes including *Schistosoma japonicum*, *S. mansoni*, S. haematobium, *Clonorchis sinensis*, *Fasciola hepatica* and *Opisthorchis viverrini*. Clinical trials are also successfully conducted in Africa among patients with schistosomiasis. A randomized, double-blind placebo-controlled trial also revealed the efficacy against schistosome infection in Cote d'Ivoire and China.

1.2 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. (Lewontin, 1973; Gardiner et al., 1986; Puecher et al., 1996; Mauria et al., 2000). 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 (Schaal et al., 1991).

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 (Lu et al., 1996). Molecular markers, unlike morphological markers are stable and have been found useful in population studies (Aitkin et al., 1994; Lakshmi et al., 1997) and phylogeny (Demeke and Adams, 1994; Adams and Demeke, 1993). 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) (Williams et al., 1990; Karp et al., 1997) and amplified fragment length polymorphism (AFLP) (Vos et al., 1995). 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 (Palacious et al., 1999; Han et al., 2000).

1.3 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 (Saiki et al., 1988) and has changed the total scenario of molecular biology and has also brought about a multitude of new possibilities in molecular marker research. Some of the PCR based marker systems used in the study have been detailed below.

1.3.1 ISSRs as molecular markers

Inter Simple Sequence Repeats (ISSR) is a type of molecular marker that can be carried out without prior knowledge of DNA sequence in the genome. Microsatellites (SSR) represent the most abundant source of polymorphism from repetitive sequences. SSR are often used as molecular markers even if this technology is time consuming and expensive. ISSR is an alternative technique to study polymorphism based on the presence of microsatellites throughout genomes (Zietkiewicz et al., 1994). ISSR markers are DNA sequences delimited by two inverted SSR sequences composed of the same units which are amplified by a single PCR primer, composed of few SSR units with or without anchored end. ISSR-PCR gives multilocous patterns which are very reproducible, abundant and polymorphic in plant genomes (Bornet et al., 2002). This approach named Inter-SSR (ISSR) employs oligonucleotides based on a simple sequence repeat anchored or not at their 5'- or 3'-end by two to four arbitrarily chosen nucleotides. This triggers site-specific annealing and initiates PCR amplification of genomic segments which are flanked by inversely oriented and closely spaced repeat sequences. The marker system called ISSRs has been developed as an anonymous RAPDs-like approach that accesses variation in the numerous microsatellite regions dispersed throughout various genomes (particularly the nuclear genome) and circumvents the challenge of characterizing individual loci that other molecular approaches require (Tautz, 1989). The resultant PCR reaction amplifies the sequence between two SSRs, yielding a multilocous marker system useful for fingerprinting, diversity analysis and genome mapping. ISSR markers are considered to be more reproducible than RAPD markers due to high annealing temperature (Bornet and Branchard,

2001; Chowdhury et al., 2002) and have been used to measure genetic diversity in potato (Bornet et al 2002), barley (Fernández et al., 2002), rice (Joshi et al., 2000), finger millet (Eleusine coracana) (Salimath et al., 1995), Sorghum (Yang et al., 1996) and Groundnut (Raina et al., 2001) as well as to identify cultivars in Maize (Kantety et al., 1995; Pejic et al., 1998), Wheat (Nagaoka and Ogihara, 1997), potato (Prevost and Wilkinson, 1999), oilseed rape (Charters et al., 1996) and bean (Métais et al., 2000). Huang et al., (2000) also tried to use ISSR markers to reveal genetic diversity and relationships in sweet potato and its wild relatives.The ISSRs have several advantages for assessing genetic diversity (Zietkiewicz et al., 1994). They are advantageous because no prior genomic information is required for their use. Abundantly polymorphic and reproducible, they are a good choice for detecting genetic diversity among crop species, germplasm characterization, establishment of genomic relation and molecular phylogeny. ISSRs analyses are more specific than RAPD analyses due to the longer SSR-based primers which enable higher-stringency amplification (Wolfe et al., 1998). The high stringency reduces the problems with reproducibility, a common criticism against the low-stringency RAPD assay (Yang et al., 1996). The shortcomings of ISSR markers like RAPDs are that most bands are scored as dominant markers giving no possibility to distinguish between homozygosis and heterozygosis directly. However, ISSR studies of natural populations have recently demonstrated the hyper variable nature of these markers and their potential use in population-level studies (Ge and Sun, 1999; Culley and Wolfe, 2001). The technique allows for dissection below the subspecies level and this gives it a good level of applicability in the study of rare or endangered plants (Blair et al., 1999).

ISSRs have been used in conjunction with RAPD data to determine the colonization history of *Olea europaea* in Macronesia along with lineages in the species complex (Hess et al., 2000); Cicer and cultivated chickpea (Iruela et al., 2002). The two techniques have also been utilized in examining the genetic diversity, varietal identification and phylogenetic relationships in peanut (*Archis hypogaea*) cultivars and wild species (Raina et al., 2001). The phylogenetic trees generated by different methods were largely congruent topologically. They are mostly dominant markers.

1.3.2 RAPDs as molecular markers

Random amplified polymorphic DNA is a PCR-based technique that has been applied to the study of populations (Williams et al., 1990). RAPDs are one of a family of techniques that produce arbitrary fragment length polymorphism and are collectively described as multiple arbitrary amplicon profiling (Caetano-Anolles, 1994). The RAPD technique utilizes single, arbitrary, decamer DNA oligonucleotide primers to amplify regions of the genome using PCR (Welsh and McClelland, 1990; Williams et al., 1990; Williams et al., 1993). Priming sites are thought to be randomly distributed throughout a genome and polymorphism in these regions results in different amplification products. The methodology is simple and has been widely used for the assessment of genetic diversity, genetic variation within species, determining relationships between closely related species and genotypes within a species to identify particular genotypes (cultivar identification). RAPD technique has also been used to study and investigate clonal and population structure (Kresovich et al., 1992). The use of RAPD for determination of genetic relationships has been demonstrated in a number of crop species like maize (Welsh et al., 1991), Sorghum (Dahlberg et al., 2002) rapeseed (Forster and Knaak, 1995), pigeon pea (Ratnaparkhe et al., 1995), aromatic rice (RayChoudhury et al., 2001) and many other crops. The simplicity of the technique and the speed of data generation have attracted many researchers, particularly those interested in either genetic fingerprinting or the patterns and levels of genetic diversity (Kazan et al., 1993 and Koller et al., 1993). In addition to the studies of genetic diversity there have been an increasing number of papers concerned with population genetics (Crochemore et al., 1996), phylogenetics (Stewart and Porter, 1995) and hybridisation/introgression (Sale et al., 1996; Comes and Abbott, 1999). The greatest attraction of this method is that it generates DNA data that are, theoretically at least, randomly scattered across the genome. Such markers are attractive for studies that involve differentiation of similar species (Zamora et al., 1996) and identification of patterns of variation (Vasconcelos et al., 1996). Rieseberg (1996) suggested that RAPDs may be useful for investigation within species or between closely related species. Many reports are available on inter and intra generic genetic diversity and molecular phylogeny using RAPDs such as, Arachis (Gimenes et al., 2000), Pistacia (Kafkas and PerlTreves, 2002) etc. The usefulness of RAPD fingerprinting was also reported for identification of Italian grape (Vitis

vinifera) varieties (Mulcahy et al., 1995) and determining the phylogenetic relationship for 28 tropical maize varieties (Parentoni et al., 2001). All the above studies confirmed the efficiency of RAPD markers for systematic investigations. Thus, it has been suggested that RAPD fingerprinting method is simple and so powerful that one primer can distinguish between different clones while the use of multiple primers reduces fingerprint similarity and resolves discrepancies. RAPD analysis is a very good starting point for studies of relationships within and among closely related species.

CHAPTER 2

Assessment of genetic diversity among *Atremisia annua* genotypes of the Leh-Ladakh region for artemisinin production using RAPD and ISSR molecular markers

2.1 Materials and methods

The present investigation was undertaken with a view to study the molecular diversity of twenty different genotypes obtained from two different regions of India. The details of materials used and methods adopted in the present investigation are described here under.

2.1.1 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 (Ladhak) Figure 1. Ten genotypes from each valley were obtained from the nursery of Defence Institute of High Altitute 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.

2.1.2 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.

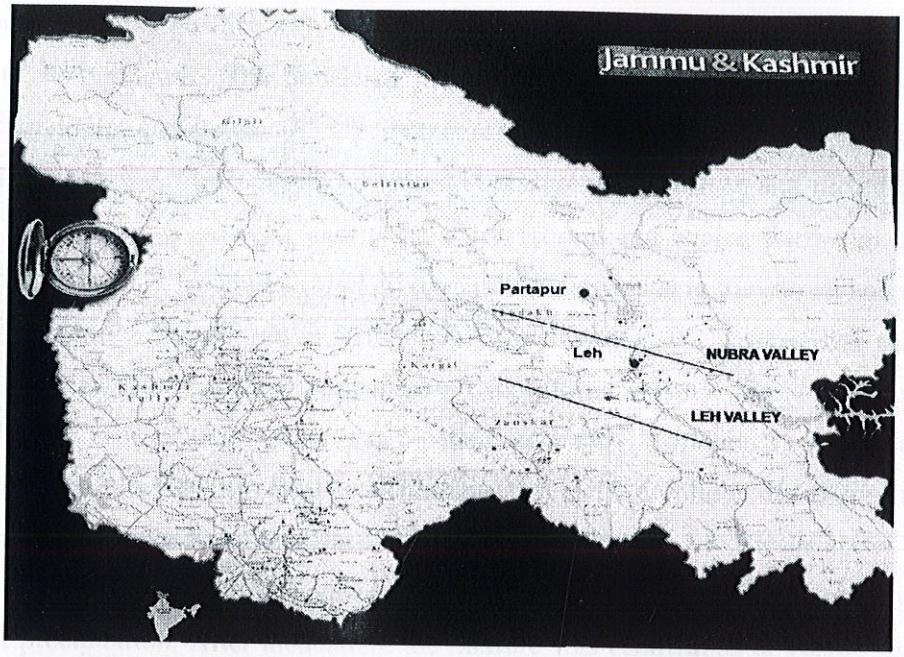

Fig. 1. Collection sites of 20 *Artemisia annua* genotypes from two valleys (Leh and Nubra) and the two collection sites (Leh and Partapur) located in Ladakh (Jammu and Kashmir, India).

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.

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 121^0C at 15 psi pressure. The stock solution 10 mg/ml of RNase A was prepared as per the user's manual (Sigma USA).

2.1.3 DNA Extraction

Fresh young plant leaves plucked form 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.

2.1.4 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. Figure 2 represents the different stages of DNA purification extracted form the leaf samples.

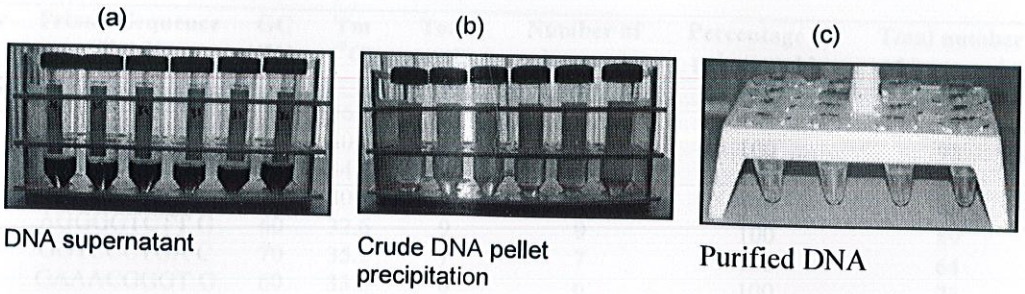

Figure 2. Stages of DNA extraction from the leaves of *Artemisia annua*.

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).

2.2 RAPD analysis

Twenty random decamer primers from IDT Tech, USA (Table 1) were used for RAPD amplification following the protocol of Williams *et al.* (1990). 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.

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

Primer	Primer Sequence (5' – 3')	GC (%)	Tm (^{0}C)	Total number of loci	Number of polymorphic loci	Percentage of polymorphic loci	Total number of fragment s amplified	Resolving power
S21	CAGGCCCTT C	70	36.4	6	6	100	81	4.5
S22	TGCCGAGCT G	70	40.7	7	7	100	93	5.17
S23	AGTCAGCCA C	60	34.3	7	7	100	81	4.5
S24	AATCAGCCA C	50	30.1	4	4	100	46	2.56
S25	AGGGGTCTT G	60	32.6	9	9	100	89	4.94
S26	GGTCCCTGA C	70	35.2	7	7	100	64	3.56
S27	GAAACGGGT G	60	33.2	6	6	100	73	4.06
S28	GTGACGTAG G	60	31.1	7	7	100	88	4.89
S29	GGGTAACGC C	70	37.4	5	5	100	72	4.0
S30	GTGATCGCA G	60	33.1	8	8	100	100	5.56
S31	CAATCGCCG T	60	36.7	5	5	100	75	4.17
S32	TCGGCGATA G	60	34.0	6	6	100	72	4.0
S33	CAGCACCCA C	70	37.7	4	4	100	55	3.06
S34	TCTGTGCTG G	60	34.3	5	5	100	69	3.83
S35	TTCCGAACC C	60	34.2	5	5	100	65	3.61
S36	AGCCAGCGA A	60	38.3	5	5	100	58	3.22
S37	GACCGCTTG T	60	35.7	7	7	100	83	4.61
S38	AGGTGACCG T	60	36.2	6	6	100	55	3.06
S39	CAAACGTCG G	60	34.2	6	6	100	77	4.28
S40	GTTGCGATC C	60	33.5	9	9	100	86	4.78
	Total	-	0	**124**	**124**	**100**	**1482**	-

2.3 ISSR analysis

Seventeen ISSR primers were obtained from 'Applied Biosciences, India' and PCR amplification was performed in reaction cocktail similar to RAPD. 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 RAPD and ISSR markers. Amplification products were electrophoresed on 1.5 % agarose gel 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).

Table 2. List of primers used for ISSR amplification, sequence, GC content, total number of loci, the level of polymorphism, size range of fragments and resolving power. Where, (Y = C, T; R = A,G).

Primer	Primer Sequence (5' – 3')	GC (%)	Tm (^{0}C)	Total number of loci	Number of polymorphic loci	Percentage of polymorphic loci	Total number of fragment s amplified	Resolving power
ISSR 1	(AG)8 T	47	47.0	4	4	100	67	3.72
ISSR 2	(GA)8 T	47	45.4	5	5	100	64	3.56
ISSR 3	(AC)8 T	47	51.4	3	2	66.7	49	2.72
ISSR 4	(TG)8 A	47	51.3	4	2	50	67	3.72
ISSR 5	(AG)8YT	47.2	49.2	6	4	66.7	87	4.83
ISSR 6	(GA)8YT	47.2	47.4	7	7	100	102	5.67
ISSR 7	(CT)8 RA	47.2	47.1	5	5	100	57	3.17
ISSR 8	(GT)8 YC	52.7	52.7	5	5	100	57	3.17
ISSR 9	(ACC)6	66.6	60.6	4	4	100	49	2.72
ISSR 10	CCG)6	10	76.8	4	4	100	42	2.33
ISSR 11	(GGC)6	10	77.3	11	11	100	97	5.39
ISSR 12	(AT)8 T	0	23.1	4	4	100	39	2.17
ISSR 13	(TA)8 RT	2.7	25.6	5	5	100	60	3.33
ISSR 14	(AT)8 YA	2.7	26.0	5	4	80	67	3.72
ISSR 15	(CT)8 T	47	45.7	4	4	100	51	2.83
ISSR 16	(TC)8 A	47	47.0	4	3	75	67	3.72
ISSR 17	(GT)8 A	47	49.4	5	5	100	65	3.61
	Total			**85**	**78**	**91.8**	**1087**	-

2.4 Data collection and analysis

The banding patterns obtained from RAPD and ISSR 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 RAPD and ISSR 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). Correlation between both the marker types used in the study was obtained by regression (R^2) analysis between similarity matrices obtained with two marker types. In this instance, the matrix regression corresponds to two independently derived dendrograms. The resolving power of the RAPD and ISSR primers was calculated according to Prevost and Wilkinson (1999). The resolving power (Rp) of a primer is: $Rp = \Sigma\ IB$ where *IB* (band informativeness) takes the value of: 1–[2* (0.5–P)], P being the proportion of the 20 genotypes containing the band.

In order to determine the utility of each of the marker systems, diversity index (DI), effective multiplex ratio (EMR) and marker index (MI) were calculated according to Powell *et al.* (1996). DI for genetic markers was calculated from the sum of the squares of allele frequencies: $DI_n = 1\text{-}\Sigma\ pi^2$ (where 'pi' is the allele frequency of the ith allele). The arithmetic mean heterozygosity, Di_{av}, was calculated for each marker class: $Di_{av} = \Sigma\ Di_n/n$, (where 'n' is the number of markers (loci) analyzed). The DI for polymorphic markers is: $(Di_{av})p = \Sigma\ Di_n/n_p$ (where 'n_p' is the number of polymorphic loci and n is the total number of loci). EMR (E) is the product of the fraction of polymorphic loci and the number of polymorphic loci for an individual assay. EMR (E) = n_p (n_p/n).

MI is defined as the product of the average diversity index for polymorphic bands in any assay and the EMR for that assay, MI = DI_{avp} * E.

2.5 Results and discussion

Many protocols were attempted for extraction of good quality DNA from *Artemisia annua* but it was not possible to obtain good quality DNA. The Artemisia species obtained from high altitude (Ladhak and Partapur) contained high amount of polysaccharides, polyphenols, essential oils & other secondary metabolites that interfere with DNA isolation. These secondary metabolites get entangled to nucleic acid during DNA isolation and hence interfere with subsequent isolation procedure. We developed a simple and efficient method of genomic DNA extraction from leaf of *Artemisia annua* by modifying CTAB protocol (Khanuja et al 1999). By using 3.5 M NaCl in extraction buffer and 80% ethanol with 2.0 M NaCl (final concentration) during precipitation and further purification with Tris saturated

phenol during purification phase the quality and quantity (200-400 ng) of DNA was improved significantly without contamination of polysaccharides and secondary metabolites. In the present protocol, the use of 3.5 M NaCl in the extraction buffer reduced 90% of polysaccharides contamination and very little or no jelly like precipitate was found during precipitation of DNA. One of the most significant steps of our protocol was the use of only Tris saturated phenol (pH 8.0), followed by chloroform: isoamyl alcohol extraction. Most of the protocols in the literature used phenol: chloroform: isoamyl alcohol (25:24:1) or chloroform: isoamyl alcohol (24:1) for protein removal (Doyle and Doyle 1990; Dellaporta et al 1983), whereas in our experiments use of either phenol: chloroform: isoamyl alcohol (25:24:1) or chloroform: isoamyl alcohol (14:1) gave yellowish pellet, which confirmed the presence of protein contamination. By using Tris saturated phenol (pH 8.0) followed by chloroform: isoamyl alcohol (24:1) extraction, protein impurities could be successfully removed, without affecting DNA yield. It was also observed that buffer to tissue ratio and incubation time were also important factors for obtaining higher yields of DNA and in case of Artemisia annua 5:1 buffer to tissue ratio and 90 min incubation at 76^0C gave best results. On agarose gel electrophoresis, DNA gave sharp bands (Figure 2).

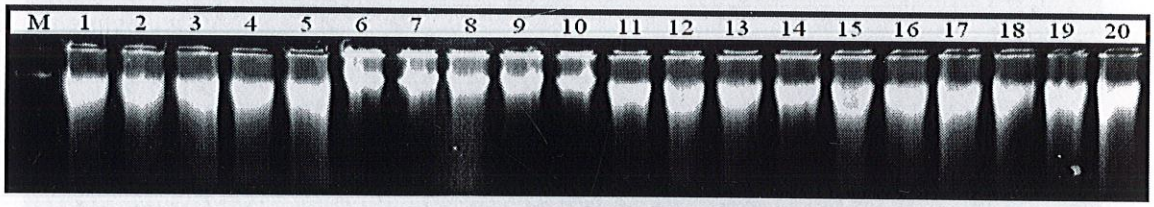

Fig.2. Agrose gel electrophoresis showing purified high molecular weight *Artemisia annua* genomic DNA of different Species, more than 200 ng genomic DNA from each genotype was electrophoresed on 0.8% agarose gel at 65 V for 2 hr and stain with ethidium bromide. Lane M, 100 ng mol. wt. uncut λ DNA. Lane 1 to 10; are the genomic DNA extracted from Leh valley and Lane 11 to 20; are from Nubra valley (Partapur).

Use of high concentration of NaCl in the extraction buffer decreased contamination of polysaccharides (Danshwar P and Sher-ullah 2004). Compared to precipitation at -20 ^{0}C (Khanuja et al 1999), we could achieve the same at room temperature without compromising the quality and quantity of DNA suitable for ligation, PCR amplification and other downstream processes necessary for DNA fingerprinting (Figure 3 & 4).

2.5.1 RAPD analysis

Genetic relationship among the 20 genotypes of *Artemisia annua* has been carried out using RAPD. Twenty primers generated reproducible, informative and easily scrabble RAPD profiles were preselected. These primers produced multiple band profiles (Figure 3) with a number of amplified DNA fragments varying from 4 to 9, with a mean number 6.2 markers per primer. All the amplified fragments varied in size from 200 – 2,500 bp. Out of 124 amplified bands, 124 were found polymorphic. The highest number of polymorphic bands (9) was observed with primers S25 and S40, with 100 % polymorphism. The average number of polymorphic bands per primer among 20 genotypes was 6.2 (Table 1). The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variation in the *Artemisia* genotypes. The resolving power of the 20 RAPD primers ranged from 2.56 for primer S24 to a maximum of 5.56 for primer S30. In addition to its high resolving power, RAPD primer S30 has the ability to distinguish all 20 *Artemisia* genotypes.

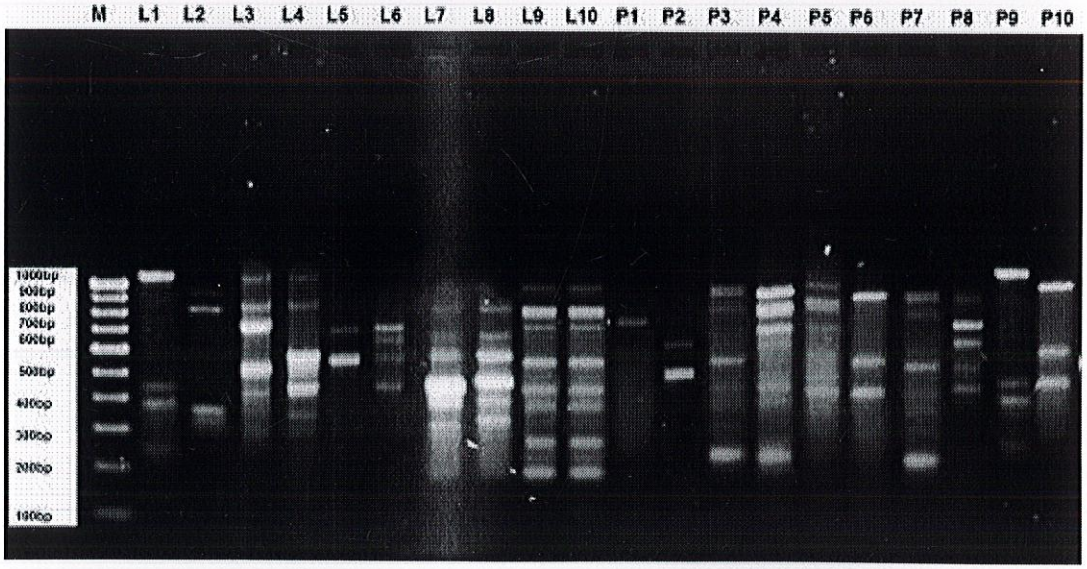

Fig. 3. RAPD amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10. are the genotypes collected form Leh valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = the size of molecular markers in base pairs using λ DNA.

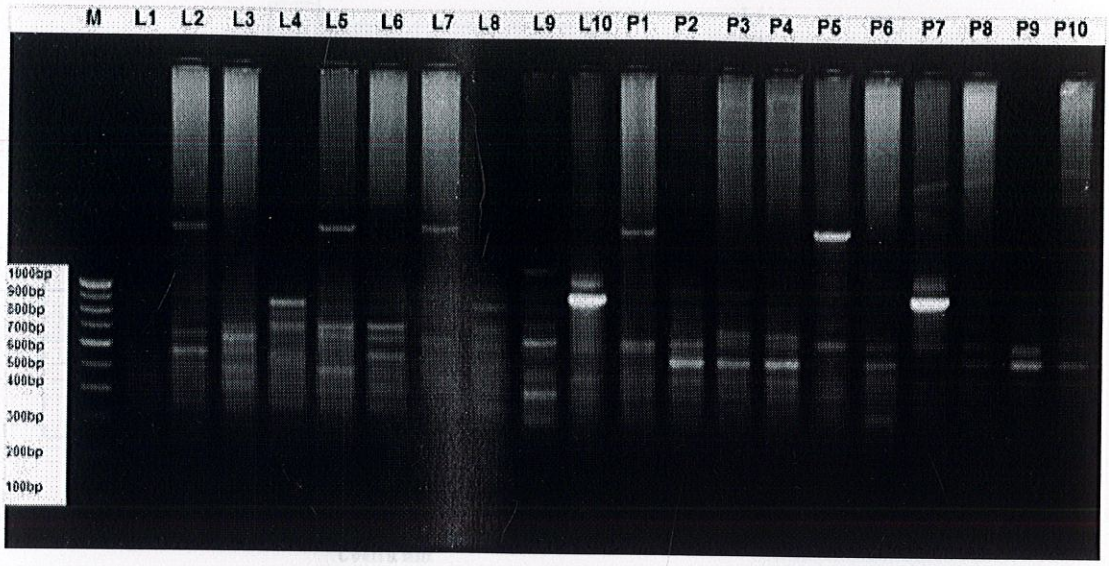

Fig. 4. ISSR amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10. are the genotypes collected form Leh valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = the size of molecular markers in base pairs using λ DNA.

2.5.2 RAPD derived dendrogram analysis

A dendrogram based on UPGMA analysis grouped the 20 genotypes into two main clusters (I and II) with Jaccard's similarity coefficient ranging from 0.55 to 0.70 (Figure 5). Cluster I represents the genotypes from Leh valley while, cluster II have all the genotypes from Nubra valley. The results of PCA analysis were comparable to the cluster analysis.

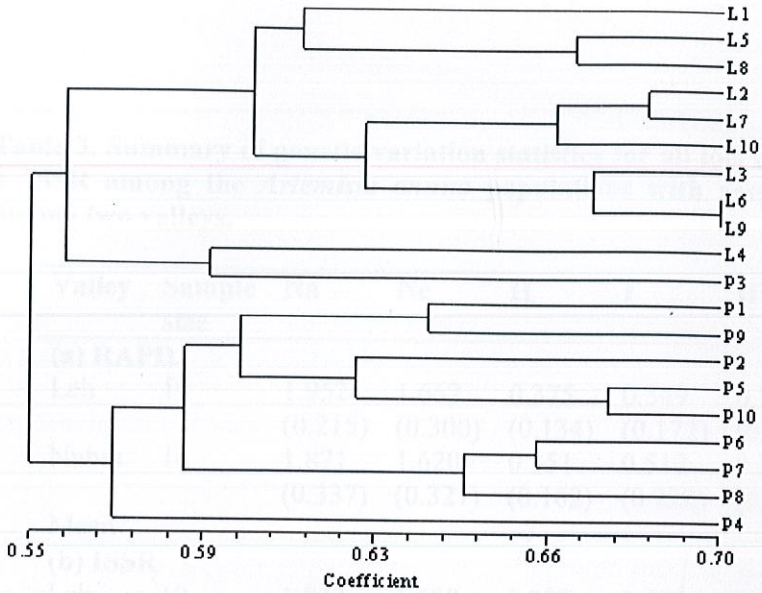

Fig.5. Dendrogram illustrating genetic relationships among 20 genotypes of *Artemisia annua* collected from 2 different valleys: Leh and Nubra Ladhak of Ladhak region. The dendrogram was generated by UPGMA cluster analysis calculated from 1482 RAPD bands produced by 20 primers.

2.5.3 Genetic Diversity Analysis

A relatively high genetic variation was detected among the *Artemisia annua* genotypes. Genetic diversity analysis in terms of Na, Ne, H, I, Ht, Hs, NPL and PPL for both the valleys (i.e. Leh and Nubra valleys) revealed higher values for Leh valley indicating more variability among the genotypes in Leh valley than in Nubra valley (Table 3). The respective values for overall genetic variability for Na, Ne, H, I, Ht, Hs, Gene flow (Nm), DI, EMR and MI across all the 20 genotypes were given in Table 4. The Gst value of 0.108 indicated that 89.2 % of the genetic diversity resided within the populations. AMOVA

Table 3. Summary of genetic variation statistics for all loci of RAPD, ISSR and RAPD + ISSR among the *Artemisia annua* populations with respect to their distributions among two valleys.

Valley	Sample size	Na	Ne	H	I	Ht	NPL	PPL
(a) RAPD								
Leh	10	1.952 (0.215)	1.662 (0.300)	0.375 (0.134)	0.549 (0.172)	0.375 (0.018)	118	95.16
Nubra	10	1.871 (0.337)	1.620 (0.327)	0.351 (0.162)	0.512 (0.222)	0.351 (0.026)	108	87.10
Mean								
(b) ISSR								
Leh	10	1.823 (0.383)	1.608 (0.370)	0.337 (0.184)	0.488 (0.253)	0.337 (0.034)	70	82.35
Nubra	10	1.812 (0.393)	1.558 (0.358)	0.318 (0.180)	0.467 (0.250)	0.318 (0.032)	69	81.18
Mean								
(c) RAPD + ISSR								
Leh	10	1.871 (0.347)	1.604 (0.341)	0.341 (0.169)	0.500 (0.230)	0.342 (0.028)	180	86.12
Nubra	10	1.861 (0.336)	1.577 (0.337)	0.332 (0.164)	0.490 (0.223)	0.332 (0.027)	182	87.08
Mean								

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.

Table 4. Overall genetic variability across all the 20 genotypes of Artemisia annua based on RAPD, ISSR and RAPD+ISSR analysis.

Marker Type	Na	Ne	H	I	Ht	Hs	Gst	NPL	PPL	Nm	DI	EMR	MI
RAPD	1.984 (0.125)	1.730 (0.258)	0.407 (0.106)	0.591 (0.124)	0.407 (0.011)	0.363 (0.012)	0.108	122	98.39	2.065	0.817	6.2	0.851
ISSR	1.906 (0.294)	1.658 (0.325)	0.367 (0.154)	0.535 (0.207)	0.367 (0.234)	0.327 (0.021)	0.109	77	90.59	2.044	0.767	4.588	0.876
RAPD+ISSR	1.962 (0.192)	1.706 (0.283)	0.394 (0.125)	0.572 (0.160)	0.394 (0.016)	0.337 (0.015)	0.145	201	96.17	1.474	-	-	-

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

Table 5. Summary of nested analysis of molecular variance (AMOVA) based on RAPD, ISSR individually and in combination, among the populations of apricot. Levels of significance are based on 1000 iteration steps.

Source of variation	Among valley			Among genotypes/valley		
df	1			18		
Marker	RAPD	ISSR	RAPD+ISSR	RAPD	ISSR	RAPD+ISSR
S.S.D	54.250	34.050	85.300	449.9	278.30	425.200
Variance component	2.962	1.859	6.168	24.994	15.461	23.622
Percentage	10	11	21	90	89	79
P-value	<0.001	<0.002	<0.001	<0.001	<0.002	<0.001

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

helps in partitioning of the overall RAPD variations among valley and among genotypes within the valley. Molecular variance among valley (10 %) and among genotypes within valley (90 %) (Table 5) revealed higher variations within the populations. All the components of molecular variations were significant ($P < 0.001$). This is helpful in making strategy for germplasm collection and evaluation. The rate of gene flow estimated using Gst value was found to be 2.065. The present study and similar studies on ginger 18, lotus 28, sweet potato 29 and *Andrographic paniculata* 30 suggest that RAPD is more appropriate for analysis of genetic variability in closely related genotypes. Moreover, the RAPD markers used here were able to differentiate *Artemisia annua* genotypes collected from Ladhak region, into two distinct clusters specific to 2 valleys. The study also indicates that *Artemisia annua* populations in the northwestern Himalayan region are genetically highly diverse.

2.5.4 ISSR analysis

The 17 ISSR primers selected generated a total of 85 ISSR bands (an average of 5 bands per primer), out of which 78 were polymorphic (91.76 %). Number of bands varied from 3 (ISSR 3) to 11 (ISSR 11), and sizes ranged from 200 – 2,500 bp. Average number of bands and polymorphic bands per primer were 5 and 4.59 respectively. Other primer amplification details are shown in Table 2. Amplification result of 17 primers seems to indicate that microsatellites more frequent in *Artemisia* contain the repeated di-nucleotides

(AG)n, (GA)n, (TG)n, (CT)n, (AT)n, (GT)nYA, and tri-nucleotides (ACC)n, (CCG)n, (GGC)n. The number of bands produced with different repeat nucleotide were more with the (GT)n, (GA)n, (CT)n, and (AC)n primers (ISSR 8, 6, 7 and 9). The primers that were based on the (GA)n, (CT)n and (GT)n motif produced more polymorphism on average (7 bands per primer) than the primers based on any other motifs used in the present investigation. The primer sequences which gave no amplification were (AT)n, (GT)nA, (TC)n, (TA)n, (CT)nT; tri-nucleotides repeats (TGC)n and the repeated tetra-nucleotides sequences (CTAG)n and (TGCA)n (Table IIb). We obtained good amplification products from primers based on (CT)n and (GT)n repeats while (AT)n and some other primers gave no amplification , despite the fact that (AT)n di-nucleotide repeats are thought to be the most abundant motifs in plant species (Martín and Sánchez-Yélamo, 2000). Similar results were obtained in grapevine (Moreno *et al.,* 1998), rice (Blair *et al.,* 1999), Vigna (Ajibade *et al.*, 2000) and wheat (Nagaoka and Ogihara, 1997). In case of non amplifying ISSR primers different PCR conditions was tried but still no amplification was observed. A possible explanation of these results is that ISSR primers based on AT motifs are self-annealing, due to sequence complementarity, and would form dimers during PCR amplification (Blair *et al.*, 1999) or it may be due to its-non annealing with template DNA due to its low Tm. Reason behind non amplification of other repeats may be their absence in the genome or due to the absence of motifs complementary to the primers in the genome. The primers with poly (GC)n and poly (GA)n motifs produced more polymorphism than any other motif. Somewhat similar result was also reported by Ajibade et al. (2000) where they found that the primer containing the CT repeats was one of those, which did not give interpretable phenotype when analyzed, while primers with GA and CA repeats revealed polymorphism in the genus Vigna. The resolving power (Rp) of the 17 ISSR primers ranged from 2.17 for primer ISSR 12 to 5.67 for ISSR 6 (Table 2).

2.5.5 ISSR derived dendrogram analysis

The complete data was based on a total of 1087 bands and Jaccard's similarity coefficient ranged from 0.57 to 0.81. The genotypes were clustered into two clusters (I and II) where, cluster I, represents all the genotypes from Leh valley while cluster II contains all

the genotypes from Nubra valley (Figure 6). The results of PCA analysis were comparable to the cluster analysis.

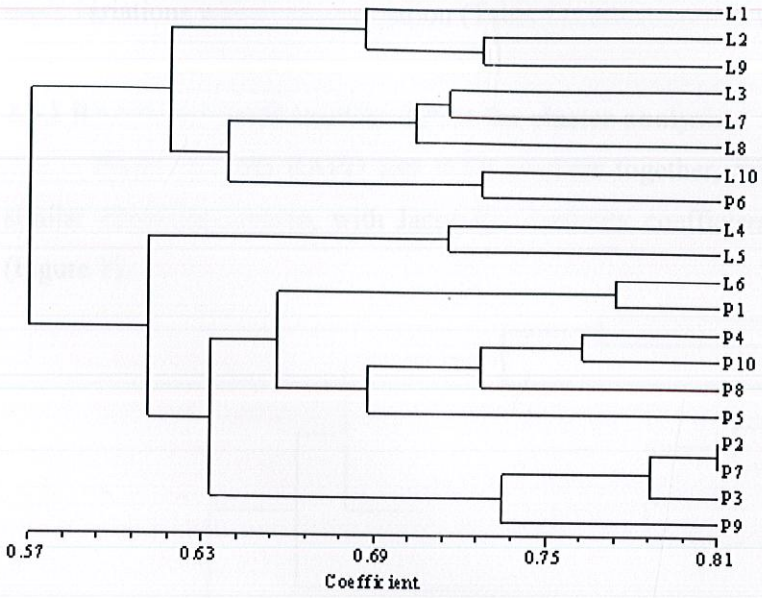

Fig.6. Dendrogram illustrating genetic relationships among 20 genotypes of *Artemisia annua* collected from 2 different valleys: Leh and Nubra Ladhak of Ladhak region. The dendrogram was generated by UPGMA cluster analysis calculated from 1087 ISSR bands produced by 17 primers.

2.5.6 Genetic Diversity Analysis

The respective values of Na, Ne, H, I, Ht, Hs, NPL and PPL were found higher for Leh valley genotypes indicating that there is more variability in Leh valley than in Nubra valley genotypes (Table 3). Gst value 0.109 indicated that 89.1 % of the genetic diversity resided within the population . The details of overall genetic variability across 20 genotypes were given in Table 4. AMOVA for among valley (11 %) and among genotypes within the valley (89 %) indicated that there are more variations within the population (Table 5). The estimated gene flow was 2.044.

2.6 Genetic variability details from RAPD + ISSR combined data

Estimated values of Na, Ne, H, I, Ht, Hs, NPL and PPL parameters for both the valleys revealed higher values among the genotypes of Leh (Table 3). Gst value 0.145

indicated that 85.5 % of the genetic diversity resided within the population (Table 4). The details of overall genetic variability of 20 genotypes are given in Table 4. AMOVA analysis for among valley (21 %) and among genotypes within valley (79 %) revealed that there are more variations within the population (Table 5). The estimated gene flow was 1.474.

2.6.1 RAPD and ISSR combined data for cluster analysis

Based on both RAPD and ISSR markers together, the dendrogram obtained gave similar clustering pattern, with Jaccard's similarity coefficient ranging from 0.57 to 0.69 (Figure 7).

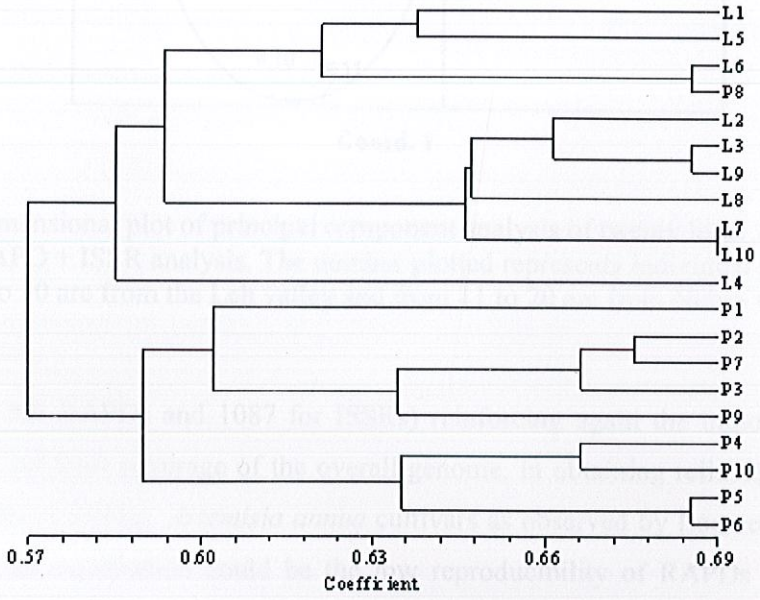

Fig.7. Dendrogram illustrating genetic relationships among 20 genotypes of *Artemisia annua* collected from 2 different valleys: Leh and Nubra Ladhak of Ladhak region. The dendrogram was generated by UPGMA cluster analysis calculated from 1482 RAPD bands+1087 ISSR bands.

All the genotypes were clustered into two major clusters (I and II) with 10 genotypes each. Cluster I represents all the genotypes from Leh valley whereas cluster II represents all the genotypes from Nubra valley. Both RAPD and ISSR clusters showed partial similarity with combined data of RAPD + ISSR. The results of PCA analysis also revealed similar clustering pattern (Figure 8) unlike that of dendrogram analysis. Other genetic variation studies were also performed on ISSR and RAPD combined data which are represented in

different tables (Tables 3 and 4). The differences found among the dendrograms generated by RAPDs and ISSRs could be partially explained by the different number of PCR products

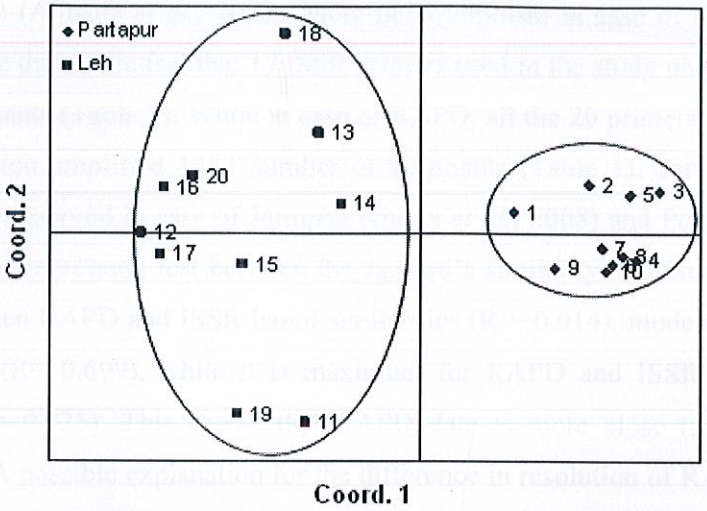

Fig. 8. Two-dimensional plot of principal component analysis of twenty local *Artemisia annua* using RAPD + ISSR analysis. The number plotted represents individual genotypes. The number 1 to 10 are from the Leh valley and from 11 to 20 are from Nubra valley (Partapur).

analyzed (1482 for RAPDs and 1087 for ISSRs) reinforcing again the importance of the number of loci and their coverage of the overall genome, in obtaining reliable estimates of genetic relationships among *Artemisia annua* cultivars as observed by Loarce *et al.* (1996) in barley. Another explanation could be the low reproducibility of RAPDs (Karp *et al.*, 1997). The genetic closeness among the Leh valley and Nubra valley cultivars can be explained by the high degree of commonness in their genotypes. Similar result has been obtained by Gaffor *et al.*, (2001) in blackgram. In all the dendrograms, genotypes from both the valleys were found clustered together. The genetic similarity of these genotypes is probably associated with their similarity in the genomic and amplified region.

2.6.2 Comparative analysis of RAPD with ISSR markers

RAPD markers were found more efficient with respect to polymorphism detection, as they detected 100 % polymorphism as compared to 91.8 % for ISSR markers. Also, the number of polymorphic loci, % polymorphic loci, diversity index, effective multiplex ratio

and marker index are more for RAPD than for ISSR markers (Table 4). This is in contrast to the results as obtained for several other plant species like wheat (Nagaoka and Ogihara, 1997) and Vigna (Ajibade *et al.*, 2000). More polymorphism in case of RAPD than ISSR markers might be due to the fact that 17 ISSR primers used in the study only amplified 1087 number of fragments (Table 2). While in case of RAPD, all the 20 primers which were used in the investigation amplified 1482 number of fragments (Table 1). Same polymorphism pattern was also observed in case of Jatropha (Gupta *et al.*, 2008) and Podophyllum (Alam *et al.*, 2009). The regression test between the Jaccard's similarity matrixes resulted in low regression between RAPD and ISSR based similarities (R = 0.014), moderate for ISSR and ISSR + RAPD (R= 0.699), while it is maximum for RAPD and ISSR + RAPD based similarities (R = 0.725). This shows that RAPD data is more close to ISSR + RAPD combined data. A possible explanation for the difference in resolution of RAPDs and ISSRs is that the two-marker techniques target different portions of the genome. With this study, we can conclude that the molecular analyses of both RAPD and ISSR markers were extremely useful for studying the genetic relationships of local *Artemisia* genotypes from the trans-Himalayan region of Ladakh. The results indicate the presence of great genetic variability among local genotypes of *Artemisia* where variation between valleys are less than within the valley variations, which should be exploited for the future conservation and breeding of *Artemisia* from this region. Geographically isolated population accumulates genetic differences as they adapt to a different environment. Genetic variation among elite genotypes of *A. annua* based on RAPD and ISSR analysis could be useful in selecting parents to be crossed for generating appropriate populations intended for both genome mapping and breeding purposes. In general, dispersal resulting in colonization and gene flow into existing populations is very important for both the persistence and genetic success of a species (Hamrick & Godt, 1996). 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).

Considering the high genetic differentiation among the wild populations of *Artemisia annua*, conservation of only a few populations may not adequately protect the genetic

variations within the species in the Himalayan region. This species, or at least a large part of its genetic diversity, may be lost in the near future, owing to its importance and consequent exploitations as a medicinal plant, if appropriate conservation measures are not adopted. Since no single, or even a few plants, will represent the whole genetic variability in *Artemisia annua*, there appears to be a need to maintain a sufficiently large population in natural habitats to conserve genetic diversity in *Artemisia annua* to avoid genetic erosion. Based on polymorphic feature, genetic diversity, genetic similarity, and gene flow among the populations of Artemisia based on RAPD and ISSR study, we recommend that any future conservation plans for this species should be specifically designed to include representative populations with the highest genetic variation for both in situ conservation and germplasm collection expeditions.

Acknowledgement

The authors are thankful to DRDO HQ for its support and funding the project.

References

Ajibade SR, Weeden NF and Chite SM (2000). Inter simple sequence repeat analysis of genetic relationships in the genus *Vigna*. *Euphytica* 111: 47–55.

Alam A, Naik PK and Mishra Gyan P (2009). Congruence of RAPD and ISSR markers for evaluation of genomic relationship among 28 populations of *Podophyllum hexandrum* from Himachal Pradesh. *Turk. J. Bot.* 33: 1–12.

Blair MW, Panaud O and Mccouch SR (1999). Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (*Oryza sativa* L.). *Theo. Appl. Genet.* 98: 780–792.

Botstein D, White R, Skolnick M and Davis R (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am. J. Genet.* 32: 314-331.

Charles DJ, Cebert E and Simon JE (1991). Characterization of the essential oil of Artemisia oil L.*J. Essential Oil Res* 3: 33–39.

Danshwar P and Sher-ullah SSK (2004). Genomic DNA extraction from Victoria amazonica. *Plant Mol Biol Rep.* 22: 195a-195j.

Dellaporta SL, Wood J and Hicks JB (1983). A plant DNA mini-preparation: version II. *Plant Mol Biol Rep.* 1: 19-21.

Dellaporta SL, Wood J and Hicks JB (1985). Maize DNA miniprep. In: Malmberg R, Messing J and Sussex I (eds), Molecular Biology of Plants, pp. 36–37. Cold Spring Harbour Laboratory Press, New York.

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

Doyle JJ and Doyle LJ (1990). Isolation of fresh DNA from fresh tissue. *Focus* 12: 13-15.

Excoffier L, Smouse PE and Quattro JM (1992). Analyses of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Ferreira JFS, Simon JE and Janick J (1996). Relationship of artemisinin content of tissue cultured greenhouse grown and field grown plants of *Artemisia annua. Planta Medica* 61: 351–355.

Gaffor A, Sharif A, Ahmad Z, Zahid MA and Rabbani MA (2001). Genetic diversity in blackgram (*Vigna mungo* L. Hepper). *Field Crop Res.* 69: 183–190.

Gupta S, Srivastava M, Mishra Gyan P, Naik PK, Chauhan RS, Tiwari SK, Kumar M and Singh R (2008). Analogy of ISSR and RAPD markers for comparative analysis of

genetic diversity among different *Jatropha curcas* genotypes. *Afr. J. Biotechnol.* 23: 4230–4243.

Higgins M (1984). Synthesis regulation of major protein seeds. *Ann. Rev. Plant Psysiol.* 35: 191-221.

Karp A, Edwards K, Bruford M, Vosman B, Morgante M, Seberg O, Kremer A, Boursot P, Arctander P, Tautz D and Hewitt G (1997). Newer molecular technologies for biodiversity evaluation: Opportunities and challenges. *Nat. Biotechnol.* 15: 625–628.

Khanuja SPS, Shasany AK, Daroker MP and Kumar S (1999). Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. *Plant Mol Biol Rep.* 17: 1-7.

Klayman DL (1993) *Artemisia annua I*: From weed to respectable antimalarial plant. In: Kinghorn AD and Balandrin MF (eds), Human medicinal agents from plants, pp. 242–255. Am Chem Soc Symp Ser Washington D.C.

Krawczak M and Schmidtke J (1994). Introduction in DNA fingerprinting Bio scientific publishers Ltd. Oxford, UK pp. 1-16.

Loarce Y, Gallego R and Ferrer E (1996). A comparative analysis of genetic relationships between rye cultivars using RFLP and RAPD markers. *Euphytica* 88: 107– 115.

Martín JP and Sánchez-Yélamo MD (2000). Genetic relationships among species of the genus *Diplotaxis* (*Brassicaceae*) using inter-simple sequence repeat markers. *Theor. Appl. Genet.* 101: 1234–1241.

Moreno S, Martin JP and Ortiz JM (1998). Inter-simple sequence repeat PCR for characterization of closely related grapevine germplasm. *Euphytica* 101: 117–125.

Nagaoka T and Ogihara Y (1997). Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet.* 94: 597–602.

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

Ovesna J, Polakova K and Lisova L (2002). DNA analysis and their applications in plant breeding. *Czech. J. Plant Breed.* 38: 29-40.

Peakall R and Smouse PE (2001). GenAlEx V5: *Genetic Analysis in Excel. Population genetic software for teaching and research.* Australian National University, Canberra, Australia, (http://www.anu.edu.au/BoZo/GenAlEx/).

Phillipson DJ and Wright CW (1991). Antiprotozoal agents from plant sources. *Plant Medica* 57 (Suppl 1): 553–59.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2: 225–238.

Prevost A and Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor. Appl. Genet.* 98: 107–112.

Rohlf FJ (1992). NTSYS-PC: Numerical taxonomy andmultivariate analysis system version 2.0. State University of New York (Stony Brook, New York).

Saghai-Maroof MA, Soliman KM, Jorgenses RA and Allard RW (1984). Ribosomal DNA spacer-length polymorphism in barley. Mendelian inheritance, chromosomal location and population dynamics . *Proc Natl Acad Sci* 81: 8014–19.

Sangwan RS, Agarwal K, Luthra R, Thakur RS and Sangwan NS (1993). Biotransformation of arteannuic acid into arteannuin-b and artemisinin in *Artemisia annua. Phytochemistry* 34: 1301–1302.

Stange C, Prehn D and Arce-Jounson P (1998). Isolation of Pinus radiate genomic DNA suitable for RAPD analysis. *Plant Mol Biol Rep*. 16: 1-8.

Stewart CN and Via LE (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other applications. *Biotechniques* 14: 748-749.

Webb DM and Knapp SJ (1990). DNA extraction from a previously recalcitrant plant genes. *Plant Mol Biol Reptr*. 8: 180–185.

Weising K, Atkinson G and Gardner C (1995). Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation. *PCR Methods and Applications* 4: 249-255.

Williams K, Kubelik A, Livak K, Rafalski J and Tingey V (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleis Acid Res.* 18: 6531-6535.

Xiao J, Li J, Yuan L, Mccouch S and Tanksley SD (1996). Genetic diversity and its relationships to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor. Appl. Genet* 92: 637-643.

Zhao WG, Zhang JQ, Wangi YH, Chen TT, Yin Y, Huang YP, Pan Y and Yang Y (2006). Analysis of genetic diversity in wild populations of mulberry from western part of Northeast China determined by ISSR markers. *J. Genet. Mol. Biol.* 7: 196–203.

Zietkiewicz E, Rafalski A and Labuda D (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176–183.