ASSESSMENT OF GENETIC DIVERSITY AND ARTEMISININ CONTENT AMONG GENOTYPES OF *A. ANNUA* AND RELATED SPECIES COLLECTED FROM THE LADAKH REGION, INDIA.

Synopsis
of
Ph. D Thesis

Submitted by

JITENDRA KUMAR

Enrollment Number: 086559

DEPARTMENT OF BIOTECHNOLOGY & BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN-173 234, (HP)

OCTOBER 2011

Introduction

Artemisia annua (Figure 1) is an important medicinal plant for the production of anti-malarial (e.g. artemisinin) and possibly antibacterial agents and natural pesticides. It grows mainly in the middle, eastern and southern parts of Europe, in the northern, middle and eastern parts of Asia and in North Africa [1,2]. In India it grows naturally in the Ladakh region. It was originally collected by the Chinese as herbal medicine and is currently processed by pharmaceutical firms for the production of artemisinin derivatives, mainly for Artemisinin-based Combination Therapies (ACTs) in the effective treatment of malaria. The artemisinin and its derivatives are effective against *Plasmodium falciparum* and *P. vivax*, including multi-drug-resistant strains. World Health Organization's (WHO) also made mandatory to give artemisinin along with presently available anti-malarial drug for treating malaria. Malaria is still severely challenging people's health. Each year, more than one million people worldwide die of malaria and more than two billion people in over 100 countries are threatened by the disease [3, 4]. In many developing countries, especially those in Africa, the mortality rate from malaria are still very high. Therefore the demand for artemisinin and its derivatives is now growing rapidly. At present, artemisinin compounds are derived from a raw substance extracted from the plant *A. annua* L. because artemisinin is very difficult to synthesize [5, 6]. Harvesting from the main source of raw material, (wild) of *A. annua* L. is causing loss of genetic diversity and habitat destruction. Cultivation of this plant is a viable alternative and can overcome cultivation difficulties of it. Unfortunately, *A. annua* produces low amount of artemisinin, ranging from 0.1 to 1% of the plant dry weight, depending on the geographical origin of the plant [7]. Enhanced production of artemisinin in the whole plant of *A. annua* L. is therefore highly desirable.

For the purpose of efficient conservation and successful breeding program, it will be prudent to study the populations of *A. annua* at genetic and molecular levels. Study within and between populations variations at the molecular level provides an efficient tool for taxonomic and evolutionary studies and for devising strategies to protect the genetic diversity of the plant. Genetic variability also can be exploited to select useful genotypes that could be utilized as cultivars to avoid batch-to-batch variations in extraction of standard drugs. Understandings of *A. annua* population biology and genetic diversity as well as its relationship with artemisinin content are hence essential. DNA-based molecular-marker techniques such as RAPD [8], ISSR [9], ILP [10] and SSR [11] have been proved powerful in genetic diversity estimations. However, the impact of environmental factors with the yield of artemisinin cannot be ruled out as the plant is very adaptable to a wide range of environmental conditions. A number of investigations have demonstrated that the quality and

quantity of several secondary metabolites have a close relationship with plant habitats [12]. Hence, it demands management of soil nutrients and optimization of climatic factors for the successful conservation of *A. annua*. Statistical methods such as artificial neural network (ANN) and multiple linear regressions (MLR) are very useful in these respects. The non-linear predictors such as ANNs were widely used to solve various problems in agriculture. For example, Alam and Naik [13] successfully predicted podophyllotoxin yield with back propagation neural network models based on environmental factors, pH and some soil nutrient elements. Furthermore, it is imperative to search for supplementary artemisinin sources to meet the future requirements for artemisinin based-combination therapies. Therefore other species of *Artemisia* should be screened as potential new sources of artemisinin for agricultural production.

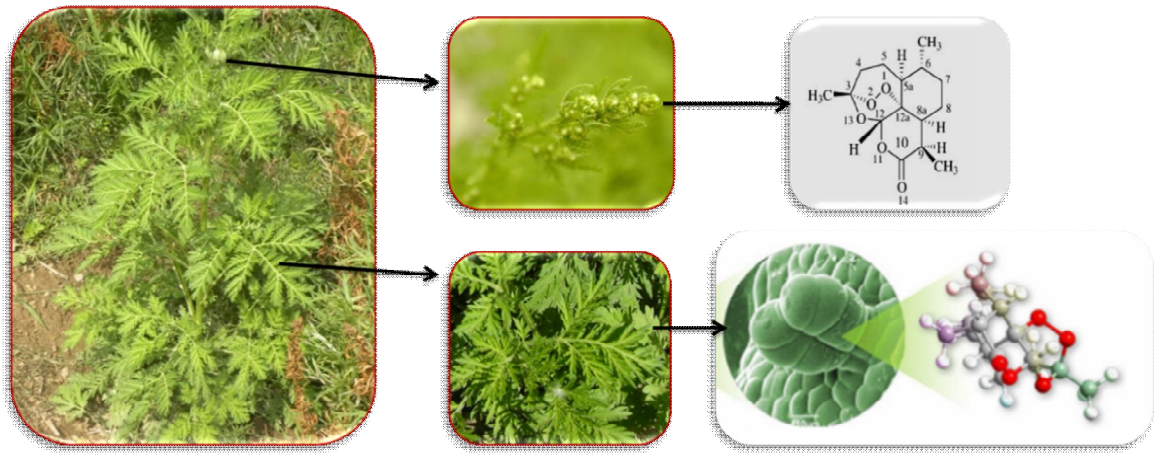


Figure 1. The plant *Artemisia annua*, its leaves and inflorescence as well as the trichomes that accumulate artemisinin.

The present work is thus originated in view of the importance of *A. annua* to help the Indian pharmaceuticals to manufacture antimalarial drugs, to understand genetic diversity among genotypes, to study the impact of environmental factors on artemisinin content, to explore the alternative source of artemisinin for commercial production of artemisinin and realizing the existing lacunae in study made so far in this aspect.

Objectives

- To ascertain genetic diversity among the genotypes and populations of *Artemisia annua* based on DNA profiling techniques utilizing dominant molecular markers such as RAPD and ISSR as well as co-dominant markers such as EST-derived SSRs and intron-flanking EST-specific markers in the Ladakh region, India.

- To optimize the extraction method and quantification of artemisinin at different developmental stages from *Artemisia annua* L. collected from Ladakh region.

- To establish the relationship between the environmental factors and artemisinin content and to develop prediction model that would be useful for selection of conservation area as well as optimization of environmental factors for commercial cultivation of *A. annua*.

- To explore an alternative source of artemisinin by screening other species of *Artemisia* growing in cold arid desert of trans-Himalayas (Ladakh, India) – experimentally validate the presence of artemisinin synthesis pathway by selective amplification of targeted genes in the metabolic pathway and sequencing.

Outline of the works

Each of these six pieces of work has distinct characteristics. At the same time they are related to one another. To clearly and coherently demonstrate the goal, results and conclusion of each piece of work, we have arranged each work chapter wise in a publishing format. The format will benefit readers to understand the idea of development, conclusion, coherence and full significance as each chapter will be a full manuscript from background to conclusion at publication stage.

Aim 1: Study of genetic variation among *Artemisia annua* genotypes collected from the Ladakh region using dominant molecular markers such as RAPDs and ISSRs.

Plant material

Twenty genotypes of *A. annua* were collected from the two valleys: Nubra valley (at altitude 3000 m) and Indus valley (at altitude 3500 m) from the cold arid desert of the trans-Himalayas (Ladakh, India). The young leaves were collected from 10 individual plants from each valley and stored in laboratory at –80 °C until further analysis. The interval between samples was 100-200 m and the pair wise distance between valleys was about 150 Km.

Extraction and quantification of genomic DNA

The plants contain high amount of polysaccharides, polyphenols & other secondary metabolites that entangle to nucleic acid during DNA isolation & interfere with subsequent isolation procedure (Zidani et al., 2005). Therefore we have modified the CTAB method [14] for total genomic

DNA extraction from frozen leaves. RNA was removed by RNase treatment. DNA was quantified by comparing with uncut λ DNA on the 1% agrose gel. The DNA sample was diluted to 15 ng/µl and used for PCR amplification. The quality of DNA that has been obtained from the samples is shown in Figure 2.

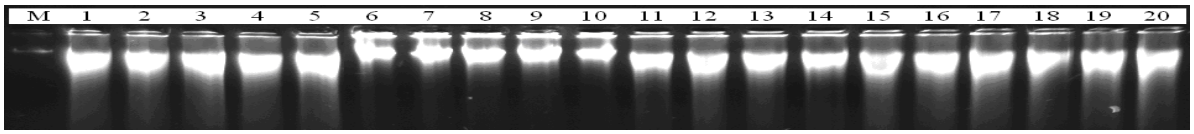


Figure 2. Agrose gel electrophoresis showing purified high molecular weight *Artemisia annua* genomic DNA from 20 genotypes. Lane M, 100 ng molecular weight uncut λ DNA. Lane 1 to 10 are the genomic DNA extracted from Indus valley and Lane 11 to 20 are from Nubra valley.

Genetic characterization using RAPD and ISSR markers

Genetic characterization of *A. annua* genotypes was performed using a set of 20 RAPDs markers (Table 1) and 17 ISSRs markers (Table 2).

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	% of polymorphic loci	Total number of fragments	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	8	88.9	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	7	87.5	100	5.56
S31	CAATCGCCG T	60	36.7	5	5	100	75	4.17
S32	TCGGCGATA G	60	34.0	6	5	83.3	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	7	77.8	86	4.78
	Total	-	-	**124**	**119**	**96.9**	**1482**	-

Amplification reactions were performed in volumes of 25 µl containing 10 mM Tris- HCl (pH 9.0), 1.5 mM $MgCl_2$, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 20 ng template DNA and 0.5 unit 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. For 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 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	% of polymorphic loci	Total number of fragments	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	5	71.4	102	5.67
ISSR 7	(CT)8 RA	47.2	47.1	5	4	80	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	8	72.7	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**	**72**	**86.02**	**1087**	**-**

Data 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 dissimilarity coefficient (J) was calculated, subjected to cluster analysis by bootstrapping and neighbor-joining method using the program DARWIN (version 5.0.158). Genetic diversity among genotypes and between two major groups (with respect to valleys) was analyzed using POPGENE software (version 1.32). 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. The RAPD and ISSR data were subjected to a hierarchical analysis of molecular variance (AMOVA) [15] using two hierarchical levels; among valleys and among genotypes. Statistically unbiased clustering of collected genotypes was performed using STRUCTURE (version 2.3.1).

Results and discussion

RAPD analysis

Twenty genotypes of *A. annua* collected from two valleys were fingerprinted using twenty preselected RAPD primers. These primers produced multiple band profiles (Figure 3) with a number of amplified DNA fragments varying from 4 to 9, with a mean of 6.2 markers per primer. All the amplified fragments varied in size from 200-1000 bp. Out of 124 amplified bands, 119 were found polymorphic (Table 1). The observed high proportion of polymorphic loci suggests that there is a high degree of genetic variation in the *Artemisia* genotypes. A dendrogram analysis based on bootstrapping and neighbor joining (NJ) method grouped all the 20 genotypes into two main clusters (with reference to their site of collection) (Figure 4a).

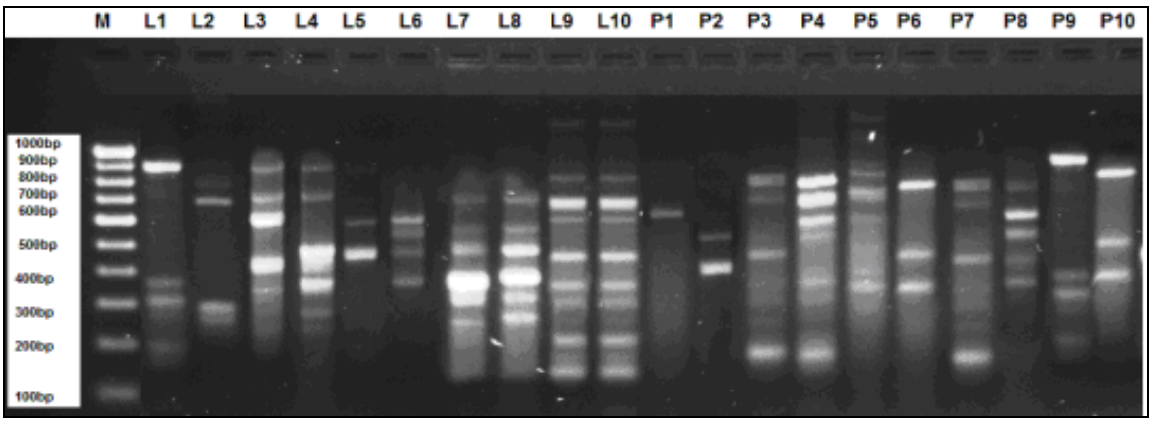


Figure 3. A representative gel picture of RAPD amplification products obtained from 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10, are the genotypes collected from Nubra valley. M = the size of molecular markers in base pairs.

An unbiased clustering of genotypes based on STRUCTURE program, without prior knowledge about the populations, clustered all the 20 genotypes into two major groups. It was found that the genotypes were more likely distributed (at high probability) with respect to their geographical distribution (Figure 4b). Genetic diversity analysis in terms of Na, Ne, H, I, Ht, and PPL with respect to both the valleys revealed higher values, indicating more variability among the genotypes (Table 3). The genetic variation was high among the genotype as measured by percentage of polymorphic bands (96.9%) and Shannon information index (I = 0.53). A relatively high genetic variation was detected among the 20 genotypes (90%), whereas the variation between the two valleys was less (10%) using AMOVA test.

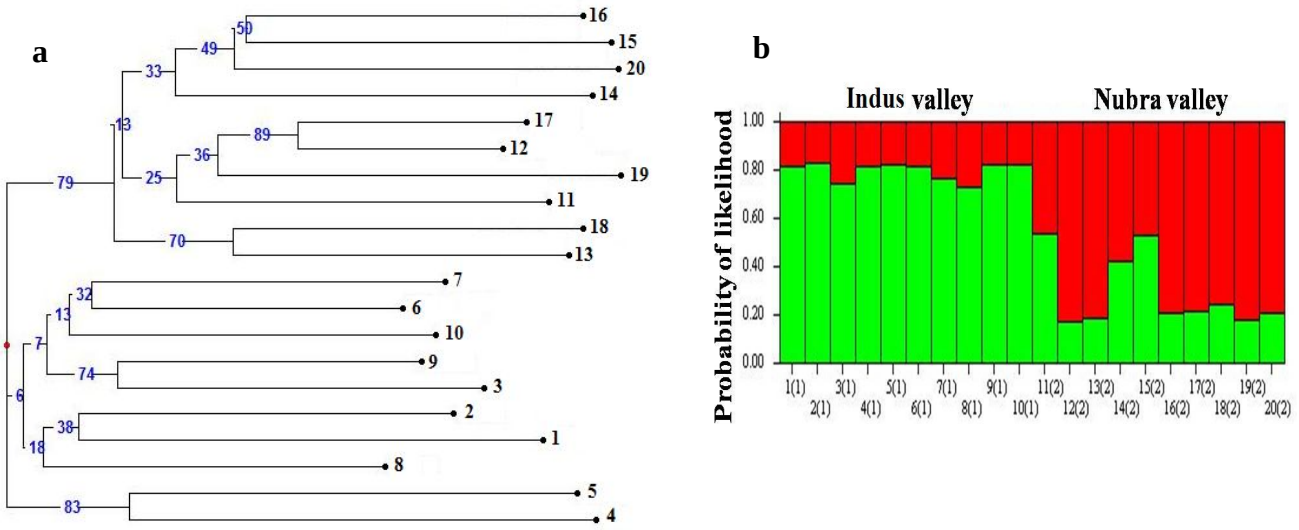


Figure 4. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on RAPD profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites. The genotypes were more likely clustered corresponding to both the valleys. The value within bracket represents the different valley (1, Indus valley and 2, Nubra valley). Genotypes from both the valleys are represented with different colours: Indus valley (green) and Nubra valley (red).

Table 3. Summary of genetic variation statistics for all loci of RAPDs among the *Artemisia annua* genotypes with respect to their distributions among two valleys.

Valley	Sample size	Na	Ne	H	I	Ht	PPL
(a) RAPD							
Indus	10	1.952 (0.215)	1.662 (0.300)	0.375 (0.134)	0.549 (0.172)	0.375 (0.018)	98.9
Nubra	10	1.871 (0.337)	1.620 (0.327)	0.351 (0.162)	0.512 (0.222)	0.351 (0.026)	94.9
Mean		1.911	1.641	0.363	0.530	0.363	96.9

Na = Observed number of alleles; Ne = Effective number of alleles; H = Nei's gene diversity; I = Shannon's Information index; Ht = Total genetic diversity; PPL = percentage of polymorphic Loci.

ISSR analysis

The 17 ISSR primers preselected in the study generated a total of 85 ISSR bands (an average of 5 bands per primer), out of which 72 were polymorphic (86.02%). Number of bands varied from 3 to 11 with sizes ranged from 200 – 1000 bp (Figure 5). Average number of bands and polymorphic bands per primer were 5 and 4.23 respectively. Amplification result of 17 primers seems to indicate that microsatellites more frequent in *Artemisia annua*, 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 complete data set of 1087 bands was used for cluster analysis based on bootstrapping and NJ method.

All the 20 genotypes were distinctly group into two major clusters as per their geographical distribution (Figure 6a). Similarly the program STRUCTURE based on unbiased clustering technique calculated the likelihood distribution of genotypes into two groups—genotypes are more likely distributed (with high probability score) with respect to their collection sites based on ISSR profiling (Figure 6B). The respective values of Na, Ne, H, I, Ht and PPL among the genotypes were found higher for Indus indicating that there is more variability in comparison to Nubra valley (Table 4). Analysis of molecular variance indicated that the source of variation among 20 genotypes was 89% which is significantly high in comparison to two valleys (11%).

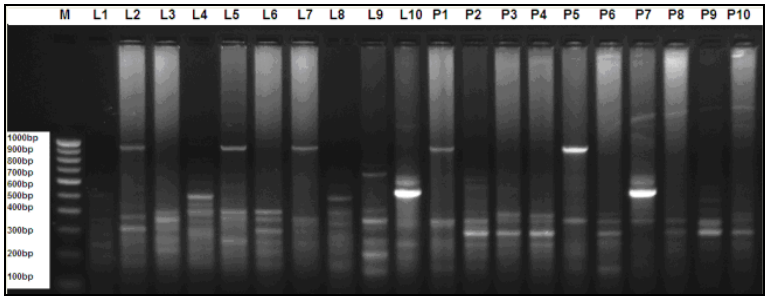


Figure 5. A representative gel picture showing ISSR profiling, obtained from 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra valley. M = the size of molecular markers in base pairs using λ DNA.

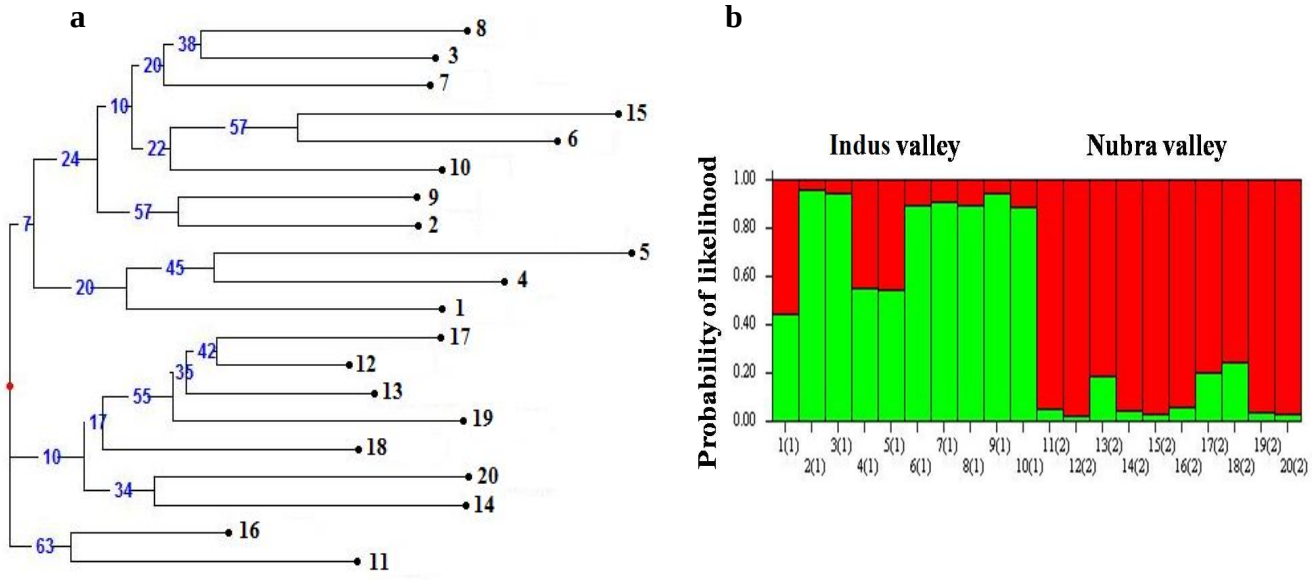


Figure 6. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on ISSR profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites. The genotypes were more likely clustered corresponding to both the valleys. The value within bracket represents the different valley (1, Indus valley and 2, Nubra valley). Genotypes from both the valleys are represented with different colours: Indus valley (green) and Nubra valley (red).

Table 4. Summary of genetic variation statistics for all loci of ISSR among the *Artemisia annua* genotypes with respect to their distributions among two valleys.

Valley	Sample size	Na	Ne	H	I	Ht	PPL
Indus	10	1.823 (0.383)	1.608 (0.370)	0.337 (0.184)	0.488 (0.253)	0.337 (0.034)	87.8
Nubra	10	1.812 (0.393)	1.558 (0.358)	0.318 (0.180)	0.467 (0.250)	0.318 (0.032)	84.3
Mean		1.817	1.583	0.327	0.477	0.327	86.1

RAPD and ISSR combined data for cluster analysis

Based on combined data set of RAPD and ISSR markers, the dendrogram obtained gave similar clustering pattern like RAPD and ISSR (Figure 7a). Cluster I represents all the genotypes from Indus valley whereas cluster II represents all the genotypes from Nubra valley. This result is also corroborative with STRUCTURE analysis. The genotypes were more likely distributed according to their collection sites (Figure 7b). Other genetic variation studies were also performed on RAPD and ISSR combined data which are represented in Table 5. The genetic closeness among the Indus valley and Nubra valley genotypes can be explained by the high degree of commonness in their genomes. In all the dendrogram, genotypes from both the valleys were found clustered distinctly. The genetic similarity of these genotypes is probably associated with their similarity in the genomic and amplified region.

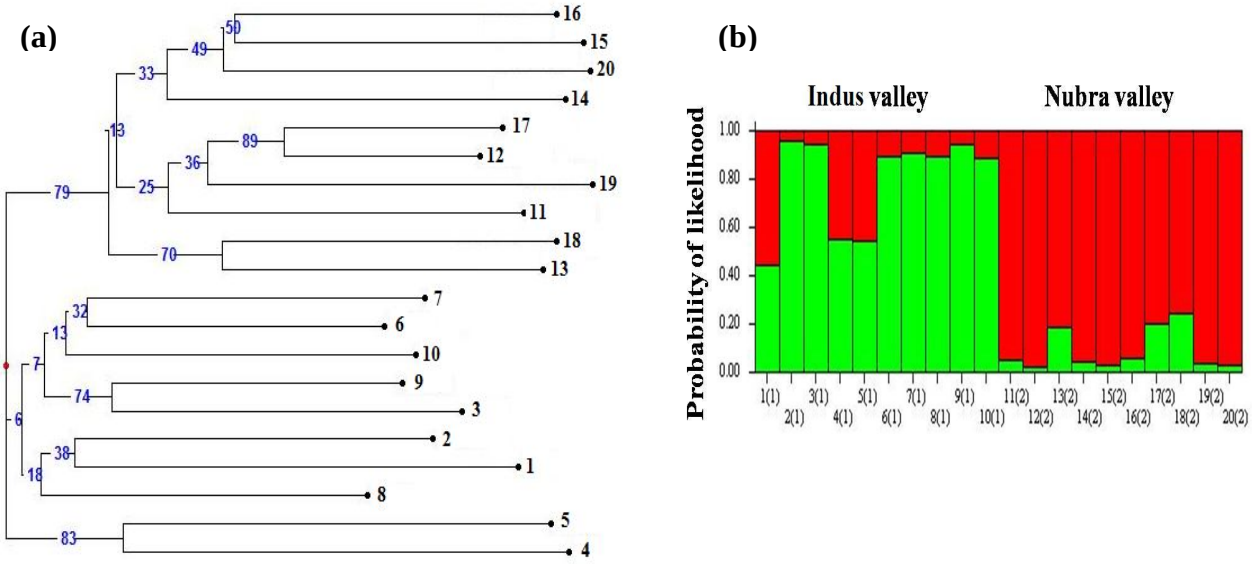


Figure 7. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on combination of RAPD and ISSR profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites. The genotypes were more likely clustered corresponding to both the valleys. The value within bracket represents the different valley (1, Indus valley and 2, Nubra valley). Genotypes from both the valleys are represented with different colours: Indus valley (green) and Nubra valley (red).

Table 5. 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	PPL
Indus	10	1.871 (0.347)	1.604 (0.341)	0.341 (0.169)	0.500 (0.230)	0.342 (0.028)	96.1
Nubra	10	1.861 (0.336)	1.577 (0.337)	0.332 (0.164)	0.490 (0.223)	0.332 (0.027)	85.8
Mean		1.866	1.591	0.336	0.495	0.337	91.0

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 indicates the presence of high genetic variability, which should be exploited for the future conservation and breeding of *Artemisia* from this region.

Aim 2: Study of genetic variation among *Artemisia annua* genotypes collected from the Ladakh region using co-dominant intron-flanking EST-specific and EST-derived SSRs molecular markers.

Introns are noncoding sequences interspersed in genes. In comparison with exons, introns are more variable because in general selective pressure in intronic regions is much less than exonic regions. Length polymorphism is the most intuitive variation in introns. So, intron length polymorphism (ILPs) has been exploited as molecular markers, which have many desirable properties, including specific, co-dominant, neutral, convenient and reliable. ILP can be conveniently detected by the PCR. To amplify introns by PCR, primers were designed from flanking exons. This approach is called exon-primed intron-crossing PCR (EPIC-PCR). The advantage of EPIC-PCR is that exon sequences are relatively more conservative and therefore the primers designed from exons may have more extensive applications than those designed from non-coding sequences.

Mining of putative intron polymorphic (PIP) markers

Putative intron polymorphic (PIP) markers were obtained utilizing the available EST sequences (~85,282) of *Artemisia annua* and by comparing with the genomic sequences of model organisms such as *Arabidopsis thaliana* and *Oryza sativa*. An in-house pipeline perl script was used to predict PIP markers by aligning EST sequences with *Oryza sativa* and *Arabidopsis* CDSs using BLASTN (Figure 8). For each of these aligned query ESTs, a pair of primers was designed using program Primer3 [16] from the query EST with 100 bp on each side of the target intron. A set of 40

primers (Table 6) were used for genetic characterization of 20 genotypes of *A. annua* collected from two valleys of the trans-Himalayan region, Ladakh, India. These primers anneal with the conserved region of exon (flanking to the intron) and amplified the introns from the genomic DNA.

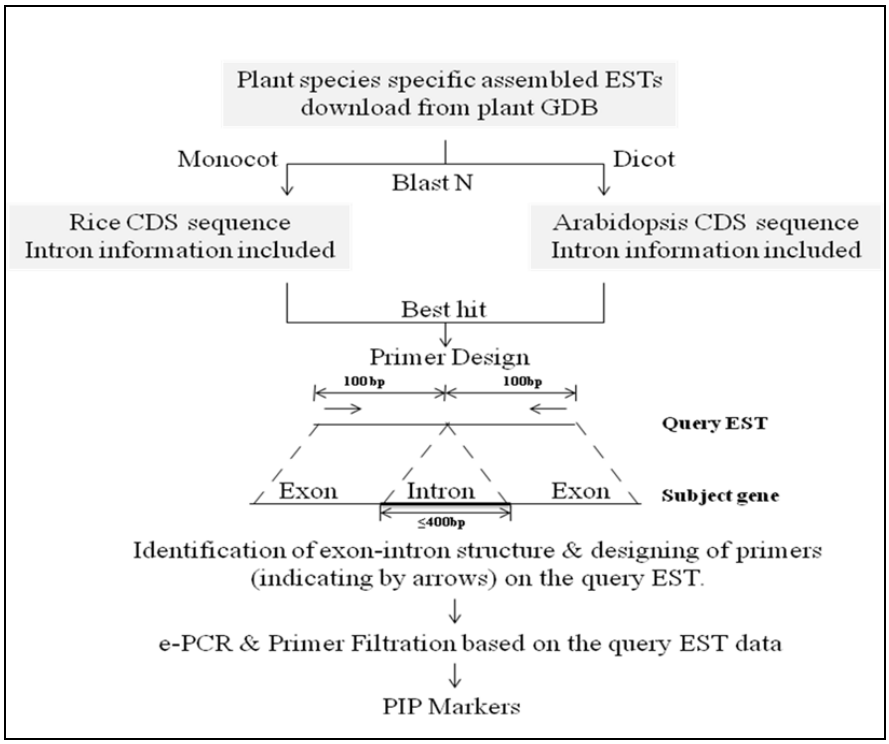


Figure 8. Flowchart demonstrating development of intron-flanking EST-specific PCR primers (PIP molecular markers).

The genomic DNA extracted from the 20 genotypes of *A. annua* (described above) were amplified using touchdown PCR using amplification conditions: initial denaturation temperature at 94 ^{o}C for 4 min, followed by 10 cycles of denaturation at 94 ^{o}C for 45 s, primer annealing at 59 ^{o}C minus 0.3 ^{o}C/cycle for 30 s and primer extension at 72 ^{o}C for 1 min. In the next 20 cycles, the annealing temperature was reduced to 56 ^{o}C for 30 s while the denaturation and primer extension time remained the same as in first 10 cycles. The last cycle consisted of only primer extension at 72 ^{o}C for 5 min. The amplification products were separated in 3% metaphor-agarose gel electrophoresis, fragments were scored as a binary unit character (present = 1; absent = 0). Only unambiguously resolved and reproducible bands were scored. The scoring matrix was used for cluster analysis as well as genetic diversity analysis as described above.

Table 6. List of primers used for ILP study, there GC content, melting temperature (Tm) and polymorphism.

Sl. No.	Primer sequence	Length of primer	% GC	Tm	Total no. of loci	NPL	PPL
1L	3' TCGTTTGGGCATTGGGATCATGC 5'	24	54	58	2	2	100
1R	5' TTGAGACGTGGTCTAACTGCCA 3'	22	50	54			
2L	3' ATTTCAGTTTTGGCCTCGTG 5'	20	45	49	2	1	50
2R	5' CGTATCACCGCGAGATTCTT 3'	20	50	51			
3L	3' CTAAGGATGCCTAACCGCCGCA 5'	22	59.10	58	2	2	100
3R	5' AGCTTGAGCCTACCTTCGCACAC 3'	23	56.52	58			
4L	3' AGGCTCAAGCTAAGATTCTGCTCG 5'	24	50	57	Not amplified		
4R	5' GAGGATCGCTAATCTCGTCTGG 3'	22	54.55	56			
5L	3' TGACACCTGCCCGGTGCTGGAA 5'	22	63.63	60	2	2	100
5R	5' TGATCGGTGCGGCGGTTAGGCA 3'	22	63.63	60			
6L	3' TGGGGAAAATCAGCCTGTTA 5'	20	40.90	49	3	3	100
6R	5' ACGAAAGTCGGCCTTAGTGA 3'	20	50	51			
7L	3' GCCTTTGCACTCGAGGGCCAA 5'	21	61.90	58	4	4	100
7R	5' GCCATCAGTGAGATACCACTCTGG 3'	24	54.17	58			
8L	3' AACCGTGGCTCCAAAGCTCTCAGC 5'	24	58.33	60	2	2	100
8R	5' CGCACTTGGTTTAGCCAACGTGCA 3'	24	50	58			
9L	3' CTCGTATTCTGCACCCATGA 5'	20	50	51	3	3	100
9R	5' TTGCACCCTCCACTACCTTT 3'	20	50	51			
10L	3' TACCAGACGGGATTGGGGTTCCA 5'	23	56.52	58	3	3	100
10R	5' GCAACATGGGAGCCGAACCCTCCA 3'	24	62.50	62			
11L	3' TCCAACCTCAAGTGGAACAC 5'	20	50	51	2	2	100
11R	5' GCATTAGATGGTCTCCTTGACGGA 3'	24	50	57			
12L	3' CTCAACAGTGATCAAGACCTCGTG 5'	24	50	57	2	2	100
12R	5' TGATAGACCGAGCTGTTTCCGGTA 3'	24	50	57			
13L	3' TCTGCAGGATTAGGGATGGCTGTG 5'	24	54.17	58	2	2	100
13R	5' CCATCACATATAGCTGCAACCGTA 3'	24	45.83	55			
14L	3' GGAGTCAACCCAGCTTCCGCTA 5'	22	59	58	2	2	100
14R	5' TGGAGACCGCTAGTGCTTCCAGAC 3'	24	58.33	60			
15L	3' AGAGAATGGTCTAATGGCTTGC 5'	22	45.45	52	Not amplified		
15R	5' TGATAGACCGAGCTGTTTCCGGTA 3'	24	50	57			
16L	3' TCTGCAGGATTAGGGATGGCTGTG 5'	24	54.17	58	Not amplified		
16R	5' CCATCACATATAGCTGCAACCGTA 3'	24	45.80	55			
17L	3' GCCATGCAACTCTGAACGGCTG 5'	22	59	58	2	2	100
17R	5' TGATAGACCGAGCTGTTTCCGGTA 3'	24	50	57			
18L	3' GTTTTGGCACGGGTCATAGC 5'	20	55	53	2	2	100
18R	5' ACAATGTCGATCCAGAACCAC 3'	21	47.60	52			
19L	3' AGAGTCCTGTGTCTCAACCATCGA 5'	24	50	57	3	2	66.7
19R	5' TTAGTTCCGGGGTCCACTAACGCA 3'	24	54.16	58			
20L	3' ACATTGTCTTGGCCGGATAG 5'	20	50	51	3	2	66.7
20R	5' TATTACACCCCTTGCCCAAA 3'	20	54.16	49			
21L	3' GAACTCCGGTCATTGTACCACCAG 5'	24	58.33	58	2	2	100
21R	5' AGGATGGCCCGACATGCGTTTACC 3'	24	50	60			
22L	3' TGCTTGATGCTTCTCACACG 5'	20	55	51	3	3	100
22R	5' GGCTCGATCTTTGCAAGCTAGG 3'	22	56.52	56			
23L	3' CTCTAAAAGCACCTCCGTGCAGG 5'	23	52.17	58	3	3	100
23R	5' TACGCCTAAGAGCACCTTCAGGA 3'	23	54.16	56			
24L	3' GTCGGGTTCTTTGAACGTCAGCGA 5'	24	56.52	58	4	3	75
24R	5' TGCGGCCATTGTATTCCGGCCAA 3'	23	52.17	58			
25L	3' ACGAAGAACCGTAAAGACCCAC 5'	22	54.17	54	3	2	66.7
25R	5' AAGATCACGGTTGACTCTGCCA 3'	22	56.52	54			
26L	3' GCGTATTCTGATGGTTGGTCTCGA 5'	24	50	57	3	3	100
26R	5' TTGGATTGCGCCTCGTCACCAA 3'	22	54.54	56			
27L	3' GCTAACAAGGCCTAAAACGGACTC 5'	24	50	57	2	2	100

27R	5' CTTGGAGCATCATCACCAGCAA 3'	22	50	54			
28L	3' GATGCTTACGTTGGTGACGAAGCA 5'	24	50	57	2	2	100
28R	5' CAACATCCCACACAGTGAAGCTG 3'	23	52.17	56			
29L	3' TTTGTGGTCGACAGCAATGACC 5'	22	50	54	2	2	100
29R	5' TGCAGTGTTGCTTGTTTTTGCCA 3'	24	41.67	53			
30L	3'GGGTCAACGTTCTTTGGGCACTG 5'	23	56.52	58	2	1	50
30R	5'TGTGGGTGGAAATCTGCCCAAACC 3'	24	54.17	58			
31L	3'TGACGACCTCTCGTGTATTGTT 5'	22	45.45	52	4	4	100
31R	5' CCATTTATTTCCCCGTGTCA 3'	20	45	49			
32L	3' GGTTCTTTGCACTGACACGA 5'	20	50	51	4	4	100
32R	5' TTGTACCCGTCTTGACACCA 3'	20	50	51			
33L	3' AGCTTTTGCGCAACTCAAGCAG 5'	22	50	54	Not amplified		
33R	5' ATCCGAGGGTGTGTGCACCTG 3'	21	61.90	58			
34L	3' AGCTTTTGCGCAACTCAAGCAG 5'	22	50	54	2	2	100
34R	5' GAACATCAGGAAAGATTGTGCTGC 3'	24	45.83	55			
35L	3' TTTGTTCAGGGGTGTGATGC 5'	20	50	51	3	3	100
35R	5' CTCAACGTTCTTTGGGCACTG 3'	21	52.38	54			
36L	3' TAGCTTTTGCGCAACTCAAG 5'	20	45	49	2	2	100
36R	5'TTGATATCATTGAGCGAGATTG 3'	22	36.36	49			
37L	3' TTGAAAACTGGTTGCAATGG 5'	20	40	47	2	2	100
37R	5'TGTGCCATCATATTCAGTGGA 3'	21	42.86	50			
38L	3' AGCTTTTGCGCAACTCAAGCAG 5'	22	50	54	Not amplified		
38R	5' ATCCGAGGGTGTGTGCACCTG 3'	21	61.90	58			
39L	3' AGCTTTTGCGCAACTCAAGCAG 5'	22	50	54	2	2	100
39R	5'GAACATCAGGAAAGATTGTGCTGC 3'	24	45.83	55			
40L	3' ACCATGGATGATCTTGGTTCCCA 5'	23	47.82	55	2	2	100
40R	5'TCTGTAAGACATGGTAGCTCACGA 3'	24	45.80	55			
Total					**88**	**81**	**92.04**

Result and Discussion

Out of the 40 primers tested, 35 primers produced PCR amplification with the success rate of 87.5% (Table 6). A representative gel demonstrating the successful amplification of introns from the genomic DNA isolated from different genotypes is shown in Figure 9. Genotypes collected from two different valleys were clearly distinguished and separated—revealed existence of higher genetic variations. All the amplified fragments varied in sizes from 200–500 bp. Out of 88 amplified bands, 81 (92.04%) were found polymorphic. The dendrogram was generated by bootstrapping and NJ cluster analysis considering 88 ILP bands amplified by 35 primers. Genotypes from both Indus and Nubra valleys were distinctly separated from each other with respect to both the valleys (Figure 10a).

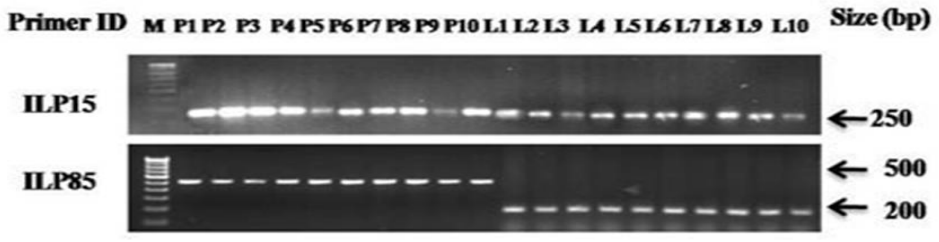


Figure 9. ILP amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10 are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = 100 bp λ DNA ladder.

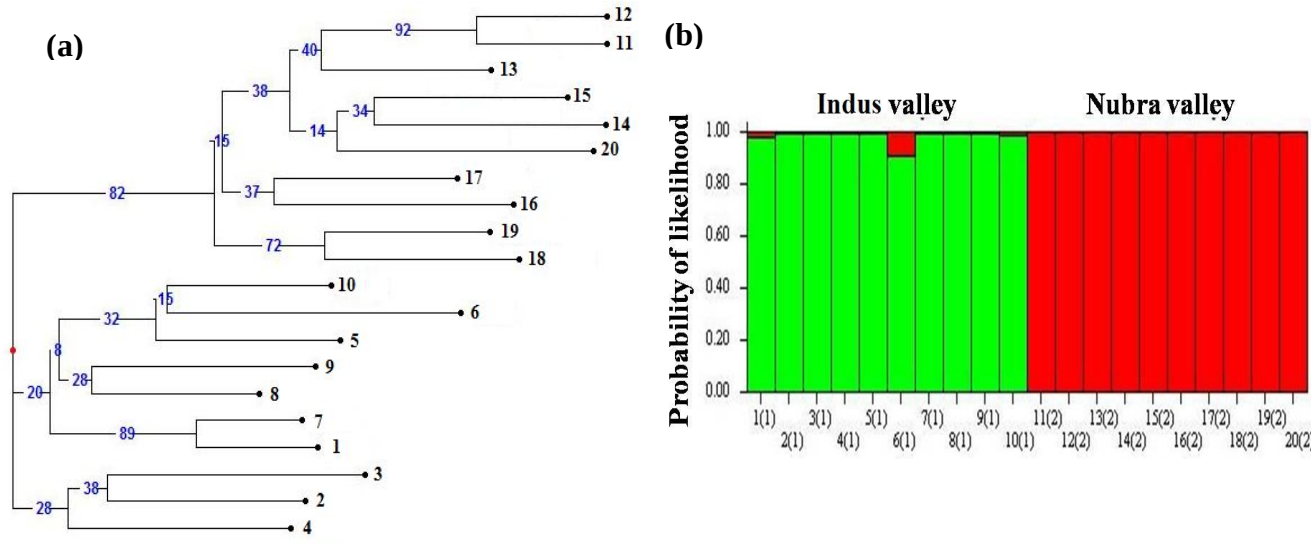


Figure 10. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on PIP markers, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites. The genotypes were more likely clustered corresponding to both the valleys. The value within bracket represents the different valley (1, Indus valley and 2, Nubra valley). Genotypes from both the valleys are represented with different colours: Indus valley (green) and Nubra valley (red).

Based on STRUCTURE analysis all the 20 genotypes were more likely distributed among two valleys with respect to their sampling (Figure 10b). The genotypes from Indus valley were more diverse in comparison to Nubra valley as mentioned in Table 7. Molecular variance among 20 genotypes was found maximum (61%) in comparison to between valleys (39%).

Table 7. Summary of genetic variation statistics for all loci of ILPs among the *Artemisia annua* genotypes with respect to their distributions among two valleys.

S.No.	Sample size	Na	Ne	H	I	Ht	NPL	PPL
Indus	10	1.7500 (0.4359)	1.5141 (0.3583)	0.2958 (0.1899)	0.4343 (0.2688)	0.2958 (0.0361)	57	75
Nubra	10	1.6184 (0.4890)	1.5171 (0.4249)	0.2792 (0.2242)	0.3975 (0.3171)	0.2792 (0.0502)	47	61.84
Mean		1.6842	1.5156	0.2875	0.4159	0.2875	52	68.72

Genetic characterization of *A. annua* genotypes using EST-derived SSRs

Microsatellites or simple sequence repeats (SSRs) DNA-based markers are short tandem repeats of 1 to 6 bp in length, evenly distributed in the genomes and small locus size [17]. The multiallelic nature, codominant inheritance and relative abundance of SSRs in the genomes facilitate high reproducible genotyping using polymerase chain reaction (PCR) [18]. Microsatellites provide high levels of polymorphism among the genotypes. The SSRs reside within the transcribed sequences

can be identified from ESTs [19]. These EST-SSRs can be rapidly developed from EST sequences by using data mining techniques. Due to their localization at the transcribed region of the genome, EST-SSRs could be used in developing gene-based maps that may help in identifying candidate genes and increase the efficiency of marker-assisted selection [20]. Flanking sequences to specific SSR loci are generally conserved within a particular species, within a genus and sometimes even across related genera which makes ease to design primers for individual SSR loci.

Data mining for SSR marker

The entire available ESTs sequences (~85,282) of *A. annua* were downloaded as FASTA format from the NCBI database (dbEST). After removing the redundant ESTs the total numbers of ESTs left were 68,974. A perl script known as MIcroSAtellite [21] was used to mine microsatellites. In the present study, SSRs were considered to contain motif one to six nucleotide in sizes. Frequency of SSRs refers to kilo-base pair of EST sequences containing one SSR.

PCR primer design and PCR amplification

A set of 16 pairs of primers (Table 8) were designed by considering only the ESTs from the gene encoding enzymes for the artemisinin biosynthetic pathway. Primer-5 program was used for designing of primers based on following core criteria: (1) melting temperature (Tm) between 57 ^{o}C and 65 ^{o}C with 60 ^{o}C as optimum; (2) product size ranging from 100 bp to 300 bp; (3) primer length ranging from 18 bp to 24 bp with amplification rate larger than 80 %; (4) GC % content between 30 % and 70 %. All these primer pairs are custom synthesized from GBiosciences (USA). Polymerase chain reactions (PCRs) were performed in a Biometra termocycler. The optimized PCR amplifying conditions used were: denaturation, 1 cycle of 3 min at 94 ^{o}C, an annealing temperature of 55 ^{o}C for 35 cycles (45 s at 94 ^{o}C, 30 s at 55 ^{o}C, 1 min at 72 ^{o}C) and an additional cycle of 7 min at 72 ^{o}C. The amplified fragments were separated on 6% metaphore IV agarose gels. The banding patterns were scanned, scored and data collected from reproducible bands. Band sizes were estimated by comparing with the molecular mass standards included in each gel. The numbers of alleles detected and amplified by the SSR markers were scored as present (1) or absent (0), each of which was treated as an independent character and used for genetic diversity analysis. The resulting matrix was used to estimate genetic dissimilarity among all genotypes and used for cluster analysis.

Table 8. List of primers used for SSR amplification, GC content, total number of loci, the level of polymorphism and PIC value.

Primer	Primer Sequence (5' – 3')	GC (%)	Tm (^{0}C)	Allele number	Product size (bp)	PIC
1 F	ATAATACGCATGAGCTGGTTAG	40.9	56.5	2	90-380	0.245
1 R	CCACTACCAATCACAATAACAG	40.9	56.5			
2 F	GAATTGAGATTGTGGTCCTTAG	40.9	56.5	2	200-400	0.444
2 R	GGTTGCTAAGAATGTCGATTG	42.9	55.9			
4 F	ATCGTATTACCTTGGTCACATC	40.9	56.5	2	300-330	0.245
4 R	TGTCATACTGACTTACACAGGG	45.5	58.4			
5 F	TAAGCCAAAGGCTCAAGTAAAC	40.9	56.5	3	180-220	0.198
5 R	GGATTGCTCATCTAGTGCTTAT	40.9	56.5			
6 F	GCATGCATTTATGTTGGATCAC	40.9	56.5	2	100-400	0.565
6 R	CAGCAGCAACAACAACAACAG	47.6	57.9			
7 F	GGAACAGATGATCTATATGCCT	40.9	56.5	2	420-450	0.340
7 R	GCATACTATGTGCAAGGTCTAGT	43.5	58.9			
8 F	TGGTAGAACTCCACCTACTAACT	43.5	58.9	2	500-550	0.426
8 R	TATAATAGTTGGGTGGTTCCTC	40.9	56.5			
9 F	GAGAAAGAGAAAGCCAAACAC	42.9	55.9	1	220	0
9 R	TAGCTCCATAGATCTCAAACCT	40.9	56.5			
10 F	GGATCATTAAGTTACGCTCCT	42.9	55.9	2	300-350	0.495
10 R	CCATGCTTTATGTTGTAGAGTG	40.9	56.5			
13 F	GTAAGTTATACCTGGTTTCCAGC	43.5	58.9	4	300-600	0.595
13 R	ACCACTACACCTTGCATTCTA	42.9	55.9			
14 F	CTCTCTTCCTCTTTGTGTGTCT	45.5	58.4	4	150-500	0.503
14 R	CAAGATGGTACGAATACTGTTG	40.9	56.5			
15 F	CGAGCAATCGGAGAGTTAGC	55.0	59.4	7	250-550	0.802
15 R	ATGCATCTCGCGAATCTTCT	45.0	55.3			
16 F	GTGTGAGGCCTCTGCTCTG	63.2	61.0	5	120-480	0.716
16 R	ACCGCCATGTCTTCTCCATA	50.0	57.3			
Total				38		

Result and Discussion

Searching for ESTs containing microsatellites

A total of 68,974 non-redundant ESTs from *A. annua* were used to evaluate the presence of SSR motifs. A complete search of dinucleotide, trinucleotide, tetranuleotide, pentanucleotide and hexanucleotide microsatellite revealed in the identification of 4,342 (6.3%) unique ESTs containing microsatellites. Out of these, 499 (11.49%) ESTs contained more than one SSR and 40 (0.92%) were compound SSRs that have more than one repeat type.

Frequency of SSRs based on motif sizes

We examined 68,974 *A. annua* ESTs comprising of 44.12 Mb DNA sequence. The frequency of occurrence for SSRs was one SSR in every 8.943 kb of EST sequences. Analysis of SSRs motifs revealed that the proportions of SSR motif sizes were not evenly distributed. Among the SSRs detected, the dinucleotides are with highest frequency of 52.19% followed by trinucleotides (42.39%). The tetra nucleotide SSRs has a much lower frequency (3.6%). The frequency of SSRs

with motif sizes of five and six are 0.6% and 1.20% respectively. The mean SSR length of each motif varied between 10 and 78 bp. The overall average SSR length was 20 bp with a maximum of 78 bp.

Distribution of SSRs based on motif types

The SSRs identified in the present study were characterised by 77 types of motifs. In general, the SSRs were found to be unevenly distributed across motif types. The motif AC/GT had the highest frequency of 23.23% followed by AT/AT (18.48%), AG/CT (9.8%), ACC/GGT (8.89%), AAC/GTT (6.89%), AAG/CTT (6.38%), AAT/ATT (5.5%) and AGT/ATC (5.57%). The other types of motifs possessed a frequency of < 5 %.

EST-derived SSRs polymorphism

DNA samples isolated from 20 genotypes of *A. annua* as templates, polymorphic DNA fragments were amplified from 13 out of 16 SSR primer pairs selected in the study. The sizes of these fragments ranged from 90 to 600 bp (Figure 11). A total of 38 alleles with the average alleles per locus of 2.92 were detected at 13 loci, with the polymorphic markers ratio of 92%. Genetic relationships were found to be very close among the genotypes from the Nubra valley in comparison to Indus valley. The genotypes from both the valleys were clustered distinctly from each other (Figure 12a). STRUCTURE analysis also predicted better distribution of genotypes with 2 major groups (Figure 12b). Genetic diversity analysis in terms of Na, Ne, H, I, Ht, NPL and PPL among both the valleys revealed higher values for Indus valley indicating more variability among the genotypes in comparison to Nubra valley (Table 9).

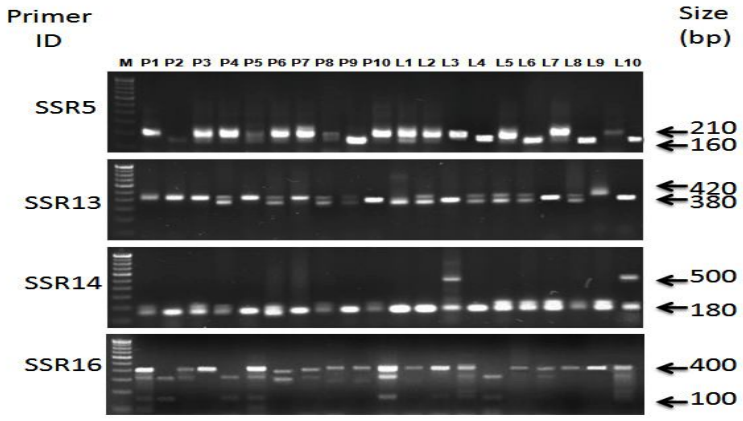


Figure 11. SSR amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = the size of molecular markers in base pairs using λ DNA.

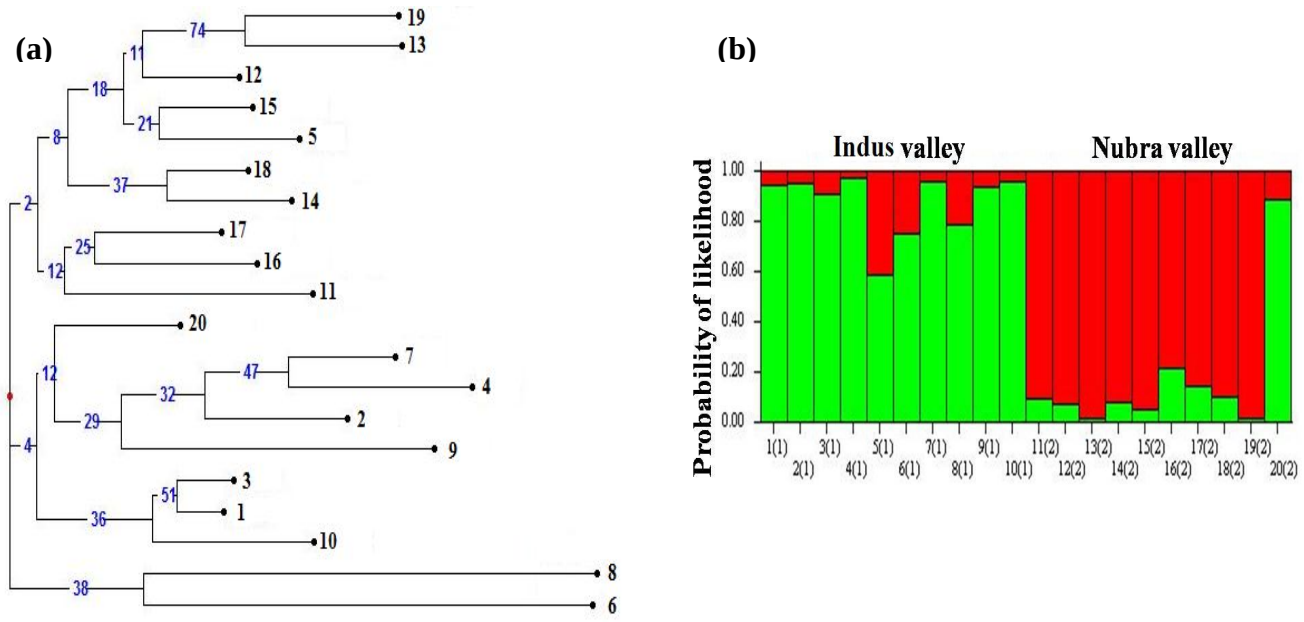


Figure 12. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on EST-derived SSRs marker, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites. The genotypes were more likely clustered corresponding to both the valleys. The value within bracket represents the different valley (1, Indus valley and 2, Nubra valley). Genotypes from both the valleys are represented with different colours: Indus valley (green) and Nubra valley (red).

Molecular variance among both the valley is predicted to be 23% and among 20 genotypes is 77% that revealed higher variation in *A. annua* genotypes because of high genome polymorphism. In summary, the successful experimental validation of the majority of the computationally predicted SSR motifs confirms the utility of mining 68,974-ESTs for genetic markers. The development of EST-derived SSR markers in *Artemisia annua* should greatly facilitate marker-assisted selection, germplasm breeding, adulterated species identification and genetic diversity studies in this medicinal species.

Table 9. 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
Nubra	10	1.5676 (0.5022	1.3544 (0.3645)	0.2097 (0.2008)	0.3127 (0.2904)	0.2097 (0.0403)	21	56.76
Indus	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	24	64.86

Genetic variation among elite genotypes of *A. annua* based on intron-flanking EST-specific and EST-derived SSR markers analysis could be useful in selecting parents to be crossed for

generating appropriate populations intended for both genome mapping and breeding purposes. Significantly, this study demonstrates that *Arabidopsis* genome sequence can be useful in developing gene-specific PCR-based markers for other non-model plant species like *A. annua* in the absence of genome sequences.

Genetic polymorphism (percentage of polymorphic loci) measured for the 20 genotypes of *A. annua* collected from two valleys using different molecular markers are found to be maximum using RAPD (96.9%), followed by ISSR (86.1%), PIP markers (68.42%) and EST-derived SSRs markers (64.86%). The variation in measuring rate of polymorphism was due to different portions of the genomic DNA amplified by these markers. Since no single, or even a few plants, will represent the whole genetic variability in *A. annua*, there appears to be a need to maintain sufficiently large populations in natural habitats to conserve genetic diversity in *A. annua* to avoid genetic erosion. Based on polymorphic feature, genetic diversity, genetic similarity, and gene flow among the populations of *Artemisia* based on RAPD, ISSR, intron-flanking EST-specific and EST-derived SSRs studies, 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.

Aim 3: Extraction and quantification of Artemisinin at different developmental stages from *Artemisia annua* L.

Presently the biotechnological approach for the commercial production of artemisinin remains disappointing, and the molecule must therefore still be extracted from *A. annua* plant grown outdoors. However, the range of artemisinin content of *A. annua* is wide and can be affected by numerous factors such as geographical conditions, harvesting time, temperature, plant variety, developmental stages of plants and extraction methods and so on. Artemisinin compounds have been predominantly found in the upper part of the *Artemisia annua* plant, with the concentration of artemisinin said to peak just before or during full flowering. The artemisinin content depends on the age of leaf, since in older leaves the glands were often found to be ruptured.

The objective of the present study was to understand the dynamics of artemisinin content in the developmental stages of plants as well as in different plant parts to facilitate artemisinin production in the field. The other objective is to use different methods of quantification of artemisinin so that a reliable method can be established for the rapid screening of cultivars for better artemisinin yield.

Plant materials and extraction of artemisinin

Leaves and flowers were collected at different developmental stages of *A. annua* during the month of March to October. All samples were collected in triplicate from Indus (3500 m) and Nubra (3000 m) valley of Ladakh region from their natural habitats and dried at 60 ^{0}C for 24 hrs in an oven and used for artemisinin estimations. Artemisinin standard solutions were prepared with concentration of 1.0 mg/ml in acetonitrile and stored at −20 ^{0}C. The standard working solutions used to generate the calibration curve were prepared by serial dilutions of a stock solution with acetonitrile.

About 100 mg dry weights of samples were taken and artemisinin was extracted with 1 ml of ethyl acetate in 2 ml centrifuge tube and vortex for 2 min at room temperature. Samples were centrifuged at 10,000 rpm for 2 min and the supernatant was transferred into a fresh 2 ml centrifuge tube and evaporated to dryness under water bath condition at 50 ^{0}C temperature. The residual was then dissolved with 1 ml of acetonitrile and extracted solution was filtered through 0.2 µm size nylon Millex-GN filters (Millipore, Bedford, MA, USA), pre-wetted with acetonitrile, and injected through disposable 3 ml syringes. Filtered aliquots from the samples were analyzed in the same day.

Quantification of artemisinin

Detection and quantification of artemisinin from *Artemisia annua* were carried out using Waters series (Waters Corporation, USA) reverse phase high performance liquid chromatography coupled with 2996 photodiode array detector (RP-HPLC-PDA) with the data collection and processing through the Empower 2 software, Delta 600 gradient quaternary pump equipped with a in-line degasser AF, 600 controller, 2707 programmed autosampler, thermostat column oven maintained at 30 ^{0}C and C18 column (5 µm; 250 mm x 4.6 mm). Alternatively artemisinin content also carried out using GC-FID. Nitrogen was the carrier gas with a column flow rate of 2mL/min, a split ratio of 3:1. The column was a HP-5 crossbond 95% dimethylpolysiloxane (Agilent) capillary column. Injector temperature was set at 240 ^{0}C; detector temperature set at 300 ^{0}C, oven temperature was programmed to start at 180 ^{0}C (1min), increasing 0.8 ^{0}C /min to 198 ^{0}C, then increasing 30 ^{0}C /min to 280 ^{0}C, then holding at that temperature for 10 min. The retention time of artemisinin was 7.629 min. Furthermore, artemisinin was estimated by spectrophotometer. This method is based on the reaction of hydrogen peroxide (H_2O_2), generated by the cleavage of endoperoxide linkage of artemisinin and its reaction with potassium iodide to liberate iodine. The liberated iodine bleaches the red colour safranin O to colourless species and is measured at 521 nm.

Result and Discussion

The dynamic variation of artemisinin content of leaf with respect to growth stages of plants is shown in Figure 13a. From the first month (March), the seedling stage to the eight month (15th October), the full flowering stage, the artemisinin content increased gradually. The highest content of artemisinin was found in 6-month old plants (at flower budding stage) and afterwards the artemisinin content was maintained more or less constant at this level (Figure 13b). Leaf samples from the Indus valley were having high artemisinin content (0.441% dry weight), whereas the samples from the Nubra valley consists of slightly less artemisinin content (0.356% dry weight).

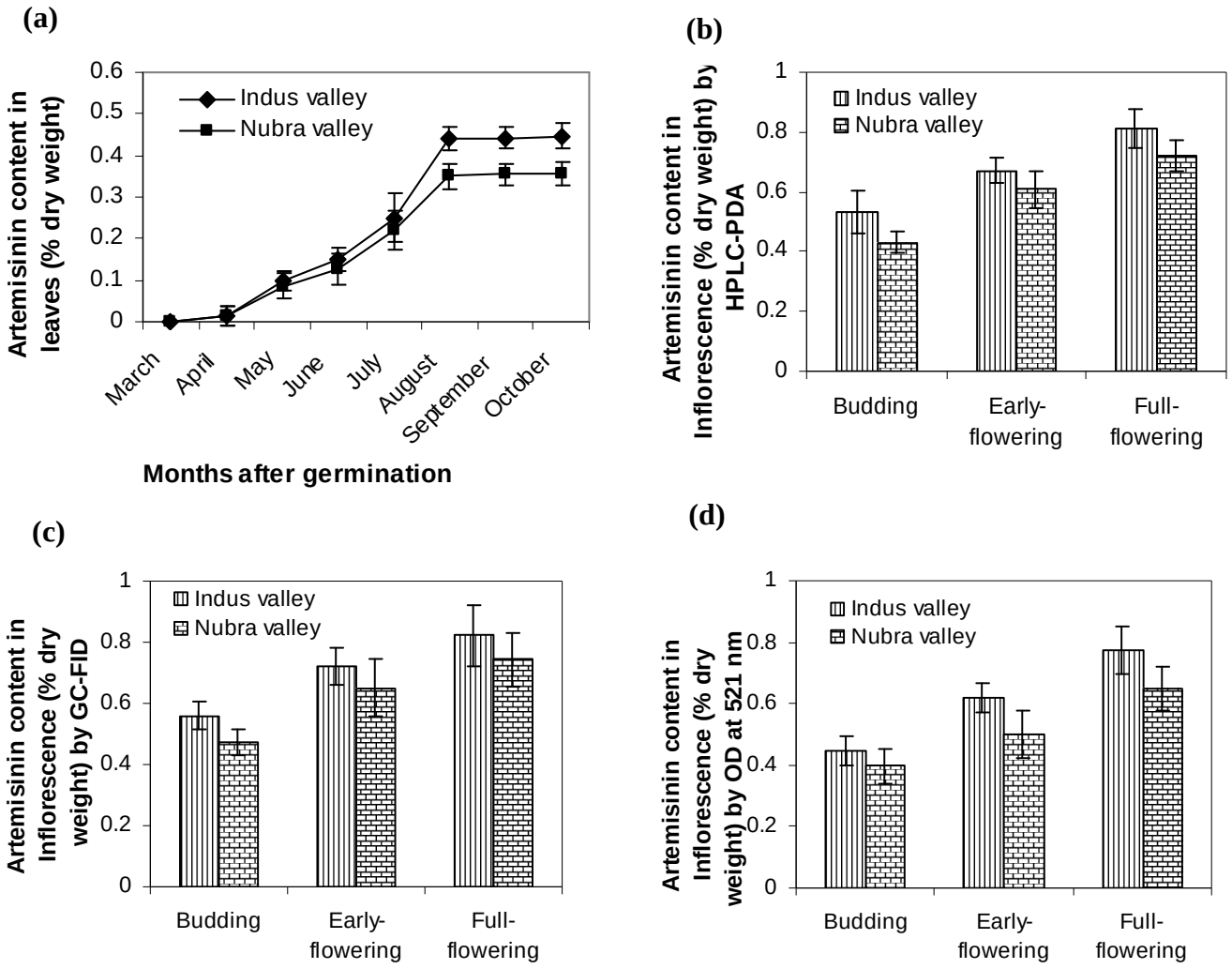


Figure 13(a-d) Dynamic variation of artemisinin content of leaves and inflorescence at different growing stages and reproductive stages of *Artemisia annua* by HPLC-PDA, GC-FID and Spectrophotometer methods.

The artemisinin content of *A. annua* was largely influenced by the development stage of the plants and there have been different reports regarding the optimal stages at which the highest production of artemisinin found. From this study, one may reasonably think that the highest artemisinin content is

closely related to reproductive growth. The artemisinin contents of inflorescence at three reproductive stages: flower budding stage, before flowering (30 days after budding) and at full flowering stage were analyzed in samples taken from both Indus and Nubra valley. The value of artemisinin content was found to be very similar to HPLC analysis using GC-FID. However, the values were found to be slightly lower based on spectrophotometric method (Figure 13c-d). The artemisinin content in inflorescence increased with the developmental stages and attained its maximum at full flowering stage.

The importance of investigating the dynamic variation of artemisinin content of the leaf and inflorescence at different stages of vegetative and reproductive growth is to define the appropriate time for harvesting. According to the result of the present study, the optimal time at which leaves should be harvested is the time when flower buds start to form. Similarly full-flowering stage is the best time of harvesting inflorescence for optimal yield of artemisinin.

Aim 4: Impact of soil nutrient and environmental factors on artemisinin content among *Artemisia annua* genotypes of Ladakh region using linear and non-linear approach

Presently the commercial source of artemisinin is *A. annua* for the pharmaceutical industry. However, the relatively low yield of artemisinin in *A. annua* plants (~ 0.6% based on dry weight) and lack of cost effective, viable synthetic protocol are the stumbling block for the commercialization of the drug [22]. Therefore, enhanced production of artemisinin in the whole plant of *A. annua* is still the only valid source. A number of investigations have demonstrated that the quality and quantity of several secondary metabolites have close relationship with plant habitats [23]. The shift of nutrition supplies in soil definitely leads to the alteration of both primary and secondary metabolism and consequently in production of secondary metabolites. However, in addition to soil factors the effect of climatic factors on artemisinin content at natural habitats cannot be ruled out. This is obvious as the plant is quite adaptable to a wide range of environmental conditions. Hence, it demands management of soil nutrients and optimization of climatic factors for the successful domestication of *A. annua* in the Ladakh region. Statistical methods such as artificial neural network (ANN) and multiple linear regressions (MLR) are very useful in this respect.

Methodology

Samples of *Artemisia annua* were collected from two valleys: Indus (3500m) and Nubra (3000m) from the trans-Himalayan of Ladakh region, India. Samples were obtained in triplicate every 15 days after germination to the stage of harvesting (~210 days). The plant emerges in the month of

March and dies in the month of October in the Ladakh region. Whole plant samples (consisting of leaves and inflorescence) were collected and were dried separately at 50 ^{0}C for 24 hrs in an oven and used for artemisinin estimations. Soil samples (root layer) from each sampling site were collected in triplicate during collecting the plant samples and brought to the laboratory for analysis of soil nutrients. From each sampling site the average data of every 3 days interval on environmental factors such as temperature and humidity were taken from March to October, 2010 and were properly documented.

The relationship between soil nutrients as well as environmental factors with artemisinin content was examined based on multiple linear regression method, using MINITAB statistical package. Further a back propagation neural-network model was created using STATISTICA (version 9.9), trained with the environmental factors and soil nutrition parameters as the inputs and the measured artemisinin reading as the output. The topological structure of this neural network model consisted of 12 input neurons in the input layer and one output neuron in the output layer to match the 12-1 input–output pattern of the training data set. One hidden layer with five neurons was the optimal topology for the neural-network model determined by a trial-and-error method.

Result and discussion

Artemisinin content estimated from the plant samples after seedling till maturity varies from 0.089 ± 0.02% to 0.666 ± 0.008% (Figure 14a). Artemisinin content was different from the samples collected from both the valleys: samples from Indus valley having high yield in comparison to Nubra valley. The variation in artemisinin content of the plants also very much related to soil nutrients as well as climatic factors in the Ladakh region (Figure 14 (b-f)). The results obtained from this study revealed that artemisinin production of *Artemisia annua* or other bioactive compound production from herbs cultivated in farms can be improved through the soil management to mimic the soil condition similar to their original ones. Our work demonstrated that high soil organic matter and nitrogen favors high artemisinin yield of *A. annua* (Figure 14b & 14c), whereas the soil pH, total potassium and total phosphorous content negatively correlated with the artemisinin content (Figure 14d-f). The temperature (both maximum and minimum) also significantly correlated with the artemisinin content, whereas the percentage of humidity is not significantly correlated (Figure 15a-d).

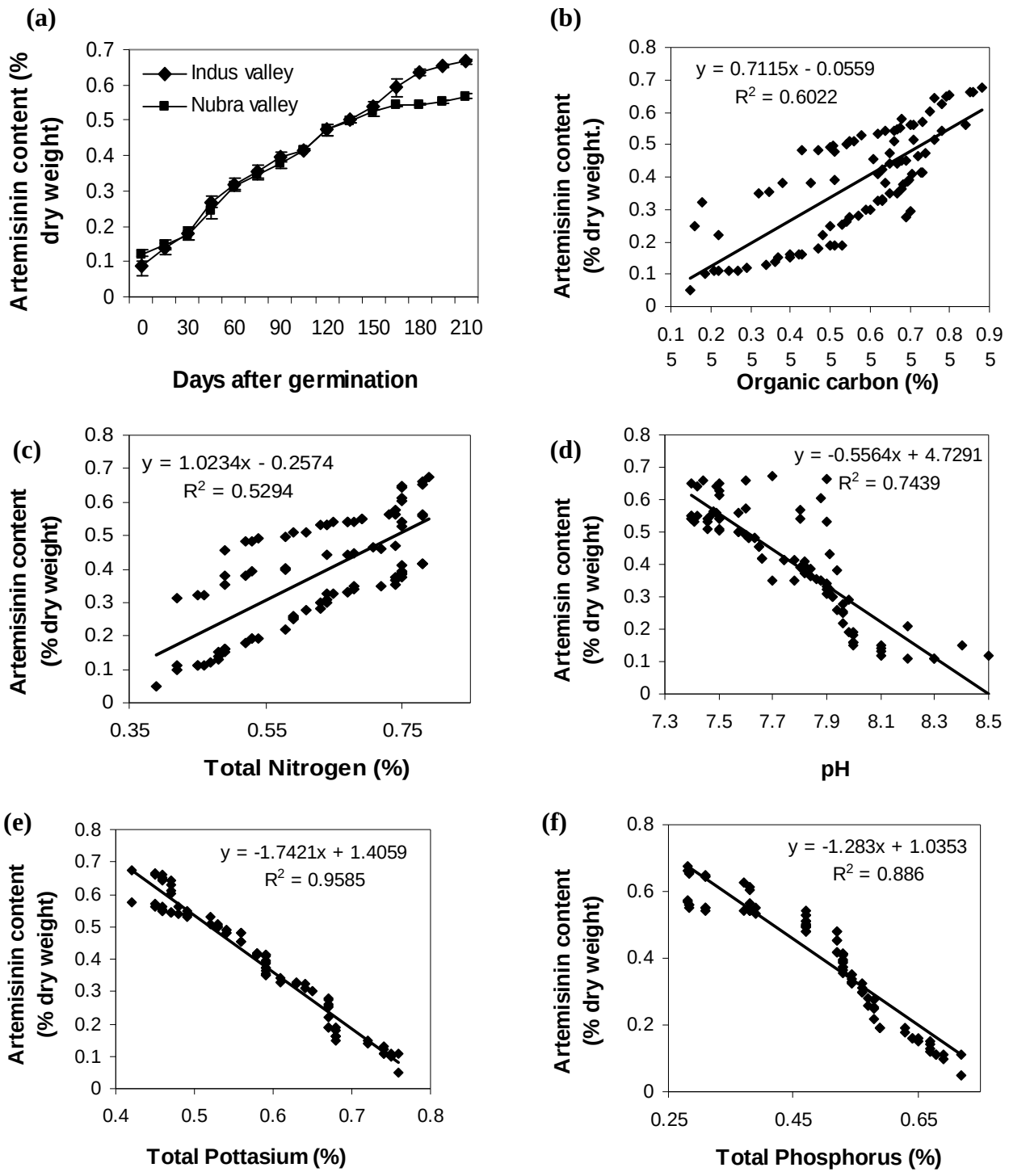


Figure 14 (a). Artemisinin content of plant samples (leaves and inflorescence) collected from both Indus and Nubra valleys. The samples were obtained every 15 days after seedling in triplicate and represented as mean ± SD. The relationship between artemisinin content with different soil parameters such as (b) soil organic carbon, (c) soil total nitrogen, (d) soil pH, (e) soil total potassium and (f) soil total phosphorus are also represented as scatter plot.

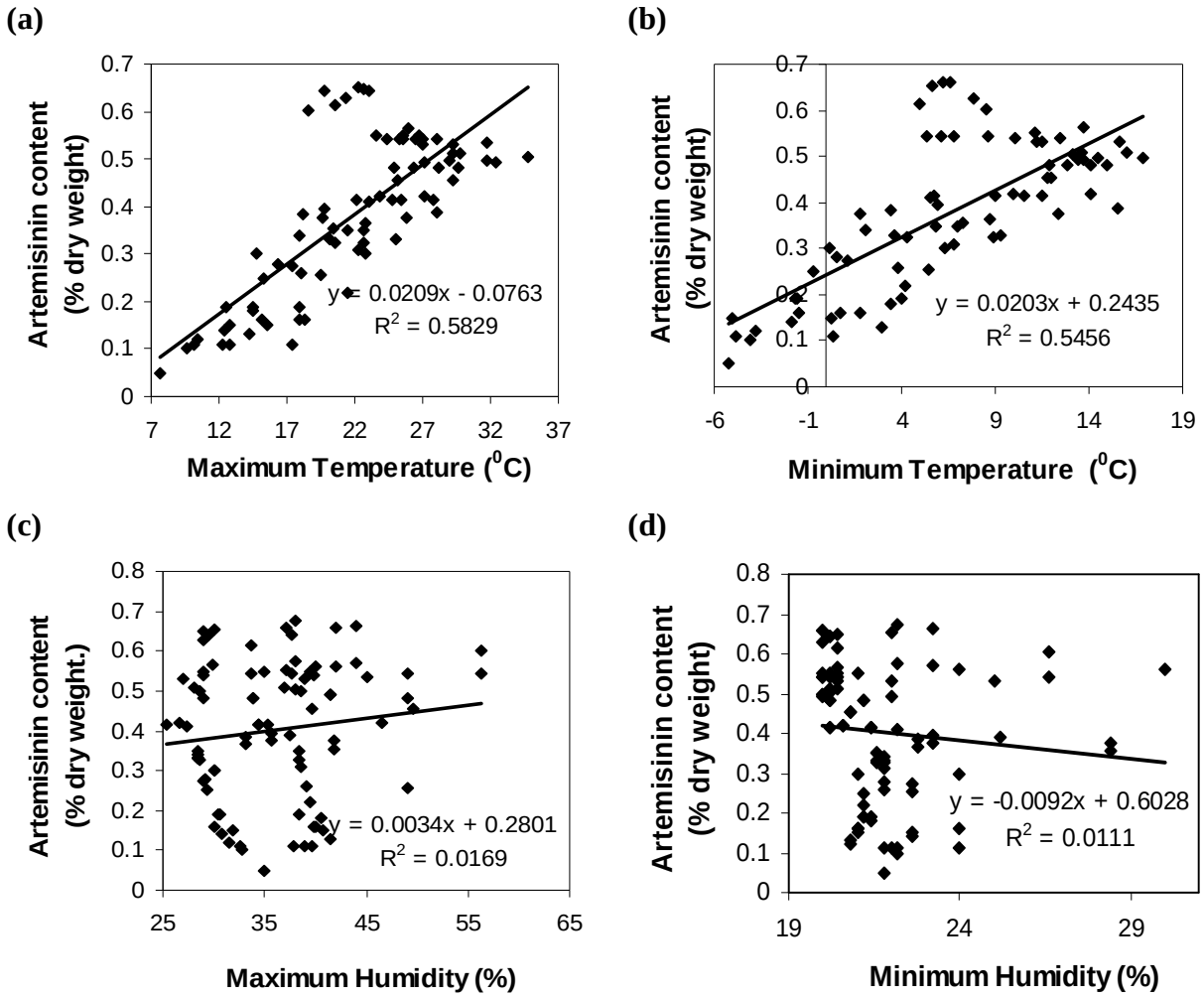


Figure 15. The relationship between artemisinin content with different climatic factors such as (a) maximum temperature, (b) minimum temperature, (c) maximum humidity, and (d) minimum humidity are represented as scatter plot.

The secondary metabolite production is influenced by the plant's own physiological age and status and other environmental factors. The prediction model like ANN and MLR developed in this study to map the effects of these factors on artemisinin yield will be helpful up to a certain extent for commercial cultivation as well as conservation of the plant. The results showed that using a combination of soil and environmental data, we were able to successfully predict artemisinin yield with ANN and MLR (Figure 16)**.**

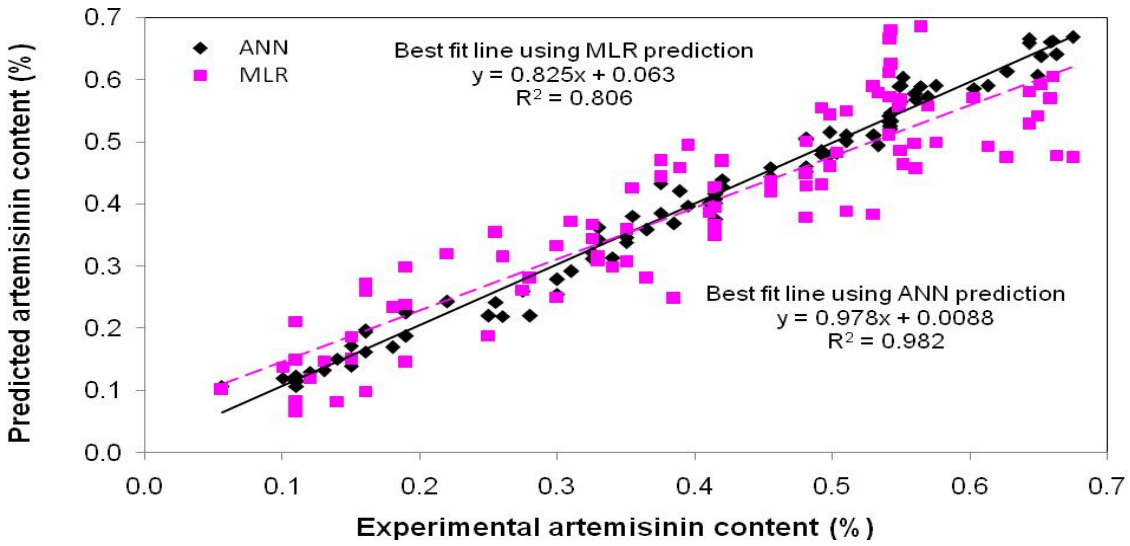


Figure 16. Comparison of estimated artemisinin content using an artificial neural-network model (—, ANN) and a 'best-fit' regression model (---, MLR). Here r^2 represent coefficient of determination.

Both the models could provide useful information regarding selection of sites, optimization of soil and environmental factors in order to increase the yield of artemisinin and thus are very important before planning any conservation strategy.

Aim 5: *Artemisia tournefortiana:* an alternative and promising source of artemisinin

In search of new antimalarial for combating the malaria worldwide, artemisinin (derived from natural *Artemisia annua* L.) and its derivatives have been recognized as a new generation of powerful antimalarial drug. World Health Organization (WHO) has recommends that all antimalarial must be coupled with an artemisinin component (artemisinin-based combination therapy; ACT) for use as the first-line treatment against malaria. Recent estimates suggest that current market for artemisinin has reached US$343 million [23] and at the current rate its supply from its known natural sources, *A. annua* and other related species [24, 25] will soon be outdone. The inadequacy of present artemisinin production at current rate to meet the ever-increasing demand by the pharmaceutical industry has generated worldwide interest in searching alternative plant species with good source of artemisinin. Ladakh, the Indian Trans-Himalayan cold desert region is the home of many medicinal plants. About 20 species of *Artemisia* were reported from this region of the Trans-Himalayan. However, only 5 species: *A. annua*, *A. tournefortiana*, *A. gmelinee*, *A. dracunculus*, and *A. sieversiana* are predominantly available (Figure 17). In order to ascertain the alternative source of artemisinin we have performed the quantitative estimation of artemisinin from these species and compared.

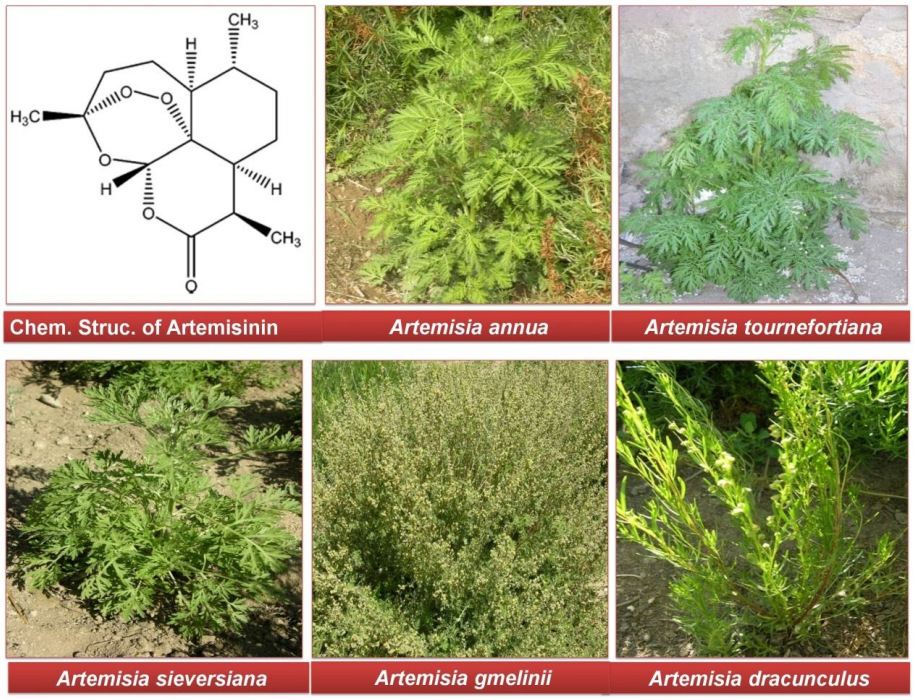


Figure 17. Five species of Artemisia: *A. annua*, *A. tournefortiana*, *A. sieversiana*, *A. gmelinee* and *A. dracunculus* that are are predominantly available in Ladakh region.

Materials and methods

Plant samples (at flowering stage) of different *Artemisia* species were collected from different location of Ladakh region from their natural habitats (altitude 3000-4000 m). These species were identified and authenticated by comparing plant descriptions with the Trans-Himalayan herbarium "Flora of Ladakh" established and maintained by Defence Institute of High Altitude Research, Indus-Ladakh, India. Extraction and quantitative estimation of artemisinin content from different parts: leaf, flower, stem and root were performed in triplicate. Detection and quantification of artemisinin from *Artemisia* species were carried out using Waters series reverse phase high performance liquid chromatography as described above.

Result and Discussion:

We report herein, *Artemisia tournefortiana* that grow in the Trans-Himalayan region of Ladakh as an alternative source of artemisinin albeit at low concentration than *A. annua*. Among the other species such as *A. dracunculus* and *A. gmelinii* artemisinin content was not detected, whereas in *A. sieversiana* only least quantity of artemisinin was detected. In case of *A. tournefortiana* artemisinin content was found to be maximum in flower (1.43%) followed by root (0.331%) (Figure 17). Further the artemisinin content in flower vary from 0.05 to 1.43% with respect to different

geographical location. Difference in artemisinin content in *A. tournefortiana* with respect to other species reported worldwide is shown in Figure 18.

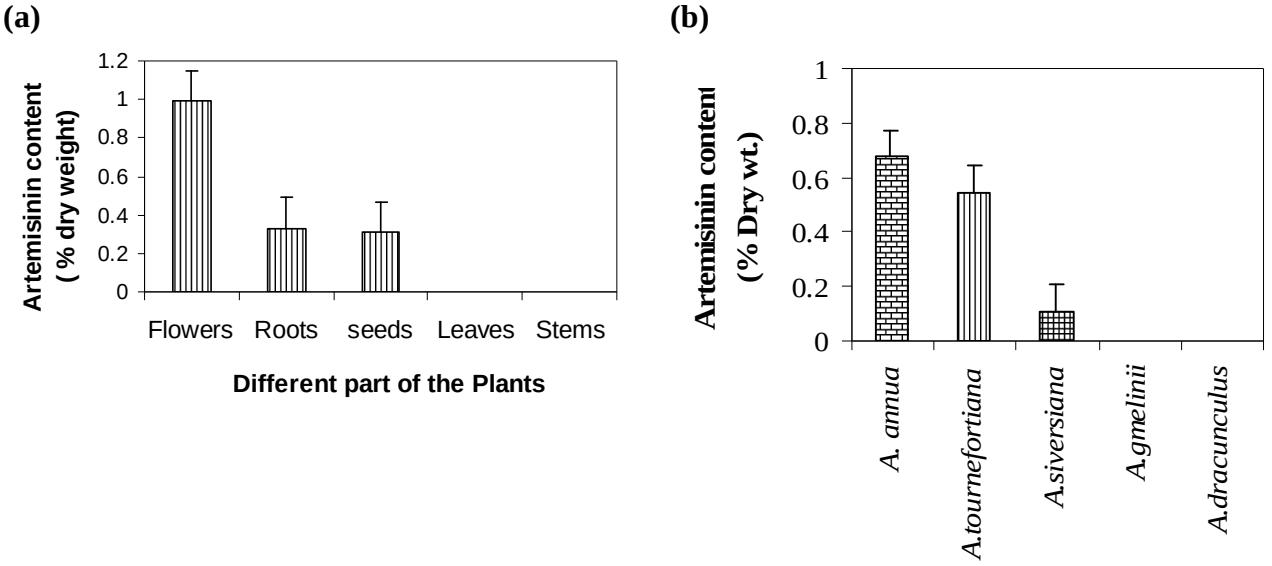


Figure 18. Comparison of artemisinin content in different plant parts of *Artemisia tournefortiana* and artemisinin content analysis from different *Artemisia* species grown in Ladakh.

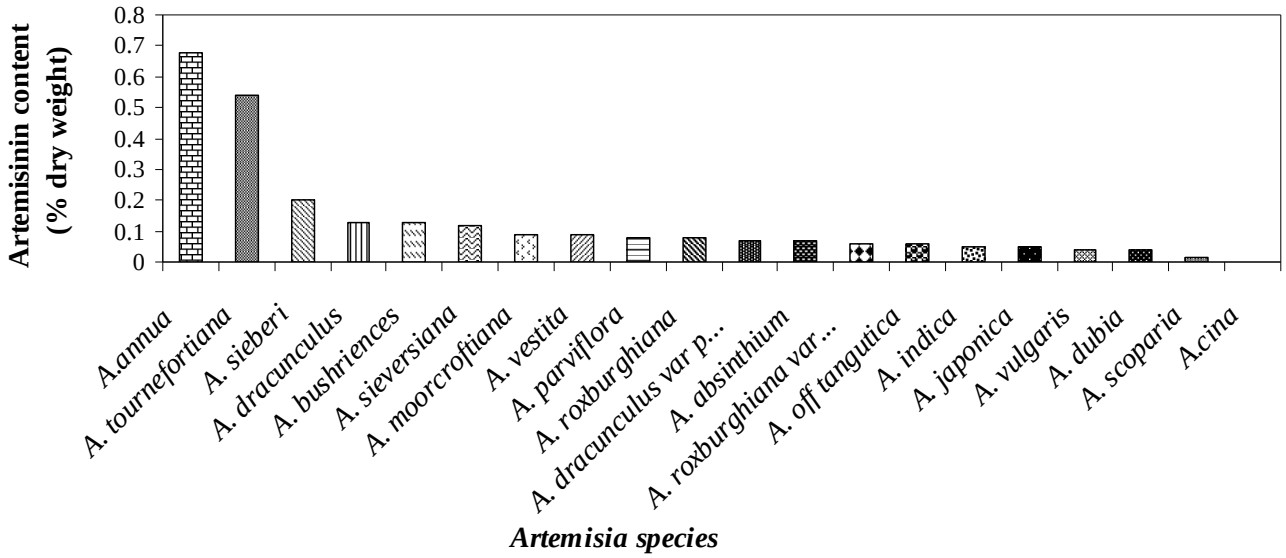


Figure 19. Comparison of artemisinin content of *Artemisia tournefortiana* with various species of *Artemisia* reported world wide.

Sequencing of targeted genes from artemisinin synthesis pathway

In order to elucidate the existence of artemisinin synthesis pathway from *A. tournefortiana* we have designed specific primers from the gene sequences involved in the synthetic pathway in *A. annua*. The synthetic pathway was partially known and we have considered only selected genes (highlighted in yellow color), for which the gene sequences are available (Figure 20). The primers

were custom synthesized and used for selective amplification from both *A. annua* and *A. tournefortiana*. The PCR amplified products were sequenced and compared pair-wise with the corresponding sequences from *A. annua* for similarity. It has been revealed that the partial gene sequences amplified using selective primers are in good agreement among both the species (Table 10). This finding further authenticated that *A. tournefortiana* is an alternative and promising source of artemisinin as alternative to *A. annua*.

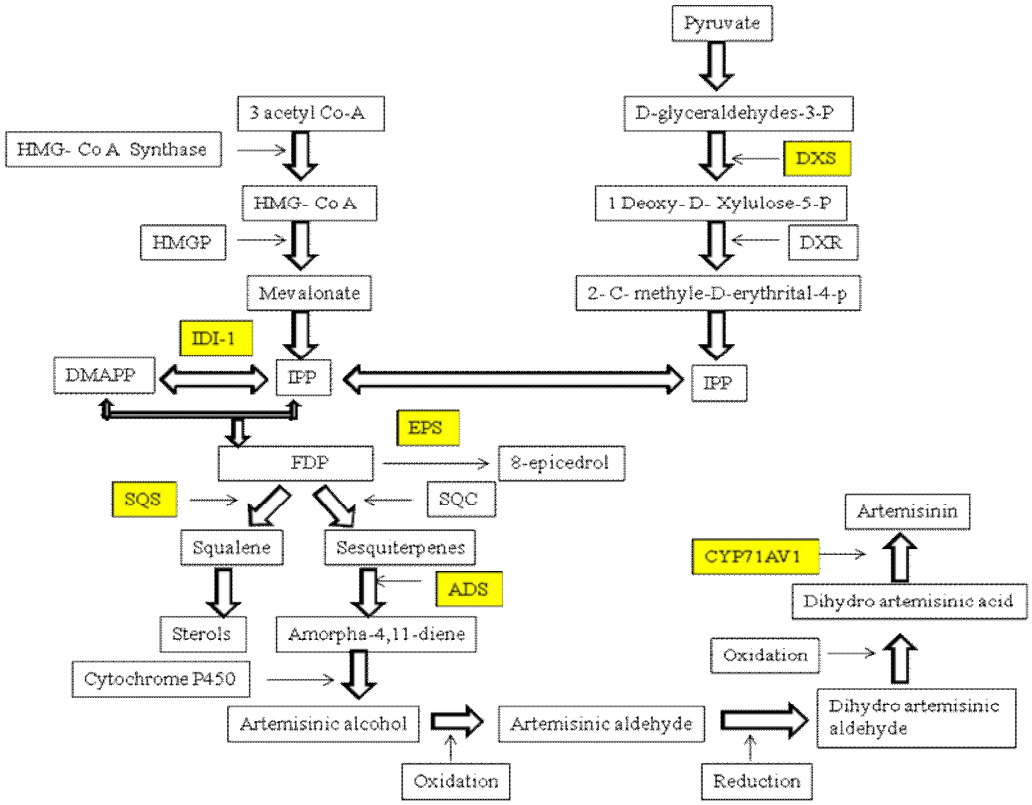
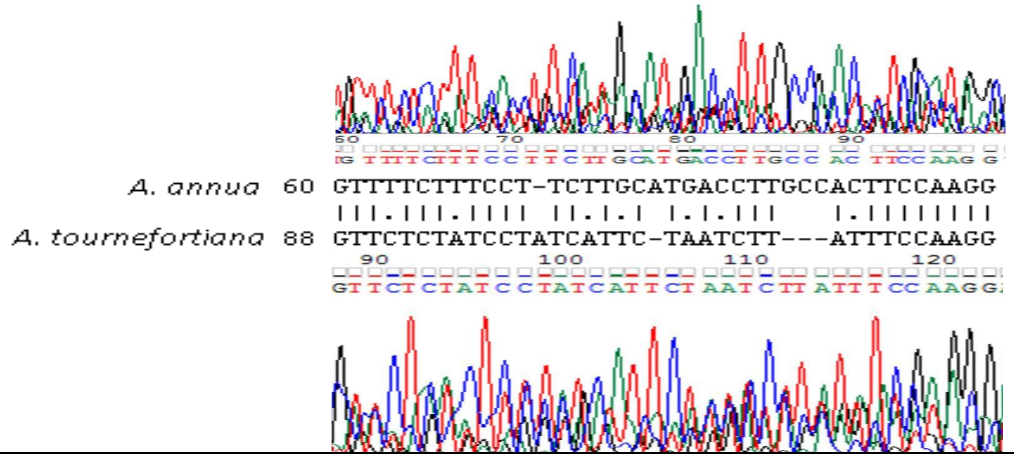


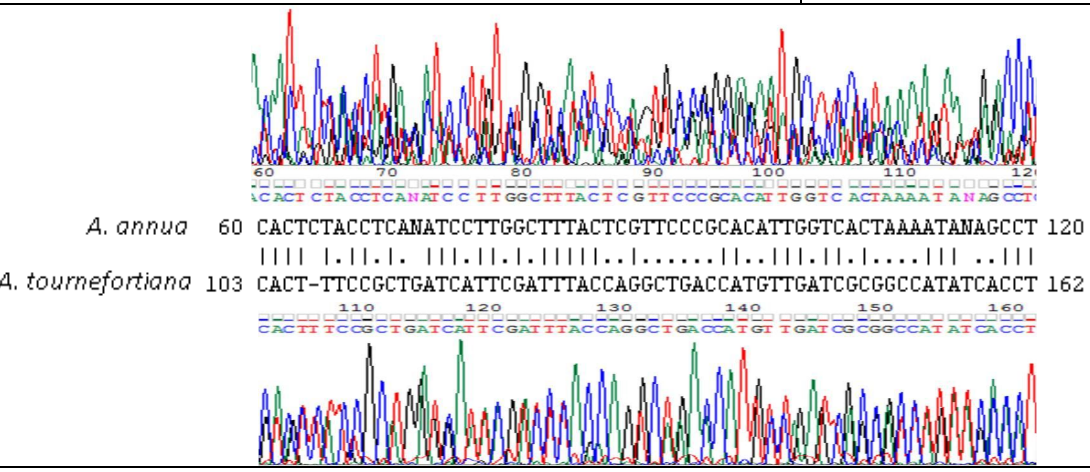
Figure 20. Artemisinin synthesis pathway elucidated in A. annua. The genes highlighted in yellow color are considered for primer designing and selective amplification in both A. annua and *A. tournefortiana.*

Table 10. Pairwise alignment of the partial gene sequences amplified using gene specific primers from both *A. annua* and *A. tournefortiana.*

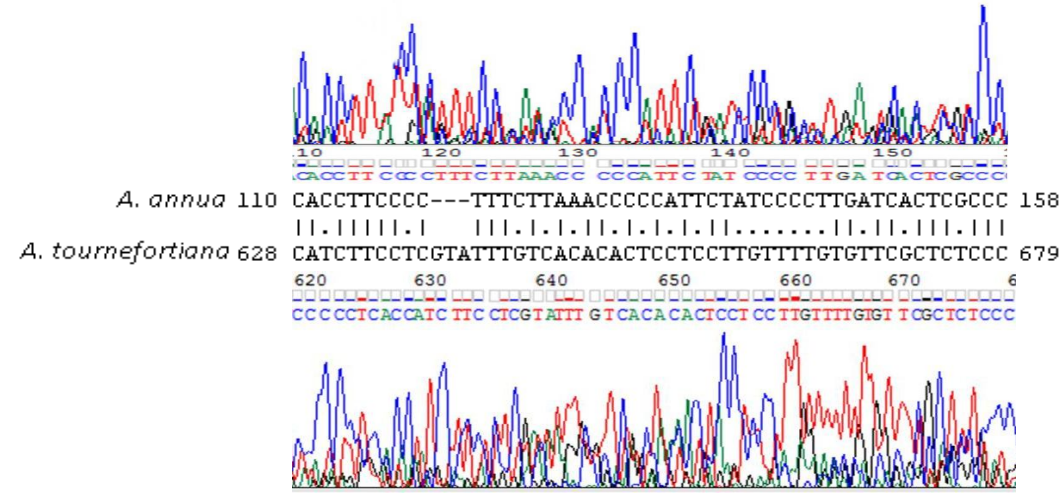
Aasqs (squalene synthase)	
ATTTCAGTTTTGGCCTCGTG	1F1_ 678-691
CAAAAACAATACCTTTTACATTTCAAG	1R1_ 902-923



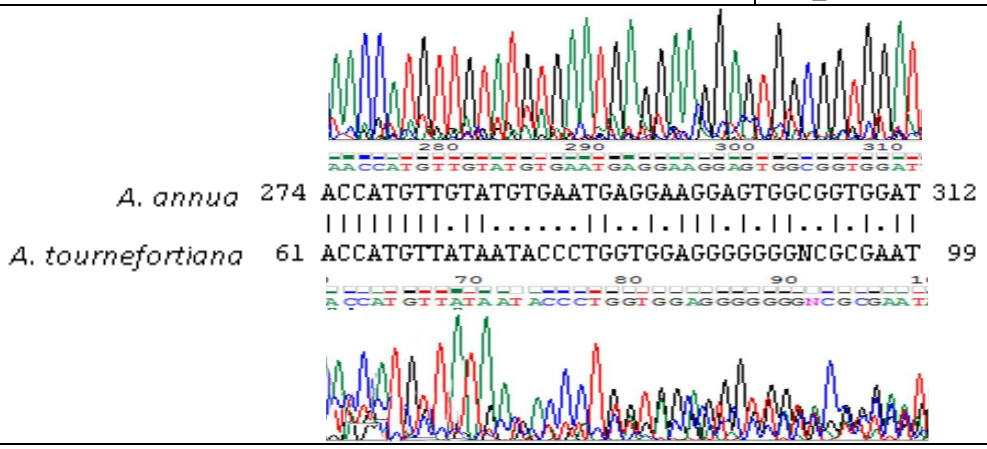
ADS (amorpha-4,11-diene Synthase)	
TAGGAAAGCGGCACATACCAGG	1F1_28-49
AGCCATCAGTGAGATACCACTCTG	1R1_944-967



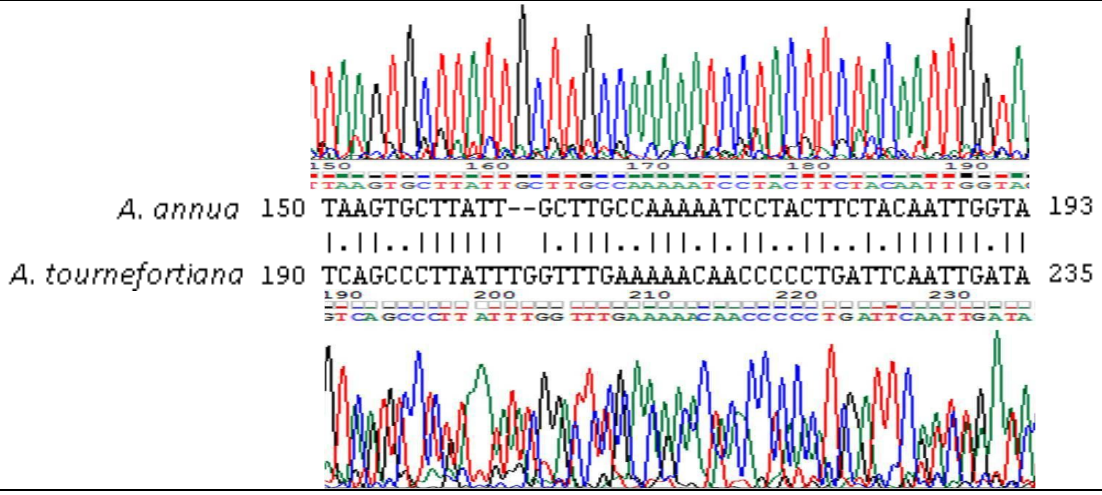
CYP71AV1(amorpha-4,11-diene mono oxygenase)	
TCATTCCAGTCGCTTCGTGAAGAA	2F1_91-114
AGGTAGTGGAGGGTGCAACCTCAA	1R1_718-741



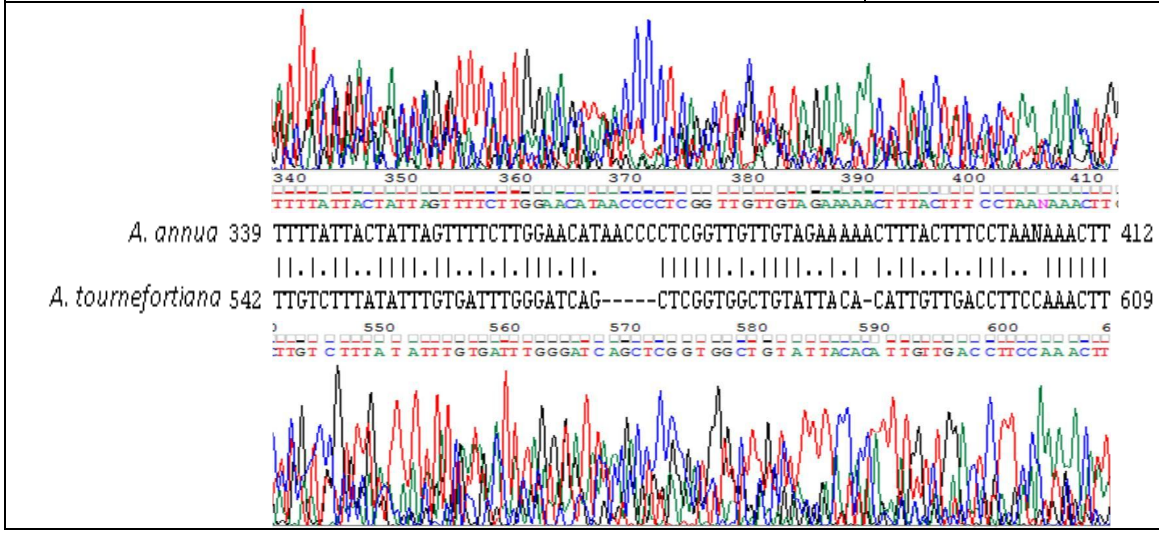
Dxs (1-deoxy-D-xylulose-5-phosphate synthase)	
AACATGGGAGCCGAACCCTCCA	1F1_25-46
CTGCAGACCAATTAGCCGAAGC	1R1_971-992



IdI-1 (Isopentyl diphosphate isomerase)	
TGTTCGCTAAGAAGGAGATGC	L_5-26
CCAAGTCCAAACGAAGGATG	R _973-995



Eps (epi-cedrol synthase)	
TGGTAGCTCACGAAGAACTGG	1F1_949-967
TGGTACCGCTGAACCATAAT	1R1_1792-1808



Conclusion

- High genetic variation measured among the genotypes of *Artemisia annua* using four different molecular markers: RAPD, ISSR, EST-derived SSRs and intron-flanking EST-specific molecular markers reveals 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.

- The artemisinin content of the plant varies significantly with respect to their geographical distribution. The artemisinin content of leaf was detected to be maximum during reproductive stage (in the month of August), whereas in the inflorescence it was detected to be highest at full flowering stages (in the month of September).

- The strong and linear relationship was detected between artemisinin content as well as soil nutrients and environmental factors suggested that further optimization of these factors is very important in the conservation and exploitation of *A. annua.*

- A new and alternative source of artemisinin was screened out form the Ladakh region—the species *A. tournefortiana* consists of even high artemisinin in the inflorescence in comparison to *A. annua* and therefore could be a great potential for artemisinin content.

List of Journal Publications

1. **Kumar J.,** Mishra G.P., Naik P.K., Murkute A.A., Srivastava R.B., *"Genomic DNA Isolation from Artemisia species Grown in Cold Desert High Altitude in India"*, African Journal of Biotechnology 2011;10(37):7303-7307.

2. **Kumar J.,** Singh H., Mishra G.P., Srivastava R.B., Naik P.K., *"Genetic analysis of Artemisia annua genotypes using RAPD and ISSR molecular markers from the trans-Himalayan (Ladakh, India) region"* Journal of Medicinal Plant Research (In Press).

3. **Kumar J.,** Singh H., Mishra G.P., Srivastava R.B. Naik P.K.,*"Utilization of Intron-flanking EST-specific Markers in the genetic characterization of Artemisia annua genotypes from Ladakh region"*. The Journal of Environmental Biology (In Press).

4. **Kumar J.,** Mishra G.P., Murkute A.A., G.P. Kumar, Naik P.K., S.B. Singh, *"Genetic relationship between Artemisia species growing in trans-Himalayan cold arid desert using RAPD markers"*, Indian Journal of Horticulture (In press).

5. **Kumar J.,** Singh H., Mishra G.P., Srivastava R.B., Naik P.K., *"Bioinformatics mining and utilization of EST-derived SSR in Artemisia annua wild population from Ladakh region"* Indian Journal of Biotechnology (under review).

6. **Kumar J.,** Tayade A.B., Mishra G.P., Srivastava R.B., Naik P.K., *"Artemisia tourenefortiana: an alternative and promising source of artemisinin"* Planta Medica. (under review)

7. **Kumar J.,** Mishra G.P., Srivastava R.B., Naik P.K., *"Impact of soil nutrient and environmental factors on artemisinin content among Artemisia annua genotypes of Ladakh region using linear and non-linear approach"*. Communication in soil science and plant analysis. (under review).

8. **Kumar J.,** Mishra G.P., Srivastava R.B., Naik P.K., *"Comparative analysis of artemisinin content among the genotypes of A.annua at different developmental stages from Ladakh region"* Phytochemistry Letters (under review).

Conference Publications

1. **Kumar J.,** Tayade A.B., Murkute A.A., Mishra G.P., Naik P.K., Singh S.B., *"Quantification of artemisinin content in Artemisia species growing in Ladakh region"* Fourth National Horticulture Congress 2010 conducted by Horticulture Society of India and National Skill Foundation of India, New Delhi, 18-21 November 2010.

2. **Kumar J.,** Mishra G.P., Srivastava R.B., Naik P.K., *"In Silico mining and utilization of EST-SSR in Artemisia annua wild population from Ladakh region"* International Conference on Emerging trends on Food and Health security in cold desert, organized by DIHAR, DRDO, Leh-Ladakh. 23-25 September 2011.

Reference

[1] Simon J.E., Cebert E., *"Artemisia annua: A Production Guide"*. Purdue University Fact Sheet. 1994.

[2] Ferreira J.F.S., Janick J. *"Immunoquantitative analysis of artemisinin from Artemisia annua using polyclonal antibodies"* Phytochemistry 41: 97–104, 1996.

[3] Bhattarai A., Ali A.S., Kachur S.P., Martensson A., Abbas A.K., Khatib R., Al- Mafazy A.W., Ramsan M., Rotllant G., Gerstenmaier J.F., et al., *"Impact of artemisinin-based combination therapy and insecticide-treated nets on malaria burden in Zanzibar"*, PLoS Med, 4, 309, 2007.

[4] World Malaria Report, Geneva, World Health Organization, 2010.

[5] Klayman D.L., *"Qinghaosu (Artemisinin): An antimalarial drug from China*", Science, 228, 1049-1055, 1985.

[6] Singh N.P., Lai H.C., *"Artemisinin induces apoptosis in human cancer cells"*, Anticancer Res, 24, 2277–2280, 2004.

[7] Wallaart E., *"Investigations on biosynthesis of the novel antimalarial drug artemisinin in the plant Artemisinin annua"* Ph. D. thesis Rjkniversiteit Groningen, 16. 2000

[8] Williams J.G.K., Kubelik A.R., Livak K.J., Rafalski J.A., Tingey, S.V. *"DNA polymorphisms amplified by arbitrary primers are useful as genetic markers"*, Nucleis Acid Research, 18: 6531-6535, 1990.

[9] Morgante M & Olivieri AM *Plant J.* 3: 175-182, 1993.

[10] Palumbi S.R., *"Nucleic acids II: the polymerase chain reaction"*, In: Hillis, D. and Moritz, C. (eds) Molecular Systematics. 2nd edn. Sinauer Associates Inc., Sunderland, MA, pp. 205–247, 1995.

[11] Scott K.D., Eggler P., Seaton G.G., Rossetto M., Ablett E.M., *et al., "Analysis of SSRs derived from grape ESTs"*, Theor Appl Genet, 100, 723-726, 2000.

[12] Endress R., *"Soil nutrient factors related to salidroside production of Rhodiola sachalinensis distributed in Chang Bai Mountain"*. Plant Cell Biotech. 13, 187-242, 1994.

[13] Alam M., Naik P.K. *"Impact of soil nutrients and environmental factors on podophyllotoxin content among 28 Podophyllum hexandrum Populations of Northwestern Himalayan Region using linear and non-linear approach"* Comm. Soil Sci. Plant Anal. 40, 2485-2504, 2009.

[14] Saghai-maroof, M.A., Soliman K.M., Jorgenses R.A., Allard R.W., *"Ribosomal DNA spacer-length polymorphism in barley. Mendelian inheritance, chromosomal location and population dynamics Proc"* National Academy of Science, 81, 8014–19, 1984.

[15] Excoffier L., Smouse P.E., Quattro J.M., *"Analyses of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data"*, Genetics, 131, 479-491, 1992.

[16] Rozen S., Skaletsky H., *"Primer3 on the WWW for general users and for biologist programmers"* Methods Mol Biology, 132, 365-386, 2000.

[17] Weising K., Atkinson G., Gardner C., *"Genomic fingerprinting by microsatellite-primed PCR: a critical evaluation"*, PCR Methods and Appl, 4, 249-255, 1995.

[18] Tautz D., Schlotterer C., *"Simple sequences"*, Curr Opin Genet Dev, 4, 832-837, 1986.

[19] Liu Z.J., Li P., Argue B.P., Dunham R.A., *"High Level of conservation at microsatellite loci among Ictalurid catfishes"*, J Hered, 90, 307-312, 1999.

[20] Gupta P.K., Rustgi S., *"Molecular markers from the transcribed/expressed region of the genome in higher plants"*, Funct Integr Genomics, 4, 139-162, 2004.

[21] MISA http://pgrc.ipk-gatersleben.de/misa/)

[22] Abdin M.Z., Israr M., Rehman R.U., Jain S.K., *"Artemisinin, a novel antimalarial drug: biochemical and molecular approaches for enhanced production"*, Planta medica, 69, 289-299, 2003.

[23] Noorden R. V., "*Demand for malaria drug soars*", Nature 466, 672, 2010.

[24] Arab H.A., Rahbari, S., Rassouli, A., Moslemi, M.H., Khosravirad, F.D.A., "*Trop Anim Health Prod"*, 38, 497, 2006.

[25] Zia M., Mannan A., Chaudhary M.F., "*Factors Affecting Agrobacterium Tumefaciens Mediated Transformation of Artemisia Absinthium L*."Pak. J. Bot. 39, 799, 2007.