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

Thesis submitted in fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

IN

BIOTECHNOLOGY & BIOINFORMATICS

by

JITENDRA KUMAR

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT,
SOLAN 173234, HIMACHAL PRADESH, INDIA
OCTOBER 2011

@ Copyright
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT
OCTOBER 2011
ALL RIGHTS RESERVED

TABLE OF CONTENTS

Abstract	**1-5**
CHAPTER 1	
INTRODUCTION	**6 - 31**
1.1 Genus *Artemisia* and Geographical distribution of *A. annua* L.	
1.2 Botany *A. annua* L.	
1.3 Artemisinin content in the plant	
1.3.1 Importance of Artemisinin and its derivatives used in malaria and other diseases.	
1.3.2 Artemisinin biosynthesis pathway	
1.3.3 Artemisinin content among *Artemisia* species	
1.3.4 Artemisia species in India and artemisinin content	
1.4 Cultivation of *A. annua* L.	
1.5 Breeding for increasing artemisinin content of *A. annua* L.	
1.6 Genetic diversity of *Artemisia annua* populations	
1.7 Use of molecular markers for studying genetic diversity	
1.7.1ISSRs as molecular markers	
1.7.2RAPDs as molecular markers	
1.7.3SSRs molecular markers	
1.7.4ILPs molecular markers	
1.8 Optimization of factors and ex situ conservation of *Artemisia annua*	
CHAPTER 2	
Genetic characterization of *Artemisia annua* using RAPD and ISSR molecular markers from the trans-Himalayan (Ladakh, India) region	**32 - 49**
Abstract	
2.1 Introduction	
2.2 Materials and Methods	
2.2.1 Plant material	
2.2.2 Isolation of DNA	
2.2.3 RAPD amplification	
2.2.4 ISSR amplification	
2.2.5 Data collection and analysis	
2.3 Results and discussion	
2.3.1 RAPD analysis	
2.3.2 ISSR analysis	
2.3.3 RAPD and ISSR combined data for cluster analysis	
2.3.4 Comparative analysis of RAPD with ISSR markers	
CHAPTER 3	
Utilization of intron-flanking EST-specific markers in the genetic characterization of *Artemisia annua* genotypes from the trans-Himalayan region of Ladakh, India	**50 - 63**
Abstract	
3.1 Introduction	
3.2 Materials and Methods	

3.2.1 Sources of sequence data
3.2.2 Development of putative intron polymorphic (PIP) markers
3.2.3 DNA extraction and PCR amplification
3.2.4 Data collection and analysis
3.3 Results and Discussion
3.3.1 Candidate ILP markers in *Artemisia annua*
3.3.2 ILP analysis
3.3.3 Phylogenetic analysis across twenty genotypes
3.3.4 Genetic diversity analysis

CHAPTER 4 **64 - 77**
Utilization of EST-derived SSR in the genetic characterization of *Artemisia annua* genotypes from the trans-Himalayan region (Ladakh, India)
Abstract
4.1 Introduction
4.2 Materials and Methods
4.2.1 Plant materials and DNA extraction
4.2.2 Data mining for SSR marker
4.2.3 PCR primer design and PCR amplification
4.2.4 Data collection and analysis
4.3 Results
4.3.1 Searching for ESTs containing microsatellites
4.3.2 Frequency of SSRs based on motif sizes
4.3.3 Distribution of SSRs based on motif types
4.3.4 Cluster analysis
4.3.5 Genetic diversity analysis
4.4 Discussion

CHAPTER 5 **78 - 90**
Extraction and quantification of Artemisinin at different developmental stages from *Artemisia annua* L. collected from Ladakh region
Abstract
5.1 Introduction
5.2 Materials and Methods
5.2.1 Plant Materials and sample preparation
5.2.2 RP-HPLC-PDA method
5.2.3 GC-FID method
5.2.4 Spectrophotometric method
5.3 Results and Discussion
5.3.1 Dynamic variation of artemisinin content of leaf during the growth stage
5.3.2 Artemisinin content of inflorescence during the reproductive stage

CHAPTER 6 **91 - 108**

Impact of Soil Nutrient and Environmental Factors on Artemisinin Content among *Artemisia annua* Populations of Ladakh Region Using Linear and Non-linear Approach

Abstract
6.1 Introduction
6.2 Materials and Methods
 6.2.1 Sample stations and plant materials
 6.2.2 Extraction of artemisinin and quantification
 6.2.3 Quantitative analysis of soil nutrition
 6.2.4 Statistical analysis
 6.2.5 Neural network data-mapping model development
6.3 Results
 6.3.1 Artemisinin content in the leaf of *A. annua*
 6.3.2 Effect of Altitude
 6.3.3 Effect of environmental factors on artemisinin content
 6.3.4 Effect of soil organic carbon (C)
 6.3.5 Effect of soil pH
 6.3.6 Effects of soil nitrogen (N)
 6.3.7 Effect of phosphorus (P)
 6.3.8 Effects of soil potassium (K)
 6.3.9 Performance measure of ANN and MLR model
6.4 Discussion

109-122

CHAPTER 7

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

Abstract
7.1 Introduction
7.2 Materials and Methods
 7.2.1 Plant Material and Chemicals
 7.2.2 Sample preparation
 7.2.3 Preparation of standard stock solution and calibration curve
 7.2.4 HPLC system, RP-HPLC-PDA conditions and statistical analysis
7.3 Results and Discussion

Conclusion

Bibliography

List of publications

DECLARATION

I hereby declare that the work reported in the Ph.D. thesis entitled **"Assessment of genetic diversity and artemisinin content among genotypes of *Artemisia annua* and related species collected from the Ladakh region, India"** submitted at **Jaypee University of Information Technology, Waknaghat, India** is an authentic record of my work carried out under the supervision of **Dr. Pradeep Kumar Naik and Dr. Gyan Prakash Mishra**. I have not submitted this work elsewhere for any other degree or diploma.

(………………….)

(Jitendra Kumar)
Department of Biotechnology and Bioinformatics
Jaypee University of Information Technology, Waknaghat , India
Date

SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled **"Assessment of genetic diversity and artemisinin content among genotypes of *Artemisia annua* and related species collected from the Ladakh region, India"**, submitted by **Jitendra Kumar** at **Jaypee University of Information Technology, Waknaghat**, **India,** is a bonafide record of his original work carried out under our supervision. This work has not been submitted elsewhere for any other degree or diploma.

(Dr. Pradeep Kumar Naik)
Jaypee University of Information Technology
Waknaghat, Solan (H.P.)
Date 2011

(Dr. Gyan Prakash Mishra)
Defence Institute of High Altitude Research
Defence Research & Development Organization
Date : 1-11-2011

ACKNOWLEDGEMENT

I would like to special thanks to my supervisors Dr. Pradeep Kumar Naik and Dr. Gyan Prakash Mishra for leading me through all these years. Their positive attitude towards research and zest for high quality research work has prompted me for the timely completion of the work. I deem it a privilege to be a scholar doing research under Dr. Pradeep Kumar Naik and Gyan Prakash Mishra who have endeared theirself to his students and scholars.

I am grateful to Defence Research & Development Organization, Ministry of Defence, Government of India, for the award of Research Fellowship and funding the project. I would like to take this opportunity to thank Dr. Ravi Bihari Srivastava (Director-DIHAR) and Dr. Shashi Bala Singh (Former Director-DIHAR) for their generosity, appreciation during my studies. I cherish the years spent in the Biotechnology Division of Defence Institute of High Altitude Research, DRDO, Indus-Ladakh.

I would like to thanks Dr. Y. Medury (COO, JUIT), Dr. Ravi Prakash (Vice Chancellor), Brigadier (Retd). Balbir Singh (Director, JUIT) and Prof. R. S. Chauhan (HOD of BT & BI) for their helping hands and providing all kinds of facilities to carry out my research work. Also, my sincere thanks goes to Dr. Harvinder Singh (Senior Lecturer), Dr. Anil Kant Thakur (Senior Lecturer) and all the faculty members of the Dept. of Biotechnology and Bioinformatics for their helping atitudes and backing me with their moral support.

Further, I would personally like to thank Mr Dipankar Sengupta, Associate Lecturer, JUIT, as a special friend who stands always with me and for his never ending helping attitude as well as moral support during research work. I have no words to express the kind of supports he has extended to me for the successful completion of my research work.

Besides, I can never ever forget my colleagues Manish, Amol, Gurucharan, Girish and my Junior colleagues Ashish, Prabodh, Vijay, Priyanka, Sunil, Prabhat, Saroj, Richa and JUIT friends Jatindra Kumar Pradhan, Nidhi Gupta, Saurav Pandit, who have always been helped me in a numerous ways. I also thank Technical staff of Department of Biotechnology and Bioinformatics Ms. Somlata Sharma, Ismail, Ravikant, Baleshwar, Mrs. Mamta for their assistants and valuable contributions. I would like to express my heartfelt gratitude to all those who have contributed directly or indirectly towards obtaining my doctorate degree.

I thank my parents for instilling in me confidence and a drive for pursuing my PhD. My greatest appreciation goes to my father who taught me the value of working hard and with honour; to my mother who taught me to never let go of a goal and to persevere in the face of adversity; and to my Bhaiya and Bhabhi who support me in all my endeavours.

Finally, I wish to thank the respondents of my study and all the individuals whose comments, reviews and suggestions have helped my project in developing an informative insight. Lastly, I extend my apologies to anyone I failed to mention.

(Jitendra Kumar)

LIST OF ABBREVIATIONS

RAPD	Random Amplified Polymorphic DNA
ISSR	Inter Simple Sequence Repeats
SSR	Simple Sequence Repeats
ILP	Intron Length Polymorphism
HPLC	High Performance Liquid Chromatography
GC	Gas Chromatography
WHO	World Health Organization
ACTs	Artemisinin based Combination Therapies
DHQ	Dihydroartemisinin
IC_{50}	Inhibitory Concentration
IPP	Isopentenyl diphosphate
DMAPP	Dimethylallyl diphospahte
MVA	Mevalonic acid
MEP	Methylerythritol Phosphate
DXS	Deoxy-D-xylulose-5-phosphate Synthase
FDPase	Farnesyl diphosphate Sythase
PCR	Polymerase Chain Reaction
ANN	Artificial Neural Network
MLR	Multiple Linear Regressions
MSL	Mean Sea Level
RFLP	Restriction Fragment Length Polymorphism
CTAB	Cetyletrimethyle Ammonium Bromide
EDTA	Ethylenediamine tetraacetic acid
PVP	Polyvinylpyrrolidone
CC	Correlation Coefficient

LIST OF FIGURES

Figure Number	Caption	Page Number
1.1	Range of Artemisia annua cultivation in the world according to latitude, and altitude	9
1.2	The plant *Artemisia annua*, its leaves and inflorescence as well as the trichome that accumulate artemisinin.	10
1.3	Scanning electron micrograph of *A. annua* leaf showing trichome that contain artemisinin	10
1.4	Molecular structure of Artemisinin and its structural derivatives	14
1.5	Simplified terpenoid biosynthetic scheme leading to artemisinin. The two arms of terpenoid biosynthesis and some of the regulatory enzymes that catalyze different reaction : HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; DXS, 1-deoxy-D-xylulose-5 phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; FDPS, farnesyl diphosphate synthase; SQC, sesquiterpene cyclase; SQS, squalene synthase; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate	16
1.6	Alternative pathways leading to synthesis of artemisinin (source: Weathers et al., 2006, Zhang et al., 2008 and Wallaart et al., 1999).	17
1.7	*1-A.aff. toggutica,* 2- *A.bravifolia,* 3- *A. desertorum,* 4- *A. dubia,* 5- *A. genipi,* 6- *A. glacialis,* 7- *A. indica,* 8- *A. japonica,* 9- *A.maritima,* 10- *A. moorcroftiana,* 11-*A. Parviflora,* 12-*A. roxburgiana,* 13- *A.vestita,* 14- *A. vulgaris,* 15- *A. absinthium,* 16-*A. Bushriences,* 17- *A. jap. Japonica,* 18- *A. rox.roxburghiana,* 19-*A. rox. Gratae,* 20- *umbelliformis*	19
1.8	Artemisisnin content of various *Artemisia* species reported from the literature	20
1.9	Geographical map representing two valleys: Indus (3500 m) and Nubra (3000 m) located in Ladakh (Jammu and Kashmir, India) that comprised of *Artemisia annua* genotypes.	30
2.1	Agarose 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	41

extracted from Indus valley and Lane 11 to 20; are from Nubra valley.

2.2 RAPD 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. 42

2.3 (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on RAPD profiling, (b) the relationship between K and ΔK, (c) unbiased clustering of genotypes between 2 groups. (c) 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). 43

2.4 ISSR 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 valley. M = the size of molecular markers in base pairs using λ DNA. 45

2.5 (a) NJ tree representing clustering of genotypes along with supported bootstrap values based on ISSR profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, the relationship between K and ΔK, (c) 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). 46

2.6 (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on combination of RAPD and ISSR profiling (1482 RAPD bands+1087 ISSR bands), (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, showing the relationship between K and ΔK, (c) 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). 48

3.1 Flowchart of developing PIP markers in query sequences. Adopted from Long Yang, et al., PIP: a database of potential intron polymorphism markers. 55

3.2 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. 60

3.3 (a) Neighbor joining (NJ) tree representing clustering of genotypes at populations' level along with supported bootstrap values based on ILP profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, showing the relationship between K and ΔK, (c) 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). 61

4.1 The distribution pattern of SSRs in terms of motif size. 71

4.2 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. 73

4.3 Dendrograms generated using unweighted pair of group method with arithmetic average analysis, showing relationships between 20 *Artemisia annua* genotypes, using SSRs data. 74

5.1 Chromatograms of artemisinin standard (A) and *Artemisia annua* extract (B) by HPLC-PDA. The concentration of artemisinin standard is 1mg/ml (artemisinin-acetonitrile). Artemisinin was extracted from 105.3 mg dry weight of inflorescence. 83

5.2 Chromatograms of artemisinin standard (A) and *Artemisia annua* extract (B) by GC-FID. The concentration of artemisinin standard is 1mg/ml (artemisinin-acetonitrile). Artemisinin was extracted from about 100 mg dry weight of inflorescence. 84

5.3 Reaction of artemisinin in acidic medium to form H_2O_2 and reaction between H_2O_2- KI System with safranin O. 84

5.4 Dynamic variation of artemisinin content of leaf in the different growing stage. 87

5.5 Correlation of artemisinin content of inflorescence of *Artemisia annua* analyzed by HPLC-PDA, GC-FID and spectrophotometric methods. Their artemisinin content varied from 0.2 % to 0.9%. 89

6.1	Architecture and connection of a feed-forward back propagating artificial neural network with multiple layers.	97
6.2	Dynamic variation of artemisinin content of whole plant (Leaf and flowers) at different stages of *Artemisia annua*.	99
6.3	The meteorological observations have been made during the course of experiment with respect to the site of collection of *Artemisia annua* populations.	100
6.4	The relationship between environmental variables humidity % (maximum and minimum), temperature ^{0}C (maximum and minimum).	101
6.5	The relationship between soil organic carbon and artemisinin contents in *Artemisia annua* in Ladakh region.	102
6.6	The relationship between soil organic carbon, pH, and artemisinin contents in *Artemisia annua* at different altitude in Ladakh region.	102
6.7	The relationship between nitrogen and artemisinin contents in the *Artemisia annua* at different altitudes in the Ladakh region.	103
6.8	The relationship between phosphorus, potassium contents and artemisinin contents in the *Artemisia annua* at different altitudes in the Ladakh region.	104
6.9	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	105
7.1	Five species of Artemisia: *A. annua*, *A. tournefortiana, A. sieversiana*, *A. gmelinee* and *A. dracunculus* that are are predominantly available in Ladakh region.	112
7.2	Plant parts (leaf, flower, root and seeds) of *Artemisia tournefortiana*, a new species for the alternative source of artemisinin.	113
7.3	Chromatograms of standards of artemisinin (A) and plant extract (B) by HPLC-PDA showing retention time at 4.796 and 4.815 minutes.	115
7.4	Artemisinin synthesis pathway gene that has been amplified in *A. tournefortiana.*	116
7.5	Analysis of artemisinin content from different *Artemisia* species grown in Ladakh.	118
7.6	Comparison of artemisinin content in different plant parts of *Artemisia*	118

tournefortiana.

7.7 Comparison of artemisinin content of *Artemisia tournefortiana* with various species of *Artemisia* reported world wide. 119

7.8 Artemisinin contentin (%) of *Artemisia annua* reported worldwide 119

LIST OF TABLES

Table Number	Content	Page Number
1.1	Artemisinin content of different plant components from greenhouse- and field- grown *Artemisia annua*, estimated by HPLC-EC.	12
1.2	In vitro antimalarial activity of constituents of *A. annua*	15
1.3	Various species of *Artemisia* reported from the Ladakh region of India	21
2.1	List of primers used for RAPD amplification, GC content, total number of loci, the level of polymorphism and resolving power.	38
2.2	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).	39
2.3	Summary of genetic variation statistics for all loci of RAPD among the *Artemisia annua* genotypes with respect to their distributions among two valleys.	43
2.4	Summary of analysis of molecular variance (AMOVA) based on RAPD analysis. Levels of significance are based on 1000 iteration steps.	43
2.5	Summary of nested analysis of molecular variance (AMOVA) based on combination of RAPD and ISSR analysis among the populations of *A. annua.* Levels of significance are based on 1000 iteration steps.	45
2.6	Summary of genetic variation statistics for all loci of ISSR among the *Artemisia annua* genotypes with respect to their distributions among two valleys.	47
2.7	Summary of analysis of molecular variance (AMOVA) based on ISSR analysis among the populations of *A. annua.* Levels of significance are based on 1000 iteration steps.	48
2.8	Summary of genetic variation statistics for the combination of RAPD + ISSR loci among the *Artemisia annua* populations with respect to their distributions among two valleys.	49
2.9	Overall genetic variability across all the 20 genotypes of *Artemisia annua* based combination of RAPD and ISSR markers.	49
3.1	Genes related to artemisinin biosynthetic pathway in *Artemisia annua*	56

3.2	List of primers used for ILP study, there GC content, melting temperature (Tm) and polymorphism	57
3.3	Summary of genetic variation statistics for all loci of ILPs among the *Artemisia annua* genotypes with respect to their distributions among two valleys.	62
3.4	Overall genetic variability across all the 20 genotypes of *Artemisia annua* based on ILPs analysis.	62
3.5	Summary of nested analysis of molecular variance (AMOVA) based on ILPs, among the populations of *Artemisia annua.* Levels of significance are based on 1000 iteration steps.	62
4.1	List of primers used for SSR amplification, GC content, total number of loci, the level of polymorphism and PIC value.	69
4.2	Summary of SSR search among *Artemisia anua* ESTs	70
4.3	Occurrence and number of repeats of 77 SSR motifs in *Artemisia annua* L	71
4.4	Summary of genetic variation statistics for all loci of SSRs among the *Artemisia annua* populations with respect to their distributions among two valleys.	75
4.5	Overall genetic variability across all the 20 genotypes of *Artemisia annua* based on SSR analysis	76
4.6	Summary of analysis of molecular variance (AMOVA) based on SSR marker, among the populations of *Artemisia annua.* Levels of significance are based on 1000 iteration steps.	76
6.1	Experimental and predicted artemisinin yield from inflorescence in A. annua corresponding to Indus valley and Nubra valley. Values are represented in mean ± standard deviation.	116
7.1	Pairwise alignment of the partial gene sequences amplified using gene specific primers from both *A. annua* and *A. tournefortiana.*	120

ABSTRACT

Artemisia annua L. is an important medicinal plant from which artemisinin was extracted to treat malaria effectively. Artemisinin is isolated from the aerial part of the *A. annua.* The recent works of several groups show that artemisinin is currently the most effective malaria drug and is the World Health Organization's currently recommended medicine in treating malaria as combination therapy with commonly used antimalarials. Artemisinin also shows promise as a potential therapeutic agent for other parasitic and viral diseases as well as for the treatment of certain cancers and the reduction of angiogenesis. *A. annua* still remain one of the major sources of drug in the traditional and modern system of medicine throughout the world. Therefore, the world market for products including artemisinin derivatives is now growing rapidly, and the demand for artemisinin is increasing. At present, artemisinin compounds are derived from a raw substance extracted from the plant *A. annua* because artemisinin is very difficult to synthesise. Harvesting from the main source of raw material, 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. Artemisinin content varies greatly between populations of *A. annua* from the Ladakh region. Therefore, it would be inefficient, environmentally destructive and economically unsound to randomly harvest *A. annua*. In order to develop strategies for commercial cultivation of *A. annua* as an economically viable alternative crop, it would be essential to evaluate the natural populations of *A. annua* for artemisinin content and agronomic traits in the Ladakh region. Appropriate methods of conservation management should be adopted including *in situ* conservation and germplasm collection from the remaining populations with great genetic variations.

In the Ladakh region *Artemisia annua* populations are distributed in restricted pockets in Indus (altitude is 3500 m) and Nubra (altitude is 3000 m) valleys. Twenty genotypes of *A. annua* (ten genotypes from each valley) were sampled from the Ladakh region for the genetic characterization using molecular markers: RAPD, ISSR, EST-derived SSRs and intron-flanking EST-specific markers. Further leaves and inflorescences at different developmental stages of the plant were sampled for quantification of artemisinin and its correlation with soil as well as environmental factors. Moreover to found out an alternative source of artemisinin, genetic profiling

and artemisinin content of various species of *Artemisia* from the Ladakh region was performed.

Genetic characterization of 20 genotypes of *A. annua* collected from two valleys *viz*. Nubra (9,600 ft) and Indus (11,500 ft) of the trans-Himalayan (Ladakh, India) region were analyzed using 37 PCR markers (20 RAPDs and 17 ISSRs). RAPD analysis yielded 124 polymorphic fragments (96.9%), with an average of 6.2 polymorphic fragments per primer. ISSR analysis produced 85 bands, of which 78 were polymorphic (86.1%), with an average of 4.58 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. The genetic diversity was high among the genotypes (Nei's genetic diversity = 0.336 and Shannon's information index = 0.495) as measured by combination of both RAPD and ISSR markers. The mean coefficient of gene differentiation (Gst) was 0.145, indicating 85.5% of the genetic diversity resided within the genotypes. RAPD markers were found more efficient with regard to polymorphism detection, as they detected 96.9% in comparison to 86.1% for ISSR markers. Genotypes collected from Ladakh region were clustered distinctly into two groups as per their sampling sites using RAPD, ISSR and in combination of RAPD+ISSR based on neighbor joining (NJ) method. A relatively high genetic variation was detected among the 20 genotypes whereas the variation between the two valleys was less using AMOVA test. It was found that the genetic diversity among genotypes from Nubra valley was narrow than that of Indus valley, suggesting the importance and feasibility of introducing elite genotypes from different origins for *Artemisia* germplasm conservation and breeding programs.

Besides RAPD and ISSR, genetic variation was assessed utilizing intron-flanking EST-specific markers, among genotypes of *Artemisia annua* collected from two sampling sites *viz*. Nubra (9,600 ft) and Indus (11,500 ft) valleys of the trans-Himalayan region, Ladakh, India. The available ESTs (~85,282) sequences of *A. annua* were aligned with the genomic sequences of *Arabidopsis* to developed 'intron-flanking' EST-specific PCR primers. These primers anneal with the conserved region of exon (flanking to the intron) and amplified the introns. Out of the 39 primers selected and tested on 20 genotypes of *A. annua*, we successfully exploited 81 codominant intron length polymorphism (ILP) markers, with an average of 2.08

markers per primer and 92.04% polymorphism detection. Clustering of genotypes revealed distribution of genotypes into 2 distinct clusters with respect to their site of collection. 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.

Further, we have used EST derived SSR markers to measure the genetic diversity among the *A. annua* germplasm collected from geographically separated Indus (11,500 ft) and Nubra (9600 ft) valleys (Ladakh, India). We investigated 68,974 non-redundant (out of ~85,282 available) ESTs of *A. annua* and developed 4,342 SSR markers. On an average, one SSR was found per 8.9 kb of EST sequence with di-nucleotide motifs in highest frequency (52.2%) followed by tri (42.4%), tetra (3.6%), hexa (1.2%) and penta-nucleotide (0.6%) repeat types. A set of 16 primer pairs were designed by considering only the SSR-containing ESTs from the artemisinin biosynthetic pathway. A total of 38 alleles were identified from 13 polymorphic SSR loci, ranging from 1-7 alleles per locus and displayed moderate genetic diversity with an average of 0.24. Genotypes from Nubra valley were less diverse in comparison to genotypes collected from Indus valley. It is, therefore, imperative for conservation planners in designing conservation strategies for wild populations of *A. annua*, to ensure that many possible separate populations are targeted for conservation rather than conserving a few selected populations.

The demand for artemisinin has increased sharply since the World Health Organization recommended its use as part of the artemisinin combination therapies in 2001. The artemisinin concentration peaks at a slightly different time in different areas and identifying this is critically important to ensure optimum content of artemisinin in *A. annua*. The present study investigated the artemisinin content in different plan parts as well as at different stages of growth of *A. annua* from the Ladakh region. Artemisinin content in the leaf gradually increased from the stage of seedling upto six months and afterward it was remain constant. The highest yield was 0.441% dry weight and 0.356% dry weight in the samples collected from Indus valley and Nubra valley. Similarly the artemisinin content in the inflorescence was found to be maximum at full-flowering stage (0.813 ± 0.065% and 0.721 ± 0.053% dry weight in the samples collected from Indus valley and Nubra valley respectively using HPLC

analysis). Therefore it is concluded that the optimal time at which leaves should be harvested is the time when flower buds start to form (in the month of August) and the inflorescence should be harvested during the full-flowering stage (in the month of October) for better yield of artemisinin from *A. annua* in the Ladakh region.

For conservation and breeding strategy for increasing artemisinin content it is very important to assess the potential of *A. annua* populations for artemisinin production. Artemisinin content in *A. annua* differs greatly in different natural habitats. In order to facilitate reasoned scientific decisions on its domestication, conservation and sustainable utilization, the effects of soil nutrients and environmental factors on artemisinin content in the leaves of *A. annua* were investigated. The artemisinin content reached up to 0.675% of dry weight when soil pH value was about 7.82, soil organic carbon was higher than 0.73% and nitrogen content was higher than 0.67% of soil dry weight. Soil with available phosphorous content higher than 0.519% and potassium content higher than 0.76% resulted in low artemisinin content. The strong and linear relationship detected between artemisinin 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.* In this regard the prediction model like artificial neural network (ANN) and multiple linear regressions (MLR) developed in this study to map the effect of these factors on artemisinin yield will be of great help. The ANN prediction model revealed better prediction of yield ($R^2 = 0.992$) than MLR prediction model ($R^2 = 0.806$). Lower level of root mean square error (RMSE) for ANN model (0.018) than MLR model (0.062) with respect to the experimental measurement establishes the ANN method as an efficient tool for optimization of soil nutrients and climatic factors for artemisinin yield.

Currently the primary source of artemisinin (the antimalarial compound) is *Artemisia annua*. Due to ever-increasing resistance of malaria parasite strains to the available antimalarial drugs and growing concern of inadequate artemisinin supply to meet the future requirements for artemisinin based-combination therapies, it is imperative to search for supplementary artemisinin sources. 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%). Further the artemisinin content in flower vary from 0.05 to 1.43% with respect to different geographical location. Our results thus indicate a great potential for the use of *A. tournefortiana* as an alternative source of artemisinin.

CHAPTER 1

INTRODUCTION

Artemisia annua L. is an important medicinal plant for the production of anti-malarial and possibly antibacterial agents and natural pesticides. It was originally collected by the Chinese as an 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. World Health Organization's also recommended ACTs for treating malaria. ACTs have been shown to have: rapid resolution to fever and parasitaemia, low toxicity, and are well tolerated. The artemisinin compounds are effective against *Plasmodium falciparum* and *P. vivax*, including multi-drug-resistant strains, although strains of *P. falciparum* with reduced sensibility have been reported from various countries. Artemisinin also shows promise as a potential therapeutic agent for other parasitic and viral diseases as well as for the treatment of certain cancers and the reduction of angiogenesis [1, 2, 3, 4]. *A. annua* still remain one of the major sources of drug in the traditional and modern system of medicine throughout the world. Harvesting from the main source of raw material, (wild) of *A. annua* 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 [5].

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 [6, 7]. 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* because artemisinin is very difficult to synthesise [8, 9, 10]. Enhanced production of artemisinin in the whole plant of *A. annua* L. is therefore highly desirable.

1.1 Genus *Artemisia* and geographical distribution of *Artemisia. annua*

The genus *Artemisia* belongs to a useful group of aromatic and medicinal plants. It is one of the largest and most widely distributed genera of the family Astraceae comprised over 450 diverse species. In relation to Astraceae, members of this family are valued in traditional and herbal medicine for treatment of a wide range

of ailments on a global scale. Its contribution to medicine and drug therapy makes it one of the top-ranking families of drug-yielding plants. A number of *Artemisia* species have a high economic value for food, medicine, forage, ornamentals, and other uses. In the literature, artemisinin has been reported in *A. annua*, *A. apiacea*, *A. lancea*, *A. cina*, *A. sieberi*, *A. absinthium*, *A. dubia* and *A. indica* [11, 12, 13]. But *A. annua* is suitable for cultivation and has been described as containing 0.5- 1.2% artemisinin in the dried plant material [14]. *A. annua* is economically the only natural botanical source for artemisinin production. The taxonomic classification of *A. annua* is as follows.

Division:	*Magnoliophyta*
Class:	*Magnoliopsida*
Sub-class:	*Asteridae*
Order:	*Asterales*
Family:	*Asteraceae*
Genus:	*Artemisia*
Species:	*annua*

A. annua originated from China and is widely distributed in the Northern Hemisphere but poorly represented in the Southern Hemisphere. 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 [15, 16]. It is also distributed in the temperate, cool temperate, subtropical zones and Mediterranean region of the world [16]. In addition, it grows widely in Canada and the United States. It has become naturalized in many countries including Argentina, Bulgaria, France, Hungary, Romania, Italy, Spain, the United States, and the former Yugoslavia etc. It is widely dispersed throughout the temperate regions of the world. It has been introduced for cultivation to Brazil, Cameroon, Ethiopia, India, Kenya, Mozambique, Myanmar, Tanzania, Thailand, Uganda, Zambia, etc. – all in high-altitude regions and/or regions with a pronounced cool period.

The plant is unadapted to the tropics because flowering will be induced when the plants are very small; with the possible exceptions of high altitude regions with a pronounced cool period such as Vietnam, where there are two months of low temperatures at the beginning of the season. In tropical zones with shorter day length variation and photoperiods, genotypes of *A. annua* will achieve only one meter or less in height before flowering and senescing, thus precluding production of sufficient

biomass per hectare to yield sufficient extract per unit land area. Reported temperature range for growth is 10-35 °C with the optimum between 13-29 °C and winter hardy down to about -15 °C. It is an annual plant with a growth cycle of about 180 days.

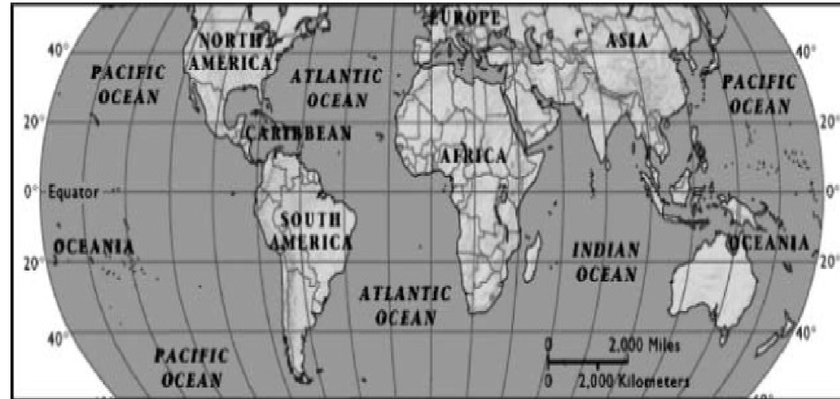


Figure 1.1. Range of *Artemisia annua* cultivation in the world according to latitude, and altitude (where known). In North America: West Lafayette, IN (40^0N, 184 m); Carbondale, IL (37^0N). Europe: Conthey, Swizerland (46^0N, ca 1330m). In Africa: Madagascar ($18^0$52'S); Calabar, Nigeria (5N, 60m); Kenya (1N); Tanzania (6S). In Asia: Chongqing, China (29N, 260m); Penang, Malasia ($5^0$30'N); Lucknow (26^0 51'N) and Kashmir (32-36^0N), India. In South America: Campinas, Brazil (23^0S,685 m); Teresina, Brazil (5^0, 75 m). In Australia: Devonport, Tasmania (43^0S). The species occurs naturally in northem parts of Chahas and Suiyuan provinces, China (40^0N). Sources: Singh *et al.*(1986), Delabays (1997), Ferreira *et al.* (1991), Laughlin *et al.* (2002).

1.2 Botany *A. Annua*

A. annua (Figure 1.2) is usually single-stemmed reaching about 2m in height with alternate branches. The plant is extremely vigorous and essentially disease and pest free. It has a short taproot and aggressive lateral fibrous roots as a result it thrives well to extreme environmental conditions. Leaves are spirally arranged, less often opposite, rarely whorled, simple, dissected or more or less compound. The foliage is fernlike. The leaves are aromatic, alternate, 2.5-5.0 cm long, bright green, deep and fine pinnately dissected into very narrow short, obtuse 5-8 lobes and of a papery texture (Fig 1.2). *A. annua* is a short-day plant and very responsive to short photoperiodic stimuli and flowers about two weeks after induction. They require about 1000 hours of sunlight per year. Annual sunlight time is a critical factor for the growth of *A. annua* [15]. *A. annua* grow in a well-drained and slightly alkaline loamy soil. Plants are longer lived, hardier and more aromatic when they are grown in a poor dry soil. Artemisinin is synthesized and sequestered in organs called glandular

trichomes [17, 18] (Figure 1.3). The glandular trichomes are more prominent in the corolla and receptacles florets than in leaves, stems, or bracts. There is strong evidence that artemisinin is sequestered in these glandular trichomes [17, 18]. Although these glands are present since the early stage of development on both leaves and inflorescences, artemisinin increases at anthesis, suggesting that it accumulates as the glands reach physiological maturity.

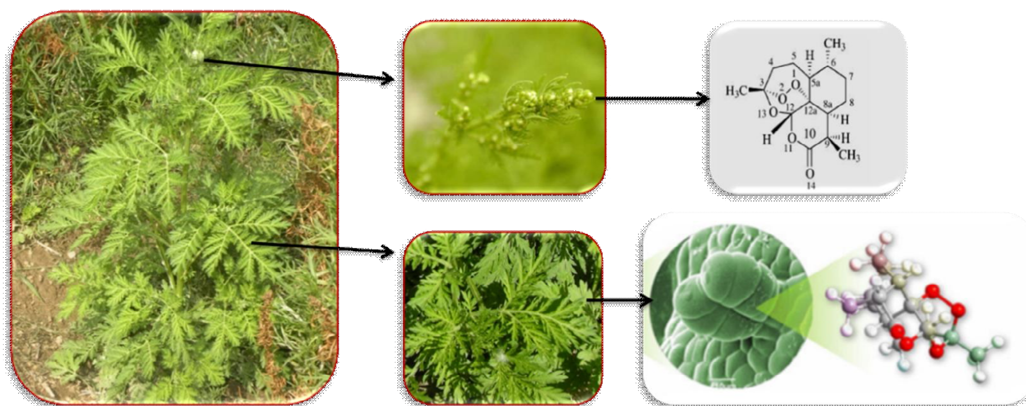


Figure 1.2. The plant *Artemisia annua*, its leaves and inflorescence as well as the trichome that accumulate artemisinin.

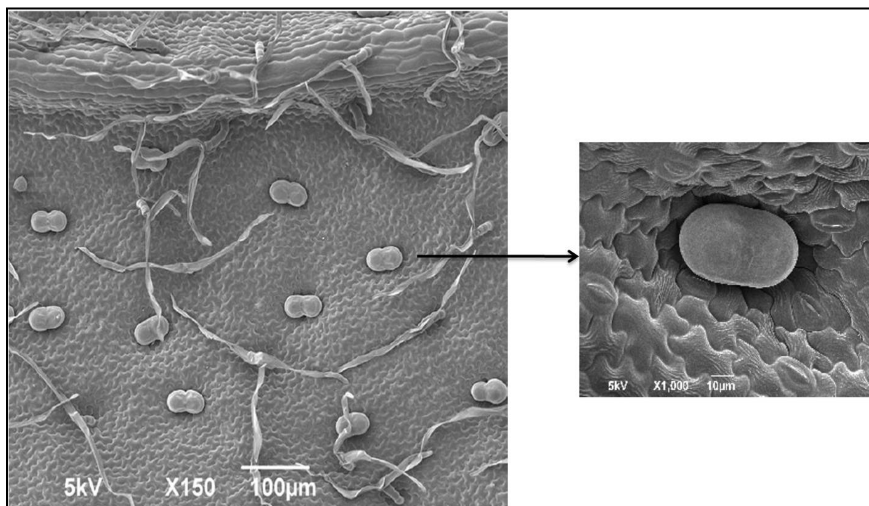


Figure 1.3 Scanning electron micrograph of *A. annua* leaf showing trichome that contains artemisinin. (source: http://www.york.ac.uk/org/cnap/artemisiaproject/index.htm).

The fruit of *A. annua* is an achene with a single seed inside. The seeds are approximately 1 mm in length, oblong, creamy endosperm with a lustrous surface marked by vertical furrows. The 1000 seed weight is approximately 0.03 g. The seeds

can be stored for an average period of 4 months if the water content is less than 13%. The seeds do not have a dormant phase and can be used in the same year or in the year following collection [16]. During the vegetative growth of *A. annua* trichome numbers increased per unit area on the adaxial leaf surface until leaf expansion ceased, at which point trichome numbers began to decline, apparently as a result of their collapse [19]. Leaves had 89% of the total artemisinin in the plant with the uppermost foliar portion of the plant containing almost double that of the lower leaves [20].

The genus *Artemisia* comprises around 500 taxa and has two basic chromosome numbers, the largely predominating x = 9 and the less extended x = 8 [21, 22, 23, 24]. A high percentage of *Artemisia* species are polyploid. Both basic chromosome numbers show polyploidy, with levels up to 12x for x = 9 and 6x for x = 8. The western North American subgenus Tridentatae based on x = 9, provides another robust example of polyploidy from diploid (2n = 18) to octoploid (2n = 72) levels. *Artemisia* chromosomes are rather small (2–8 µm) (Torrell et al., 2003). *A. annua* is a diploid plant (2n=2x 18). Genetic improvement of it is difficult because the flowers are around 1.0 mm. *A. annua* dies after seed set, thus plants with high artemisinin content cannot be maintained after flowering, and their seeds will be only the result of open pollination.

1.3 Artemisinin content in the plant

Artemisinin is one of the secondary metabolites in *A. annua* [25]. It is a sesquiterpene lactone with an endoperoxide bridge. Artemisinin is mostly extracted from the aerial parts of *A. annua* [26]. Roots and pollen were found to be deficient in artemisinin. It is present in main stems, side stems, leaves, and flowers of *A. annua* [16]. Table 1.1 shows the artemisinin content of different organs and structures of greenhouse- and field- grown *Artemisia annua*, determined by HPLC-EC. The association of artemisinin with glandular trichomes sequestration explains why artemisinin was not detected in parts of the plant that do not bear glands, such as pollen or roots. The leaves from the same plant may have different artemisinin contents, according to their localization along the stem: upper leaves contain significantly more artemisinin than middle and lower ones. The artemisinin content in plant also varies during the season. Under Swiss conditions, highest artemisinin

concentrations are found at the end of August, independently of the developmental stage of the plant. Furthermore, the genetic basis and environmental factors such as temperature or nutrient availability further influence the artemisinin content in the plant [27]. Results from experimental fields of CIMAP, subtropical north India, show that the yield of oil obtained from leaves and inflorescence vary according to the season of cultivation. This indicates that in the biosynthesis of the oil constituents of *A. annua*, temperature plays an important role. A significant difference in temperature of winter and monsoon months has been observed in the subtropical north Indian plains. The average temperature usually ranges between 10-26 °C in the winter and 25-33 °C during monsoon season. The results also show that as the plants matured, the oil content decreased in leaves from 0.32-0.14% and increased in inflorescence from 0.35-0.42% on a fresh weight basis [28].

Table 1.1 Artemisinin content in different component of *Artemisia annua* plant from greenhouse and field- grown estimated by HPLC-EC.

	Artemisinin (% dry weight	Artemisinin (% dry weight)
Organ / structure	Greenhouse	Field
Leaves	0.03-0.3	0.06-0.6
Main stems	0-0.03	0.04-0.07
Side stems	0	0.04-0.14
Roots	0	0
Flowers	0.12-0.42	0.104-0.264
Pollen	0	ND
Seed husks	ND	0.0116
Seeds	0.036	0.081

1.3.1 Importance of artemisinin and its derivatives

Artemisinin and its derivatives have been recognized as a new generation of powerful antimalarial drugs for combating the most popular infectious disease malaria worldwide. Artemisinin and its derivatives induce more rapid reduction of parasitemia [29], decreasing the number of parasites faster than any other known drug. As a consequence they are of special interest for severe malaria [30]. The fast decline in the number of parasites is also beneficial in combination therapies. Artemether and artesunnate were approved by the Chinese authority and collected in the "Essential Medicine List" by WHO. These derivatives have been successfully applied to remedy

several million malaria-suffering patients since their advent. Artemisinin has been used as a traditional medicine for at least 2000 years in China [31].

Antimalarial drugs derived from natural *Artemisia annua* have many advantages: quick reduction of fevers, fast clearing parasites in blood (90% of malaria patients recovered within 48 hrs) and no significant side effects. Experimental and clinical studies reveal that artemisinin, artemether and artesunnate are not only the potent antimalarial drugs but also the useful agents for other disease, especially as antiparasitic agent. In 1970s, artemether and artesunnate were confirmed to be more active than artemisinin in animal models. As a response to increasing levels of antimalarial resistance, WHO recommends that all countries experiencing resistance to conventional mono therapies should use combination therapies preferably those containing artemisinin derivatives (Figure 1.4) for effective treatment of malaria. Artemisinin based Combination Therapies (ACT) is a scientific approach to tackle this problem. It is a combination of artemisinin derivative drug with one or more long acting antimalarial drug having different modes of action and different drug targets. Artemisinin drugs have a short half-life of 1–4 h or so, but because of their strong anti-plasmodial activity, they reduce biomass of the existing parasites approximately 95% at each dosage of administration, and also kill the sexual stages of the malaria parasite. Residual parasites, if any, and the recrudescence are eliminated by the long acting antimalarial and the host immunity. WHO currently recommends the following combination therapeutic options :

- Artemether
- artesunate + maodiaquine
- artesunate + sulphdoxine/ pyremethamine
- artesunate + mefloquine
- amodiquine + sulphaodxine-pyrimethamine

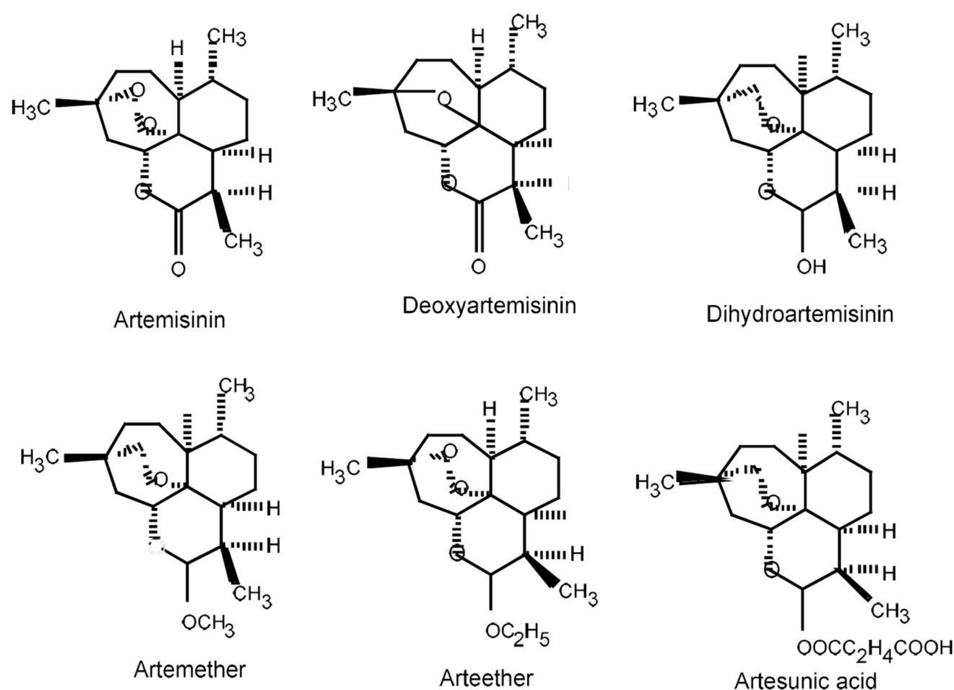


Figure 1.4 molecular structures of artemisinin and its structural derivatives.

The white needle crystals of artemisinin are hardly soluble in water or oil therefore formulations other than oral and rectal are not in clinical use. However, since the peroxide bridge is stable under certain chemical reactions, several more soluble artemisinin derivatives, arteether, artemether, sodium artesunate, sodium artelinate and dihydroartemisinin (DHQ) have been synthesized for the treatment of malaria. It is now universally accepted that this family of compounds is among the most powerful antimalarial drugs ever discovered. The pharmacological and clinical evidence is well documented [32, 33].

In addition, *A. annua* produces at least 36 flavonoids. Many of these have antimalarial activity *in vitro*, although the inhibitory concentration (IC_{50}) is much higher than that of artemisinin (Table 1.2). Five of these: artemetin, casticin, chrysoplenetin, chrysosplenol-D, and cirsilineol, have been shown selectively to potentiate the *in vitro a*ctivity of artemisinin against *P. falciparum* [34] Casticin, at a concentration of 5 mmol/l, induced a three- to five- fold reduction in the IC_{50} for artemisinin [35]. Chrysosplenol-D has the strongest potentiating effect, and this is also the most abundant flavone in plant material [34]. Interestingly, the flavones do not potentiate the antimalarial activity of chloroquine [35]. The effect of all the

flavones in combination with artemisinin has not been investigated. Other flavones, and indeed other components of *A. annua*, may have a similar effect; they have not been tested because it is difficult to purify them. The antimalarial properties of the traditional preparation of *A. annua* most probably reside in the combination of many constituents, not just artemisinin.

Table 1.2 In vitro animalarial activity of constituents of *A. annua*.

Constituent	IC_{50} (μ*M*)	Constituent	IC_{50} (μ*M*)
Artemisinin	0.03	Chrysosplenol-D	32
Artemetin	26	Cirsilineol	36
Casticin	24	Eupatorin	65
Chrysoplenetin	23		

Artemisinin and its derivatives were shown to be therapeutic against schistosomiasis disease caused by the protozoan species *Schistosoma japonicum, S.mansoni,* and *S.haematobium*. This disease is known to cause about 1.5 million disabilities each year [36]. Artemisinin is active against other protozoan infectious diseases such as leishmaniasis. In addition, artemisinin can prove lethal to cancer cell lines, fungi, and bacteria [37]. Also, it has been used to eliminate necrotic material from the body such as from wounds, intestines of people who have ulcerative colitis, and from Crohn's disease [38].

1.3.2 Artemisnin biosynthesis pathway

Artemisinin falls under the family of terpenoids. The primary precursor of terpenoid is isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphospahte (DMAPP). There are two pathways in which IPP is synthesized. One occurs in the cytosol, hence called the cytosolic mevalonic acid (MVA) pathway, and the other one occurs in the plastid, thus called the plastidial methylerythritol phosphate (MEP) pathway [39]. In the MVA pathway, IPP is derived from acetyl CoA. The conversion of 3-hydroxyl-3-methylglutaryl CoA (HMG-CoA) into mevalonic acid is catalyzed by an enzyme called HMG-CoA reductase. For the MEP pathway, IPP originates from the condensation of pyruvate with D-glyceraldehyde-3-phosphate [39]. This condensation reaction is catalyzed by the enzyme deoxy-D-xylulose-5-phosphate synthase (DXS), which represents the first step of the MEP pathway leading to the product 1-deoxy-D-xylulose-5-phosphate. IPP can be exchanged between the two

compartments, the cytosol and plastid [40] shown in Figure 1.5. A key enzyme responsible for catalyzing the step that produces many sesquiterpenes, including artemisinin, is farnesyl diphosphate sythase (FDPase). This enzyme catalyzes the condensation of the C5 substrate DMAPP with two molecules of IPP to the C15 product farnesyl diphosphate (FDP). FDP can then be cycled to sesquiterpene by a variety of sesquiterpene cyclase or form sterols.

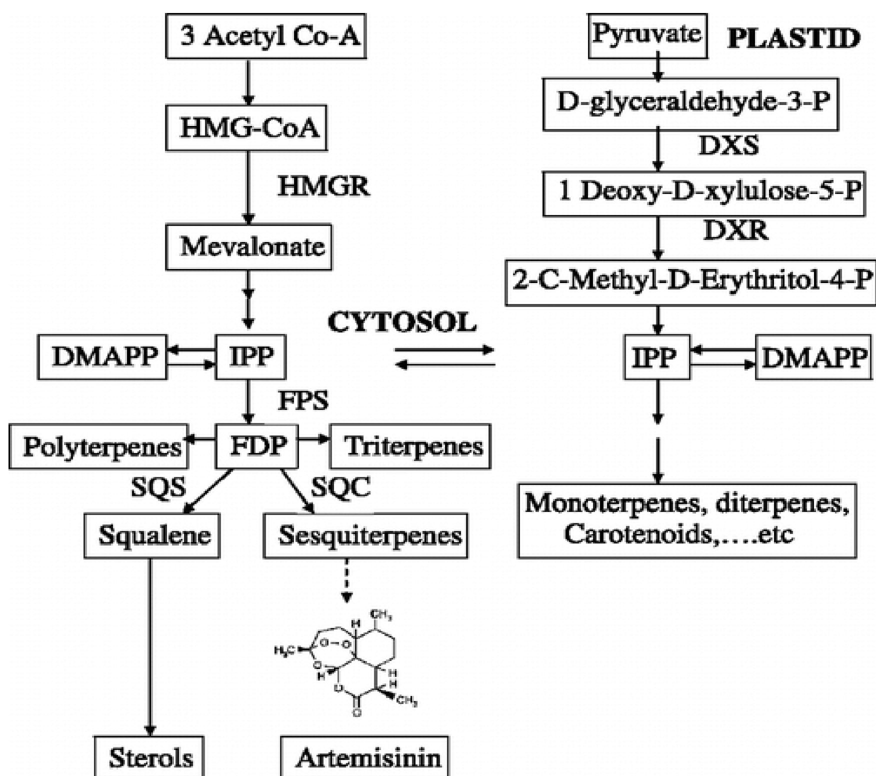


Figure 1.5 Simplified terpenoid biosynthetic schemes leading to artemisinin. The two arms of terpenoid biosynthesis and some of the regulatory enzymes that catalyze different reaction : HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; DXS, 1-deoxy-D-xylulose-5 phosphate synthase; DXR, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; FDPS, farnesyl diphosphate synthase; SQC, sesquiterpene cyclase; SQS, squalene synthase; IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate.

The committed step of artemisinin synthesis is the cyclization of farnesyl diphosphate (FDP) into amorpha-4, 11-diene by the enzyme amorphadiene synthease (ADS). The formation of artemisinic alcohol is catalyzed by the enzyme cytochrome p450 (CYP 71AV1). Artemisinic alcohol is oxidized by either CYP 71AV1 or artemisinic alcohol dehydrogenase into artemisinic aldehyde [41]. Artemisinic

aldehyde has two fates, and each reaction is catalyzed via a different enzyme as shown in Figure 1.6. CYP 71AV1 action leads to the formation of artemisinic acid, and artemisinic aldehyde hydrogenase (DBR2) activity leads to the formation dihydroartemisinic aldehyde. Dihydroartemisinic aldehyde forms dihydroartemisinic acid (DHAA) in a reaction step catalyzed by the enzyme dihydroartemisinic aldehyde dehydrogenase (AldH1). DHAA has only one possible route in the biochemical pathway. This route leads to the formation of artemisinin in a hypothesized non-enzymatic photo-oxidative step. On the other hand, artemisinic acid leads to the formation of arteannuin B [42].

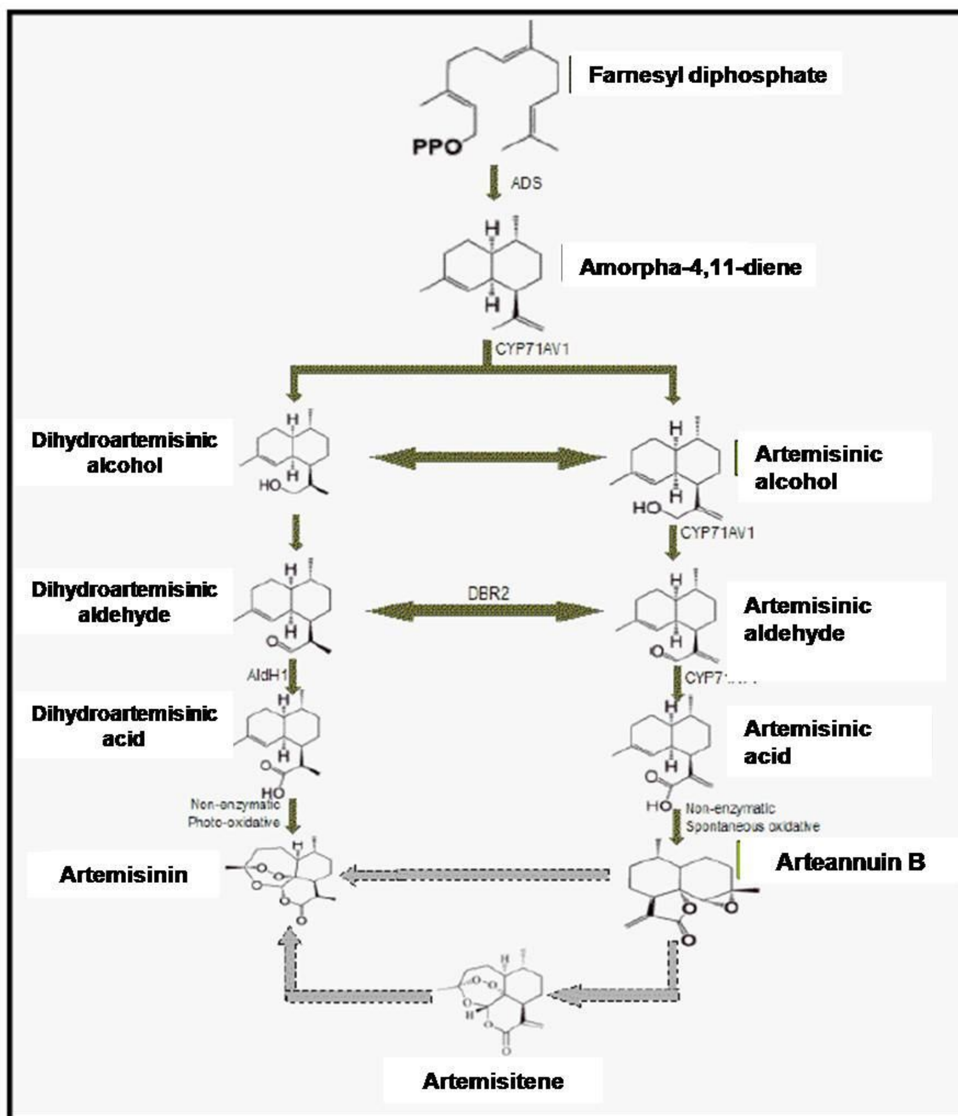


Figure 1.6. Alternative pathways leading to synthesis of artemisinin (source: Weathers et al., 2006, Zhang et al., 2008 and Wallaart et al., 1999).

1.3.3 Artemisinin content among *Artemisia* species

For a pharmaceutical industry supplying over 100 million treatments per annum [43], there is a growing concern that the artemisinin supply will be unable to meet future requirements [44]. Access of patients to artemisinin based combination therapy was inadequate in all countries surveyed during 2007 and 2008. Many researchers have reported that the yields of extracted artemisinin is very poor [10, 20] and much effort must be made to increase artemisinin content of *A. annua*. Because the plant material in wild is typically variable in its artemisinin content and plant biomass, this has an impact on drug extraction. Efforts are being made to increase its production in many ways such as plant tissue culture systems, [45] biotechnological approaches [46], agronomical practices [47]. But artemisinin was not found to be accumulated in callus and cell suspension cultures, it is presumed that the biosynthesis of artemisinin to be restricted to the green part of the plant [48]. Microbial-based systems to synthesize artemisinin precursors for chemical conversion have been reported [49, 50, 51]. However, at the present time, 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. The recent works showed that, the biosynthesis of artemisinin is almost completely elucidated [52], but production through synthetic chemistry is not economically feasible. As a result the population of *Artemisia annua* is under severe pressure [27]. As an alternative source of artemisinin, other species of *Artemisia* have been screened as potential new sources for agricultural production. Artemisinin has been reported in *Artemisia apiacea* and *Artemisia lancea* [53], *Artemisia cina* [51] and in aerial parts of *Artemisia sieberi* [54]. Presence of artemisinin has also been reported in *Artemisia absinthium* [55], *Artemisia dubia* and *Artemisia indica* [49] albeit a low concentration.

About 20 species of *Artemisia* were reported in the literature, showing variations morphologically (Figure 1.7). Many of these species have a long history of use as folk therapeutic plants. Also the artemisinin content among various species of *Artemisia* were reported to be significantly different (Figure 1.8).

Figure 1.7. List of various species of *Artemisia*: 1. *A. toggutica,* 2. *A.bravifolia,* 3. *A. desertorum,* 4. *A. dubia,* 5. *A. genipi,* 6. *A. glacialis,* 7. *A. indica,* 8. *A. japonica,* 9. *A.maritima,* 10. *A. moorcroftiana,* 11. *A. Parviflora,* 12. *A. roxburgiana,* 13. *A. vestita,* 14. *A. vulgaris,* 15. *A. absinthium,* 16.*A. Bushriences,* 17. *A. Japonica,* 18. *A. roxburghiana,* 19. *A. rox Gratae* and 20. *Umbelliformis.*

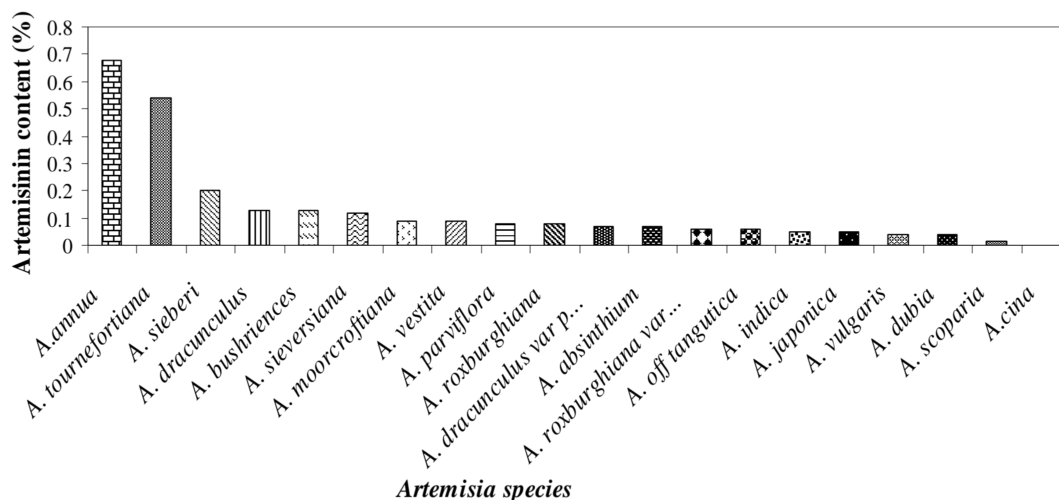


Figure1.8. Artemisisnin content of various *Artemisia* species reported from the literature [226].

1.3.4 *Artemisia* species in India and artemisinin content

India has a very rich plant biodiversity due to prevailing very highly divergent ecosystem and significant altitudinal variations ranging from sea level to the highest ranges of the Himalayas and allied factors. Cold deserts are usually confined to high altitudes and circumpolar regions of the Universe. The major portions under cold desert trans-Himalaya belong to Ladakh (Jammu & Kashmir) (Figure1.9). A very harsh and characteristic climate prevails in trans-Himalayan cold desert where the summer temperature touches up to 40 ^{0}C while the winter temperature drops below – 35 ^{0}C and are characterized by barren mountains, nail biting winters, low humidity and sparse vegetation. The cold arid regions of Jammu & Kashmir has been divided into five major geographical valleys for exploration and enlisting the plant biodiversity viz. Indus, Nubra, Changthang, Suru, and Zanskar. These cold arid regions are characterized with harsh environmental conditions as follows.

- Coarse, porous and immature sandy soils prone to wind and water erosions,
- Heavy influx of infra red and ultraviolet radiations
- Low air density – reduced oxygen levels
- Short agriculture (cropping) season 5 months (May-Sep)
- Very long freezing winters 5 to 7 months (Nov-April)
- Large variation in seasonal temperature : 40°C to -40°C

- Large diurnal variations in daily atmospheric temperature during cropping season ranging from 0°C to 35°C
- Fast blowing winds 40 to 60 km per hour mainly in afternoon hours,
- Precipitation mostly in the form of snow mainly during winters
- Irrigation water mostly from snow melts, major portion of which is unharvested
- Low relative humidity during growing season
- Soil moisture remains frozen during winters
- Very little to no rainfall.

Like other parts of the Himalayas, these regions are considered treasure of medicinal, aromatic and other important plants including *Artemisia*. Different species of *Artemisia* have been explored from this region of India as mentioned in Table 1.3. However, only *Artemisia annua* is so far use as the primary source of artemisinin.

Table 1.3. Various species of Artemesia reported from the Ladakh region of India.

SN.	*Artemisia sp.*	Location (Altitude)
1	*A. brevifolia*	Changthang (4,620 m), Suru valley (3,850 m)
2	*A. capilaris*	Pangong Tso (3,580 m)
3	*A. dracunculus*	Suru valley (3,850m) ,Khaltse (3,350 m)
4	*A. gmelinii*	Zanskar valley (5,091 m), Indus valley (3,450 m)
5	*A. minor*	Changla (4,710 m)
6	*A. macrocephala*	Khaltse (3,350 m)
7	*A. moorcroftiana*	North pullu (5,320 m)
8	*A. parviflora*	Kargil (2,860 m)
9	*A. salsoloides*	Pangong (3,750 m)
10	*A. stricta*	Changthang (4,350 m), Suru valley (3,850 m)
11	*A. tournefortiana*	Indus valley (3,450 m), Nubra valley (3050 m)
12	*A. wallichiana*	Tsottak (4,680 m)
13	*A. annua*	Indus valley (3,400 m), Nubra valley (3,050 m)
14	*A. sieversiana*	Indus valley (3,400 m), Nubra valley (3,050 m)
15	*A. biennis*	Changthang (3,380 m), Suru valley (3,015 m)
16	*A. desertorum*	Zanskar valley (3,605 m)
17	*A. laciniata*	Changthang valley (3,870 m)
18	*A. persica*	Suru valley (2,980 m)
19	*A. scoparia*	Changthang valley(3,810 m)
20	*A. stracheyi*	Changthang valley (4,560 m)

The important medicinal plants including *Artemisia annua* have been in use by the local tribes for day to day needs and this long dependence on plant wealth had enriched their knowledge about multifarious uses of plant biodiversity. Defence Institute of High Altitude Research of Defence Research & Development Organisation, Indus – Ladakh has carried out detailed ethnobotanical exploration covering far flung areas and higher passes of Ladakh and documented the plant wealth. During the ethno-botanical survey, the authors have observed that the frequency of some of important plants especially medicinal plants of Indian cold desert Himalaya with which it was earlier abound, have considerably declined due to their unscientific exploitation, natural calamities, road construction, uprooting for fuel, overgrazing and other activities. There is a need to conserve genetic diversity of this prized medicinal plant which may become extinct if its reckless exploitation continues.

1.4 Cultivation of *A. annua*

Chemical synthesis of artemisinin is not economically viable. Hence, artemisinin is still extracted from *A. annua* aerial parts. Therefore, the science of commercial cultivation of *A. annua*, to maximize artemisinin yields, should be well developed [56]. The overharvesting of *A. annua* from wild may restrict the ability for the plant to cross-pollinate and reseed naturally, eventually limiting the gene pool and genetic variability, which is critical to the development of improved seed lines.

In order to maximize the yield of artemisinin, the critical factor is day length, because the plant usually grows in the long summer days at high latitudes. The concentration of artemisinin peaks around the time of flowering, although in some cases this may be just before flowering and in other cases during full flowering [18, 56]. The artemisinin concentration peaks at a slightly different time in different areas, and identifying this will help to maximize the yield of artemisinin in harvested plants [57]. In the Tropics, where days are shorter than in northern summers, flowering occurs earlier, reducing the biomass achieved. However, yields can be maximized at higher altitudes and with late-flowering varieties [56], or by artificially lengthening hours of daylight to over 13.5 hours [18]. In wild-type plants, the greatest concentration of artemisinin is found in the inflorescence, although it occurs in all other aerial parts of the plant, except the seed [14]. Although artemisinin content is

affected by climate and time of harvesting, the main influence is genetic variation. Ferreira et al. [18] evaluated the same 23 clones of *A. annua*, which varied from 0.001 to 0.35% artemisinin, under tissue culture, greenhouse, and field conditions. Broad sense heritability analyses indicated that artemisinin was mainly under genetic, not environmental, control. Delabays et al. [27] confirmed that genes outplay the environment by studying different varieties, which yielded from 0.02 to about 1.4% artemisinin. Efforts have been made to increase the artemisinin content as far as possible, by exploring the natural variability. This has been achieved in a hybrid (*A. annua* var. *Artemis*, seeds available from www.anamed.org) and in a nonhybrid strain collected from Vietnam. However artemisinin yield depends not only on its concentration, but also on the total number of leaves and branches.

A. annua has been grown in a wide diversity of soils and latitudes, showing its potential for adaptation. If the *A. annua* cultivar and geographic region allow for a long vegetative cycle, more than one harvest can be performed to increase the final yield of leaves and artemisinin. In a study by Kumar et al. [58] with the *A. annua* cultivar Jeevanraksha, carried out in a subtropical climate in India for 3 years the crops were harvested once, twice, three times and four times, respectively, during a 1-year growth cycle. Very little published work exists on the vegetative growth responses of *A. annua* to the nitrogen, phosphorus and potassium or of their effects on the concentration of artemisinin and related compounds. Significant increase of total plant and leaf dry matter (1–3 t/ha) was obtained in Mississippi, USA, where a complete fertilizer mixture containing 100 kg N, 100 kg P and 100 kg K/ha was broadcast and worked uniformly through the soil [7]. In China, a range of growing media and micronutrients were tested for their effect on the synthesis of artemisinin. There were no effects on artemisinin from any of these treatments [59].

A. annua can be grown under a wide range of soil pH (5.0–8.0), depending on the plant origin, but there are only a few studies on the effect of soil pH on the vegetative growth and artemisinin concentration in *A. annua* L. It has been shown that some strains of *A. annua* are sensitive to soil pH below 5.0–5.5 [60]. Research is also needed on the development of new cultivation practices methods to robust yield of new breeding lines. Such methods will require collaboration between Agronomists,

breeders, farmers, and social scientists in developing techniques for initiating and sustaining farmer participation in yield improvement.

1.5 Breeding for increasing artemisinin content of *A. annua*.

A new breeding strategy of comprehensive integration of biotechnology and DNA marker applications with conventional backcross breeding techniques for *Artemisia* improvement should be developed. In designing breeding programs, germplasm collections are important reservoirs of genetic diversity and this diversity should exploit for artemisinin improvement. Artemisinin enhancing alleles may exist in wild *Artemisia* germplasms. Furthermore, this technique enables us to select potential parent lines of *A. annua.* Breeders should propose strategies for introgressing artemisinin alleles into elite cultivars. The scientific studies have shown that artemisinin content can vary widely among different cultivars or ecotypes of *A. annua* from different origins [61], so breeding for this trait may be feasible.

1.6 Genetic diversity of *Artemisia annua* populations

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. Considerable variations in morphological characters such as plant height, leaf characteristics, number of branches, seed weight and colour, etc. and in biochemical characters such as artemisinin content in leaves have been reported in *A. annua* plants from Central Institute of Medicinal and Aromatic Plants [62].

1.7 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. The drawbacks of biochemical markers led many workers to shift to nuclear DNA markers such as RFLPs, RAPDs, ISSRs, AFLPs, SSRs, ILPs, CAPs etc. DNA-based molecular-marker techniques have been proved powerful in genetic diversity estimations [63]. 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) [64, 65] and amplified fragment length polymorphism (AFLP) [66]. 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, [67].

1.7.1 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 [68]. 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 [69]. 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 [70]. 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 [71, 72] and have been used to measure genetic diversity in many plants.

The ISSRs have several advantages for assessing genetic diversity [68]. 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 [73]. The high stringency reduces the problems with reproducibility, a common criticism against the low-stringency RAPD assay [74]. 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 [75, 76]. 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 [77].

1.7.2 RAPDs as molecular markers

Random amplified polymorphic DNA (RAPD) is a PCR-based technique that has been applied to the study of populations [64]. RAPDs are one of a family of techniques that produce arbitrary fragment length polymorphism and are collectively described as multiple arbitrary amplicon profiling [78]. The RAPD technique utilizes single, arbitrary, decamer DNA oligonucleotide primers to amplify regions of the genome using PCR [79, 64]. 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 [80].

The use of RAPD for determination of genetic relationships has been demonstrated in a number of crop species like maize [81], Sorghum [82] rapeseed [83], pigeon pea [84] 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 [85,86]. In addition to the studies of genetic diversity there have been an increasing number of papers concerned with population genetics [87], phylogenetics [88] and hybridisation/introgression [89, 90]. 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 [91] and identification of patterns of variation [92]. Rieseberg [93] 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* [94], *Pistacia* [95] etc. The usefulness of RAPD fingerprinting was also reported for identification of Italian grape (*Vitis vinifera*) varieties [96] and determining the phylogenetic relationship for 28 tropical maize varieties [97]. 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. Only one report was so far available on RAPD characterization of *A. annua* from Central Institute of Medicinal and Aromatic Plants, Lucknow, India. [62].

1.7.3 SSRs molecular markers

Simple sequence repeats (SSR) or microsatellite from an important class of molecular markers for genomic and plant breeding applications. SSR markers quickly became the markers of choice for plant and animal genomes during the last decade because of the small sample size (genomic DNA) requirement for their analysis and

their suitability for automation and high-throughput analysis [98]. The presence of SSR in the transcript of genes suggests that they may have a role in gene expression or function [99]. While di-, tri- or tetra-nucleotide SSR are most commonly used for the construction of linkage maps of nuclear genomes, single nucleotide repeats have been used in the population genetic analysis of chloroplast genomes.

SSR can be assayed using PCR technology and can be screened using high-throughput platform for molecular genetic linkage and population studies [100]. The number of repeats at a locus can change by mutation and the rate of mutation depends on the number of tandem units within the repeat [101]. Another area where SSR markers are extremely valuable and are increasingly becoming popular is comparative genomics where SSR markers developed from one species could be utilized in a related or heterologous species towards genetic mapping, characterization, gene cloning, diversity, evolutionary studies of genetic variation, linkage mapping, gene tagging, establishment of genetic maps, integration of physical and genetic maps, determination of evolutionary relationships and comparative genome analyses [100, 102]. SSRs, once developed are extremely valuable, although their development is time consuming, laborious and expensive. SSRs derived from ESTs essentially represent expressed gene sequences and hence are potential candidates for markers for comparative genomic studies. ESTs are particularly attractive for marker development since they represent coding regions of the genome and are also being developed at an extremely faster pace for many genomes.

1.7.4 ILPs molecular markers

Intron length polymorphisms (ILPs) have been used as genetic markers in some studies. However, a systematic investigation and large-scale exploitation of ILP markers has not been reported. Introns are non-coding sequences in a gene that are transcribed but spliced out of the precursor mRNA [103, 104]. Introns are widespread and abundant in eukaryotic genomes. Generally speaking, introns have little functional significance, although some introns may influence the level of gene expression [105]. Therefore, introns are more variable than coding sequences. Intron polymorphisms can also be exploited as genetic markers. They have been successfully utilized in population genetics surveys [106, 108] and gene mapping [108]. There could be various polymorphisms in introns, but intron length polymorphism (ILP) is

the most easily recognizable type. It can be conveniently detected by the PCR. To amplify introns by PCR, primers can be designed in flanking exons. This approach is called exon-primed intron-crossing PCR (EPIC-PCR) [109]. ILP has many similar advantages to SSR including specific (being a STS marker), codominant (providing complete information of genotypes), neutral (no phenotypic effect), convenient (detectable by PCR) and reliable (result stable). In addition, ILP has a special advantage, namely, it directly reflects variation within genes. Therefore, the genetic maps constructed with ILP markers would be more valuable for genetic studies because they are similar to conventional maps consisting of morphological markers. Moreover, ILP marker would be more useful for marker-assisted breeding because it allows us to trace a gene directly as long as an ILP can be found in the gene.

1.8 Optimization of factors and ex situ conservation of *Artemisia annua*

An alternative approach to protecting the genetic diversity of *A. annua* is to standardize the synthetic protocol for artificial synthesis of artemisinin in the laboratories. However, the total synthesis of artemisinin is complicated and has not been established. Although attempt has been made to produce artemisinin in genetically modified microorganisms the success rate is not significant. Hence, different species of *Artemisia* are presently the commercial source of artemisinin for the pharmaceutical industry. The demand for the compound continues to increase and thus encourages ex situ conservation of *A. artemisia* in the Himalayan region. In an attempt at ex situ conservation Field Research Laboratory of Defence Research & Development Organisation, Indus – Ladakh has already growing *A. annua* as well as other species of Artemisia at two different valleys: Indus (3500 m) and Nubra (3000 m) (Figure 1.8) that are geographically distinct.

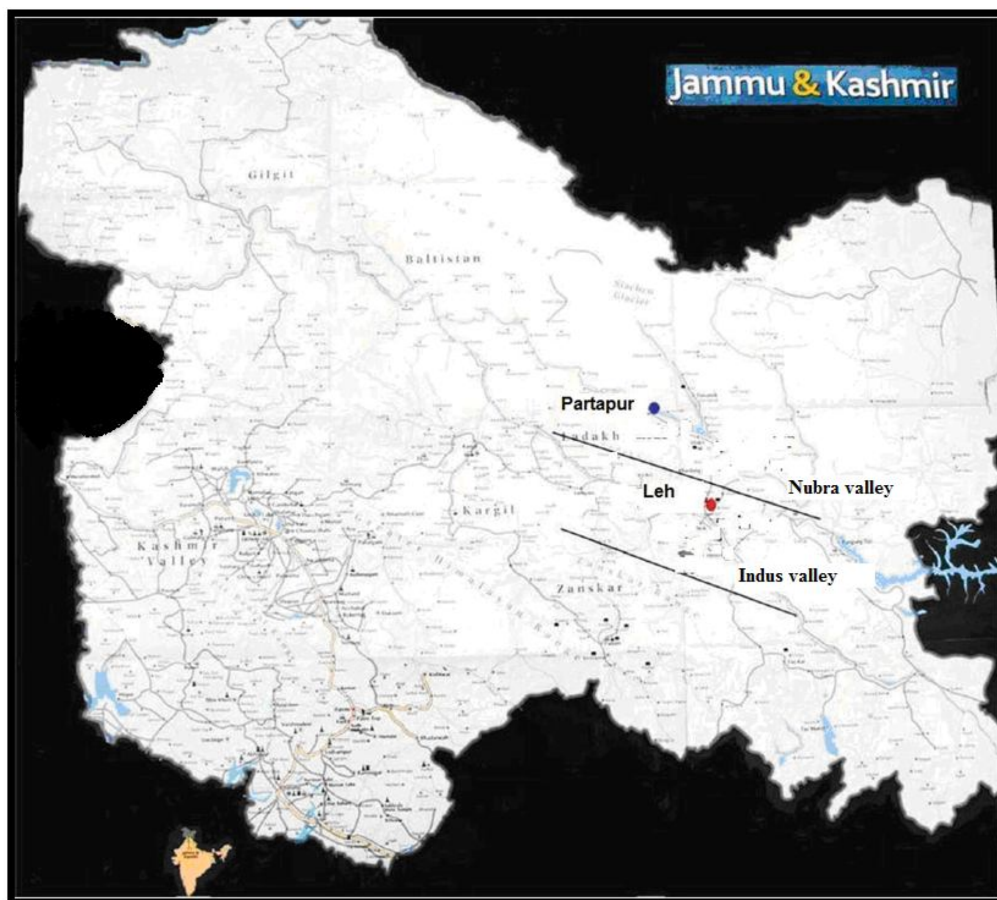


Figure 1.9. Geographical map representing two valleys: Indus (3500 m) and Nubra (3000 m) located in Ladakh (Jammu and Kashmir, India) that comprised of *Artemisia annua* genotypes.

The extreme environmental conditions in Ladakh may influence the artemisinin content of the plant. This is obvious as the plant is very adaptable to a wide range of environmental conditions. It can survive under varying conditions and adapt well from the extreme low winter temperature of the Northern climates to the high summer temperatures and at the altitudes. A number of investigations have demonstrated that the quality and quantity of several secondary metabolites have a close relationship with plant habitats [100]. Hence, it demands management of soil nutrients and optimization of climatic factors for the successful conservation of *A. annua*. We are interested in whether the soil nutrition and climatic factors contribute to the high yield of artemisinin in the natural habitats. No work has been done previously in these aspects nor has any prediction model been developed for selection of conservation area for successful domestication of *A. annua.*

Statistical methods such as artificial neural network (ANN) and multiple linear regressions (MLR) are very useful in these respects. MLR has been used to explain the spatial variations in soil nutrients and its impact on crop yield at field scale [111]. However, MLR requires a normal distribution of the input variables which is not always the case [112]. The non-linear predictors such as ANNs were widely used to solve various problems in agriculture. For example, Sudduth et al., [113] successfully predicted corn yield with back propagation neural network models based on soil texture, topography, pH and some soil nutrient elements. This prediction model was superior to those of the nonparametric statistical benchmark methods.

Each of these five 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 for background to conclusion at publication stage.

CHAPTER 2

Genetic characterization of *Artemisia annua* using RAPD and ISSR molecular markers from the trans-Himalayan (Ladakh, India) region

Abstract

Artemesia annua is an important medicinal palnt valued all over the world. Genetic characterization of 20 genotypes of *A. annua* collected from two valleys *viz.* Nubra (9,600 ft) and Indus (11,500 ft) of the trans-Himalayan (Ladakh, India) region were analyzed using 37 PCR markers (20 RAPDs and 17 ISSRs). RAPD analysis yielded 124 polymorphic fragments (96.9%), with an average of 6.2 polymorphic fragments per primer. ISSR analysis produced 85 bands, of which 78 were polymorphic (86.1%), with an average of 4.58 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. The genetic diversity was high among the genotypes (Nei's genetic diversity = 0.336 and Shannon's information index = 0.495) as measured by combination of both RAPD and ISSR markers. The mean coefficient of gene differentiation (Gst) was 0.145, indicating 85.5% of the genetic diversity resided within the genotypes. RAPD markers were found more efficient with regard to polymorphism detection, as they detected 96.9% in comparison to 86.1% for ISSR markers. Genotypes collected from Ladakh region were clustered distinctly into two groups as per their sampling sites using RAPD, ISSR and in combination of RAPD+ISSR based on neighbor joining (NJ) method. A relatively high genetic variation was detected among the 20 genotypes whereas the variation between the two valleys was less using AMOVA test. It was found that the genetic diversity among genotypes from Nubra valley was narrow than that of Indus valley, suggesting the importance and feasibility of introducing elite genotypes from different origins for *Artemisia* germplasm conservation and breeding programs.

2.1 Introduction

The herb *Artemisia annua* L. (*Asteraceae*) is one of the most important medicinal plants of the modern world for the production of anti-malarial and possibly antibacterial agents and natural pesticides. It was originally collected by the Chinese as an herbal medicine and is currently processed by pharmaceutical firms for the production of artemisinin for artemisinin-based combination therapies (ACT) in the treatment of malaria. The artemisinin is effective against *Plasmodium falciparum* and *P. viva*, including multi-drug resistant strains [114, 115]. It is native to Asia, the center of origin is most probably China; wild populations occur both in China and Vietnam. However, due to its high demands it has been introduced to Brazil, Cameroon, Ethiopia, India, Kenya, Mozambique, Myanmar, Tanzania, Thailand, Uganda, Zambia, Italy, Spain, Uninet States, France, Hungary, Argentina, etc. all in high altitude regions and/or regions with a pronounced cool period [115]. It is one of the most important medicinal plants of India, especially Ladakh region. It thrives well in dry cold region of Ladakh having marginal rocky and sandy soils due to its aggressive fibrous root system. 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 or in modified micro-organisms have so far not been very fruitful. Thus, the naturally growing as well as the cultivated plants is the sole source of the drug [116]. This has generated worldwide interest in studying the genetic diversity of *A. annua* populations, clonal variants, chemotypes, and ecotypes, and in the synthesis of pure-line cultivars. Over the years, the *Artemisia* populations in the Ladakh region have developed considerable variability pertaining to genetic characterization of this plant from this region. No reports are available so far for the genetic characterization of this plant from Ladakh region and thus necessitate detail investigation. It is prerequisite towards effective utilization and protection of plant genetic resources [117], identification of molecular markers linked to agronomic traits and to achieve rational conservation.

Unlike the morphological and biochemical markers which may be affected by environmental factors and growth practices [118, 119], 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 decades. Restriction fragment length polymorphism (RFLP) was the first molecular marker [120], developed for genome analysis and mapping. The developments of the polymerase chain reaction (PCR) technology has introduced a considerable number of useful molecular markers, e.g., random amplified polymorphic DNA (RAPDs) [64] and inter simple sequence repeats (ISSR) [68] which are independent of environmental factors and unaffected by developmental stages of the plant. These markers have been used both for DNA fingerprinting [121] and population genetic studies [122]. The only requirement is good quality of DNA. Currently, there are several protocols available for isolation of DNA, which are variants of the few principal protocols (123, 124, 125, 126]. Experience has shown that extraction 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. Particularly, medicinal and aromatic plants like *A. annua* are rich in the myriad of natural products and they require additional efforts to find an appropriate procedure for reliable and consistent results.

In this study we have tailored the DNA extraction protocols extensively in order to get clean and intact DNA from *A. annua*, suitable for PCR amplification. The objective of this study was to compare the effectiveness of both the PCR-based molecular approaches to determine the genetic relationships among several genotypes of *A. annua* from the trans-Himalayan (Ladakh, India) region.

2.2 MATERIALS AND METHODS

2.2.1 Plant materials

Twenty genotypes of *A. annua* were collected from two valleys at different altitudes; 3000 m (Nubra) and 3500 m (Indus) 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 100-200m and the pair wise distance between valleys was 50–250 Km.

2.2.2 DNA extraction and PCR amplification

We modified the CTAB protocol [125] for extraction of DNA using fresh plant materials: (i) incubation time of buffer and tissue mixture at 65 ^{0}C, (ii) buffer to tissue ratio and (iii) extraction with phenol:chloroform:isoamyl alcohol vs. **t**ris saturated phenol followed by chloroform:isoamyl alcohol in extraction and purification phases. All the experiments were repeated 3-4 times to check reproducibility.

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

DNA EXTRACTION

Leaves samples were rinsed with distilled water and blotted gently with soft tissue paper. About 0.1 g of this tissue were taken and ground to fine powder using liquid nitrogen, with a precooled mortar and pestle along with 2% of extraction buffer and 10 mg of PVP (Sigma). The powdered tissue was scraped into a 2.0 ml microcentrifuge tube containing preheated (65 ^{0}C) extraction buffer (1:5). After that β-mercaptoethanol was added to the final concentration of 0.2 M and mixed well. The mixture was incubated in water bath at 65 ^{0}C 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 for second time. The aqueous phase was pipetted out gently, avoiding the interface. To the above solution, 5 M NaCl (to final concentration of 2M) and 0.6 volume of isopropanol was added and incubated at room

temperature for 1 h. Two volumes of 80% ethanol was added and incubated again for 10 min at room temperature in order to precipitate the DNA. 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.

Purification phase

The extracted DNA was then treated with 20 µL of 10 mg/ml of RNase A and incubated at 37 ^{0}C 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 centrifugation 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 volume of chloroform:isoamyl alcohol (24:1), twice. The DNA was precipitated by adding 0.6 volume of isopropanol and 2.0 M NaCl. 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. The yield of the extracted DNA and purity was checked by running the sample on 0.8% agarose gel along with standard (non restriction enzyme digested) lamda DNA marker (Biogene, USA).

2.2.3 RAPD analysis

Twenty random decamer primers from IDT Tech, USA (Table 2.1) were used for RAPD amplification following the protocol of [64]. PCR 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.

2.2.4 ISSR analysis

Seventeen ISSR primers were obtained from Applied Biosciences, India (Table 2.2) and PCR amplification was performed in reaction cocktail similar to RAPD. Initial denaturation for 4 min at 94 °C was followed by next 40 cycles of denaturation at 94 °C for 45 second, 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. PCR products were stored at 4 °C before analysis.

Table 2.1. 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	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**	-

Table 2.2 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	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**	**-**

The amplification for each primer was performed twice independently with same procedure in order to ensure the fidelity of RAPD and ISSR markers. Amplified PCR 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}). 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 sizes of amplicons were estimated using a 1 Kb DNA ladders ('Bangalore Genei, India).

2.2.5 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 dissimilarity coefficient (J) was calculated, subjected to cluster analysis by bootstrapping and neighbor-joining method using the program DARWIN (version 5.0.158). Statistically unbiased clustering of collected genotypes was performed using STRUCTURE (version 2.3.1) [127]. POPGENE software was used to calculate Nei's unbiased genetic distance among genotypes. 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 [128]. Within species diversity (Hs) and total genetic diversity (Ht) [129] 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) [130], using two hierarchical levels; among valleys and among genotypes. Correlation between both the marker types used in the study was obtained by regression 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 [131]. 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. [132]. *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, $MI = DI_{avp} * E$.

2.3 RESULTS AND DISCUSSION

The *Artemisia* species from high altitude (Indus and Nubra valleys) 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 modified the CTAB protocol of genomic DNA extraction and obtained good quality DNA from the leaf sample. The modification includes addition of 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 [133] 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.

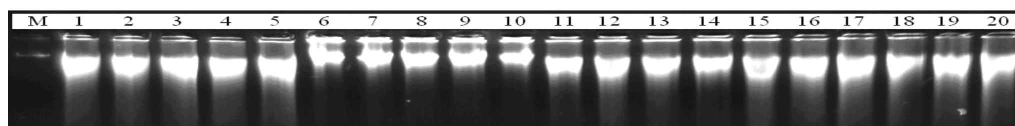


Figure 2.1 Agarose gel electrophoresis showing purified high molecular weight *Artemisia annua* genomic DNA of different Species, more than 200 ng genomic DNA from each genotype was electrophoresed on 0.8% agarose gel at 65 V for 2 hr and stain with ethidium bromide. Lane M, 100 ng mol. wt. uncut λ DNA. Lane 1 to 10; are the genomic DNA extracted from Indus valley and Lane 11 to 20; are from Nubra valley.

Most of the protocols in the literature used phenol: chloroform: isoamyl alcohol (25:24:1) or chloroform: isoamyl alcohol (24:1) [123,134], is used for removing protein. 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 *A. annua* 5:1 buffer to tissue ratio and 90 min incubation at 65^0C gave best results. Figure 2.1 represents the quality of the DNAs that have been extracted.

RAPD analysis

Twenty RAPD primers generated reproducible, informative and easily scorable RAPD profiles were preselected. These primers produced multiple band profiles (Figure 2.2) with a number of amplified DNA fragments varying from 4 to 9, with a mean of 6.2 bands per primer. All the amplified fragments varied in size from 200-1000 bp. Out of 124 amplified bands, 119 were found polymorphic (Table 2.1). 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. 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 2.3a). An unbiased clustering of genotypes based on STRUCTURE program, without prior knowledge about the populations, clustered all the 20 genotypes into two major groups. Under the admixed model, STRUCTURE calculated that the estimate of likelihood of the clustering of data was greatest when K = 2, ΔK reached its maximum value when K = 2 (Figure 2.3b), suggesting that all the populations fell into one of the 2 clusters (with respect to valleys). It was found that the genotypes were more likely distributed (at high probability) with respect to their geographical distribution albeit small interference (Figure 2.3c).

A relatively high genetic variation was detected among the genotypes. Genetic diversity analysis in terms of Na, Ne, H, I, Ht, Hs, and PPL for both the valleys (Indus and Nubra) revealed higher values for Indus, indicating more variability among the genotypes in comparison to Nubra valley (Table 3). Analysis of molecular variance among valley (10%) and among genotypes within valley (90 %) (Table 5) revealed higher variations within the population. 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 [135], *Podophyllum hexandrum* [122] and *Andrographic paniculata* [136] suggested that RAPD is more appropriate for analysis of genetic variability in closely related genotypes. It indicates that *A. annua* populations in the northwestern Himalayan region are genetically highly diverse.

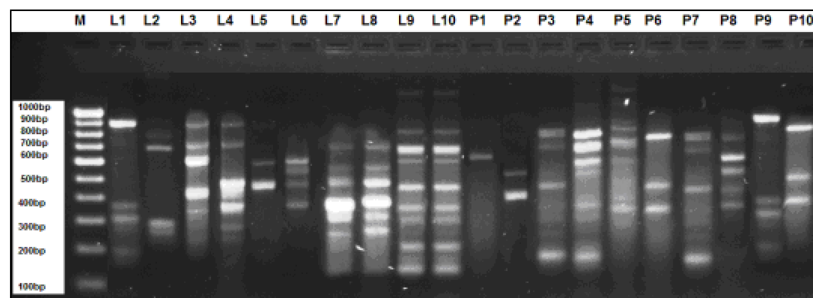


Figure 2.2. RAPD amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10, are the genotypes collected from Nubra (Partapur) valley. M = the size of molecular markers in base pairs using λ DNA.

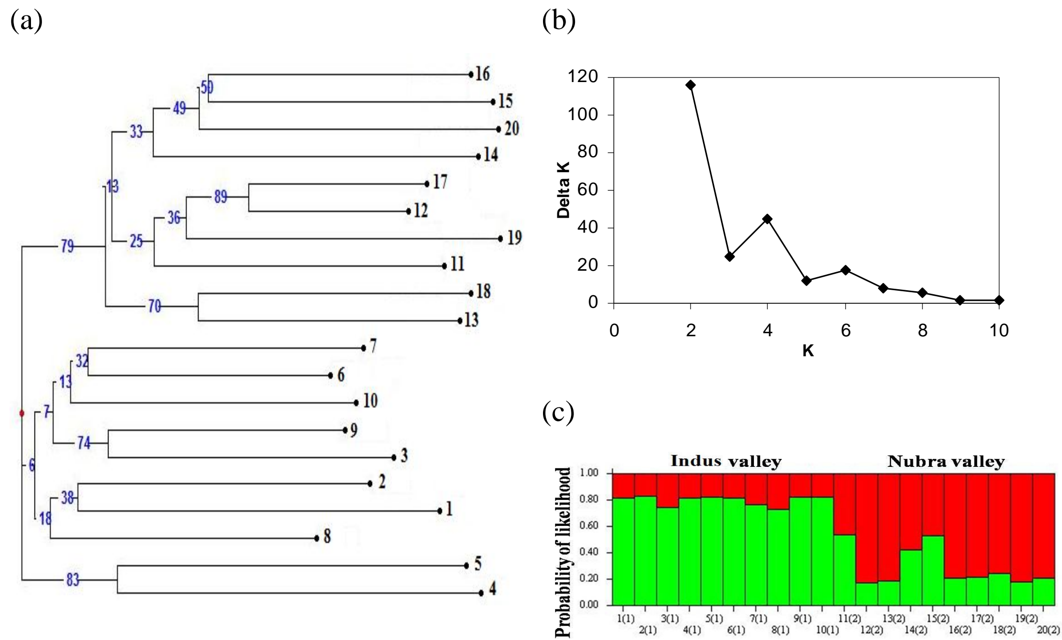


Figure 2.3. (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on RAPD profiling, (b) the relationship between K and ΔK, (c) unbiased clustering of genotypes between 2 groups. (c) Statistically unbiased clustering of 20 genotypes as per 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 2.3. Summary of genetic variation statistics for all loci of RAPD 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.

Table 2.4. Summary of analysis of molecular variance (AMOVA) based on RAPD analysis. Levels of significance are based on 1000 iteration steps.

Source of variation	Degree of freedom	Variance component	Percentage of variation	P-value
Among valley	1	2.962	10	< 0.001
Among genotypes	18	24.994	90	< 0.001

2.3.2 ISSR analysis

The 17 ISSR primers selected 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 2.4). Average number of bands and polymorphic bands per primer were 5 and 4.23 respectively, other primer amplification details are shown in Table 2.2. 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. 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. 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 [137]. Similar results were obtained in grapevine [138], rice [139], *Vigna* [140] and wheat [141]. 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 [139] or it may be due to its-non annealing with template DNA due to its low Tm. 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. [140], 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 to 5.67 (Table 2.2).

The complete data set of 1087 bands was used for cluster analysis based on bootstrapping and NJ method. The genotypes were clustered into two clusters (with respect to their site of collection) where, cluster I represents all the genotypes from Indus valley while cluster II contains all the genotypes from Nubra valley (Figure 2.5a). The estimated likelihood of the clustering of data using STRUCTURE was found to be optimal when K = 2, ΔK reached its maximum value when K = 2 (Figure

2.5b), suggesting that all the populations fell into one of the 2 clusters (with respect to valleys). It was found that the genotypes were more likely distributed (at high probability) with respect to their geographical distribution albeit small interference (Figure 2.5c).

A relatively high genetic variation was detected among the genotypes. Genetic diversity analysis in terms of Na, Ne, H, I, Ht, Hs, and PPL for both the valleys (Indus and Nubra) revealed higher values for Indus, indicating more variability among the genotypes in comparison to Nubra valley (Table 3). AMOVA for among valley (11%) and among genotypes within the valley (89%) indicated that there are more variations within the population (Table 2.7). The estimated gene flow was 2.044.

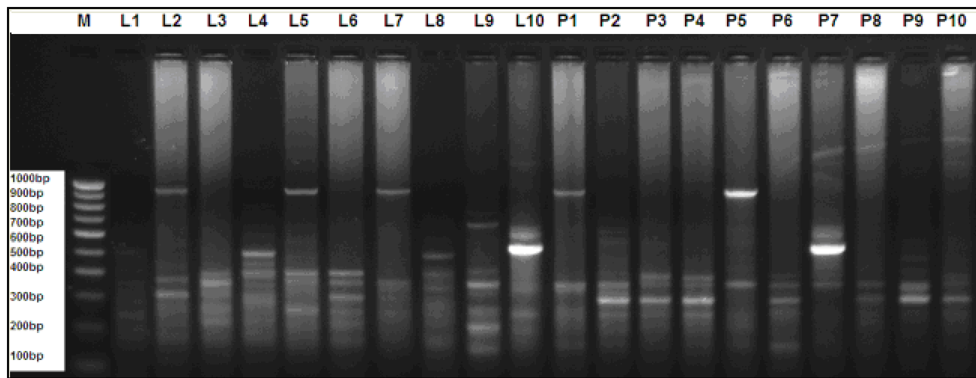


Figure 2.4. ISSR amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra valley. M = the size of molecular markers in base pairs using λ DNA.

Table 2.5. Summary of nested analysis of molecular variance (AMOVA) based on combination of RAPD and ISSR analysis among the populations of *A. annua.* Levels of significance are based on 1000 iteration steps.

Source of variation	Degree of freedom	Variance component	Percentage of variation	P-value
Among valley	1	6.168	21	< 0.001
Among genotypes/valley	18	23.622	79	< 0.001

P-value = probability of null distribution.

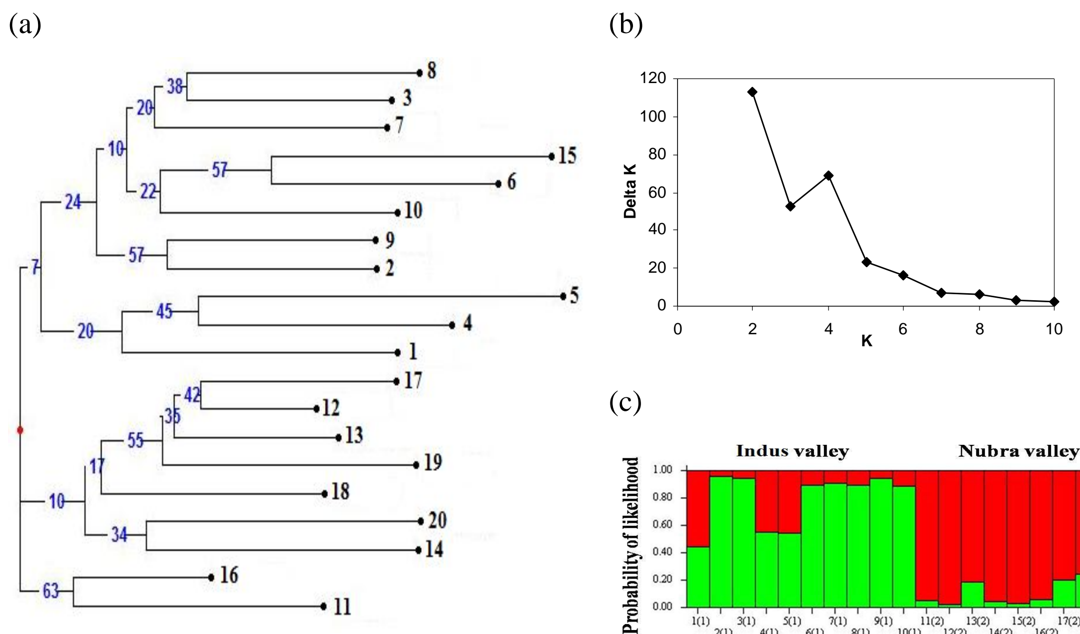


Figure 2.5. (a) NJ tree representing clustering of genotypes along with supported bootstrap values based on ISSR profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, the relationship between K and ΔK, (c) The genotypes were more likely clustered corresponding 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).

2.3.3 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 2.7a). 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 (Figure 2.7b and 2.7c). Other genetic variation studies were also performed on RAPD and ISSR combined data which are represented in different tables (Tables 2.8 and 2.9). The differences found among the dendrograms generated by RAPDs and ISSRs could be partially explained by the different number of PCR products 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 as observed by Loarce et al. [142] in barley. Another explanation could be the low reproducibility of RAPDs [143]. The genetic closeness among the Indus valley and Nubra valley genotypes can be explained by the high degree of commonness in their genomes. Similar result has been obtained by

Gaffor et al. [144] in blackgram. In all the dendrograms, 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.

Table 2.6. 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

2.3.4 Comparative analysis of RAPD with ISSR markers

RAPD markers were found more efficient with respect to polymorphism detection, as they detected 96.9% polymorphism as compared to 86.02% for ISSR markers. This is in contrast to the results obtained for several other plant species like wheat [141] and *Vigna* [140]. 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 2.2). While in case of RAPD, all the 20 primers which were used in the investigation amplified 1482 number of fragments (Table 2.1). Similar polymorphism pattern was also observed in case of *Jatropha* [12] and *Podophyllum* [122]. The regression test between the Jaccard's similarity matrix resulted in low regression between RAPD and ISSR based similarities (R = 0.014), moderate for ISSR and RAPD+ISSR (R= 0.699), while it is maximum for RAPD and RAPD+ISSR based similarities (R = 0.725). This shows that RAPD data is more close to RAPD+ISSR 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. The diversity index, effective multiplex ratio and marker index are more for RAPD than for ISSR markers (Table 2.9). 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. Marker index of ISSR was more (0.876) in comparison to RAPD (0.851), indicating ISSR is the powerful molecular marker for genetic characterization of *A. annua* genotypes. AMOVA for among valley (21%) and

among genotypes (79%) indicated that there are more variations within the population.

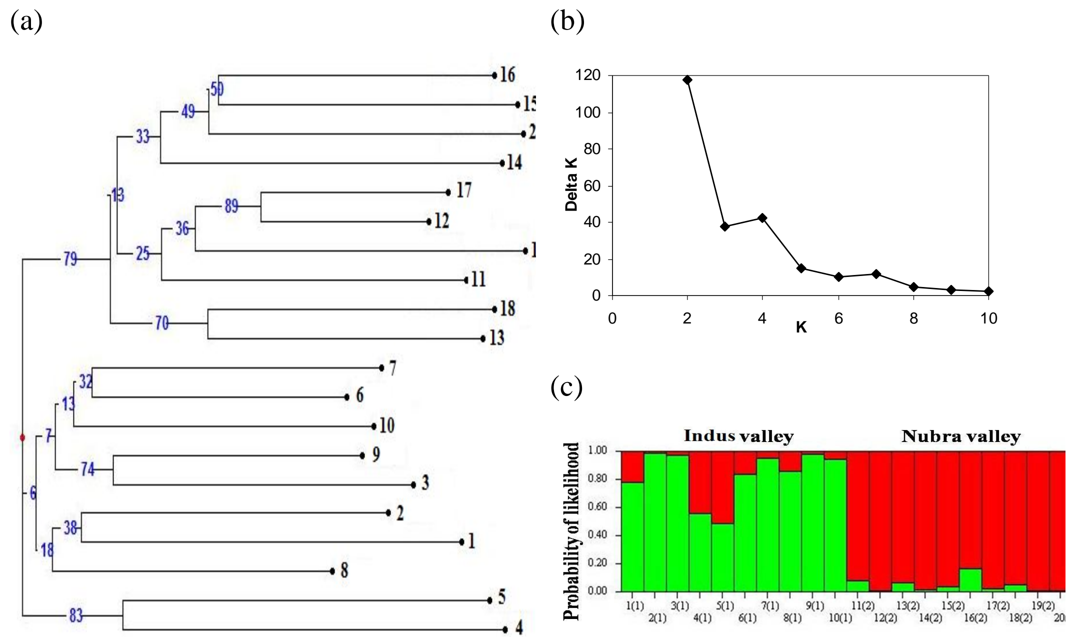


Figure 2.6 (a) NJ tree representing clustering of genotypes at populations' level along with supported bootstrap values based on combination of RAPD and ISSR profiling (1482 RAPD bands+1087 ISSR bands), (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, showing the relationship between K and ΔK, (c) 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 2.7. Summary of analysis of molecular variance (AMOVA) based on ISSR analysis among the populations of *A. annua*. Levels of significance are based on 1000 iteration steps.

Source of variation	Degree of freedom	Variance component	Percentage of variation	P-value
Among valley	1	1.859	11	< 0.001
Among genotypes	18	15.461	89	< 0.001

P-value = probability of null distribution.

Table 2.8 Summary of genetic variation statistics for the combination of RAPD + ISSR loci 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

Table 2.9. Overall genetic variability across all the 20 genotypes of *Artemisia annua* based combination of RAPD and ISSR markers.

Marker Type	Na	Ne	H	I	Ht	Hs	Gst	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	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	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	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.

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. 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.

CHAPTER 3

Utilization of intron-flanking EST-specific markers in the genetic characterization of *Artemisia annua* genotypes from the trans-Himalayan region of Ladakh, India

---

ABSTRACT

Genetic variation was assessed utilizing intron-flanking EST-specific markers among genotypes of *Artemisia annua* collected from two sampling sites *viz*. Nubra (9,600 ft) and Indus (11,500 ft) valleys of the trans-Himalayan region, Ladakh, India. The available ESTs (85,282) sequences of *A. annua* were aligned with the genomic sequences of *Arabidopsis* and rice to developed 'intron-flanking' EST-PCR based primers. These primers anneal with the conserved region of exon (flanking to the intron) and amplified the introns. Out of the 39 primers selected and tested on 20 genotypes of *A. annua*, we successfully exploited 81 codominant intron length polymorphic (ILP) markers, with an average of 2.08 markers per primer and 92.04% polymorphism detection. Clustering of genotypes revealed distribution of genotypes into 2 distinct clusters with respect to their site of collection. Significantly, this study demonstrates that *Arabidopsis* and rice 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.

3.1 INTRODUCTION

Molecular markers are powerful tools for genetic research and breeding. Many types of molecular marker have been developed since 1980, such as restriction fragment length polymorphism [145] random amplified polymorphic DNA [64], amplified fragment length polymorphism [66], simple sequence repeat polymorphism (SSR; Becker and Heun, 1995), single-nucleotide polymorphism [147] and intron length polymorphism [148]. Depending on the classification purpose, functional as opposed to non-functional DNA sequences may be preferred because the former should be more closely related to actual physiological and/or structural differences [149, 150]. Depending on the depth of phylogenetic analysis and taxon relatedness, degree of resolution may be reduced or lost with the use of slowly evolving DNA sequences/genes, whereas it may be artificially increased using fast-evolving sequences. For example, microsatellite/SSR analysis, although efficient for discriminating between closely related individuals and material with a narrow genetic base, may be less effective in determining relations at higher taxonomic level [74]. Introns, originally thought of as junk DNA, and thereof presumed to evolve under minimal constraints in a fashion consistent with the neutral theory of sequence evolution ([151], introns have now become of increasing interest as sources of DNA polymorphism. Recent reports have shown intron length polymorphism to be a convenient and reliable molecular marker with high interspecies transferability and it can be exploited for the construction of genetic maps because it directly reflects variations occurring within genes [105].

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. For example, the average number of SNPs per 1000 bp in introns (12.1) is over three times as high as that in exons (3.6) among eight varieties in rice (*Oryza sativa* L.)[152]. Length polymorphism is the most intuitive variation in introns. So, ILPs have been exploited as molecular markers, which have many desirable properties, including specific, codominant, neutral, convenient and reliable. However, ILP has not been widely utilized because it is newly developed as a kind of molecular marker and the number of ILP markers having been exploited in plants is still very limited. To date, studies of exploiting ILP markers have been restricted to a few species [148, 153]. ILP can be conveniently

detected by the PCR. To amplify introns by PCR, primers can be designed in 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 in exons may have more extensive applications than those designed in non-coding sequences. Using this approach, Bierne *et al.* [154], developed several ILP markers in penaeid shrimps. Obviously, the key point of developing ILP markers is to identify suitable introns. A general method for identifying introns is to compare cDNA/EST sequences with the genome sequence. Therefore, introns can be easily identified in many model organisms. However, this method is not applicable to most other organisms because they have only cDNA/EST sequences available. Fortunately, studies have indicated that the exon-intron structures are largely conserved among homologous genes from different species [155]. Therefore, the joints between adjacent exons after splicing (or termed 'intron positions' for short) in a cDNA/EST of an organism can be deduced according to the homologous genes from related model organisms. This provides a way of developing ILP markers in any organisms. Now a day, many organisms have got a large number of cDNA/EST sequences available in public databases. Therefore, large-scale exploitation of ILP markers in various organisms becomes possible.

Artemisia annua is a vigorous, annual, aromatic, herbaceous plant reaching 1-3 m height and 1 m in width. It has many bioactive compounds including anti-malarial, antibacterial agents and natural pesticides. It was originally collected by the Chinese as an herbal medicine and is currently processed by pharmaceutical firms for the production of artemisinin for Artemisinin-based Combination Therapies (ACTs) in the treatment of malaria. Artemisinin has proved to be a dramatically effective antimalarial against multi-drug resistant *Plasmodium* species [156, 157]. In addition to antimalarial effects, artemesinin was found to have promise in treating the parasitic diseases schistosomiasis and clonorchiasis (common in China and Africa, affecting over 200 million people each year) caused by trematodes (blood flukes). Many other effects of *A. annua* products have been described in various sources: antiperiodic, antiseptic, digestive, febrifuge; an infusion of the leaves is used internally to treat fevers, colds, diarrhoea etc.; externally, the leaves are poulticed onto nose bleeds, boils and abscesses. Genetic resources are threatened as harvesting of natural wild plant populations is done by cutting whole plants before flowering for the extraction

of drugs. Therefore it will be prudent to study the genotypes of this prize worthy plants at genetic and molecular levels for efficient conservation and management of genetic diversity. Genetic variability also can be exploited to select useful genotypes that could be utilized as cultivars to avoid batch-to-batch variation in extraction of standard drugs.

To our knowledge, there is no published information on the use of ILP markers for the characterization of genetic diversity in *A. annua*. Hence, the objective of the present study was to use ILP marker for the molecular characterization of genetic variability of *A. annua* genotypes from the cold arid deserts of the trans-Himalayas. We performed a genome-wide search for ILPs and a large-scale exploitation of candidate ILP markers via electronic -PCR based on the released EST sequence data in *A. annua*. We developed a set of ILP markers selected from the candidates and investigated their characteristics by experiment.

3.2 MATERIALS AND METHODS

3.2.1 Sources of sequence data

The dicot model plant *Arabidopsis* and monocot model plant rice were taken as subject species; whereas the available EST sequence data from *Artemisia annua* were taken as query species. The genome, cDNA and CDS (coding sequence) data of rice (*Oryza sativa* L. ssp. japonica cv. Nipponbare) and *Arabidopsis* (*Arabidopsis thaliana* ecotype Columbia) were downloaded from http:// www.tigr.org/ and http://www.arabidopsis.org/, respectively. The EST sequences of *A. annua* (~85,282) were downloaded from the website http://www.plantgdb.org/. The redundant ESTs sequences were removed using BLASTCLUST (http://www.ncbi.nlm.nih.gov/ BLAST) with sequence similarity 60% and 100% sequence coverage. After removing the redundant ESTs the total numbers of ESTs left were 68,974.

3.2.2 Development of putative intron polymorphic (PIP) markers

An in-house pipeline perl script was used to predict putative intron polymorphic (PIP) markers from *Artemisia* ESTs. We identify possible introns in *A. annua* by aligning EST sequences with *Oryza sativa* and *Arabidopsis* CDSs using BLASTN (Figure 3.1). A query EST was thought to be homologous to a subject CDS only if there were at least 100 bp overlapping and 80% similarity between them. The

corresponding position and length of identified introns from the subject species were obtained from the PIP database [158]. For each of these query ESTs, a pair of primers was designed using program Primer3 [159] from the query EST with 100 bp on each side of the target intron. The designed primers were tested by electronic PCR (e-PCR) on the EST sequences of *Artemisia annua*. A putative intron was taken as a PIP marker if the e-PCR yielded the unique product as expected. To increase the quality and usability of the *in silico* exploited ILP markers, we required exact matches between primers and templates and set a high margin value on the product size for the e-PCR. We took a putative ILP locus as a candidate ILP marker when it was successfully and uniquely detected by the e-PCR. For the experimental validation of putative ILP markers we considered only those ESTs containing possible positions of introns from the genes of artemisinin synthesis pathway (Table 3.1) for primer designing. We selected 40 candidate ILP markers (Table 3.2) and were synthesized by GBiosciences Inc.

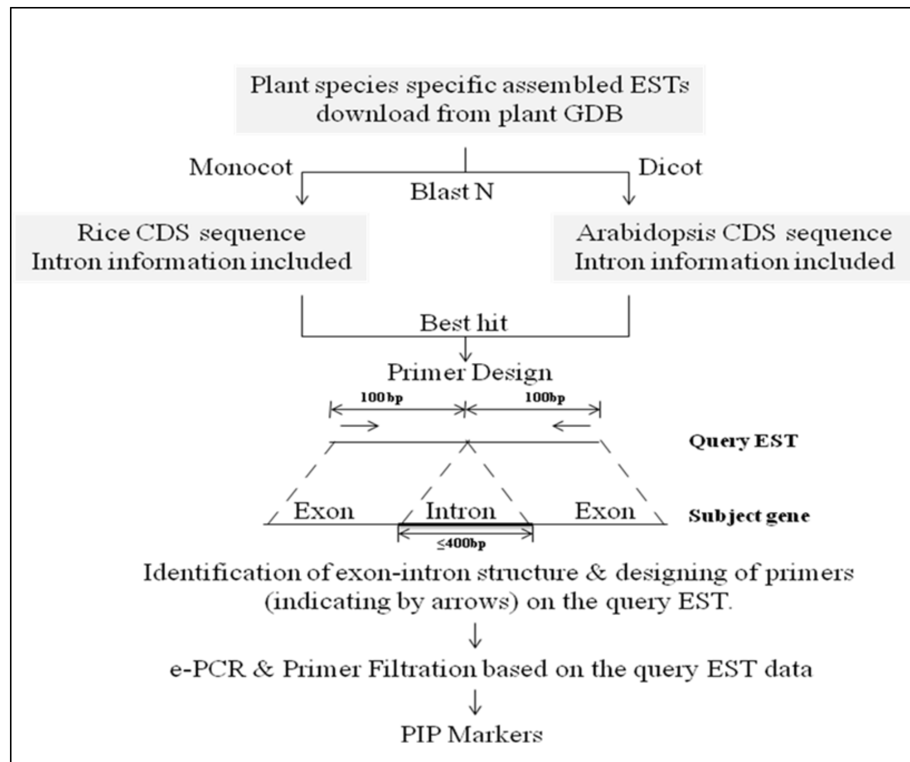


Figure 3.1 Flowchart of developing PIP markers in query sequences. Adopted from Long Yang, et al., PIP: a database of potential intron polymorphism markers.

Table 3.1 The enzymes and genes involved in artemisinin biosynthetic pathway in *A.annua*.

Enzyme	Gene	Function	Genbank accession number
Deoxyxylulose synthase	Dxs	1-deoxy-D-xylulose 5 phosphate synthase	AF182286
Deoxyxylulose reductase	Dxr	Isomerase and oxydoreductase	AF182287
3-hydroxy 3-methyleglutaryl coenzyme reductase	Hmg	Catalyses the two step redction of (S) 3-hydroxy 3-methyleglutaryl coenzyme-A(S-HMG-Co A) into R mevalonate	AF 142473
Famesyl diphosphate synthase sesquiterpene cyclases	Fps	Synthesis of FDP Catalysis cyclisation of FDP to:	AF112881
Epicedrol synthase	Eps	8-Epicedrol	AJ001539
Amorphadiene synthase	Ads	Amorpha-4,11 diene	AJ251751
2-caryophyllene synthase	cs(qhsl)	2-Carophyllene	AF 472361
2- famesene synthase	Fs	2-Farnesene	Ay 835398
Putative sesquiterpene cyclase	cascl 25	Isoprenoid biosynthesis and lyase activity	AJ271792
	case 34	Isoprenoid biosynthesis and lyase activity	AJ2717193
	Ses	Reaction product not determined	AAD39832
Squalene synthase	Aasqs	Famesyl- diphosphate famesyle transferase activity	AY445506
Squalene synthase fragment	Sqsl	Transferase activity	AF182286
CYP71AVl	Na	Catalyzes 3 steps post ADS	DQ315671
PsbA fragment	Psba	Act as barcode for flowering plants	DQ006143
Ribulose -1,5 bi phosphate carboxylase/oxygenase	Rbcl	Carbon dioxide fixation , barcoding for flowering plant	DQ006057
Peroxidase 1	Podl	Favored the bioconversion of artemisinis acid to artemisinin	AY 208699
Beta –pinene synthase	Qh6	Circadian pattern of expression	AF276072
(3R)-linalool synthase	Qhl	Lyase activity	AF154125
Isopentenyl transferase	Ipt	Biosynthesis of the cytokinin phytohormones	M91610

Table 3.2 List of primers used for ILP study, there GC content, melting temperature (Tm) and polymorphism.

Sl. No.	Gene Name	Primer sequence	Length of primer	% GC	Tm	No. of loci	NPL	PPL
1L	Aasqs	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	Ads	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	cyp71av1	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	Dxs	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	Dxraa	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	idi1	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	pod1	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	Eps	3' ACCATGGATGATCTTGGTTCCCA 5'	23	47.82	55	2	2	100
40R		5'TCTGTAAGACATGGTAGCTCACGA 3'	24	45.80	55			
Total						88	81	92.04

NPL, Number of polymorphic loci and PPL, Percentage of polymorphic loci

3.2.3 DNA extraction, PCR amplification and analysis of data

Genomic DNA from twenty genotypes of *A. annua* was isolated and purified as previously described. Amplification reactions were performed by ‘touchdown PCR’ in volumes of 25 µl reaction mixture containing 50 ng template DNA, 0.5 mM primer, 200 mM of each dNTPs, 15 mM $MgCl_2$, 0.1% Triton X-100, and 1unit of Taq polymerase. To the reaction mixture 1.5 ml of 10X PCR buffer was added and incubated at 94 ^oC for 4 min to denature the template DNA followed by 10 cycles of denaturation at 94 ^oC for 45 s, primer annealing at 59 ^oC minus 0.3 ^oC/cycle for 30 s and primer extension at 72 ^oC for 1 min. In the next 20 cycles, the annealing temperature was reduced to 56 ^oC 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 ^oC for 5 min. The amplification products were separated in 3% metaphor-agarose gel electrophoresis (100 V, 2 hr) and band sizes were determined against ‘1 kb plus DNA ladder’ (Invitrogen Life Technologies, Carlsbad, CA, USA). The banding patterns were scored as 1 (if it is present) or 0 (if it is absent)

and this binary file was used for genetic diversity analysis using various programs as described above.

3.3 RESULTS AND DISCUSSION

3.3.1 Candidate ILP markers in *Artemisia annua*

ILP markers were applied to characterize and compare genetic diversity among two populations of *A. annua*. ESTs sequences of plant species are valuable sources of information for PCR-based gene-specific markers studies and gene mapping. However, as a result of unavailability of genomic DNA sequence in many non-model plants, the potential utility of PCR-based EST markers (in terms of their ability to generate high frequency of band polymorphism) has not been fully exploited. It is generally believed that intron regions are more divergent than exons (Choi *et al.*, 2004). Therefore the EST–PCR primers that are designed to anneal to exons, amplify across intron regions and resulted in relatively higher band polymorphism. In the current study EST-PCR primers were designed based on the predicted exon/intron junction sites in *A. annua*. These splice sites were identified by aligning the ESTs of *A. annua* with the genomic sequence of *Arabidopsis*. Using BLASTN, out of ~85,282 ESTs sequences from *A. annua* we detected 40,439 ESTs homologs with *Arabidopsis.* We successfully obtained 19,885 (73.86%) e-PCR products as putative ILPs in *A. annua* based on the EST-PCR primers designed. However, we failed to acquire e-PCR products from ~1/3 putative ILPs in *Artemisia*, probably due to the several constraint conditions set during primer designing and e-PCR. It is noted that the 19,885 ILP loci detected by e-PCR in *Artemisia* were located only on 13,254 ESTs. This result also reflects the nonrandom distribution of ILPs in *Artemisia* genome.

3.3.2 ILP analysis

Forty intron-flanking EST-primers were used for PCR amplification and subsequent polymorphism detection. Out of the 40 primers tested, 39 primers produced PCR amplification with the success rate of 98% (Table 1). To increase PCR specificity in the amplification of ILP loci, we adopted a Touchdown-PCR (Td-PCR) program. Most of the primers (23) amplified well by the Td-PCR with the default initial annealing temperature (59 ^{o}C). Some other primers (8) required a lower (57 ^{o}C) or higher (60 ^{o}C) initial annealing temperature. The remaining 8 primers produced

good amplification at a constant annealing temperature of 54 °C. A representative gel demonstrating the successful amplification of introns from the genomic DNA isolated from different genotypes is shown in Figure 3.2. The genotypes collected from two different valleys 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.

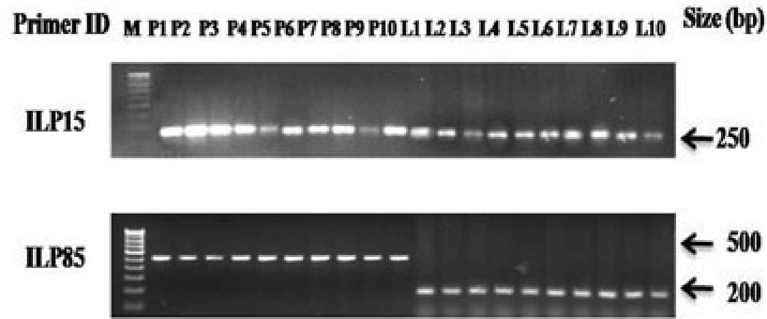


Fig 3.2. ILP amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10 are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = 100 bp λ DNA ladder.

3.3.3 Phylogenetic analysis across twenty genotypes

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). The STRUCTURE analysis predicted number of cluster for better grouping of data at K = 2 (Figure 10b), suggesting that 20 genotypes were better clustered with respect to both the valleys. It was found that the genotypes were more likely distributed (at high probability) with respect to their geographical distribution albeit small interference (Figure 10c).

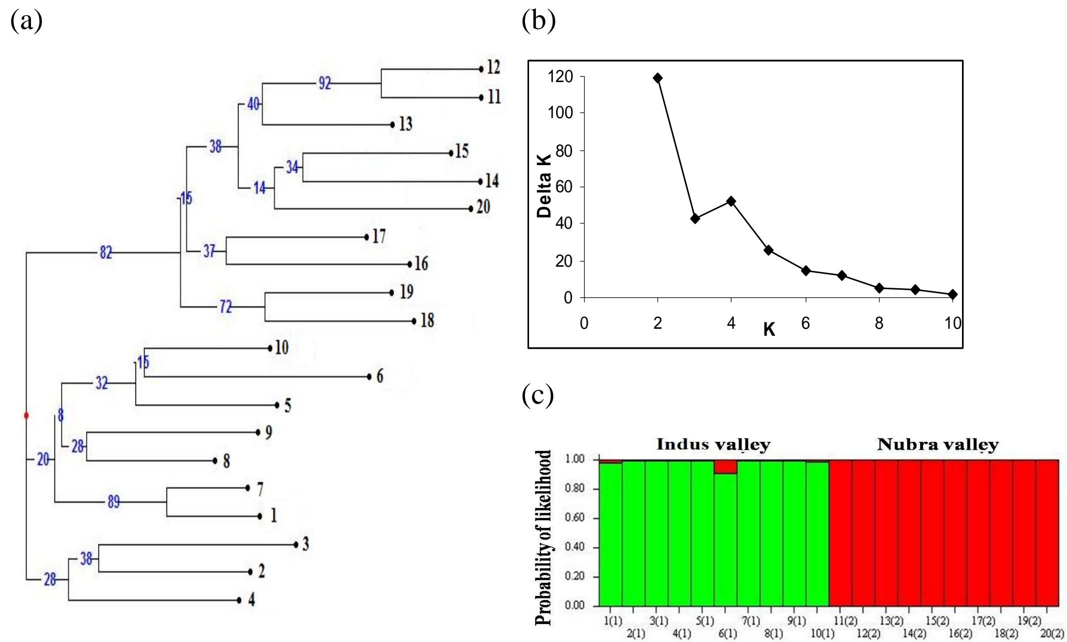


Figure 3.3. (a) Neighbor joining (NJ) tree representing clustering of genotypes at populations' level along with supported bootstrap values based on ILP profiling, (b) Statistically unbiased clustering of 20 genotypes as per their sampling sites, showing the relationship between K and ΔK, (c) the genotypes 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).

3.3.4 Genetic diversity analysis

Relatively high genetic variation was detected among the *A. annua* genotypes. Genetic diversity analysis in terms of Na, Ne, H, I, Ht, Hs, NPL and PPL between both the valleys, revealed higher values for Indus valley indicating more variability among the genotypes in comparison to Nubra valley (Table 2). The respective values for overall genetic variability for Na, Ne, H, I, Ht, Hs and Gene flow (Nm) across all the 20 genotypes are given in Table 3. AMOVA helps in partitioning of the overall ILP variations among valley and among genotypes. Molecular variance revealed higher variations among genotypes (61%) in comparison to among valley (39 %). All the components of molecular variations were significant ($P < 0.01$). The occurrence of more genetic variance of *A. annua* among genotypes as reported in our study is a pattern often described for mountain plants [160], for endemic plants from Tibet [161] and from central Asian desert plants [162] and several other studies of out crossing endemic species [163]. Similar reports have been made in ISSR studies of Populations of *R. crenulata* from Hengduan Mountain Region, China, where among genotypes

diversity was 55.14% while among population diversity was 44.8 [164]. This is helpful in making strategy for germplasm collection and evaluation. The rate of gene flow estimated using Gst value was found to be 0.612. 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 [165]. 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 [166].

Table 3.3 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	Gst
Indus	10	1.7500 ±0.4359	1.5141 ±0.3583	0.2958 ±0.1899	0.4343 ±0.2688	0.2958 ±0.0361	57	75	1
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	1
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; Gst = genetic diversity between population; NPL = number of polymorphic Loci; PPL = percentage of polymorphic Loci.

Table 3.4. Overall genetic variability across all the 20 genotypes of *Artemisia annua* based on ILPs analysis.

Na	Ne	H	I	Ht	Hs	Gst	NPL	PPL	Nm
1.934 (0.250)	1.747 (0.296)	0.405 (0.138)	0.581 (0.183)	0.405 (0.019)	0.288 (0.033)	0.290	71	93.42	0.612

Hs = genetic diversity in population; Nm= estimate of gene flow from Gst; Nm = 0.25 (1-Gst)/Gst.

Table 3.5 Summary of nested analysis of molecular variance (AMOVA) based on ILPs, among the populations of *Artemisia annua.* Levels of significance are based on 1000 iteration steps.

Source of variance	d.f	SSD	Variance component	Percentage	P-value
Among valley	1	89.25	7.711	39	< 0.01
Among genotypes/valley	18	218.5	12.139	61	< 0.01

The study indicates that *A. annua* populations in the northwestern Himalayan region are genetically highly diverse which should be exploited for future conservation and breeding of *Artemisia* from this region. Genetic variation among elite genotypes of *A. annua* based on ILP analysis could be useful in selecting parents to be crossed for generating appropriate populations intended for both genome mapping and breeding purposes. Considering the high genetic differentiation among the wild populations of *A. annua*, conservation of only a few populations may not adequately protect the genetic variations within the species in the north-western 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 will not be adopted. 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 a sufficiently large population 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 ILP 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.

CHAPTER 4

Utilization of EST-derived SSR in the genetic characterization of *Artemisia annua* genotypes from the trans-Himalayan region (Ladakh, India)

ABSTRACT

Artemisia annua is an important medicinal plant that produces substantial quantities of artemisinin (antimalarial agent). In India it is grown wild in the Ladhak region and has developed considerable variability over the year. We have used EST derived SSR markers to measure the genetic diversity among the *A. annua* germplasm collected from geographically separated Indus (11,500 ft) and Nubra (9600 ft) valleys (Ladakh, India). We investigated 68,974 non-redundant (out of 85,282 available) ESTs of *A. annua* and developed 4,342 SSR markers. On an average, one SSR was found per 8.9 kb of EST sequence with di-nucleotide motifs in highest frequency (52.2%) followed by tri (42.4%), tetra (3.6%), hexa (1.2%) and penta-nucleotide (0.6%) repeat types. A set of 16 primer pairs were designed by considering only the SSR-containing ESTs from the artemisinin biosynthetic pathway. A total of 38 alleles were identified from 13 polymorphic SSR loci, ranging from 1-7 alleles per locus and displayed moderate genetic diversity with an average of 0.24. It was found that the genetic diversity among genotypes from Nubra valley was narrow than that of Indus valley, suggesting the importance and feasibility of introducing elite genotypes from different origins for *Artemisia* germplasm conservation and breeding programs.

4.1 INTRODUCTION

Artemisia annua L. 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 from cold arid regions of India especially in Ladakh region. It is well adapted to high altitudes of 9600-11500 ft above mean sea level (MSL) and thrives well to the temperature, nutrient and environmental stress they are exposed. Over the year, the *Artemisia* populations in the Ladakh region have developed considerable variability; necessitate the study of genetic characterization. The plant synthesizes and accumulates substantial quantities of many derivatives of a cadiene skeleton including artemisinin (endoperoxide seco sesquiterpene lactones). Artemisinin is currently the most effective agent against multidrug resistant strains of *Plasmodium* species, the malarial parasites (114, 115]. Due to this discovery, *A. annua* is now rated as one of the top ten industrial medicinal plants of the modern world.

Chemical synthesis of the artemisinin is commercially non-viable, and efforts to produce artemisinin in engineered yeast cells as well as in tissue cultured cells have so far not been very fruitful. Thus the sole source of the drug is from the widely grown or cultivated plants [116]. Since the production of artemisinin varies among the genotypes, it has generated worldwide interest in studying the genetic diversity of *A. annua* populations, cloned variants, chemotypes, ecotypes and in the synthesis of pure-line cultivars. No reports are available so far regarding the genetic characterization of this plant from Ladakh region and detailed investigation is required. For effective utilization and protection of plant genetic resources the analysis of genetic diversity and relatedness between or within different genotypes is a prerequisite. It also helps in developing DNA based molecular markers for identification of genotypes with better traits. Traditionally the genetic diversity of the plant is assessed by various DNA fingerprinting methods using specific and non-specific DNA-based markers. 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 [117]. The multiallelic nature, codominant inheritance and relative abundance of SSRs in the genomes facilitate high reproducible genotyping using polymerase chain reaction (PCR) [70]. Microsatellites provide high levels of polymorphism among the genotypes.

The SSRs reside within the transcribed sequences can be identified from ESTs [167] Recently, EST derived SSRs (EST-SSRs) have received much attention due to the availability of increasing amounts of ESTs from various plants [168, 169]. 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 [170]. In addition, EST-SSRs show a higher level of transferability to closely related species than genomic SSR markers [168] and can be served as anchor markers for comparative mapping and evolutionary studies [169]. Similar advantages of EST-SSRs have been reported for a number of plant species, such as grape [168], Medicago species [171], Soybean [172], Sugarcane [173], maize [174] , rice [175], rye [176] and wheat [177], indicating that EST-SSR markers have potential for use in *A. annua* genetic diversity studies. 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.

The development or selection of genetically superior genotypes of *Artemisia* is a high priority of various departments and funding agencies. However, lack of essential genome resources such as availability of a marker enriched genetic map, identification of high artemisinin content genotypes, non-existence of molecular markers are the major stumbling blocks in realizing the full potential of molecular approaches in genetic improvement of *A. annua*. The objective of this study is to determine the genetic relationships among several genotypes of *A. annua* from Ladhak region (the trans-Himalayan region, India). This information will helpful in enriching the available genetic linkage map of *A. annua* and could help further in the development of high yielding genotypes.

4.2 MATERIALS AND METHODS

4.2.1 Plant materials and DNA extraction

Twenty genotypes of *A. annua* were collected from two valleys as mentioned above from the Ladakh region, India. Ten individual plants (~ three-week-old) from each valley were selected and leaf samples were collected. Leaf samples were stored

in laboratory at –20 °C until further analysis. DNA was extracted from the leaf sample as mentioned above.

4.2.2 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). The redundant ESTs were removed using BLASTCLUST (http://www.ncbi.nlm.nih.gov/BLAST) with sequence similarity 60% and 100% sequence coverage. After removing the redundant ESTs the total numbers of ESTs left were 68,974. A perl script known as MIcroSAtellite (MISA http://pgrc.ipk-gatersleben.de/misa/) 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.

4.2.3 PCR primer design and PCR amplification

A set of 16 pairs of primers (Table 4.1) were designed by considering only the ESTs from the gene encoding enzymes for the artemisinin biosynthetic pathway. Primer-5 program (Whitehead Institute for Biomedical Research, Cambridge, Mass) was used for designing of primers based on following core criteria: (1) melting temperature (Tm) between 57 °C and 65 °C with 60 °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 custome synthesized from GBiosciences (A Geno Technology Inc. USA). Polymerase chain reactions (PCRs) were performed in a MyCycler™ thermocycler (Bio-Rad Laboratories, USA). The volume of PCR solution was 15 µl, containing 75 ng of template DNA, 1x PCR buffer (Mg^{2+} free), 0.5 U of *Taq* DNA polymerase, 300 µmol/L of deoxynucleotide triphosphates (dNTPs), 2.25 mmol/L of Mg^{2+}, and 0.75 µmol/L of forward and reverse primers. The optimized PCR amplifying conditions used were: denaturation, 1 cycle of 3 min at 94 °C, an annealing temperature of 55 °C for 35 cycles (45 s at 94 °C, 30 s at 55 °C, 1 min at 72 °C) and an additional cycle of 7 min at 72 °C. Each primer pair was screened twice to confirm the repeataility of the observed bands in each genotype. The amplified fragments were seperated on 6 % metaphore IV agarose gels. The electrophoresis was performed at 100 V for 2 h in 1x TBE [Tris-borate-ethylene diamine tetra acetic acid (EDTA)] buffer. 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.

Table 4.1 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 range (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		

The numbers of alleles detected and amplified by the SSR markers were scored as present (1) or absent (0) and used for genetic diversity analysis as described previously. The genetic diversity of the samples as a whole was estimated based on the number of alleles per locus (total number of alleles/number of loci), the percentage of polymorphic loci (number of polymorphic loci/total number of loci analyzed) and polymorphism information content (PIC). The polymorphism was determined according to the presence or absence of the SSR locus. The value of PIC was calculated using the formula:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2$$

where p_i is the frequency of an individual genotype generated by a given EST-SSR primer pair and summation extends over n alleles.

4.3 RESULTS AND DISCUSSION

4.3.1 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 (Table 4.2).

Table 4.2 Summary of SSR search among *Artemisia anua* ESTs.

Total number of ESTs analyzed	68974
Total size of examined sequences (bp)	44124870
Number of ESTs containing SSRs	4342 (6.3%)
Total number of identified SSRs	4934
Number of ESTs having more than 1 SSR	499
Compound SSRs	40
Number of ESTs containing Bi-type	2344
Number of ESTs containing Tri-type	1872
Number of ESTs containing Tetra-type	172
Number of ESTs containing Penta-type	30
Number of ESTs containing Hexa-type	56

4.3.2 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 the most frequent with a frequency of 52.19% followed by trinucleotides (42.39%). The tetranucleotide SSRs have a much lower frequency (3.6%). The frequency of SSRs with motif sizes of five and six are 0.6% and 1.20% respectively (Figure 4.1). 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.

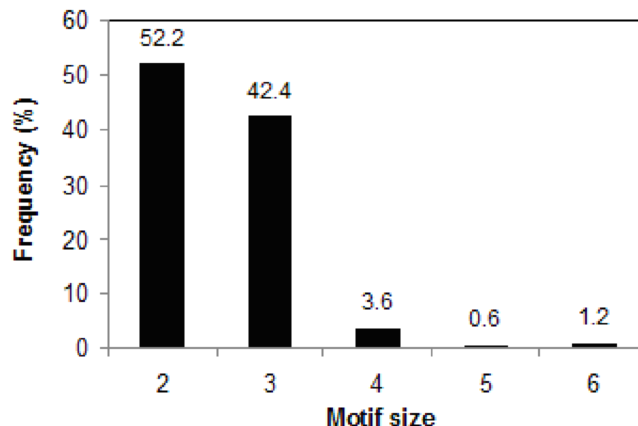


Figure 4.1 The distribution pattern of SSRs in terms of motif size.

4.3.3 Distribution of SSRs based on motif types

The SSRs identified in the present study were characterised by 77 types of motifs (Table 4.3). 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 %.

Table 4.3 Occurrence and number of repeats of 77 SSR motifs in *Artemisia annua*

	Number of repeat unit										Total	%
Repeats	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**Above**	**repeat**	
AC/GT	617	195	103	107	47	22	12	13	4	26	1146	23.23
AG/CT	276	83	57	33	8	3	4	4	1	16	485	9.83
AT/AT	663	115	66	26	13	5	2	1	3	18	912	18.48
CG/CG	27	2	1	1	1					0	32	0.65
AAC/GTT	220	51	28	13	15	5	2		1	5	340	6.89
AAG/CTT	226	53	21	8	1		3			3	315	6.38
AAT/ATT	165	62	20	14	2	2	1	2		11	279	5.65
ACC/GGT	269	137	27	7	1	2				0	443	8.98
ACG/CTG	56	11	6	4	2			2		0	81	1.64
ACT/ATG	124	31	6	7	1	1	1	2		1	174	3.53
AGC/CGT	69	15	11	3	1					0	99	2.01
AGG/CCT	42	17	4		1	2				0	66	1.34
AGT/ATC	188	55	15	8	2	3	1	2		1	275	5.57
CCG/CGG	15	4	1							0	20	0.41
AAAC/GTTT	11	6							1	0	18	0.36
AAAG/CTTT	8	1	1							0	10	0.20
AAAT/ATTT	44	26	12	6	7			1		0	96	1.95
AACC/GGTT	13									0	13	0.26
AACT/ATTG	1	1								0	2	0.04
AAGG/CCTT	2									0	2	0.04
AATC/AGTT	8	2								0	10	0.20
AATT/AATT	9	3								0	12	0.24

ACAT/ATGT	3								0	3	0.06
ACCC/GGGT	2								0	2	0.04
AGAT/ATCT	3								2	5	0.10
AGGG/CCCT	1								0	1	0.02
AGGT/ATCC	1								0	1	0.02
AGTC/AGTC			1						0	1	0.02
CCCG/CGGG	2								0	2	0.04
AAAAC/GTTTT		3		2					0	5	0.10
AAAAG/CTTTT	4	3							0	7	0.14
AAAAT/ATTTT	2	3	2						0	7	0.14
AAAGT/ATTTC	1								0	1	0.02
AAATC/AGTTT	2								0	2	0.04
AAATG/ACTTT	1								0	1	0.02
AACAG/CTTGT	1								0	1	0.02
AAGAC/CTGTT	1								0	1	0.02
AAGGT/ATTCC	1								0	1	0.02
ACACT/ATGTG			1						0	1	0.02
ACATC/AGTGT			1						0	1	0.02
AGCGT/ATCGC	1								0	1	0.02
AGGCC/CCGGT	1								0	1	0.02
AAAAAC/GTTTTT	1								0	1	0.02
AAAAAG/CTTTTT	1								0	1	0.02
AAAAAT/ATTTTT	1	1							0	2	0.04
AAAACC/GGTTTT	1								0	1	0.02
AAACAC/GTGTTT	1								0	1	0.02
AAACCC/GGGTTT	1								0	1	0.02
AAACTG/ACTTTG	2								0	2	0.04
AAAGAT/ATTTCT	2								0	2	0.04
AACAAG/CTTGTT	1								0	1	0.02
AACACC/GGTTGT	1								0	1	0.02
AACACT/ATTGTG	1								0	1	0.02
AACCAC/GGTGTT	2								0	2	0.04
AACCAG/CTTGGT	2								0	2	0.04
AACCCC/GGGGTT	1								0	1	0.02
AACCCT/ATTGGG	4		1						0	5	0.10
AAGACC/CTGGTT	4		1						0	5	0.10
AAGCCC/CGGGTT	3								0	3	0.06
AAGGTT/AATTCC		2							0	2	0.04
AAGTGG/ACCTTC			1						0	1	0.02
AATAGT/ATCATT	1								0	1	0.02
AATGAT/ACTATT	1								0	1	0.02
AATGGG/ACCCTT	1								0	1	0.02
AATGGT/ACCATT	3								0	3	0.06
ACCACT/ATGGTG	1								0	1	0.02
ACCATC/AGTGGT	1		1						0	2	0.04
ACCGCC/CGGTGG	2								0	2	0.04
ACCTCG/AGCTGG	1								0	1	0.02
ACGAGC/CGTGCT		3							0	3	0.06
ACGATG/ACTGCT	1								0	1	0.02
ACGGGG/CCCCTG	1								0	1	0.02
ACGGTG/ACTGCC	1								0	1	0.02
ACGTCC/AGGTGC		1							0	1	0.02
ACTCGG/AGCCTG	2								0	2	0.04
AGCAGT/ATCGTC	1								0	1	0.02
AGGCTC/AGTCCG	2								0	2	0.04
Total	**3128**	**886**	**388**	**239**	**102**	**45**	**26**	**27**	**10**	**83**	**4934**	

4.3.4 EST-SSR polymorphism

Using 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. A total of 38 alleles with the average alleles per locus of 2.92 were detected at 13 loci. More than one allele was detected at 12 out of 13 loci, with the polymorphic markers ratio of 92%. The maximum of 3-7 alleles were detected with five markers, followed by 2 alleles detected by 7 markers and one allele by only 1 marker (Table 1). Figure 3 shows the representative amplified products from 3 different SSRs marker.

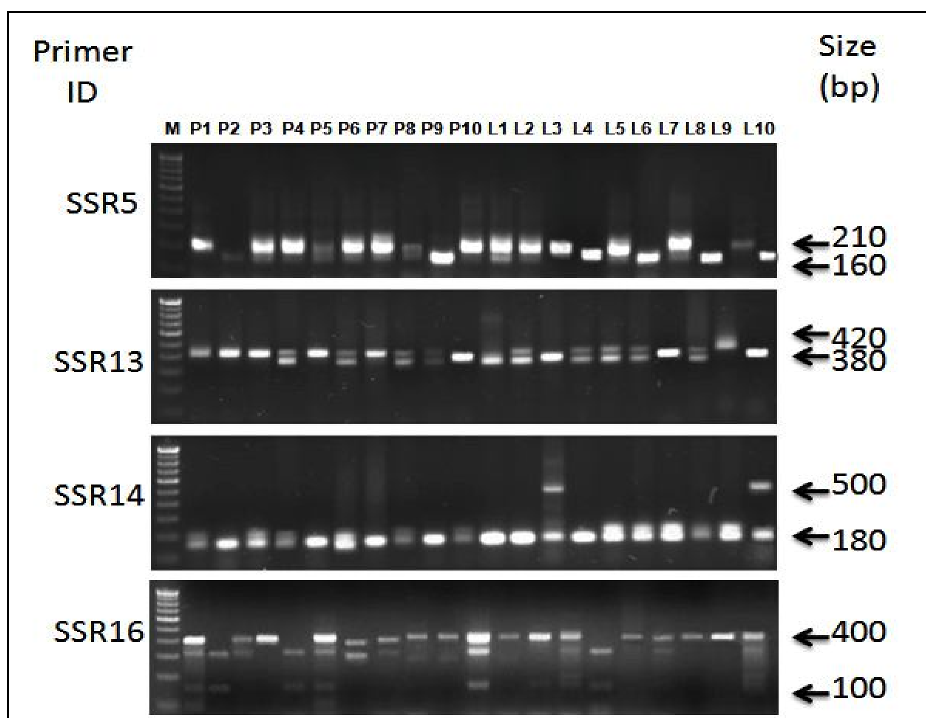


Figure 4.2 SSR amplification products obtained from the 20 genotypes of *Artemisia annua* studied. L1 to L10, are the genotypes collected form Indus valley and P1 to P10 are the genotypes collected from Nubra (Partapur) valley. M = the size of molecular markers in base pairs using λ DNA.

Polymorphism information content (PIC) refers to the value of a marker for detecting polymorphism within a population or set of genotypes by taking into account not only the number of alleles that are expressed but also the relative frequencies of alleles per locus. PIC was calculated for the markers which generated 2 or more alleles. As evident, SSR marker '15' showed the highest level of polymorphism with PIC value of 0.802 followed by SSR marker 16 (0.716), SSR marker 13 (0.595) and SSR marker 6 (0.565). The PIC values for rest of the SSR

markers were in the range of 0.198-0.495. In case of non amplifying SSR primers, different annealing temperatures (± 5 °C of Tm) in combination with different PCR reactions were tried but no amplification was observed.

4.3.5 Cluster analysis

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 bawed on neighbour joining clustering techniques (Figure 12a). STRUCTURE analysis also predicted better distribution of genotypes with 2 major groups (Figure 12b). It was found that the genotypes were more likely distributed (at high probability) with respect to their geographical distribution (Figure 12c). Therefore in order to avoid the potential risks associated with too little genetic diversity, the adoption of elite genotypes from different origins used as parental lines is highly recommended for any genetic conservation and breeding program in *Artemisia annua*.

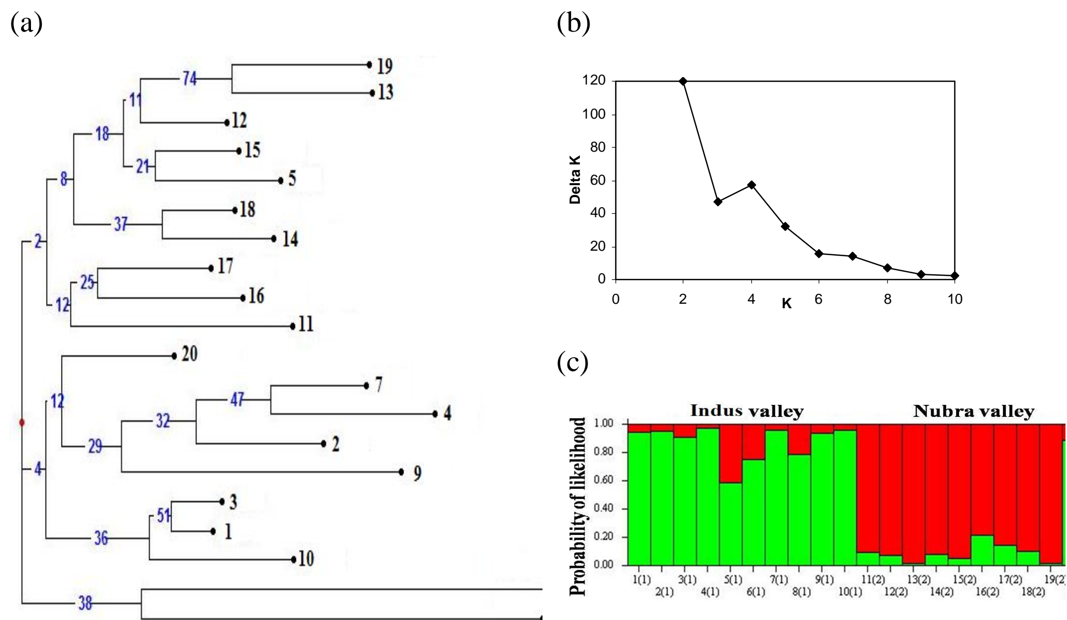


Figure 4.3. Dendrogram generated using Neighbor joining method with bootstrap supporting values, showing relationships between 20 *Artemisia annua* genotypes, using SSRs data. (b) the relationship between K and ΔK based on unbiased clustering demonstrating optimal distribution of genotypes into two groups, (c) the likelihood distribution of genotypes collected from Indus valley (green colour) and Nubra valley (red colour).

4.3.5 Genetic diversity analysis

A relatively high genetic variation was detected among the *A. annua* genotypes. 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 4). The respective values for overall genetic variability for Na, Ne, H, I, Ht, Hs, Gst, NPL, PPL and Gene flow (Nm) across all the 20 genotypes were also given in Table 5. Gst value of 0.1808 indicated that 81.92 % of the genetic diversity resided within the population. AMOVA helps in partitioning of the overall genetic variations among groups and among genotypes. Molecular variance among valley is 23 % and among genotypes is 77 % (Table 6) which revealed higher variation among the genotypes in *A. annua*. All the components of molecular variation were significant ($P < 0.001$).

Microsatellites have been used as a resource for random candidate markers in population genetics studies. To better understand the natural diversity of *A. annua* genotypes and to develop strategies for its sustainable utilization, we identified SSR motifs in a dataset of 68,974-unique EST sequences (about 44.12 Mb). A total of 4,934 putative SSRs motif from 4,342 (6.3 %) unique ESTs were identified. Some of the ESTs does not consists of any SSRs motifs whereas some ESTs consists of more then one SSR motifs. The incidence of SSRs (6.3 %) within ESTs was lower in comparison to other plant such as apple (20 %) [178], but it was similar to that of some dicotyledonous species (ranging from 2.65 % to 16.82 %) [179].

Table 4.4 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

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

Table 4.5. Overall genetic variability across all the 20 genotypes of *Artemisia annua* based on SSR analysis.

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

Hs = Genetic diversity in population; Gst = Genetic diversity between population; Nm = Estimate of gene flow from Gst; Nm = 0.25 (1-Gst)/Gst

Table 4.6 Summary of analysis of molecular variance (AMOVA) based on SSR marker, among the populations of *Artemisia annua.* Levels of significance are based on 1000 iteration steps.

Source of variation	df	S.S.D.	Variance component	Percentage	P-value
Among valley	1	19.35	1.448	23	<0.001
Among genotypes	18	87.70	4.872	77	<0.001

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

In this work, the frequency of occurrence for EST-derived SSRs was one in every 8.94 kb. This is in accordance to the earlier findings with many species such as: one EST-SSR occurs every 13.8 kb in *Arabidopsis thaliana*, 3.4 kb in rice, 8.1 kb in maize, 7.4 kb in soybean, 11.1 kb in tomato, 20.0 kb in cotton and 14.0 kb in poplar [180]. Among the SSRs, dinucleotides (2,344) were the most abundant repeat units followed by trinucleotides (1,872), tetranucleotides (172), hexanucleotides (56) and pentanucleotides (30). The frequency of occurrence of SSRs types in *A. annua* is different from that of other medicinal plant *E. Sagittatum* [181] and cereal species [169] where trinucleotide repeat units are the most dominant SSR, followed by di- and tetra-nucleotide repeat units [169, 181]. The relative frequency of repeats with different dinucleotide compositions was also biased towards one of four possible repeat classes (Table 3). Among the dinucleotide repeat classes, (AC/GT)n repeats were the most common dimer motif (23.23 %), followed by (AG/CT)n, (AT/AT)n and (CG/CG)n with a frequency of 9.83 %, 18.48 % and 0.65 % respectively. This was in agreement with recent studies in cultivated peanut (*Arachis hypogaea* L.)[182] and wild *Arachis* species [183]. Among the tri-nucleotide repeats, (ACC/GGT)n, (AAC/GTT)n, (AAG/CTT)n, (AAT/ATT)n and (AGT/ATC)n were the largest repeat class followed by (ACT/ATG)n, (AGC/CGT)n, (ACG/CTG)n, (AGG/CCT)n and (CCG/CGG)n (Table 2). In other plant species, the most frequent tri-nucleotide repeat motifs were (AAC/TTG)n in wheat, (AGG/TCC)n in rice, (CCG/GGC)n in maize,

(AAG/TTC)n in soybean, and (CCG/GGC)n in barley and sorghum [169, 184, 185]. These unique putative transcript-derived SSR markers that were generated in the present study provide a valuable genetic resource for future studies of *Artemisia annua* and other related species.

Some of the EST-SSRs may tightly link with functional genes that control synthesis of bioactive compound, artemisinin. Therefore, we have considered ESTs derived from the genes belonging to artemisinin biosynthetic pathway for experimental analysis and polymorphism study among 20 genotypes of *A. annua.* The SSRs confirmed in this analysis may be valuable for screening of high yielding genotypes. In this study, 38 alleles were detected with 13 SSR loci with an average of 2.92 alleles per locus. This relatively small number is probably due to the limited number of genotypes studied and also relatively high genetic similarity that exists within the investigated group of *Artemisia* germplasm. Genetic relationships were found to be very close among the genotypes from the Nubra valley in comparison to Indus valley. The fact that eight genotypes from the Nubra valley were clustered into the same subgroup gave the strong indications of the narrow genetic background in *Artemisia* germplasm. The genotypes from Indus valley investigated in this study, on the other hand, showed a broader genetic diversity. Therefore in order to avoid the potential risks associated with too little genetic diversity, the adoption of elite genotypes from different origins used as parental lines is highly recommended for any genetic conservation and breeding program in *Artemisia annua*. AMOVA analysis revealed higher genetic variation among genotypes in *A. annua*. This is helpful in making strategy for germplasm collection and evaluation. The estimated gene flow was 1.1325. In population genetics, a value of a gene flow (Nm) < 1.0 (less than one migrant per generation into a population) or, equivalently, a value of gene differentiation (Gst) > 0.25 is generally regarded as the threshold quantity beyond which significant population differentiation occurs [166]. Overall the study indicates that *Artemisia annua* populations in the trans Himalayan region are genetically highly diverse.

CHAPTER 5

Extraction and quantification of Artemisinin at different developmental stages from *Artemisia annua* L. collected from Ladakh region, India

ABSTRACT

The demand for artemisinin has increased sharply since the World Health Organization recommended its use as part of the artemisinin combination therapies in 2001. The artemisinin concentration peaks at a slightly different time in different areas and identifying this is critically important to ensure optimum content of artemisinin in *A. annua*. The present study investigated the artemisinin content in different plan parts as well as at different stages of growth of *A. annua* from the Ladakh region. Artemisinin content in the leaf gradually increased from the stage of seedling upto six months and afterward it was remain constant. The highest yield was 0.441% dry weight and 0.356% dry weight in the samples collected from Indus valley and Nubra valley. Similarly the artemisinin content in the inflorescence was found to be maximum at full-flowering stage (0.813 ± 0.065% and 0.721 ± 0.053% dry weight in the samples collected from Indus valley and Nubra valley respectively using HPLC analysis). Therefore it is concluded that the optimal time at which leaves should be harvested is the time when flower buds start to form (in the month of August) and the inflorescence should be harvested during the full-flowering stage (in the month of October) for better yield of artemisinin from *A. annua* in the Ladakh region.

5.1 INTRODUCTION

At the present time, in the 21st century, malaria is still severely challenging people's health. Each year, more than one million people around the globe die of malaria and more than two billion people in over 100 countries are threatened by the disease [6, 7]. In many developing countries, especially those in Africa, the mortality from malaria are still very high. Therefore the world market for products including artemisinin derivatives is now growing rapidly, and the demand for artemisinin is increasing. At present, artemisinin compounds are derived from a raw substance extracted from the plant *A. annua* L. because artemisinin is very difficult to synthesise [8, 9, 10]. Enhanced production of artemisinin in the whole plant of *A. annua* L. is therefore highly desirable.

Many researchers have reported that the yields of extracted artemisinin is very poor [10, 20] and much effort must be made to increase artemisinin content of *A. annua*. Because the plant material in wild stands is typically variable in its artemisinin content and plant biomass, this has an impact on drug extraction. Efforts are being made to increase its production in many ways such as plant tissue culture systems, [45] biotechnological approaches [46], agronomical practices [47]. But artemisinin was not found to be accumulated in callus and cell suspension cultures, it is presumed that the biosynthesis of artemisinin to be restricted to the green part of the plant [48]. At the present time, 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. Due to physico-chemical properties of artemisinin (low thermal and chemical stability of the endoperoxide function, low polarity and, hence, poor solubility in water and good solubility in organic solvent), its extraction with non-polar solvent is necessarily complicated by simultaneous extraction of essential oil,

chlorophylls and waxes. Therefore, the extraction steps must be followed by separation of artemisinin from liquor.

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.

5.2 MATERIALS AND METHODS

5.2.1 Plant materials and sample preparation

Leaves were collected in triplicate from the plant every month after seedling, during the period of March to October from both the valleys (Indus and Nubra). Seedling established in the month of March and the plant die in the month of October. Similarly the infloresences were collected in triplicate at three different developmental stages: flower budding stage, early flowering stage (30 days after budding) and at full flowering stage of *A. annua* from their natural habitats (Indus and Nubra valleys). Artemisinin standard was purchased from Sigma Aldrich, USA. Artemisinin stock 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*.

All samples were dried at room temperature. 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 by vortex for 2 min at room temperature. Samples were centrifuged at 10000 rpm for 2 min and the supernatant was transferred into a fresh 2 ml centrifuge tube and evaporated to dryness under steam nitrogen at room 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.

5.2.2 Artemisinin estimation using RP-HPLC-PDA method

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 auto sampler, thermostat column oven maintained at 30°C and C18 column (5 µm; 250 m 4.6 mm).

Five replicates were used for each level of concentration in order to check the reproducibility of the detector response at various concentrations. Exactly 10 µl of the standard sample was injected using auto sampler. The condition on the PDA was set as follows: mobile phase consisting of acetonitrile to achieve maximum separation and sensitivity at 229 nm; flow rate of 0.6 ml/min with stop and rest time of 12 and 5 min respectively. The peak of artemisinin was observed at the retention time of 4.796 min (Fig. 5.1A). Calibration curve was generated by plotting a graph of peak area against the concentrations used to quantify the artemisinin content in different samples. Excellent linearity was retained for artemisinin between peak areas and concentrations in the range of 0.1-100 µg/ml ($r^2 = 0.998$).

Extracted artemisinin from the plant samples were analyzed by injecting 10 µl of samples into HPLC system using auto sampler. Artemisinin content was analyzed in five replicates and was expressed as the mean ± standard deviation. Figure 5.1B shows a chromatogram of an extract sample, the retention time of artemisinin is 4.750 minutes.

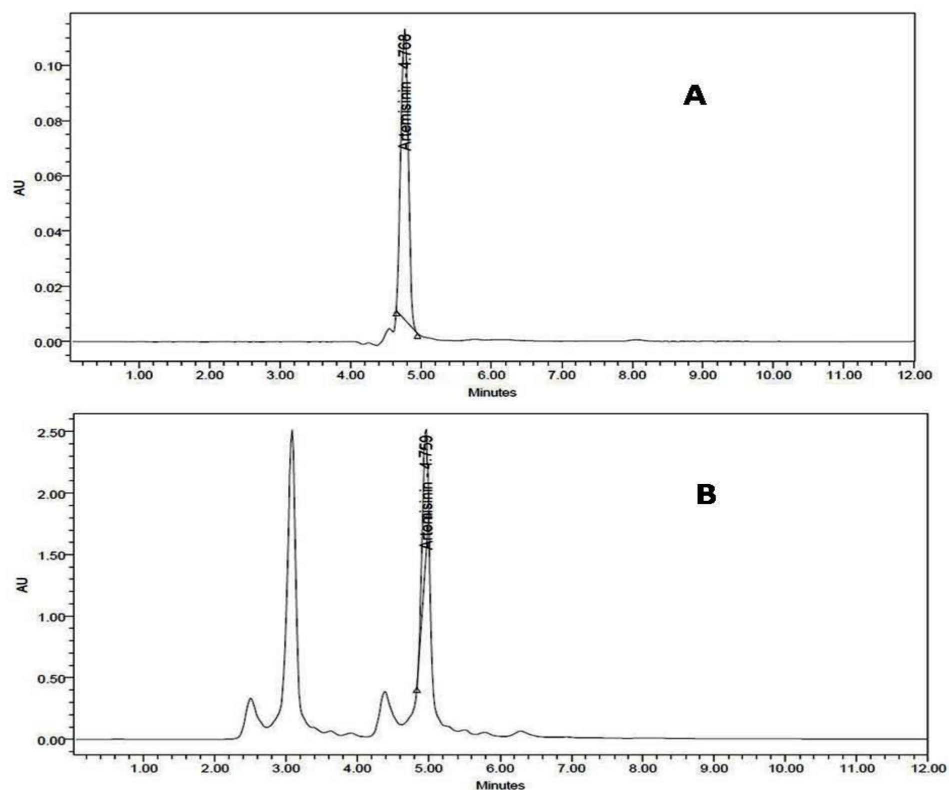


Figure 5.1. Chromatograms of artemisinin standard (A) and *Artemisia annua* extract (B) by HPLC-PDA. The concentration of artemisinin standard is 1mg/ml (artemisinin-acetonitrile). Artemisinin was extracted from 105.3 mg dry weight of inflorescence.

5.2.3 Artemisinin estimation using GC-FID method

Analysis of artemisinin was performed by GC using an Agilent (Santa Clara, CA, USA) GC-6820N system with a FID detector and an Agilent data collection system. Nitrogen was the carrier gas with a column flow rate of 2 mL/min, a split ratio of 3:1. The column was a HP-5 crossbond 95% dimethylpolysiloxane (Agilent), (30 m ×0.32 mm I.D. 0.25 mm film thickness). Injector temperaturewas set at 240 ˚C, and detector temperature set at 300 ˚C. Oven temperature was programmed to start at 180 ˚C (1min), increasing 0.8 ˚C/min to 198 ˚C, then increasing 30 ˚C/min to 280 ˚C, then holding at that temperature for 10 min. Possible problem during estimation was monitored before and throughout the analysis by running an artemisinin standard of known concentration (normally 1 mg/mL) twice in the beginning, once every 10 samples, and at the end to check if changes in detector sensitivity and response would occur. The representative chromatograms that were obtained using GC-FID analysis was shown in Figure 5.2a (the artemisinin standard) and Figure 5.2b (the sample).

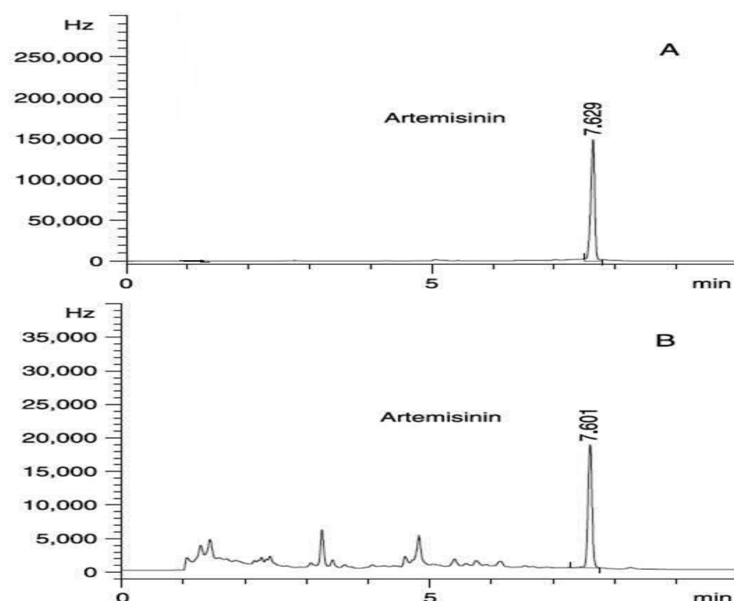


Figure 5.2 Chromatograms of artemisinin standard (A) and *Artemisia annua* extract (B) by GC-FID. The concentration of artemisinin standard is 1mg/ml (artemisinin-acetonitrile). Retention time of artemisinin is 7.629 ± 0.5 min.

5.2.4 Artemisinin estimation using spectrophotometric method

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.

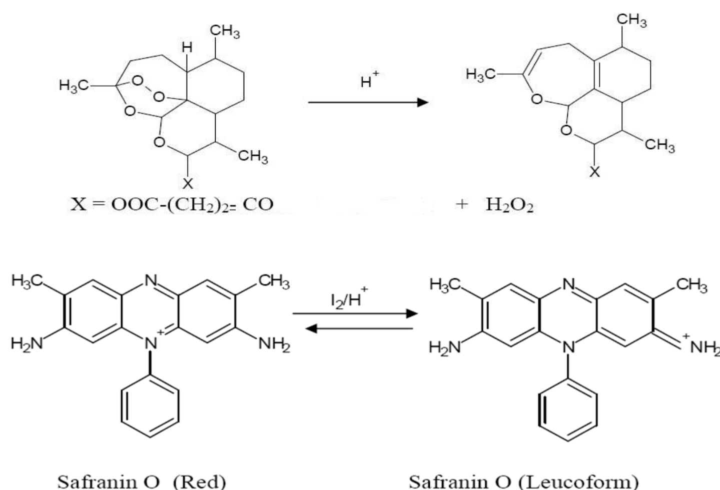


Figure 5.3 Reaction of artemisinin in acidic medium to form H_2O_2 and reaction between H_2O_2- KI System with safranin O.

All solutions were prepared with double distilled water. Safranin O (0.01%) was prepared by dissolving 0.01 g of safranin O in 50% ethanol and making up to 100 mL. Different aliquots (10-100 µg mL^{-1}) of artemisinin were added in a series of 10 mL calibrated flasks consisting of potassium iodide and 5 M HCl and the mixture was gently shaken until the appearance of yellow colour, indicating the liberation of iodine. To this system, 0.5 ml of 0.01% safranin O was added followed by 2 mL of 2 M sodium acetate solution and the reaction mixtures were shaken for 5 min. The contents were diluted to the mark with distilled water and mixed well. The absorbance of each solution was measured at 521 nm against the corresponding reagent blank. Reagent blank was prepared by replacing the analyte (artemisinin) solution with distilled water. The absorbance corresponding to the bleached colour, which in turn corresponds to the (analyte) artemisinin concentration, was obtained by subtracting the absorbance of the blank solution from that of test solution.

5.3 RESULTS AND DISCUSSION

Liquid extractions with toluene, n-hexane, chloroform, ethyl acetate, petroleum ether, or dichloromethane are currently the most commonly used technique for artemisinin extraction. Duke et al. [17] and Woerdenbag et al. [186] previously found that artemisinin can be rapidly extracted from the fresh leaves of *A. annua* by suitable solvent at room temperature. This may be due to artemisinin present in foliar tissue are localized entirely in the subcuticular space of glandular trichomes [17], which are easily released into the solvent. Van Nieuwerburgh et al. [187] and Woerdenbag et al. [188] effectively and rapidly extracted artemisinin in fresh *A. annua* leaves by chloroform and toluene, respectively, under room temperature. However, both chloroform and toluene are highly toxic. Shuoqian Liu et al. [189] reported efficiency of various extraction methods in different solvents, such as ethyl acetate, hexane, chloroform, toluene and petroleum ether. They have found extraction recoveries in toluene (about 65%), petroleum (67%) and hexane (70%) after 2 min vortex. In contrast, extraction in ethyl acetate and chloroform showed high recovery [187]. Hence ethyl acetate is relatively low toxic and was a suitable organic solvent for artemisinin extraction. In this study we have used ethyl acetate as solvent for extracting artemisinin from the sample effectively.

Artemisinin content from *A. annua,* was analyzed using HPLC-PDA, GC-FID and spectrophotometric method. Standard of artemisinin was used to calibrate HPLC-PDA, GC-FID, and spectrophotometric method. Standards at concentrations of 1, 0.1, 0.01, 0.001 mg/ml were freshly prepared and used in instruments. Correlation coefficient (r^2) in a linear plot was $r^2 = 0.9727$ for the HPLC-PDA standard curve and $r^2 = 0.998$ for GC-FID standard curve and $r^2 = 0.8573$ for spectrophotometer. Regarding precision and accuracy, a standard of 1 mg/ml of artemisinin was injected (10µl for HPLC and 1 µl for GC) five times consecutively in each instruments. The retention time of artemisinin was observed at 4.796 min using HPLC-PDA (Figure 5.1). It is known that GC-FID degrades artemisinin and that artemisinin is measured through its major degradation peaks. The retention time (RT) for GC-FID of individual standards at 1 mg/ml was 7.629 minute (Figure 5.2).

5.3.1 Dynamic variation of artemisinin content of leaf during the growth stages

The dynamic variation of artemisinin content of leaf with respect to growth stages of plant is shown in Figure 5.4. 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. 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). The artemisinin content of *A. annua* was largely influenced by the developmental stages of the plants and there have been different reports regarding the optimal stages at which the highest production of artemisinin was found. It was previously reported that the artemisinin content reaches a peak before flowering [190, 191], whereas others have indicted highest content at the full-flowering stage [216, 18]. However, some previous reports also mentioned that the highest content is present during the vegetative growth stage, just before the formation of flower buds with the highest leaf yield [192]. From these studies, one may reasonably think that the highest artemisinin content is closely related to reproductive growth.

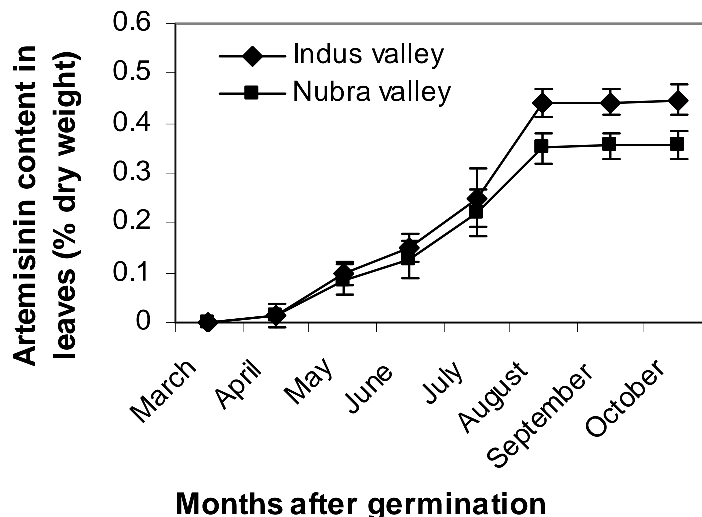


Figure 5.4. Dynamic variation of artemisinin content of leaf at different growing stages.

5.3.2 Artemisinin content of inflorescence during the reproductive stages

The artemisinin contents of inflorescence at three reproductive stages: 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 artemisinin content in inflorescence increased with the developmental stages and attained its maximum at full flowering stage (Figure 5.5a-c). Based on HPLC quantification the artemisinin content was found to be 0.531 ± 0.071 at budding stage, 0.672 ± 0.044 at early flowering stage (30 days after budding) and 0.813 ± 0.065 at full flowering stage in the samples collected from Indus valley. Similarly the artemisinin content in the samples collected from Nubra valley were 0.43 ± 0.035, 0.61 ± 0.062 and 0.721 ± 0.053 at budding, pre-flowering and full-flowering stage respectively (Figure 5.5a). The artemisinin content was also estimated using GC-FID and spectrophotometric methods from the above samples. The value of artemisinin content was found to be very similar to HPLC analysis using GC-FID (Figure 5.5b). However, the values were found to be slightly lower based on spectrophotometric method (Figure 5.5c). The quality of fit for the estimation of artemisinin between HPLC-PDA and GC-FID methods is represented in Figure 5.6a (R^2 = 0.9415), between HPLC and spectrophotometirc methods is represented in Figure 5.6b (R^2 = 0.9258) and similarly between GC-FID and spectrophotometric methods is represented in Figure 5.6c (R^2 = 0.9114). The high value of R^2 for the comparative analysis revealed that all the three

techniques used in this study for the quantification of artemisinin are robust and useful.

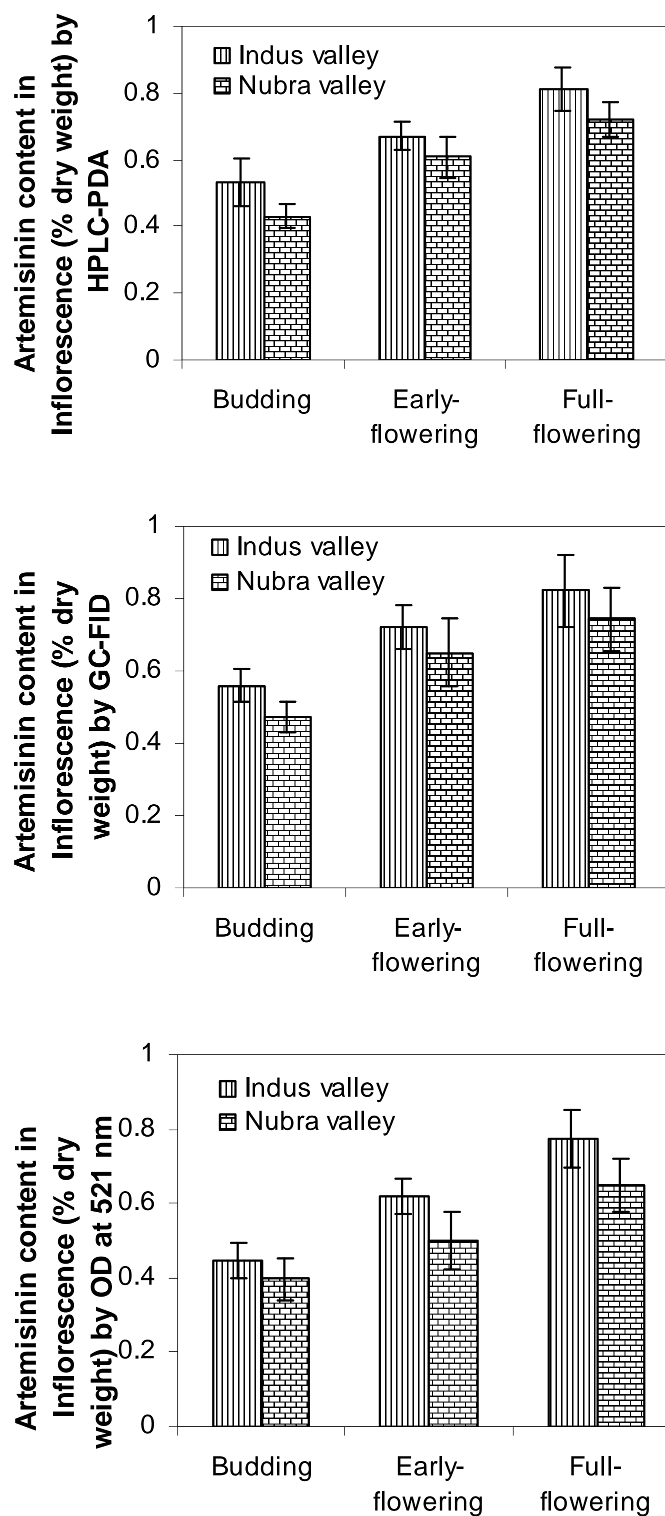
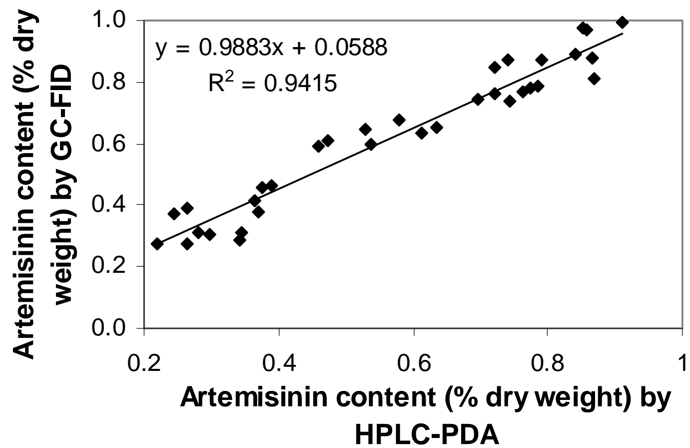
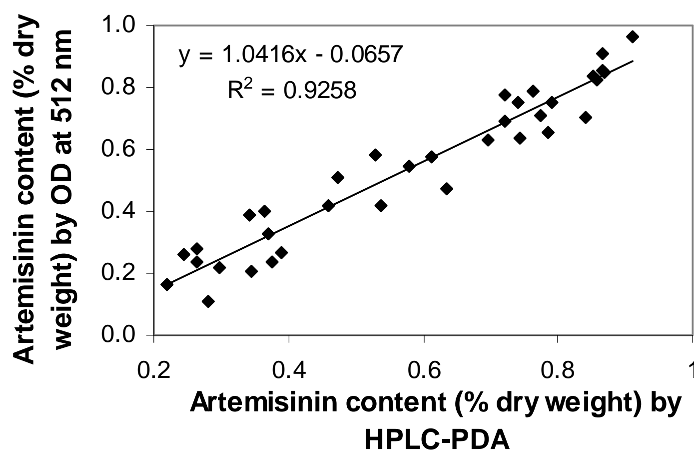


Figure 5.6 (a-c) Artemisinin content of inflorescence during reproductive stages of *Artemisia annua* by HPLC-PDA, GC-FID and spectrophotometric methods.





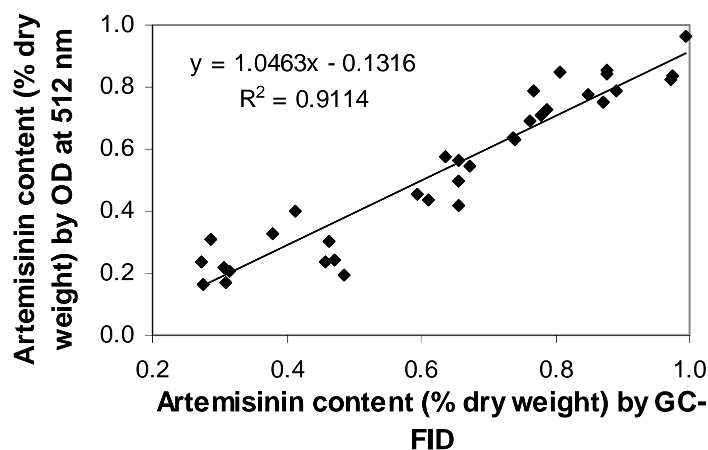


Figure 5.5 Correlation of artemisinin content of inflorescence of *Artemisia annua* analyzed by HPLC-PDA, GC-FID and spectrophotometric methods. Their artemisinin content varied from 0.2 % to 0.9%.

Artemisia annua is the main source of artemisinin production for pharmaceutical processing in the absence of viable synthetic protocol. 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.

CHAPTER 6

Impact of Soil Nutrient and Environmental Factors on Artemisinin Content among *Artemisia annua* Populations of Ladakh Region Using Linear and Non-linear Approach

---

ABSTRACT

Artemisia annua is a promising medicinal herb and is currently processed by pharmaceutical firms for the production of anti-malarial compound, artemisinin for Artemisinin-based Combination Therapies (ACTs) in the treatment of malaria. Artemisinin content in *A. annua* differs greatly in different natural habitats. In order to facilitate reasoned scientific decisions on its domestication, conservation and sustainable utilization, the effects of soil nutrients and environmental factors on artemisinin content in the leaves of *A. annua* were investigated. The artemisinin content reached up to 0.675% of dry weight when soil pH value was about 7.82, soil organic carbon was higher than 0.73% and nitrogen content was higher than 0.67% of soil dry weight. Soil with available phosphorous content higher than 0.519% and potassium content higher than 0.76% resulted in low artemisinin content. The strong and linear relationship detected between artemisinin 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.* In this regard the prediction model like artificial neural network (ANN) and multiple linear regressions (MLR) developed in this study to map the effect of these factors on artemisinin yield will be of great help. The ANN prediction model revealed better prediction of yield (R^2 = 0.992) than MLR prediction model (R^2 = 0.806). Lower level of root mean square error (RMSE) for ANN model (0.018) than MLR model (0.062) with respect to the experimental measurement establishes the ANN method as an efficient tool for optimization of soil nutrients and climatic factors for artemisinin yield.

6.1 INTRODUCTION

The Ladakh region is home to numerous highly valued medicinal herbs including *Artemisia annua.* It is distributed in very restricted pockets in the Ladakh region at altitudes ranging from 3000 to 3500 m from the sea level. It is recognized for its antimalarial, antibacterial and anti-cancer properties [1]. The flowers and leaves of *A. annua* contain anti-malarial compound known as artemisinin, which has been used as a remedy for chills and fevers for more than 2000 years [8, 30]. Artemisinin is currently the most effective malaria drug and is the World Health Organization's currently recommended medicine in treating malaria [6]. Its derivatives are found effective against *P. falciparum* and *P. vivax*, including multi-drug resistant strains [193]. In addition to potent anti-malarial activity, artemisinin posses properties like anti-cancer [1], herbicidal [194, 17], anti-hepatitis B [4], anti-HIV [195], anti-leishmanial [196] and anti-schistosomiatic activities [197]. Moreover, there is increasing demand of this compound for Artemisinin-based Combination Therapy (ACTs) in the treatment of malaria. Therefore the world market for products including artemisinin derivatives is now growing rapidly, and the demand for artemisinin is increasing.

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 [198]. Attempts have been made worldwide to enhance its production employing various approaches such as conventional breeding, biochemical, physiological, molecular and genetically engineered microbes [199, 200, 201, 202, 203, 204, 205]. Although these approaches show potential for future development, at present time, the biotechnological approach for the commercial production of artemisinin remains disappointing. Therefore, enhanced production of artemisinin in the whole plant of *A. annua* is still the only valid source. In order to maximize artemisinin yield, Laughlin [60] reviewed the agronomic aspects such as the time of establishment, the density of plantation, the weed control, the fertilization and the stage of harvesting. However the range of artemisinin content of *A. annua* is wide and can be affected by numerous factors such as geographical conditions, harvesting time, climatic conditions, and fertilizer so on. A number of investigations have demonstrated that the quality and

quantity of several secondary metabolites have close relationship with plant habitats [110]. Soil is essential for the growth and metabolism of plants as it provides nutrients (nitrogen (N), phosphorus (P), potassium (K), sulphur (S)) and metal elements. The shift of nutrition supplies in soil definitely leads to the alteration of both primary and secondary metabolism and consequently results in changes in productivity of the 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. It can survive under varying growing conditions and adapt well from the extreme low winter temperature of the northern climates to the high summer temperatures and at altitudes ranging from 3000 to 3500 m. 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. MLR has been used to explain the spatial variations in soil nutrients and its impact on crop yield at field scale [11]. However, MLR requires a normal distribution of the input variables which is not always the case [112], the non-linear predictors such as ANNs have been used to solve various problems in agriculture. For example, Sudduth et al., [113] successfully predicted corn yield with back propagation neural network models based on soil texture, topography, pH and some soil nutrient elements. This prediction model was superior to those of the nonparametric statistical benchmark methods.

Artificial neural networks can be used to develop empirically based agronomic models. ANNs find relationships by observing a large number of input and output examples to develop a formula that can be used for predictions [206]. Non-linear relationships overlooked by other methods can be determined with a little a priori knowledge of the functional relationship [207]. A minimum of three layers is required in an ANN: the input, hidden and output layers. The input and output layers contain nodes that correspond to input and output variables respectively. Data move between layers across weighted connections. A node accepts data from the previous layer and calculates a weighted sum of all its inputs, t:

$$t_i = \sum_{j=1}^{n} w_{ij} x_j \qquad (1)$$

Where n is the number of inputs, w is the weight of the connection between node i and j and x is the input from node j. A transfer function is then applied to the weighted value, t, to calculate the node output, o_i..

$$o_i = f(t_i) \qquad (2)$$

The most commonly used transfer function is a sigmoidal function for the hidden and output layers and a linear transfer function is commonly used for the input layer. The number of hidden nodes determines the number of connections between inputs and outputs and may vary depending on the specific problem under study.

The objective of present study is to examine the contents of soil nutrient factors such as soil organic matter, pH value, total N, total P and total K as well as different climatic factors like high temperature, low temperature, high humidity and low humidity to analyze their relationships with artemisinin contents of *A. annua* populations in the Ladakh region. The results obtained and the prediction model developed would guide the soil management and optimization of environmental and geographical factors for the domestication, conservation and sustainable utilization of *A. annua* at commercial scale from the Ladakh region.

6.2 MATERIALS AND METHODS

6.2.1 Sampling of plant materials and soil

Samples of *Artemisia annua* were collected from two valleys: Indus (3500 m) and Nubra (3000 m) from the trans-Himalayan of Ladakh region, India in 2010. 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. Samples were dried 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. The soil is mostly rocky mixed with humus. However, the texture of the soil may be sandy loam, mica shift, loamy or rock clay state. Since the sampling was done from the wild there was no human intervention. 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.

6.2.2 Extraction of artemisinin and quantification

Dried samples were ground to powder in a mortar and pestle. Artemisinin was extracted following the procedure of Liu et al. [189] with minor modification and estimated using HPLC-PDA as described previously. All the experiments on extraction of artemisinin and HPLC analysis were repeated three times.

6.2.3 Quantitative analysis of soil nutrition

About 200g of soil was collected from the root level and was air-dried to a constant weight and then sieved through a 2 mm-mesh. The fine soil (particles <2 mm) was used for nutrient analysis. Soil water pH was determined by dissolving 5g of air-dried soil sample into 5ml of SMP buffer (12.9 mM paranitrophenol, 15.4 mM of K_2CrO_4, 0.361 M of $CaCl_2.2H_2O$, 12.6 mM of $Ca(OAc)_2$, adjusting to pH 7.50 with 15 % NaOH) and measured pH value with pH meter [208], Soil and Plant Analysis Council, 1992). The soil organic matter was determined by measuring organic carbon content according to the wet-oxidation procedure described by Mebius [209]. Total nitrogen was estimated by Kjeldahl digestion through steam distillation. The resulting ammonium was converted into boric acid and titrated with 0.10 N or 0.02 N HCL to pH 4.6 using automatic titrator. This method measures both organic and inorganic forms of nitrogen which were reported as dry weight percent [201, 211]. Total phosphorus (organic & inorganic) was determined by perchloric acid digestion [212]. Total potassium was analyzed by sodium hydroxide digestion and estimated by atomic absorption spectrometry.

6.2.4 Statistical analysis

The correlation and regression analysis between the artemisinin content and soil nutrients, environmental factors were examined by using MINITAB statistical package.

6.2.5 Neural network data-mapping model development

In this study, a back propagation neural-network model was created using Statistica dataminer (version 8.5) and trained using the environmental factors and soil

nutrition parameters as the inputs and the measured corresponding 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 (Figure 6.1).

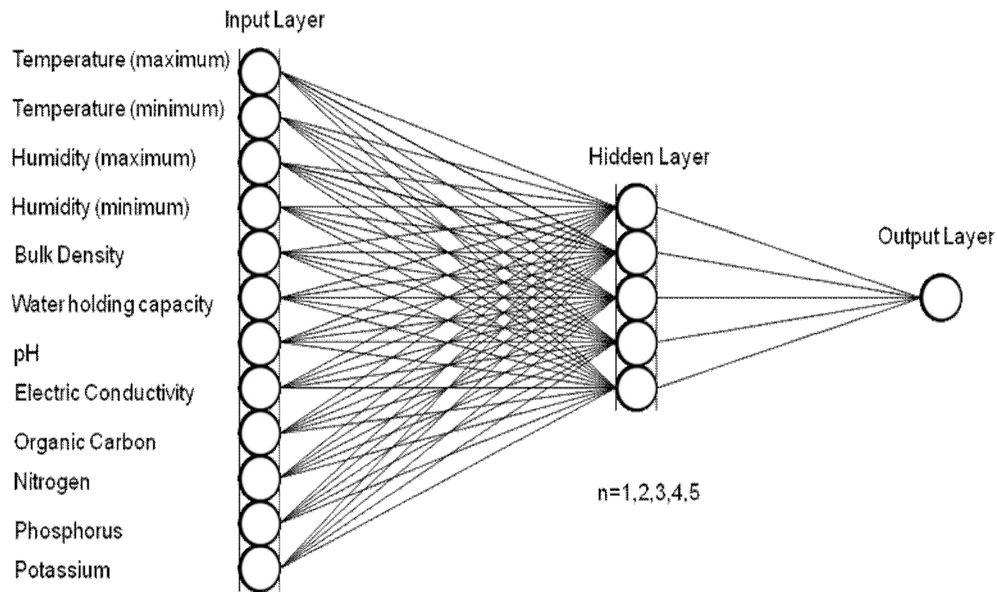


Figure 6.1 Architecture and connection of a feed-forward back propagating artificial neural network with multiple layers.

The evaluation criterion for determining the optimal topology was the best correlation value of the training set. The neural-network model was trained in an iterative training process using the obtained training set as follows:

D_{T},i = {0.007, 0.032, 0.22, 0.406,0.482, 0.507, 0.256, 0.758, 0.005, 0.075, 0.007, 0.024, 0.675}

The first four values referred to the climatic factors and the next eight values belong to soil nutrition parameters and the last number is the corresponding artemisinin content obtained for the plant sample. To avoid possible bias, the order of input–output data pair in a training data set was randomized before the training process. During the training process, the back propagation (BP) training algorithm compares the estimated output value with the target value (namely the measured

value). Then it tunes weighting values connecting all the neurons to minimize the difference between the estimated and the target values until the error is smaller than a predefined level or the number of the iteration reaches a preset maximum number. The constructed model was trained with the input data for an epoch of 10,000 with 0.1 learning rate. After completing the training process all weighting indices that describe the interconnection strengths between neighboring neurons are fixed and the neural-network model will then be capable of mapping input variables to an estimated output promptly and accurately.

The neural-network model developed here was applied to sigmoid transfer function to compute the strength of interconnection between each pair of neurons. The input variables in this model were normalized basing on their possible ranges to avoid data saturation with the help of the following equation:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where x, x_{min}, x_{max} and x_{norm} are the real-valued input variables, the minimum and maximum possible value of the input variables and their normalized value respectively. The output from this neural-network model is an indexed value that corresponds to the input variables. To get the real-valued output, the indexed output value needs to be denormalized according to the following equation:

$$y = y_{norm}\,(y_{max} - y_{min}) + y_{min} \quad (4)$$

where y, y_{min}, y_{max} and y_{norm} are the real-valued output variable, the minimum and maximum possible value of the real-valued output and the indexed output value from the neural-network model.

6.3 RESULTS AND DISCUSSION

6.3.1 Artemisinin content in *A. annua*

Artemisinin was extracted from the whole plant of *A. annua* in triplicate during different stages of growth, every 15 days after germination, corresponding to each valley and analyzed. The artemisinin content gradually increased after seedling stage till full flowering stage. It was found that the artemisinin content in the *A. annua* plants obtained from Indus valley was comparatively more than that of the plant samples collected from Nubra valley (Figure 6.2). At the time of harvesting (at ~ 210

days after germination) the artemisinin content was found to be 0.666 ± 0.008% on dry weight basis from the Indus valley, whereas the artemisinin content was 0.568 ± 0.007% from the Nubra valley.

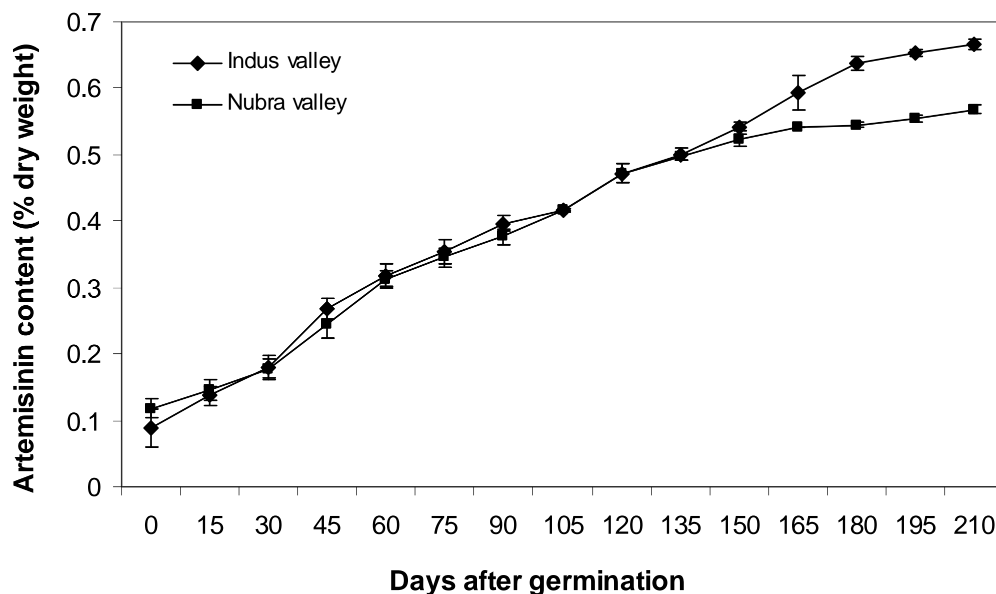


Figure 6.2 Dynamic variation of artemisinin content of whole plant (Leaf and flowers) at different stages of *Artemisia annua*.

6.3.3 Effect of environmental factors on artemisinin content

The environmental factors recorded during the course of the experiment has revealed a wide range in climatic factors like temperature (minimum & maximum) and humidity (minimum & maximum) among the sites from where the samples were collected. At these sites the minimum temperature ranges from 2 ^{0}C to -10 ^{0}C, maximum temperature ranges from 12 ^{0}C to 35 ^{0}C as well as relative humidity varies from 20% to 56% (Figure 6.3).

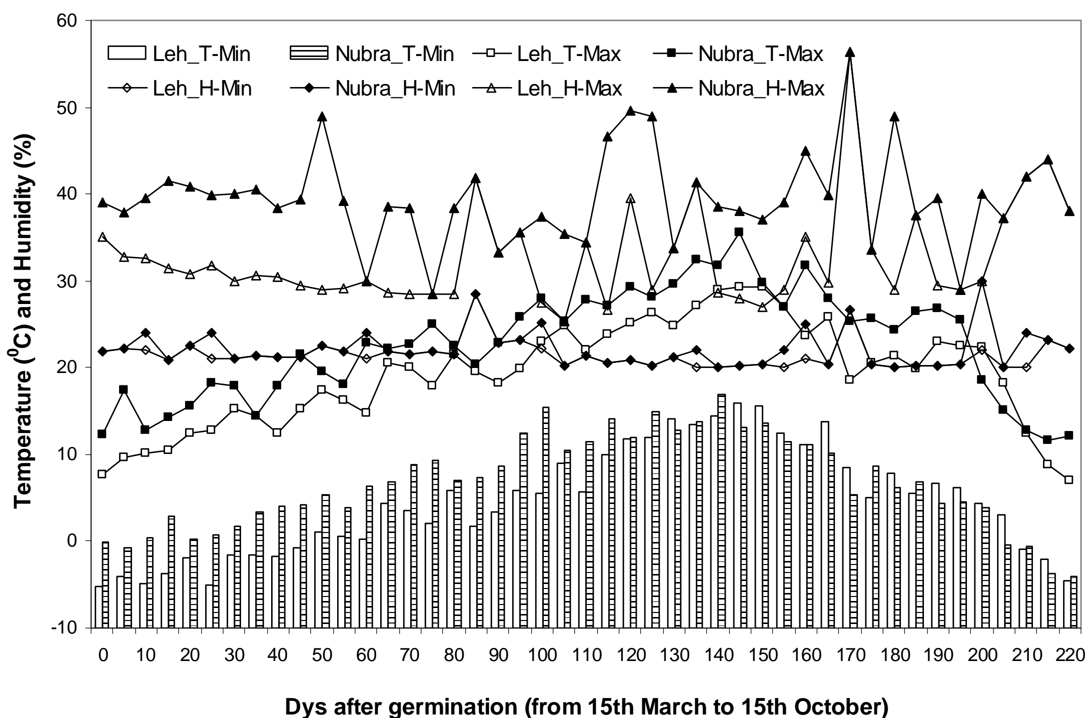


Figure 6.3. The meteorological observations have been made during the course of experiment with respect to the site of collection of *Artemisia annua* populations.

It was seen that the variation in artemisinin content in the *A. annua* is highly dependent on these climatic factors. Reported temperature range for growth is 10-35 °C with the optimum between 13-29 °C. The variation in artemisinin content was seen to be positively related to maximum humidity; r^2 = 0.016 and negatively correlated with minimum humidity; r^2 = 0.011(Figure-6.5a,b). However, the linear correlation coefficient (r^2) was 0.580 for maximum temperature (significant at $P < 0.001$) (Figure 6.4c) and 0.545 ($P < 0.001$) for minimum temperatures (Figure 6.4d) with artemisinin content.

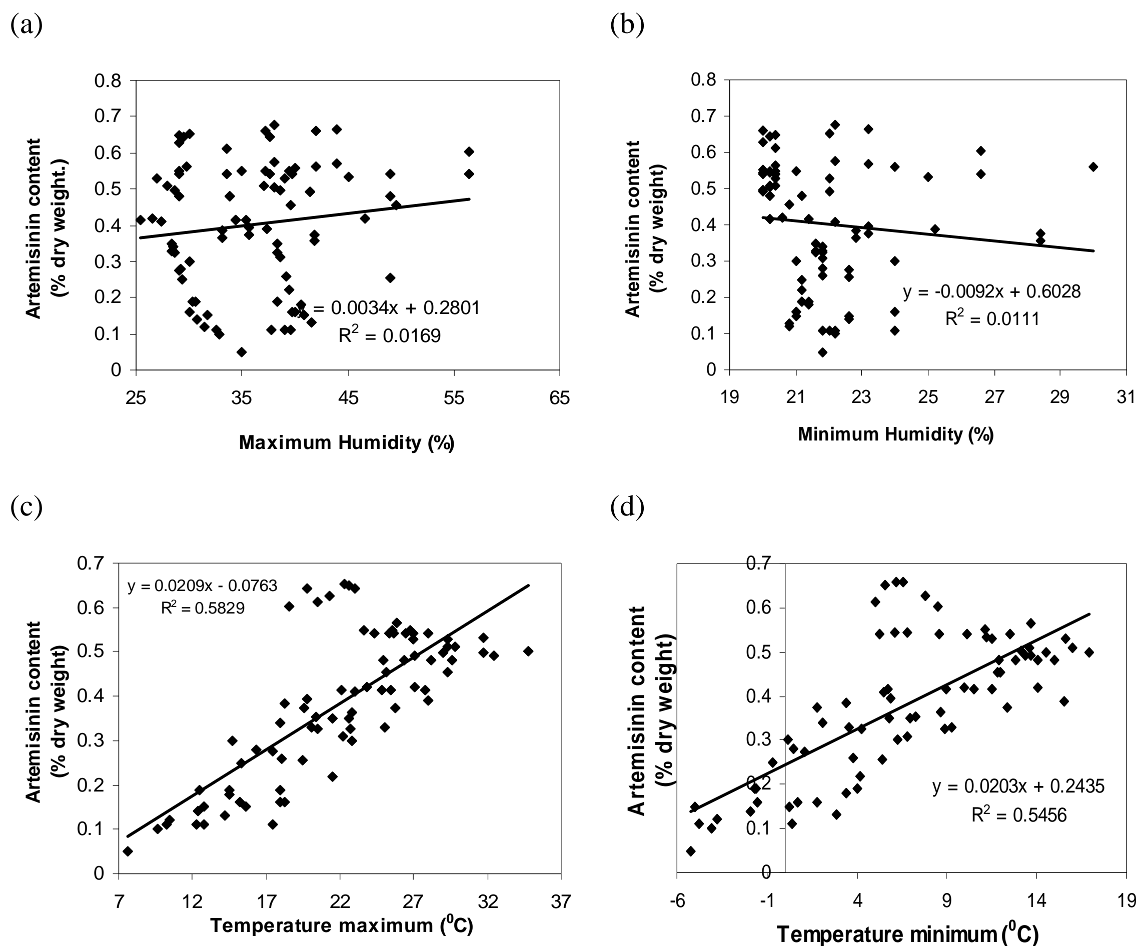


Figure 6.4 (a-d): The relationship between environmental variables humidity % (maximum and minimum), temperature ^{0}C (maximum and minimum).

6.3.4 Effect of soil organic carbon (C)

According to our investigation, soil organic carbon of respective sites ranged between 0.2 to 0.78% from the Nubra valley and 0.21 to 0.93% from the Indus Valley during the entire duration of experiment. Figure 6.5a shows the linear regression between soil organic carbon and artemisinin content in the sample of *Artemisia annua* collected from both the valleys. The artemisinin content reached 0.382% on an average in the soil organic carbon content of 0.616% (on average). Artemisinin content in the plant is positively correlated with the soil organic carbon with significant correlation coefficient R^2 = 0.602. The results demonstrated that high soil organic carbon significantly favored artemisinin production in the *A. annua* found at the altitude higher than 3000 m in Ladakh.

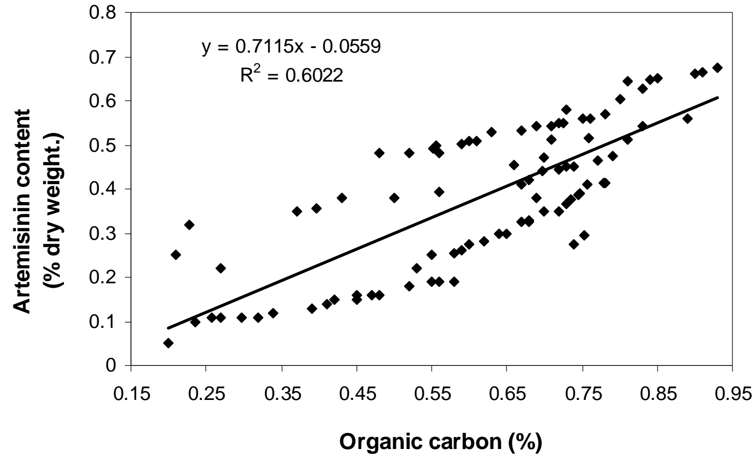


Figure 6.5. The relationship between soil organic carbon and artemisinin contents in *Artemisia annua* in Ladakh region.

6.3.5 Effect of soil pH

Figure 6.6 shows that the soil water pH value from both Indus and Nubra valleys ranges from 7.4 to 8.5 during the period of experiment. It reveals that the soil from the Ladakh region is slightly alkaline in nature. The statistical analysis demonstrated that artemisinin contents in *A. annua* grown at varying soil pH values had negatively correlated with artemisinin content. Artemisinin contents reached about 0.382% (on average) when the soil pH value was 7.769% (on average). Artemisinin contents of the plant is negatively correlated with the soil pH with significant correlation coefficient ($R^2 = 0.743$). This revealed that slightly acidic pH in the soil in Ladakh region may favor the artemisinin content in the *A.annua*.

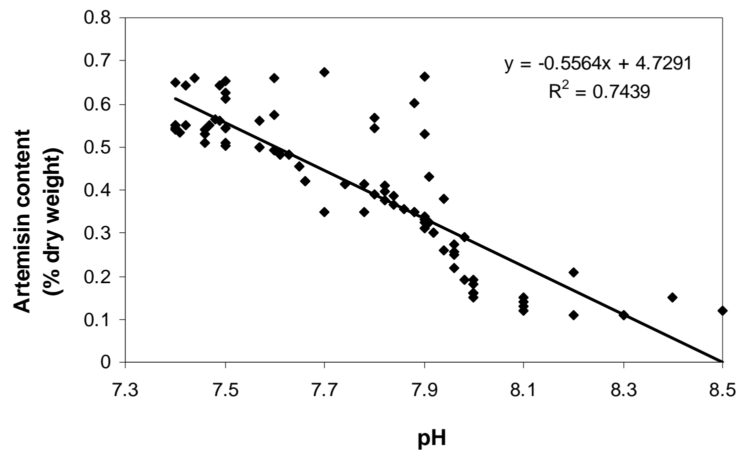


Figure 6.6. The relationship between soil organic carbon, pH, and artemisinin contents in *Artemisia annua* at different altitude in Ladakh region.

6.3.6 Effects of soil nitrogen (N)

In Ladakh from the altitude of 3000 m to 3500 m the total soil nitrogen contents varies in between 0.39% and 0.79%. Soil nitrogen in respective sites ranged between 0.39 to 0.59% from the Nubra valley and 0.61 to 0.79% from the Indus Valley during the entire duration of experiment. The linear correlation coefficient (R^2) was 0.529 reached the statistical significance level ($P < 0.05$). The artemisinin content reached 0.382% (on average) when the soil nitrogen content was 0.624% (on average). The results demonstrated that high soil nitrogen content significantly favored artemisinin production in the *A. annua* found at the altitude higher than 3000 m in Ladakh.

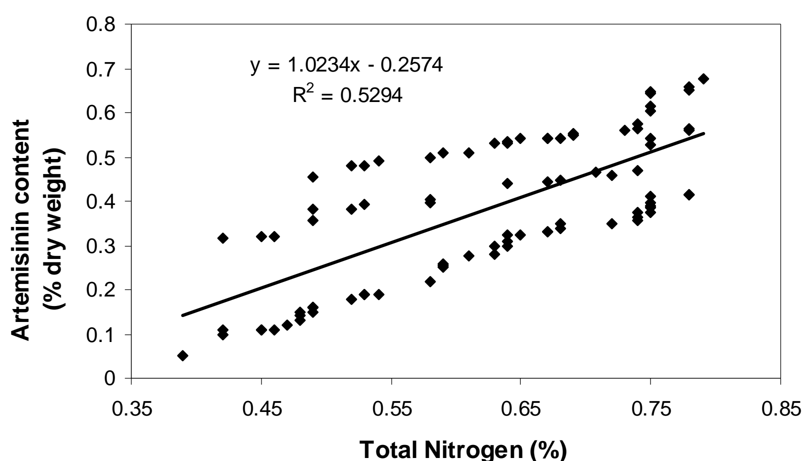


Figure 6.7. The relationship between nitrogen and artemisinin contents in the *Artemisia annua* at different altitudes in the Ladakh region.

6.3.7 Effect of phosphorus (P)

Total phosphorus content in the soil ranged from 0.28% to 0.719% (on average 0.494%) in the Ladakh region. Figure 6.8a shows the relationships between soil phosphorus content and artemisinin content in *A. annua*. The linear correlation coefficient (r^2) was 0.886 and the value was statistically significant ($P < 0.001$). Phosphorus content of soil is negatively correlated with artemisinin content. The results demonstrated that with the increase of soil phosphorous content above 0.71% significantly reduce yield of artemisinin in the plant.

6.3.8 Effects of soil potassium (K)

Total soil potassium contents from the Ladakh region varies between 0.421% and 0.759%. The artemisinin content in *A. annua* is negatively correlated with soil

potassium content (Figure 6.8b). The correlation coefficient (R^2) was 0.958 and the value was statistically significant ($P < 0.001$). This indicates that total soil potassium content also had a significant relationship with the artemisinin production in *A. annua.*

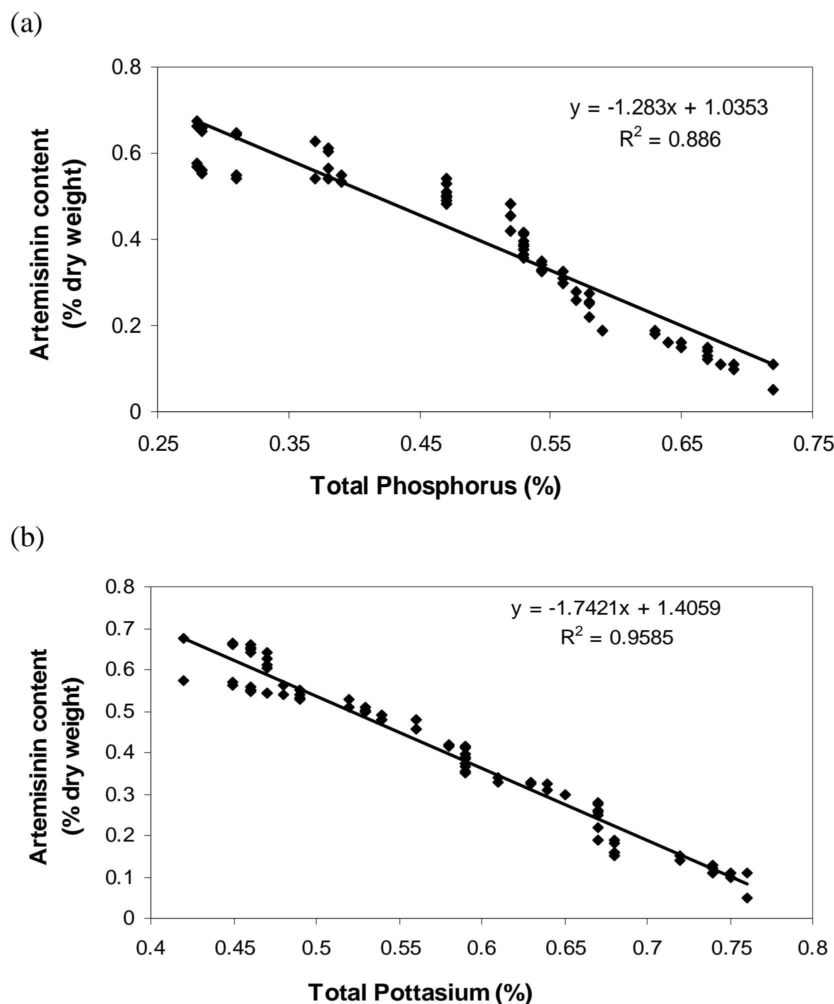


Figure 6.8. The relationship between (a) soil phosphorus and artemisinin content; (b) soil potassium and artemisinin content in the *Artemisia annua* in the Ladakh region.

6.3.9 Performance measure of ANN and MLR model

The ANN and MLR models developed in this study was statistically best fitted and consequently used for management of soil and environmental factors for enhancement of artemisinin yield of *A. annua* in the Ladakh region. The relationships between predicted (using both ANN and MLR models) and the corresponding experimental values is shown in Figure 6.9. The R^2 value of 0.982 and 0.806 as well as root mean square error (RMSE) for the models are 0.018 and 0.062 based on ANN and MLR, which are an indicator of the quality of fit between experimental and

predicted artemisinin yield. The average predicted and experimental values of artemisinin yield in *A. annua* corresponding to two valleys from the Ladakh region are given in Table 6.2. Comparing the results (the R^2 value and RMS error) it is revealed that the neural-network model can provide more accurate estimations of artemisinin values than a 'best-fit' regression model. More importantly, a well trained neural-network model can be implemented in real-time to estimate the required artemisinin content in terms of changed environmental and soil factors with a minimal computational load.

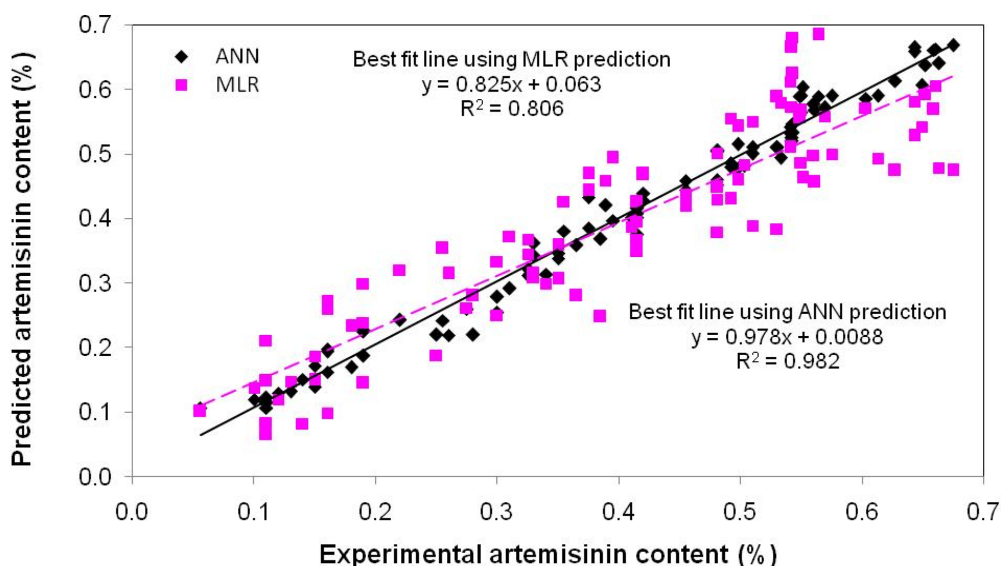


Figure 6.9. Comparison of estimated artemisinin content using an artificial neural-network model (—, ANN) and a 'best-fit' regression model (---, MLR). Here R^2 represent coefficient of determination.

Table 6.1. Experimental and predicted artemisinin yield in *A. annua* corresponding to Indus valley and Nubra valley. Values are represented in mean ± standard deviation.

	Artemisinin yield (% dry weight)		
	Experimental	Predicted artemisinin yield	
		ANN	MLR
Indus valley	0.415 ± 0.185	0.413 ± 0.181	0.391 ± 0.171
Nubra valley	0.389 ± 0.156	0.390 ± 0.156	0.398 ± 0.142

6.4 DISCUSSION

The existing variations in artemisinin content of *A. annua* population from the Ladakh region are very much dependent on environmental variables and soil nutritional factors. Plants from the Indus valley consists of slightly more artemisinin

content in comparison to Nubra valley. The altitude ranges from 3000 m (Nubra Valley) to 3500 m (Indus valley) – there is a wide variation in environmental factors that led to variation in artemisinin content among the populations used in the study. The present study revealed that artemisinin production has close relationship with high soil carbon, slightly alkaline soil pH, high soil nitrogen, low concentration of soil phosphorus and potassium as well as environmental variables such as temperature and humidity.

Artemisinin content in the plant samples collected from these regions was very much influenced by temperature. Similar observation was made by various workers previously. For an example, Wallaart et al. [31] recorded increased levels of artemisinin (58%) in the levels of Vietnamese *A. annua* cultivar after a night frost period. This is possibly a stress-induced phenomenon and needs to be confirmed under controlled environment conditions. Wallaart et al. [31] stated that the increased artemisinin after a frost is consistent with the hypothesis that stress will trigger the conversion of dihydroartemisinic acid to artemisinin, the final step in the biochemical pathway. It is also assumed that the presence of high levels of dihydroartemisinic acid may be an adaptation to stress conditions (e.g. night frost), during which relatively high levels of O_2 are formed, and that dihydroartemisinic acid might protect the plant by reacting with these reactive oxygen species and yielding artemisinin as a stable end-product. Marchese [215] found high artemisinin content (0.41%) in *A. annua* plants from Brazil at temperatures ranging from 18–28 ^{0}C in comparison to plants submitted to 11–20 ^{0}C (0.36%). Ferreira et al. [18] also indicated decrease in artemisinin content (0.04 to 0.01%) in a Chinese clone of *A. annua*, under greenhouse conditions with higher temperatures observed from April to July, compared to cooler greenhouse temperatures observed from January to March. Singh et al. [216] also found artemisinin to be higher in *A. annua* plants grown in the temperate climates of Kashmir than in the subtropical climate (0.1% versus 0.06%) of Lucknow, India. Although water stress and temperature regimen cannot be controlled under field conditions, these results indicate that *A. annua* plants can tolerate some changes in water availability and temperature without a drastic decrease in artemisinin content. However, it is important to note that variations in artemisinin content may depend on the origin of *A. annua* cultivar and on the regional environmental conditions.

Among soil nutritional factors, soil organic matter and nitrogen are most significantly correlated with the artemisinin production. Our work demonstrated that the soil in the Ladakh region is slightly alkaline (pH ranges from 7.4 to 9.5) and *A. annua* has not only adapted well to it but it also produces more artemisinin in slightly alkaline soil. Other researches also reported that pH value influenced secondary metabolites production. It prefers light to medium textured soils (sandy and loamy soils). It requires well drained or dry soils and thrives in fertile soils but will grow in nutritionally poor soils. The plants are longer lived, hardier and more aromatic when they are grown in a poor dry soil. *Artemisia annua* can be grown under wide range of soil pH (7.4–8.5), depending on the plant origin, but there are only a few studies on the effect of soil pH on the vegetative growth and artemisinin concentration in *A. annua* [56]. Yugoslavian strains of *A. annua* grew well at pH 5.4–7.4. However, the Chinese strain was more tolerant of both high (8.2) and low (5.0) pH conditions than the Yugoslavian strain [60]. Hydrogen ions in soil change the membrane permeability of the root cell and thus soil pH directly affects the growth of plant. It also affects indirectly the uptake of soil nutrients by plants [110]. Therefore, the high artemisinin yield in alkaline soil may be caused by the influence of the alkaline soil on the availability and uptake of soil elements such as N, P and K.

Soil organic matter provides plants with NPK and essential metal co-factors for metabolism. High soil organic matter content can uniformly supply the nutrition to plants, guarantee the plants a good growth and metabolic status and enhance the resistance of the plants to stresses. Our work demonstrated that high soil organic matter favors high artemisinin yield of *A.annua.* Other soil factor significantly affecting the yield of artemisinin is Nitrogen (N). Soil N has close positive relationship with plant growth and metabolism because N is a structural component of amino acids from which proteins are synthesized and wide metabolisms take place. Organic N might be more important to high artemisinin yield than that of inorganic nitrogen. Therefore, how N nutrient and which forms of N take part in biosynthesis of artemisinin is the essential work for improving artemisinin yield in the future study.

High phosphorus and potassium content in soil inhibited the artemisinin yield. The reduced phosphate content in the culture medium was also reported to increase secondary metabolite accumulation in other plant species [217, 100]. The mechanism

of P and K affecting the artemisinin yield in the cellular bioprocess remains unclear. It might be due to the fact that both elements regulate the activity of certain enzymes involved in artemisinin biosynthesis.

Many Indian medicinal herbs have higher medicinal productivity in their original habitat than in cultivated lands. Soil nutrient characters and environmental factors similar to original habitats must be most suitable for the active compound production. According to the results in this research, 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. The first approach is to measure the key soil nutrient factors such as pH, organic matter, nitrogen, phosphorous and potassium content when choosing a farm for the cultivation of *A.annua*. The second approach is the balanced fertilization. It is important and necessary to increase and maintain soil organic matter and N content at high levels during the period of the plants' growth and harvesting season. Organic N fertilizers are suggested to be used for the cultivation of *Artemisia annua*.

The secondary metabolite production is influenced by the plant's own physiological age and status and other environmental factors. Therefore, the effects of soil on artemisinin production of *A. annua* are far complicated beyond those mentioned above. Impact of these topological and climatic factors with artemisinin content is not studied so far. However, it is assumed that the variation in artemisinin content is dependent on these factors which call for further research. 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 conservation of the plant. The results showed that using a combination of topographic soil and environmental data, we were able to successfully predict artemisinin yield with ANN and MLR**.** 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.

CHAPTER 7

Artemisia tournefortiana: an alternative and promising source of artemisinin

ABSTRACT

Currently the primary source of artemisinin (the antimalarial compound) is *Artemisia annua*. Due to ever-increasing resistance of malaria parasite strains to the available antimalarial drugs and growing concern of inadequate artemisinin supply to meet the future requirements for artemisinin based-combination therapies, it is imperative to search for supplementary artemisinin sources. 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%). Further the artemisinin content in flower vary from 0.05 to 1.43% with respect to different geographical location. Our results thus indicate a great potential for the use of *A. tournefortiana* as an alternative source of artemisinin.

7.1 INTRODUCTION

Malaria is one of the world's most important parasitic diseases threatening half of the global population [7]. It has been estimated that malaria transmission occurs in 109 countries putting 3.3 billion people at risk [43]. Furthermore, the disease is estimated to be responsible for an average annual reduction of 1.3% in economic growth for countries with the heaviest malaria burden [218]. Over many decades, different variants of quinine such as chloroquine (4-amino-quinoline), mefloquine (quinoline methanol), primaquine have been used for malaria treatment [219]. But the anti-malarial efficacy of each of these has been far from satisfactory due to evolution of resistant strains of the parasite within the last two decades due to indiscriminate usage of the drugs [220]. 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. These compounds induce more rapid reduction of parasitemia [29], decreasing the number of parasites faster than any other known drug. As a consequence they are of special interest for severe malaria [30]. Recently, World Health Organisation (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. Due to this discovery, *A. annua* is now rated as one of the top ten industrial crops of the modern world. Recent estimates suggest that current market for artemisinin has reached US$343 million [221], and at the current rate its supply from its known natural sources, *A. annua* and other related species [54, 222] will soon be outdone. 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 [58]. 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 situated at an altitude of 2700-6100 m above mean sea level between latitudes 30^0 and 36^0 N and longitudes 76^0 and 79^0 E. The region is characterized by harsh climatic conditions but is the home of many medicinal plants [223]. Due to diverse and complex land formations the flora of this region is isolated from its neighboring ecological regions

[223]. Further the available biodiversity, pharmacological importance and chemical characterization of the plants from this region is mostly unexplored. 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. All these species show morphologically different from each other (Figure 7.1).

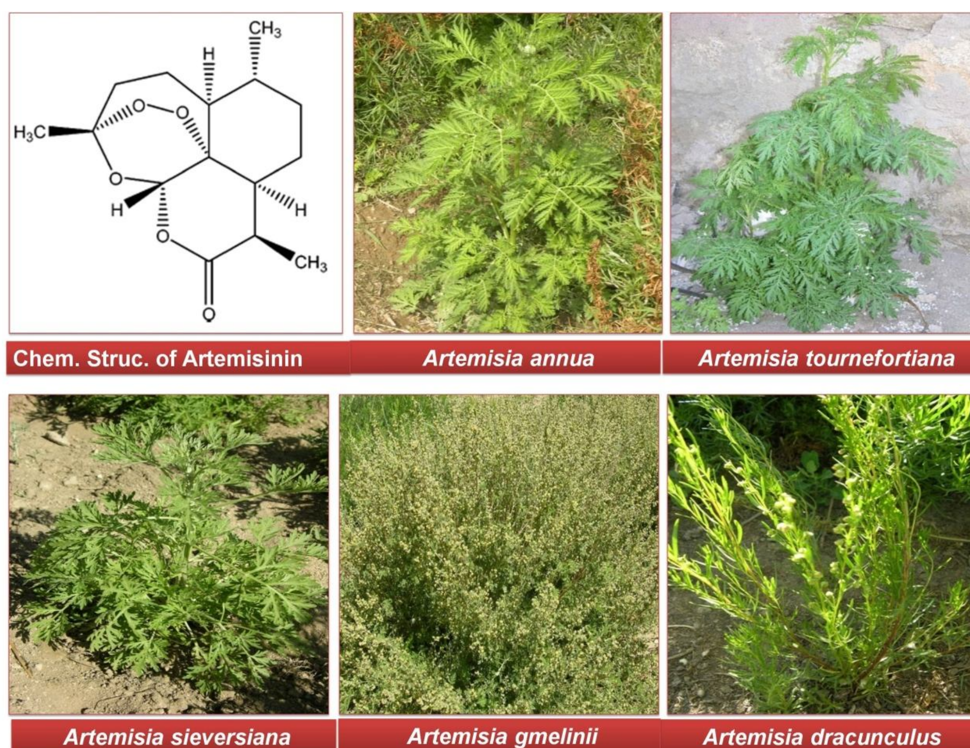


Figure 7.1 Five species of Artemisia: *A. annua*, *A. tournefortiana, A. sieversiana*, *A. gmelinee* and *A. dracunculus* that are are predominantly available in Ladakh region.

In order to ascertain the alternative source of artemisinin we have performed the quantitative estimation of artemisinin from these species and compared. We reported here another source for artemisinin production; *Artemisia tournefortiana* that grow in the Trans-Himalayan region of Ladakh (Figure 7.2 includes the different plant parts) with artemisinin content albeit at low concentration in comparison to *A. annua*. Further we compared the artemisinin content of *A. tournefortiana* with the artemisinin content from various species reported worldwide in the literature. The findings thus indicate a great potential for the use of *A. tournefortiana* as an alternative source of artemisinin.

Figure 7.2 Plant parts (leaf, flower, root and seeds) of *Artemisia tournefortiana*, a new source for the alternative source of artemisinin.

7.2 MATERIALS AND METHODS

7.2.1 Plant material and chemicals

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. HPLC grade standard artemisinin was purchased from Sigma-Aldrich, Co. (India). All the solvents used were of HPLC grade purchased from MERCK Specialties Private Limited (India).

7.2.2 Sample preparation

We have used a simple, fast and micro-scale analysis procedure for quantification of artemisinin with slight modification as suggested by Liu et al.[189]. In brief, 100 mg of plant sample (*viz.* root, stem, leaf and flower) was taken and artemisinin was extracted with 1 ml of ethyl acetate in 2 ml centrifuge tube and by vortexed for 2 min at room temperature. Then the samples were centrifuged at 10000g

for 2 min and the supernatant was transferred into a fresh 2 ml centrifuge tube and evaporated to dryness under steam nitrogen at room temperature. The residual was then dissolved with 1 ml of acetonitrile and used for HPLC analysis.

7.2.3 Preparation of standard stock solution and calibration curve

Artemisinin standard was obtained from Sigma Aldrich, USA, and stock solution with concentration of 1 mg/ml in acetonitrile was prepared. A desire range of concentrations of artemisinin were prepared from the stock solution by serial dilution with acetonitrile. Five replicates were used for each level of concentration in order to check the reproducibility of the detector response at various concentrations. Exactly 10.0 µl of the sample was injected using autosampler (Waters Inc.). Calibration curve was generated by plotting a graph of peak area against the concentrations used and has been used to quantify the amount of artemisinin in different samples under analysis. Excellent linearity was retained for artemisinin between peak areas and concentrations in the range of 0.1-100 µg/ml (regression equation, $Y = 10.0031 X + 14.8232$; $r^2 = 0.99908$).

7.2.4 HPLC system, RP-HPLC-PDA conditions and chemical analysis

Detection and quantification of artemisinin from *Artemisia* species 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°C and C18 column (5 µm; 250 mm x 4.6 mm). Extracted artemisinin from the plant samples were analyzed by injecting 10 µl of samples into HPLC system using autosamplers (Waters Inc.). The condition on the PDA was set as follows: mobile phase consisting acetonitrile to achieve maximum separation and sensitivity at 229 nm; flow rate of 0.6 ml/min with stop and rest time of 12 and 5 min. The peak of artemisinin was observed at the retention time of 4.796 min (Figure 7.1a). Similarly Figure 7.1b shows the representative chromatogram of a plant sample. Artemisinin content asanalyzed in five replicates and was expressed as the mean ± standard deviation.

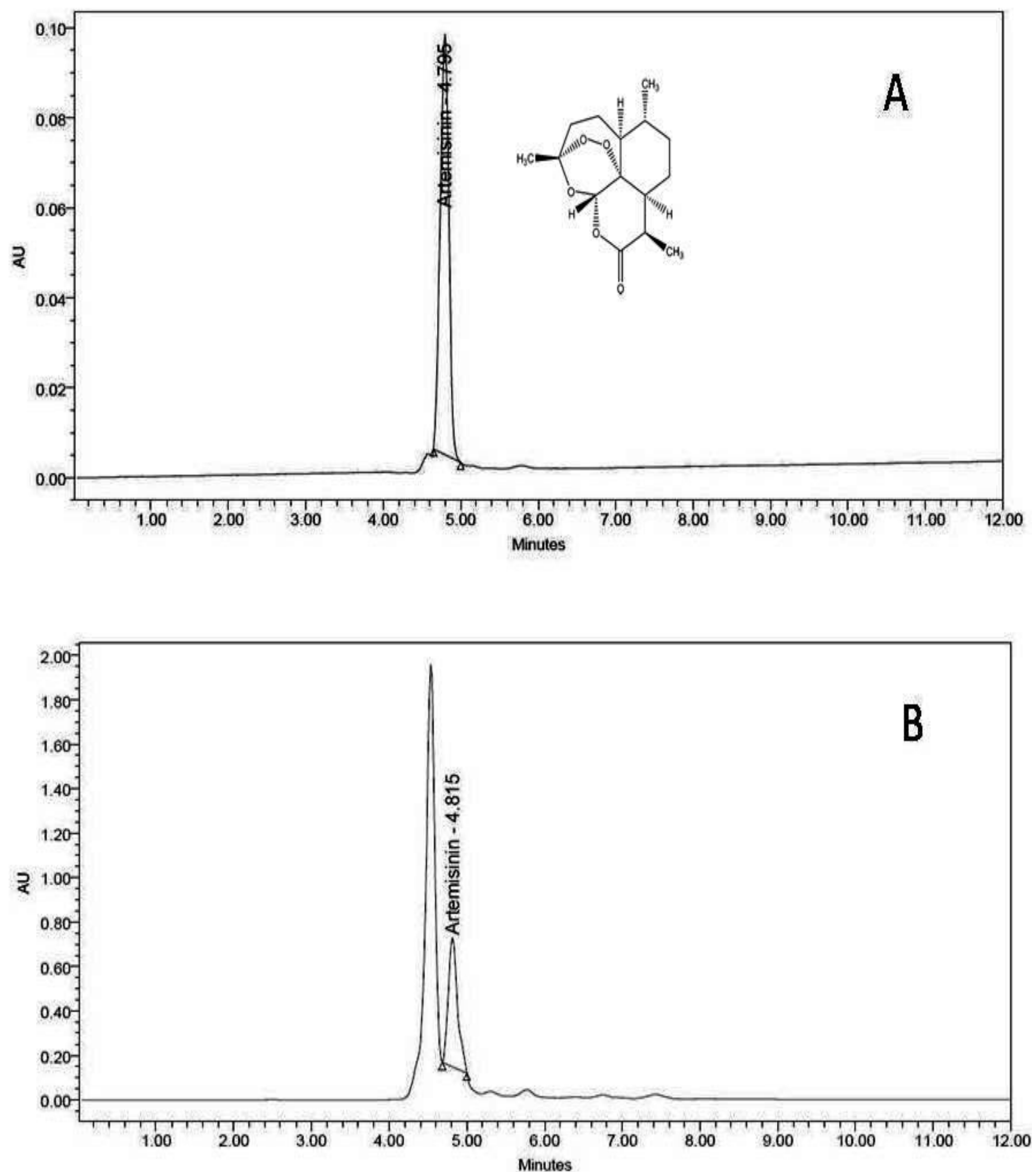


Figure 7.3. Chromatograms of standards of artemisinin (A) and plant extract (B) by HPLC-PDA showing retention time at 4.796 and 4.815 minutes.

7.2.5. Amplification of genes involved in artemisinin biosynthetic pathway and sequencing

Gene specific primers were designed considering the gene sequencing of the enzymes involved in the artemisinin biosynthetic pathway from *A. annua*. Some of the gene sequences from *A. annua* are available (highlighted in yellow color in Figure 7.2) and were taken from the GeneBank and specifc primers were designed using Primer3. These primers were custome synthesized from Xcelris Genomics

Laboratory. The genes were amplified using Biometra PCR machine, amplification reactions were performed in volumes of 50 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, 40 ng template DNA and 1.0 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.

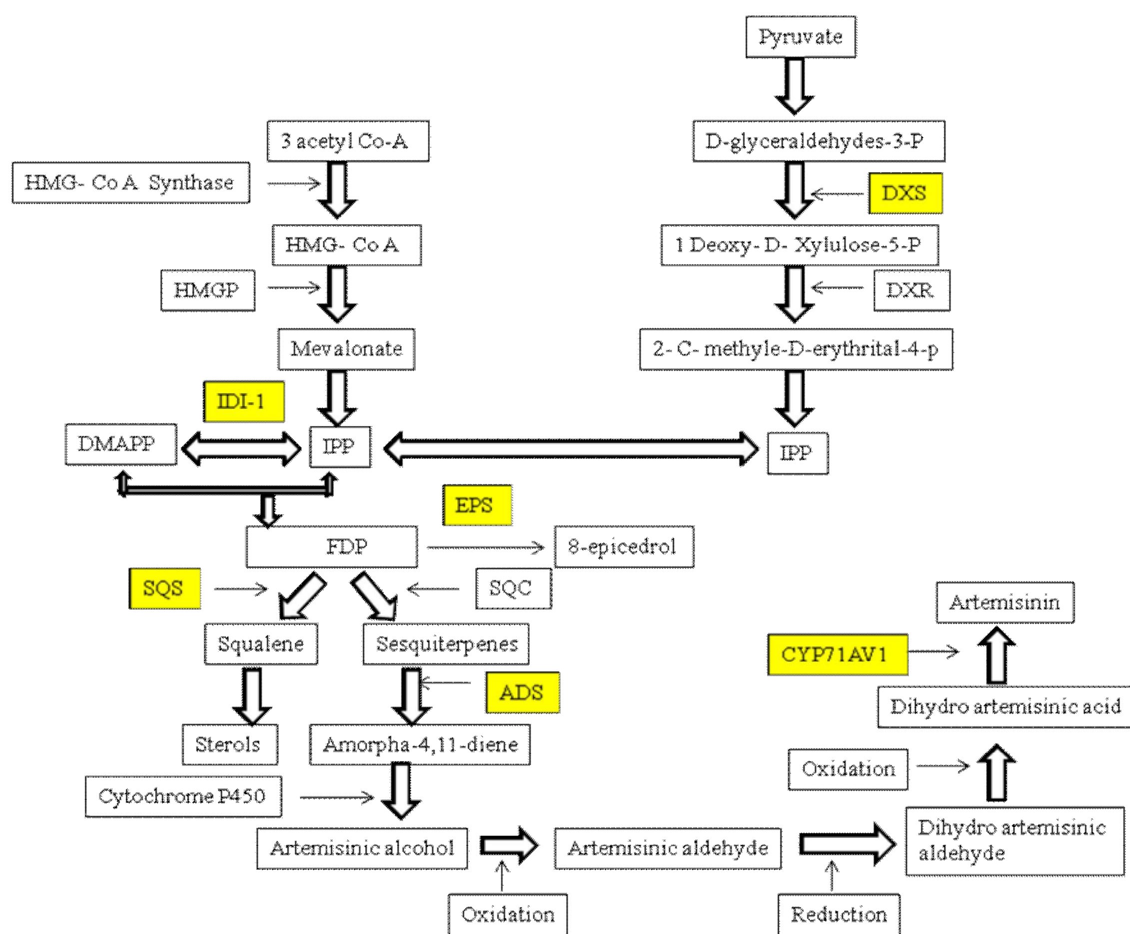


Figure 7.4. Artemisinin synthesis pathway gene that has been amplified in *A. tournefortiana.*

The amplified products were gel purified and sequenced directly from Xcelris Genomic Laboratories, Ahmedabad, India. DNA sequencing was performed by using the Sanger dideoxy-chain terminator method on an automated sequencer using Dye

Terminator chemistry. The assembly of contigs among *Artemisia* genotypes was done by Consed software. Its operations are integrated with the use of the Phred, Phrap and Consed programs [224] and together these tools can generate a high throughput system for detecting DNA assembly by large scale fluorescence-based resequencing.

7.3 RESULTS AND DISCUSSION

Artemisinin content from five species: *A. annua, A. tournefortiana*, *A. sieversiana*, *A. dracunculus* and *A. gmelinii,* all of which are endemic to the region of Ladakh were analyzed using RP-HPLC-PDA system. Artemisinin content was found to be highest in *A. annua* (0.675 ± 0.1% dry weight) followed by *A. tournefortiana* (0.39 ± 0.08% dry weight) and *A. sieversiana* (0.11 ± 0.04% dry weight), whereas artemisinin was not detected for the other two species, *A. gmelinii* and *A. dracunculus* (Figure 7.3). Artemisinin content of *A. tournefortiana* was highest in flower (0.05−1.43% dry weight) in comparison to root (0.14−0.33% dry weight) whereas it was not detected in leaves and stems (Figure 7.4). Even the artemesinin content in the flowers of *A. tournefortiana* is more than *A. annua* (0.675 % dry weight) in the Ladakh region. Difference in artemisinin content in *A. tournefortiana* with respect to other species reported worldwide is also shown in Figure 7.4. Further the artemisinin content that has been reported worldwide from the *A. annua* vary from 1.16% −1.38% among high yielding variety [225] whereas the wild plant consists of 0.02% −1.09% (Figure 7.5). Therefore *A. tournefortiana* emerged as a substitute and most promising source of artemisinin—the first report from the Trans Himalayan region, Ladakh. Moreover *A. tournefortiana* grow rapidly into large plants and reached an average height of about 3-3.5 m in just 7-8 months after germination. The biomass of *A. tournefortiana* is approximately 40 times more in comparison to *A. annua*. and composed of more inflorescences compared to *A. annua.*

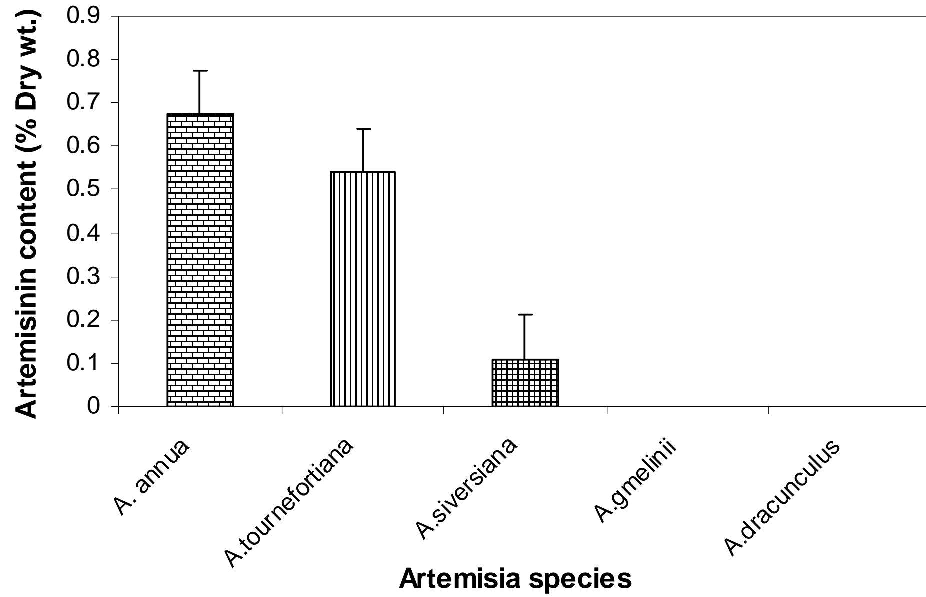


Figure 7.5. Analysis of artemisinin content from different *Artemisia* species grown in Ladakh.

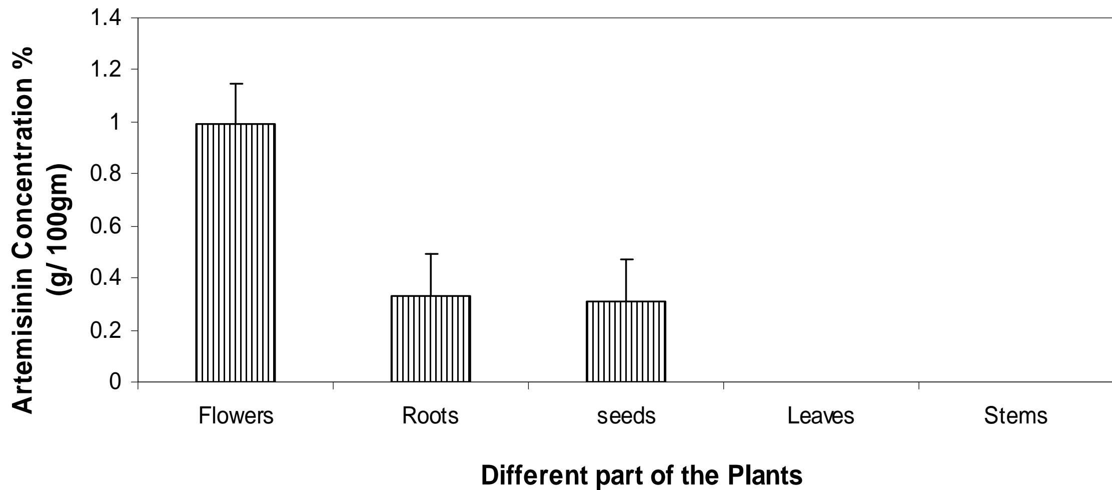


Figure 7.6. Comparison of artemisinin content in different plant parts of *Artemisia tournefortiana.*

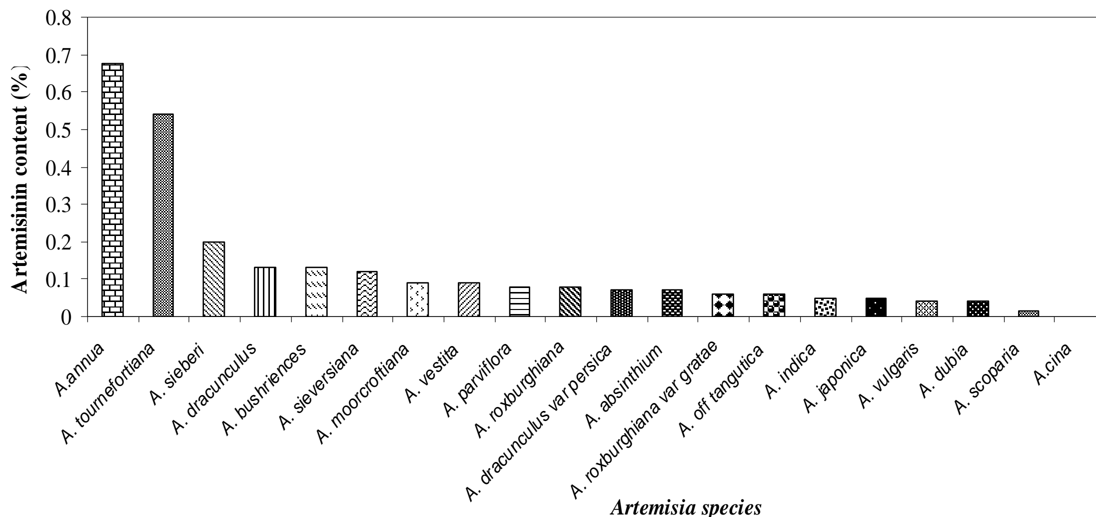


Figure 7.7. Comparison of artemisinin content of *Artemisia tournefortiana* with various species of *Artemisia* reported world wide.

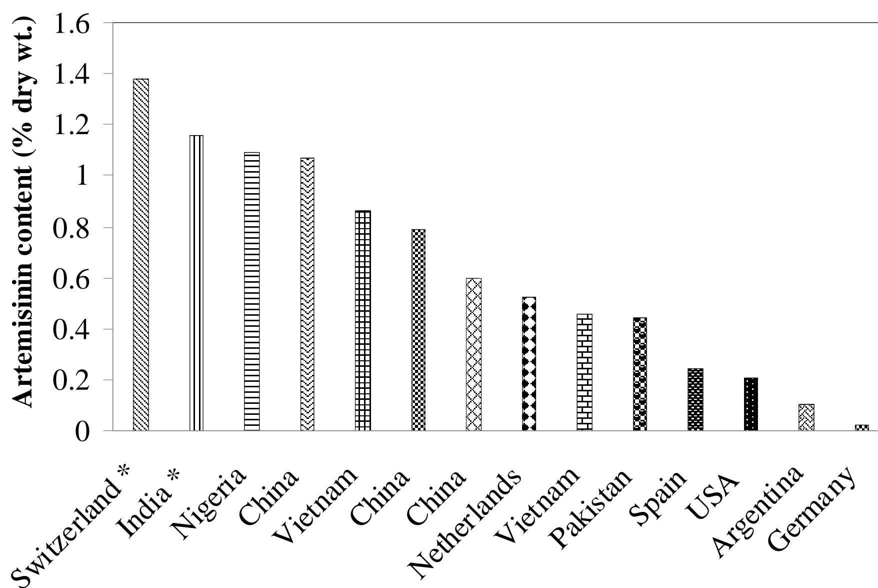
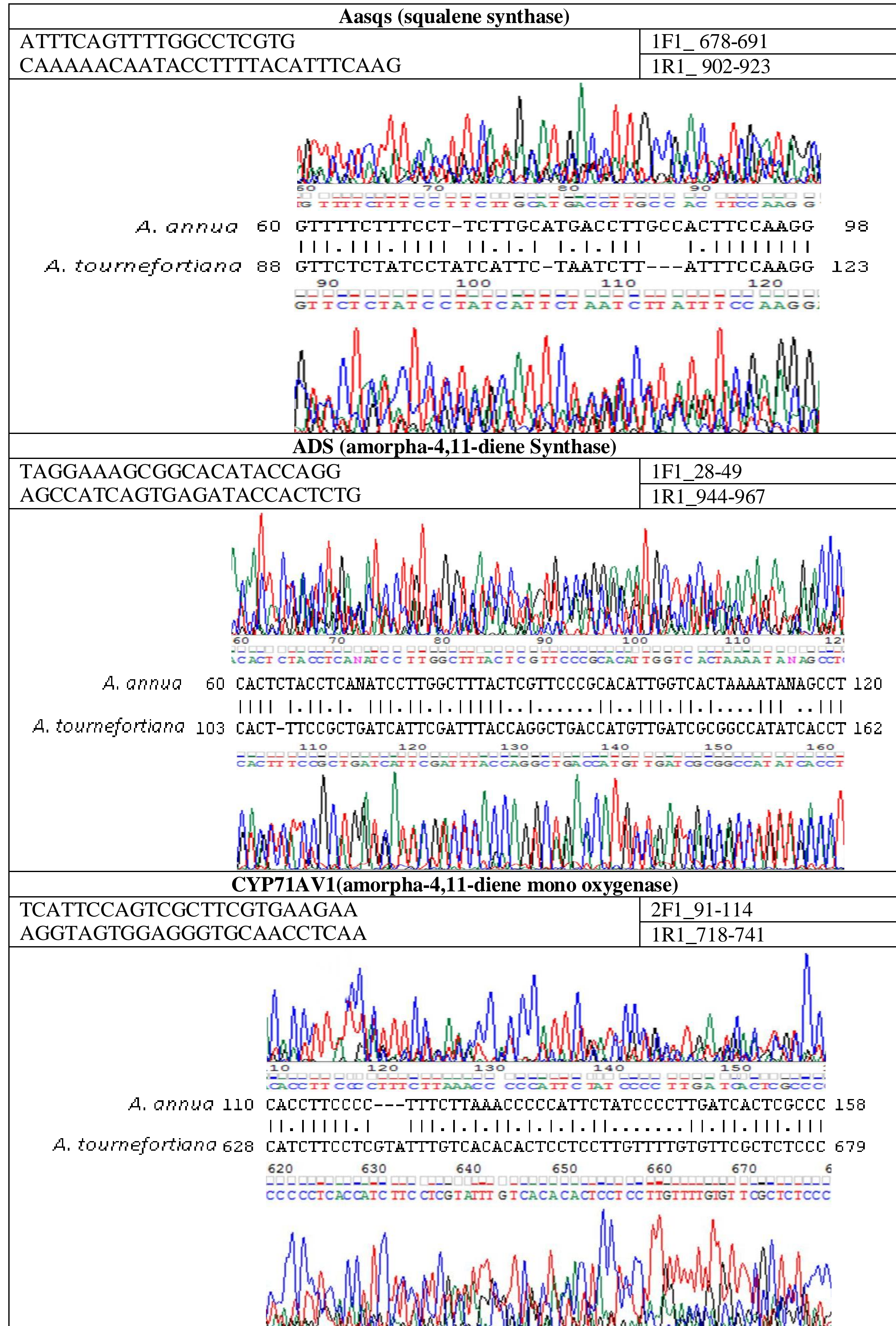


Figure 7.8. Artemisinin contentin (%) of *Artemisia annua* reported worldwide.

To further validate the existence of artemisinin biosynthetic pathway in *A. tournefortiana* we have compared the gene sequences that have been amplified from both *A. tournefortiana* and A. annua. The partial pairwise sequence similarity using EMBOSS package is included in Table 7.1. The close similarity between the gene sequences revealed that similar synthetic pathway of artemisinin is also exist in *A. tournefortiana.*

Table 7.1. 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
A. annua 60 GTTTTCTTTCCT-TCTTGCATGACCTTGCCACTTCCAAGG 98 *A. tournefortiana* 88 GTTCTCTATCCTATCATTC-TAATCTT---ATTTCCAAGG 123	
ADS (amorpha-4,11-diene Synthase)	
TAGGAAAGCGGCACATACCAGG	1F1_28-49
AGCCATCAGTGAGATACCACTCTG	1R1_944-967
A. annua 60 CACTCTACCTCANATCCTTGGCTTTACTCGTTCCCGCACATTGGTCACTAAAATANAGCCT 120 *A. tournefortiana* 103 CACT-TTCCGCTGATCATTCGATTTACCAGGCTGACCATGTTGATCGCGGCCATATCACCT 162	
CYP71AV1(amorpha-4,11-diene mono oxygenase)	
TCATTCCAGTCGCTTCGTGAAGAA	2F1_91-114
AGGTAGTGGAGGGTGCAACCTCAA	1R1_718-741
A. annua 110 CACCTTCCCC---TTTCTTAAACCCCCATTCTATCCCCTTGATCACTCGCCC 158 *A. tournefortiana* 628 CATCTTCCTCGTATTTGTCACACACTCCTCCTTGTTTTGTGTTCGCTCTCCC 679	

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

```
A. annua          274 ACCATGTTGTATGTGAATGAGGAAGGAGTGGCGGTGGAT 312
                      ||||||||.||......||..|.|||.|.||..|.|.||
A. tournefortiana  61 ACCATGTTATAATACCCTGGTGGAGGGGGGGNCGCGAAT  99
```

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

```
A. annua          150 TAAGTGCTTATT--GCTTGCCAAAAATCCTACTTCTACAATTGGTA 193
                      |.||..||||||  |.|||..|||.|.||..||..|.||||||.||
A. tournefortiana 190 TCAGCCCTTATTTGGTTTGAAAAACAACCCCCTGATTCAATTGATA 235
```

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

```
A. annua          339 TTTTATTACTATTAGTTTTCTTGGAACATAACCCCTCGGTTGTTGTAGAAAAACTTTACTTTCCTAANAAACTT 412
                      ||.|.||..||||.||..|.|.|||.||.     ||||||.|.||||..|.| |.||..|..|||.. ||||||
A. tournefortiana 542 TTGTCTTTATATTTGTGATTTGGGATCAG-----CTCGGTGGCTGTATTACA-CATTGTTGACCTTCCAAACTT 609
```

Till date, to the best of our knowledge, out of over 400 *Artemisia* species growing worldwide [226], only 32 species, other than *A. annua*, have been screened for the artemisinin content. Among these, just 19 species: *A. dracunculus var. persica, A. off tangutica, A. absinthium, A. japonica, A. dubia, A. sieversiana, A. roxburghiana var. roxburghaiana, A. dracunculus var. dracunculus, A. bushriences, A. moorcroftiana, A. vestita, A. vulgaris, A. indica, A. roxburghiana var. gratae, and A. parviflora* from Pakistan [49] *A. apiacea* and *A. lancea* from China [53], *A. scoparia* from India [50], *A. cina* from Indonesia [51], and *A. sieberi* from Iran [54] were reported to have fewer amount of artemisinin ranging from 0.0006% − 0.2% [49]. These data revealed that till date there is no alternate and economically feasible source of artemisinin other than *A. annua.* However, we reported here that *A. tournefortiana* indicates a great potential as an alternative source of artemisinin. Considering the economic viability of extraction of artemisinin from *A. tournefortiana*, its scientific conservation and biotechnological intervension may even increase the yield of artemisinin.

It has been concluded that *A. tournefortiana* could be served as and alternative and promising source of artemisinin rather than *A. annua*. This Artemisia species is widely and easily grows in high altitude cold desert Trans-Himalayan region of Ladakh. However, its full geographic distribution is beyond the scope of this study. *A. tournefortiana* yield comparable amount of artemisinin to that of *A. annua*. It has been used as a folk therapeutic plant for various purposes by the traditional healers associated with Amchi System of Medicine [223] and can be screened further for the economically feasible extraction of artemisinin.

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.

Bibliography

[1] Efferth T., Dunstan H., Sauerbrey A., Miyachi H., Chitambar C.R., *"The anti-malarial artesunate is also active against cancer",* International Journal of Oncology, 18, 767-763, 2001.

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

[3] Utzinger J., Keiser J., *"Schistosomiasis and soil-transmitted helminthiasis: common drugs for treatment and control",* Expert Opinion of Pharmacotherapy, 5, 263–285, 2004.

[4] Romero M.R., Efferth T., Serrano M.A., Castano B., Macias R.I.R., Briz O., Marin J.J.G., *"Effect of artemisinin/artesunate as inhibitors of hepatitis B virus production in an "in vitro" replicative system"* Antiviral Research, 68, 75-83, 2005.

[5] Wallaart E, Pras N and Quax., *"WJ Seasonal variations of artemisinin and its biosynthetic precursors in tetraploid Artemisia annua plants compared with diploid wild-type".* Planta Medica 65: 723–728, 1999.

[6] 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 Medicine, 4, 309, 2007.

[7] *World Malaria Report 2010*. Geneva, World Health Organization, 2010.

[8] Klayman *D.L.,"Qinghaosu (artemisinin): an antimalarial drug from China*", Science, 228, 1049-1055, 1985.

[9] Klayman D.L., Lin A.J., Acton N., Scovill J.P., Hock J.M., Milhous W.K., and Theoharides A.D., *"Isolation of artemisinin (qinghaosu) from Artemisia annua growing in the United States"*, Journal of Natural Products 47, 715-717, 1984.

[10] Singh A., Vishwakarma R.A., Husain, A., "*Evaluation of Artemisia annua strains for higher artemisinin production".* Planta Medica, 64, 475–476, 1988.

[11] Tan R.X., Zheng W.F., Tang H.Q., *"Biologically active substances from the genus Artemisia",* Planta Medica, 64, 295-302. 1998.

[12] Iranshahi M., Emami S. A., Soltani M.M., *"Detection of Sesquiterpene Lactones in Ten Artemisia Species Population of Khorasan Provinces",* Iranian Journal of Basic Medical Sciences, 10, 183–188, 2007.

[13] Temraz A., El-Tantawy W.H., *"Characterization of antioxidant activity of extract from Artemisia vulgaris",* Pakistan Journal of Pharmaceutical Sciences, 21, 321-326, 2008.

[14] Ferreira J.F.S., Simon J.E., Janick J., *"Artemisia annua: botany, horticulture, pharmacology",* Horticulture Review, 19, 319–371, 1997.

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

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

[17] Duke M.V., Paul R.N., Elsohly H.N., Sturtz G., Duke S.O., *"Localization of Aremisinin and Artemisitene in Foliar tissue of glanded and glandless biotype of Artemisia annua L.",* International Journal of Plant Science, 155, 365–372, 1994.

[18] Ferreira J.F.S., Simon J.E., Janick J. *"Relationship of artemisinin content of tissue-cultured, greenhouse-grown, and field-grown plants of Artemisia annua."* Planta Medica 61, 351–355, 1995.

[19] Lommen W.J., Schenk E., Bouwmeester H.J., Verstappen F.W., *"Trichome dynamics and artemisinin accumulation during development and senescence of Artemisia annua leaves".* Planta Medica, 72, 336, 2006.

[20] Charles D.J., Simon J.E., Wood K.V., Heinstein P., *"Germplasm variation in artemisinin content of Artemisia annua using an alternative method of artemisinin analysis from crude plant extracts"* Journal of Natural Products. 53, 157-160, 1990.

[21] Torrell M., Cerbah M., Siljak-Yakovlev S. and Valle´s J., 2003 — *Molecular cytogenetics of the genus Artemisia (Asteraceae, Anthemideae): fluorochrome banding and fluorescence in situ hybridization. I. Subgenus Seriphidium and related taxa*. Plant Systematic Evolution, 239: 141-153.

[22] McArthur R.H., Horn H.S., *"Foliage profile by vertical measurements"* Ecology 50, 802-804, 1969.

[23] Kawatani T. and Ohno T.*,"Chromosome numbers in Artemisia*". Bull. Natl. Inst. Hyg. Sci. (Tokyo), 82: 183-193. 1964.

[24] Valles J., Siljak-Yakovlev S., **"***Cyiogenetic studies in the genus Artemisia L.: fluorochrome banded karyotypes of five taxa, including the Iberian endernic species Artemisia barrelieri Besser"* Canndian Journal of Botany, 75, 595-606, 1997.

[25] Bhandari A., Gupta P., Singh B., Kaul V.K., *"Simultaneous densitometric determination of artemisinin, artemisinic acid and arteannuin-B in Artemisia annua using reversed-phase thin layer chromatography".* Journal of Separation Science, 28, 2288-2292, 2005.

[26] Webster H.K., Indusnert E.K., *"chemistry of artemisinin: an overview",* Trans

Transaction of Royal Society of Tropical Medicine and Hygine 88, 27-29, 1994.

[27] Delabays N., Simmonet X., Gaudin M. *"The genetics of artemisinin content in Artemisia annua L. and the breeding of high yielding cultivars"* Current Medicinal Chemistry 8, 1795-1801, 2001.

[28] Haider F., Dwivedi P., Singh S., Naqvi A.A., Bagchi G., *"Influence of transplanting time on essential oil yield and composition in Artemisia annua plants grown under the climatic conditions of sub-tropical North India"* Flavour & Fragrance Journal, 19, 51-53, 2004.

[29] Van Agtmael M.A., Eggelte T.A., Van Boxtel C.J., *"Artemisinin drugs in the treatment of malaria: from medicinal herb to registered medication"*. Trends Pharmacological Science, 20, 199-205, 1999.

[30] Hien T.T., White N.J., *"Qinghaosu"*, Lancet, 341, 603-608, 1993.

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

[32] Wright C., *"Artemisia. Medicinal and Aromatic Plants"* - Industrial profiles. London, Taylor and Francis. Ed. 2002.

[33] Wilairatana P., Looaresuwan S., *"The clinical use of artemisinin and its derivatives in the treatment of malaria"*. In *Artemisia*, Wright, C.W., Ed. Taylor & Francis, London, 2002.

[34] Liu K.C.C., Lang S., Roberts M.F. *"Antimalarial activity of Artemisia annua flavonids from whole plant and cell cultures"*. Plant Cell Reports., 11, 637-640, 1992.

[35] Elford B.C., Roberts M.F., Phillipson J.D., Wilson R.J.M., *"Potentiation of the antimalarial activity of qinghaosu by methoxylated flavones"*, Transaction of Royal Society of Tropical Medicine and Hygien, 81, 434–436, 1987.

[36] Shuhua X., Chollet J., Weiss N.A., Bergquist R.N., Tanner M., *"Preventive effect of artemether in experimental animals infected with Schistosoma mansoni"* Parasitology International, 49, 19-24, 2000.

[37] Kumar T., Khanuja S., Jain D., Srivastava S., Bhattacharya A., Sharma R., Kumar S.A., *"Simple microbiological assay for the stereospecific differentiation of alpha and beta isomers of arteether"* Phytotherapy Research, 14, 644-646, 2000.

[38] Rowen RJ "Artemisinin: From Malaria to Cancer Treatment", Townsend Letter Press, USA. http://www.townsendletter.com/Dec2002/artemisinin1202.htm

[39] Croteau R., Kutchan T., Lewis N., *"Natural products (secondary metabolites). In Biochemistry and Molecular Biology of Plants",* American Society of Plant Biologists, Rockville, MD, 1250–1268, 2000.

[40] Mahmoud S., Croteau R., *"Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase*", Proceeding of Natl Academyof Science USA, 98, 8915–8920, 2001.

[41] Bouwmeester H.J., Wallaart T.E., Janssen M.H., van Loo B., Jansen B.J., Posthumus M. A., Schmidt C.O., De Kraker J.W., Knig W.A., Franssen M.C., "*Amorpha-4,11-diene synthase catalyzes the first probable step in artemisinin biosynthesis*". Phytochemistry 52: 843–854; 1999.

[42] Zhang Y., Teoh K.H., Reed D.W., Maes L., Goossens A., Olson D.J.H, Ross A.R.S., Covello P.S., *"The molecular cloning of artemisinic aldehyde Δ11 (13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in Artemisia annua.* Journal of Biological Chemistry, 283, 21501- 21508, 2008.

[43] Graham I.A., Besser K., Blumer S., Branigan C.A., Czechowski T., Elias L., Guterman I., Harvey D., Isaac P.G., Khan A.M., Larson T.R., Li Y., Pawson T., Penfield T., Rae A.M., Rathbone D.A., Reid S., Ross J., Smallwood M.F., Segura V., Townsend T., Vyas, D., Winzer T., Bowles D., *"The genetic map of Artemisia annua L. identifies loci affecting yield of the antimalarial drug artemisinin"*, Science, 327, 328-331, 2010.

[44] Dellicour S., Hall S., Chandramohan D., Greenwood B.D., *"The safety of artemisinins during pregnancy: a pressing question",* Malaria Journal, 6, 1-10, 2007.

[45] Jose Kudakasseril **"***Effect of Sterol Inhibitors on the Incorporation of 14C-Isopentenyl Pyrophosphate into Artemisinin by a Cell-Free System from Artemisia annua Tissue Cultures and Plants"* Planta medica. 53, 280-284, 1987.

[46] Herman J.W., Bos R., Salomoms M.C., Hendrik H., Pras N., Malingre T.M. *"volatile constituents of Artemisia annua L. (Asteraceae)"* Flavour and Fragrance. Journal, 8, 131-137, 1993.

[47] Jaziri M., Shimomura K., Yoshimatsu K., Faucconier M.L., Marlier M., Homes J., *"Establishment of normal and transformed root cultures of Artemisia annua L. for artemisinin production"*, Journal of Plant Physiology, 145,175-177, 1995.

[48] Fulzele D.P., Sipahimalani A.T., Heble M.R., *"Tissue cultures of Artemisia annua: organogenesis and artemisinin production",* Phytotherapy Research, 5,149-153, 1991.

[49] Mannan A., Ahmed I., Arshad W., Asim M.F., Qureshi R.A., Hussain I., et al. *"Survey of artemisinin production by diverse Artemisia species in northern Pakistan",* Malaria Journal, 9, 310-319, 2010.

[50] Singh A., Sarin R., *"Artemisia scoparia – A new source of artemisinin",* Bangladesh Journal of Pharmacology, 5, 17-20, 2010.

[51] Aryanti B.M., Ermayanti T.M., Mariska I., *"Production of antileukemic agent in untransformed and transformed root cultures of Artemisia cina",* Annales Bogorienses 8, 11-16, 2001.

[52] Covello P.S., Teoh K.H., Polichuk D.R., Reed D.W., Nowak G., *"Functional genomics and the biosynthesis of artemisinin",* Phytochemistry, 68, 1864–1871, 2007.

[53] Hsu E., *"The history of qing hao in the Chinese materia medica",* Trans R Soc Trop Med Hyg 100, 505-508, 2006.

[54] Arab H.A., Rahbari S., Rassouli A., Moslemi M.H., Khosravirad F.D.A., "Tropical Animal Health and Production", 38, 497, 2006.

[55] Zhou W.S., Xu X.X., *"In: Studies in natural product chemistry",* (Ed.): Atta-ur-Rahman. 3, 494. 1989.

[56] Laughlin J.C., *"Post-harvest drying treatment effects on antimalarial constituents of Artemisia annua L."* Acta Horticulturae, 576, 315–320, 2002.

[57] Chang H.M. and But P.P.H. (1986). *Pharmacology and Applications of Chinese Materia Media*, Vol. 1. World Scientific Publishing, Singapore.

[58] Kumar S., Gupta S.K., Singh P., Bajpai P., Gupta M.M., Singh D., et al., *"High yields of artemisinin by multi-harvest of Artemisia annua crops",* Industrial Crops and Products, 19, 77-90, 2004.

[59] Chen D., Ye H., Li G., *"Expression of a chimeric farnesyl diphosphate synthase gene in Artemisia annua. L. transgenic plants via Agrobacterium tumefaciens-mediated transformation"* Plant Science, 155, 179-185, 2000.

[60] Laughlin J.C., *"Agricultural production of artemisinin—a review".* Transactions of the Royal Society of Tropical Medicine and Hygiene", 88, 21–22, 1994.

[61] Wallaart T.E., Bouwmeester H.J., Hille J., Poppinga L., Maijers N.C., *"Amorpha-4,11-diene synthase: cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel anti-malarial drug artemisinin".* Planta 212,460-465, 2001.

[62] Sangwan R.S., Sangwan N.S., Jain D.C., et al., *"RAPD profile based genetic characterization of chemotypic variants of Artemisia annua L."* Biochemical and Molecular Biololy, 47, 935- 944, 1999.

[63] Lu J.J., Knox M.R., Ambrose M.J., Brown J.K.M., Ellis T.H.N., *"Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR based methods"*. Theortical and Applied Genetic, 93, 1103-1111, 1996.

[64] Williams K., et. al., *"DNA polymorphisms amplified by arbitrary primers are useful as genetic markers"*, Nucleis Acid Research, 18: 6531-6535, 1990.

[65] Karp A., et al., *"Newer molecular technologies for biodiversity evaluation: Opportunities and challenges"*, Nature Biotechnology, 15, 625–628, 1997.

[66] Vos P.R., Hogers M., Bleeker M., Van D.L., Reijans T., Hornes M., Fritjers A., Pot J., Peleman J., Kuiper M., Zabeau M., *"AFLP: a new concept for DNA fingerprintin,* Nucleic Acids Research*"* 23, 4407–4414, 1995.

[67] Welsh J., McClelland M., *"Fingerprinting genomes using PCR with arbitrary primers",* Nucleic Acids Research, 18, 17213-17218, 1990.

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

[69] Bornet B., Goraguer F., Joly G., Branchard M., *"Genetic diversity in European and Arentinian cultivated potatoes (Solanum tuberosum) subsp. tuberosum detected by inter-simple sequence repeats (ISSRs)"*, Genome, 45, 48-484, 2002.

[70] Tautz D., Schlotterer C., *"Simple sequences",* Current Opinion of Genetics Development, 4, 832-837, 1986.

[71] Bornet B., Branchard M., *"Non-anchored inter-simple sequence repeat (ISSR) markers: reproducible and specific tools for genome fingerprinting"*, Plant Molecular Biology Reporter, 22, 427–432, 2001.

[72] Chowdhury M.A., Vandenberg B., Warkentin T., *"Cultivar identification and genetic relationship among selected breeding lines and cultivars in chickpea (Cicer arietinum L.)"*, Euphytica, 127, 317–325, 2002.

[73] Wolfe A.D., Xiang Q.Y. Kephart S.R. *"Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands."* Molecular Ecology, 7, 1107-1126, 1998.

[74] Yang W., de Oliveria A.C., Godwin I., Schertz K.F. Bennetzen J.L., *"Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums",* Crop Science, 36, 1669-1676, 1996.

[75] Ge X.J., Sun M., *"Reproductive biology and genetic diversity of a cryptoviviparous mangrove* Aegiceras corniculatum *(Myrsinaceae) using allozyme and intersimple sequence repeat (ISSR) analysis"*, Molecular Ecology, 8, 2061-2069, 1999.

[76] Culley T.M., Wolfe A.D., *"Population genetic structure of the cleistogamous plant species Viola pubescens Aiton (Violaceae), as indicated by allozyme and ISSR molecular markers"*, Heredity 86, 545-556 , 2001.

[77] Blair M.W., Panaud O., Mccouch S.R., *"Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (Oryza sativa L.)",* Theoretical and Applied Genetics, 98, 780–792, 1999.

[78] Caetano A.G., Bassam B.J., Bresshoff P.M., *"DNA amplification fingerprinting: A strategy for genome analysis",* Plant Molecular Biology Reporter, 9, 294-307, 1991.

[79] Welsh J., Honeycutt R.J., McClelland M., Sobral B.W.S., *"Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR)"*, Theoretical and. Applies Genetic, 82, 473-476, 1991.

[80] Kresovich S., Williams J.G.K., McFerson J.R., Routman E.J., Schaal B.A., *"Characterisation of genetic identities and relationship of Brassica oleracea L. via a random amplified polymorphic DNA assay",* Theoretical and. Applies Genetics, 85, 190-196, 1992.

[81] Welsh J., Honeycutt R.J., McClelland M., Sobral B.W.S., *"Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR)"*, Theoretical and. Applies Genetics, 82, 473-476, 1991.

[82] Dahlberg J.A., Zhang X., Hart G.E., Mullet J.E., *"Comparative Assessment of Variation among Sorghum Germplasm Accessions Using Seed Morphology and RAPD Measurements"*, Crop Science, 42, 291-296, 2002.

[83] Forster J., Knaak C., *"Estimation of the genetic distance of 21 winter rapeseed varities by RAPD analysis in comparison to RFLP results",* Proc. 9th Int. rape seed Congr. Cambridge, UK., 1184-1186, 1995.

[84] Ratnaparkhe M.B., Gupta V.S., Murthy M.R.V., Ranjekar P.K. *"Genetic fingerprinting of pigeon pea (Cajanus cajan (L.) Millsp.) and its wild relatives using RAPD markers",* Theoretical and Applied Genetics, 91, 893-898, 1995.

[85] Kazan K., Manners J.M., Cameron D.F., *"Genetic variation in agronomically important species of Stylosanthes determine using random amplified polymorphic DNA markers",* Theoretical and Applied Genetics, 85, 882-288, 1993.

[86] Koller B., Indusmann A., McDermott J.M., Gessler C., *"Identification of apple cultivars using RAPD markers",* Theoretical and Applied Genetics, 85, 901-904, 1993.

[87] Crochemore M.L., Huyghe C., Kerlan M.C., Durand F., Julier B., *"Partitioning and distribution of RAPD variation in a set of population of the Medicago sativa complex",* Agronomie, 16, 421-432, 1996.

[88] Stewart C.N., Porter D.M., *"RAPD profiling in biological conservation: an application to estimating clonal variation in rare and endangered Iliamna in Virginia"*, Biological Conservation, 74, 135-142, 1995.

[89] Sale M.M., Potts B.M., West A.K. Reid J.B., *"Molecular differentiation within and between Eucalyptus risdonii, E. amygdalina and their hybrids using RAPD markers",* Australian Journal of Botany, 44: 559-569, 1996.

[90] Comes H.P., Abbott R.J., *"Reticulate evolution in the Mediterranean species complex of Senecio sect. Senecio: uniting phylogenetic and population-level approaches",* Molecular systematics and plant evolution, 12, 171-198, 1999.

[91] Zamora R., Jamilena M., Ruiz R.M., Blanca G. *"Two new species of the carnivorous genus Pinguicula, (Lentibulariaceae) from Mediterranean habitats"* Plant Systmatic Evolution, 200, 41-60, 1996.

[92] Vasconcelos M.J.V., De Barros E.G., Moreira M.A., Vieira C., *"Genetic diversity of the common bean Phaseolus vulgaris L. determined by DNA-based molecular markers",* Brazelian Journal of Genetics, 19: 447-451, 1996.

[93] Rieseberg L.H., *"Homology among RAPD fragments in interspecific comparisons",* Molecular Ecology, 5, 99-105, 1996.

[94] Gimenes M.A., Lopes C.R., Galgaro M.L., Valls J.F.M., Kochert G., *"Genetic variation and phylogenetic relationships based on RAPD analysis in section Caulorrhizae, genus Arachis (Leguminosae)"*, Euphytica, 116, 187-195, 2000.

[95] Kafkas S., Perl-Treves R., *"Inter-specific relationships in the genus Pistacia L. (Anacardiaceae) based on RAPD fingerprints",* Horticulture Science, 37, 168-171, 2002.

[96] Mulcahy D.L., Cresti M., Linskens H.F., Intrieri C., Silvestroni O., Vignani R., Pancaldi, M., *"DNA fingerprinting of Italian grape varieties: a test of reliability in RAPDs"*, Advance Horticulture Science, 9: 185–187, 1995.

[97] Parentoni S.N., Magalhaes J.V., Pacheco C.A.P., Santos M.X., Abadie T., Gama E.E.G., Guimaraes P.E.O., Meirelles W.F., Lopes M.A., Vasconcelos M.J.V. Paiva E., *"Heterotic groups based on yield-specific combining ability data and phylogenetic relationship determined by RAPD markers for 28 tropical maize open pollinated varieties",* Euphytica, 121: 197-208, 2001.

[98] Hearne C.M., Ghosh S., Todd J.A. *"Microsatellites for linkage analysis of genetic traits",* Trends in Genetics, 8, 288-94, 1992.

[99] Kashi Y. and King D.G., "*Simple sequence repeats as advantageous mutators in evolution",* Trends in Genetics, 22, 254-257, 2006.

[100] Karlin S., Campbell A.M., Mrazek J., *"Comparative DNA analysis across diverse genomes"*, Annual Review of Genetics, 32, 185-225, 1998.

[101] King D.G., *"Triplet repeats DNA as highly mutable regulatory mechanism"*, Science 263, 595-596, 1994.

[102] Katti M.V., Ranjeker P.K., Gupta V.S., *"Differential distribution of simple sequence repeats in eukaryotic genome sequences",* Molecular Biology and Evolution, 18, 1161–1167, 2001.

[103] Berget S.M., Moore C., Sharp P.A., *"Spliced segments at the 50 terminus of adenovirus 2 late mRNA"*, Proceeding of Natural Academy of Science. USA, 74, 3171–3175. 1977.

[104] Chow L., Gilinas R., Broker T., Roberts R., "*An amazing sequence arrangement at the 50 ends of adenovirus massenger RNA"*, Cell, 12, 1–8, 1977.

[105] Wang D.G., Fan J.B., Siao C.J., et al., *"Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome",* Science, 280, 1077–1082, 1998.

[106] Lessa E.P., *"Rapid survey of DNA sequence variation in natural populations",* Molecular Biology and Evolution, 9, 323-330, 1992.

[107] Daguin C., Borsa P., *"Genetic characterisation of Mytilus galloprovincialis Lmk. in North West Africa using nuclear DNA markers",* Journal of Experimental Marine Biology and Ecology, 235, 55-65, 1999.

[108] Wydner K.S., Sechler J.L., Boyd C.D., et al., "*Use of an intron length polymorphism to localize the tropoelastin gene to chromosome 5 in a region of linkage conservation with human chromosome 7",* Genomics, 23, 125–131, 1994.

[109] 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.

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

[111] Sudduth K., Fraisse C., Drummond S., Kitchen N., *"Analysis of spatial factors influencing crop yield",* Proceedings of 3rd International Conference on Precision Agriculture. Madison, WI. 129-140. 1996.

[112] Atkinson P., Tatnall A., *"Neural networks in remote sensing",* International Jourournal of Remote Sensing, 18 (4), 699-709, 1997.

[113] Sudduth K., Fraisse C., Drummond S., Kitchen N., *"Integrating spatial data collection, modeling and analysis for precision agriculture",* First Int. Conf. on Geospatial Information in Agriculture and Forstery, Lake Buena Vista, Florida. 1-3, Jun. 2, 66-173, 1998.

[114] Phillipson D.J., Wright, C.W., *"Antiprotozoal agents from plant sources",* Plant Medica, 57, 553–59, 1991.

[115] Klayman D.L., "*Artemisia annua I: From weed to respectable antimalarial plant",* In: Human medicinal agents from plants (eds.) A D Kinghorn, M F Balandrin, (Am. Chem. Soc. Symp. Ser. Washington D.C.) 242–255, 1993.

[116] Sangwan R.S., Agarwal K., Luthra R., Thakur R.S., Sangwan S.M., *"Biotransformation of arteannuic acid into artennuin-b and artemisinin in Artemisia annua"*, Phytochemistry, 34, 1301-1302, 1993.

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

[118] Xiao J., Li J., Yuan L., McCouch S.R., Tanksley S.D., *"Genetic diversity and its relationships to hybrid performance and heterosis in rice as revealed by PCR-based markers",* Theoretical and Applied Genetics, 92, 637-643, 1996.

[119] Ovesna J., Polakova K., Lisova L., *"DNA analysis and their applications in plant breeding",* Czech. Journal Plant Breeding, 38, 29-40, 2002.

[120] Bostein D., White R., Skolnick M., Davis R., *"Construction of a genetic linkage map in man using restriction fragment length polymorphism",* American Journal of Genetics, 32, 314-331, 1980.

[121] Gupta S., et al. *"Analogy of ISSR and RAPD markers for comparative analysis of genetic diversity among different Jatropha curcas genotypes",* African Journal of

Biotechnology, 23, 4230–4243, 2008.

[122] Alam A., Gulati P., Gulati A., Mishra G.P., Naik P.K., *“Assessment of genetic diversity among Podophyllum hexandrum genotypes of the North-western Himalayan region for podophyllotoxin production”,* Indian Journal Biotechnology, 8, 391-399, 2009.

[123] Dellaporta S.L., Wood J., Hicks J.B., *“A plant DNA manipreparation: version II”*, Plant Molecular Biology Reporter, 1, 19-21, 1983.

[124] Doyle J.J., Doyle J.L., *“A rapid DNA isolation procedure for small quantities of fresh leaf tissue”*, Phytochemical Bulletin, 19, 11–15, 1987.

[125] 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.

[126] Webb D.M., Knapp S.J., *“DNA extraction from previously recalcitrant plant genes”*, Plant Molecular Biology Reporter, 8, 180–185, 1990.

[127] Evann G., Regnaut S., Goudet J. *“Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study”* Molecular Ecology, 14, 2611–2620, 2005.

[128] Zhao W.G., Zhang J.Q., Wangi Y.H., Chen T.T., Yin Y., *et al.*, *“Analysis of genetic diversity in wild populations of mulberry from western part of Northeast China determined by ISSR markers”,* Journal of Genetics and Molecular Biology, **7,** 196-203, 2006.

[129] Nei M., *“Estimation of average heterozygosity and genetic distance from a small number of individual”,* Genetics, 89, 583-590, 1978.

[130] 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.

[131] Prevost A., Wilkinson M.J., *“A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars”,* Theoetical and. Applied Genetics, 98: 107-112, 1999.

[132] Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S. and Rafalski A., *“The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis”,* Molecular Breeding. 2: 225-238, 1996.

[133] Danshwar P., Sher-ullah S.S.K., "*Genomic DNA extraction from Victoria amazonica",* Plant Molecular Biology Reporter, 22, 195-205, 2004.

[134] Doyle J.J., Doyle L.J., *"Isolation of fresh DNA from fresh tissue",* Focus, 12, 13-15, 1990.

[135] Nayak S., et al. *"Assessment of genetic diversity among 16 promising cultivars of ginger using cytological and molecular markers"*, *Science Asia* 60: 485-492, 2005.

[136] Padmesh P., Sabu K.K., Seeni S., Pushpaangadan S., *"The use of RAPD in assessing genetic variability in Andrographis paniculata Nees, a hepatoprotective drug"*. Current Science, 76, 833-835, 1999.

[137] Martín J.P., Sánchez-yélamo M.D., *"Genetic relationships among species of the genus Diplotaxis (Brassicaceae) using inter-simple sequence repeat markers"*. Theoretical and Applied Genetics, 101, 1234–1241, 2000.

[138] Moreno S., Martin J.P., Ortiz J.M., *"Inter-simple sequence repeat PCR for characterization of closely related grapevine germplasm",* Euphytica, 101, 117–125, 1998.

[139] Blair M.W., Panaud O., Mccouch S.R., *"Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (Oryza sativa L.)",* Theoretical and Applied Genetics, 98, 780–792, 1999.

[140] Ajibade S.R., Weeden N.F., Chite S.M., **"***Inter simple sequence repeat analysis of genetic relationships in the genus Vigna",* Euphytica, 111, 47–55, 2000.

[141] Nagaoka T., Ogihara Y., *"Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers",* Theoretical and Applied Genetics, 94, 597–602, 1997.

[142] Loarce Y., Gallego R., and Ferrer E.A., *"Comparative analysis of genetic relationships between rye cultivars using RFLP and RAPD markers",* Euphytica, 88, 107-115, 1996.

[143] Karp A., et al. *"Newer molecular technologies for biodiversity evaluation: Opportunities and challenges",* Nature Biotechnology, 15: 625–628, 1997.

[144] Gaffor A., et al. *"Genetic diversity in blackgram (Vigna mungo* L. *Hepper)",* Field Crop Research, 69: 183–190, 2001.

[145] Botstein D., White, K.L. Skolnick M. and Davis R.W. *"Construction of a genetic linkage map in man using restriction fragment length polymorphisms",* 1980.

[146] Becker J., Heun M., *“Barley microsatellites: allele variation and mapping”*, Plant Molecular Biology, 27, 835–845, 1995.

[147] Kruglyak L., *“The use of a genetic map of biallelic markers in linkage studies”*. Nature Genetics, 17, 21–24, 1997.

[148] Choi H.K. et al., *“A sequence-based genetic map of Medicago truncatula and comparison of marker colinearity with M. sativa”*. Genetics 166:1463–1502, 2004.

[149] Koebner R.M.D., Summers R.W., *“21st century wheat breeding: Plot selection or plate detection?”,* Trends Biotechnology, 21, 59–63, 2003.

[150] Martins-Lopes P., Zhang H., Koebner R., *“Detection of single nucleotide mutations in wheat using single strand conformation polymorphism gels”*, Plant Molecular Biology Reporter, 19, 159–16, 2001.

[151] Kimura M., *“The Neutral Theory of Molecular Evolution”,* Cambridge University Press, Cambridge, 1983.

[152] Feltus F.A., Wan J., Schulze S.R., et al. *“An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments”* Genome Research, 14, 1812–1819, 2004.

[153] Wei H., *“Intron-flanking EST-PCR markers from genetic marker development to gene structure analysis in Rhododendron”,* Theoretical and Applied Genetics, 111, 1347–1356, 2005.

[154] Bierne N., Indusnert S.A., Bedier E., et al., *“Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR”*, Molecular Ecology, 9, 233–235, 2000.

[155] Batzoglou S., et al., *“Human and mouse gene structure: comparative analysis and application to exon prediction”,* Genome Research, 10, 950–958, 2000.

[156] Bown D., *“Encyclopedia of Herbs and Their Uses (RHS)”,* DK Publishing, Dorling Kindersley, London. 1995.

[157] Chevallier A., *“The Encyclopedia of Medicinal Plants”* Dorling Kindersley. London, 9-780751-303148, 1996.

[158] Yang L., Jin G., Zhao X., Zheng Y., Xu Z., Wu W., *“PIP: a database of potential intron polymorphism markers”* Bioinformatics, 23 , 2174–2177, 2007.

[159] Rozen S., Skaletsky H., *“Primer3 on theWWWfor general users and for biologist programmers”,* Methods Mol. Biol 132, 365–386, 2000.

[160] Cotrim H.C., Chase M.W., Pais M.S., *"Silene rothmaleri (Caryophyllaceae), a rare, fragmented but genetically diverse species",* Biodiversity Conservation., 12,1083-1098, 2003.

[161] Chen S., Xia T., Chen S., Zhou Y., *"RAPD profiling in detecting genetic variation in endemic Coelonema (Brassicaceae) of Qinghai-Tibet Plateau of China"*, Biochemcal Genetics, 43, 189-201, 2005.

[162] GE X.J., Sun M., *"Population genetic structure of Ceriops tagal (Rhizophoraceae) in Thailand and China",* Wetlands Ecololgy Management, 9, 203-209, 2001.

[163] Juan A., Crespo M.B., Cowan R.S., *"Patterns of variability and gene flow in Medicago citrine, an endangered endemic of islands in the western Mediterranean, as revealed by amplified fragment length polymorphism (AFLP)",* Molecular Ecology, 13, 2679-2690, 2004.

[164] Lei Y., Gao H., Tsering T., Shi S., Zhong, Y., *"Determination of genetic variation in Rhodiola crenulata from the Hengduan Mountains Region, China using inter-simple sequence repeats"*, Genetics and Molecular Biology, 29, 339-344, 2006.

[165] Hamrick J.L., Godt M.J.W., *"Conservation genetics of endemic plant species",* Conservation genetics-case histories from nature. Chapman & Hall, London, 281–304 (1996).

[166] Slatin M., *"Gene flow and geographic structure of natural populations"*, Science, 236, 787-792, 1987.

[167] Liu Z.J., Tan G., Li P., Dunham R.A., *"Transcribed dinucleotide microsatellites and their associated genes from channel catfish, Ictalurus punctatus",* Biochemcal and Biophysics Research Communication, 259,190-194, 1999.

[168] Scott K.D., Eggler P., Seaton G., Rossetto M., Ablett E.M., Lee L.S., Henry R.J., *"Analysis of SSRs derived from grape ESTs"*, Theoretical and Applied Genetics 100,723-726, 2000.

[169] Varshney R.K., Thiel T., Stein N., Langridge P. Graner A., *"In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species",* Cell and Molecular Biology Letters, **7,** 537-546, 2002.

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

[171] Eujayl I., Sledge M.K., Wang L., May G.D., Chekhovskiy, K., *"Medicago truncatula EST-SSRs reveal cross-species genetic markers for Medicago spp.",* Theoretical and Applied Genetics, 108, 414-422, 2004.

[172] Gao L., Tang J., Li H., Jia J., *"Analysis of microsatellites in major crops assessed by computational and experimental approaches",* Molecular Breeding, 12, 245-261, 2003.

[173] Cordeiro G.M., Casu R.E., McIntyre C.L., Manners J.M., Henry R.J., *"Microsatellite markers from sugarcane (Saccharum spp.) ESTs cross transferable to erianthus and sorghum"*, Plant Science, 160, 1115-1123, 2001.

[174] Varshney R.K., Graner A., Sorrells M.E., *"Genic microsatellite markers in plants: features and applications"*, Trends Biotechnology, 23, 48-55, 2005.

[175] Kantety R.V., Rotal M.L., Matthews D.E., Sorrells M.E., *"Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat",* Plant Molecular Biology, 48, 501-510, 2002.

[176] Kuleung C., Baenziger P.S., Dweikat I., *"Transferability of SSR markers among wheat, rye, and triticale",* Theoritical and Applied Genetics, 108, 1147-1150, 2004.

[177] Gadaleta A., Mangini G., Mulè G., Blanco A., *"Characterization of dinucleotide and trinucleotide EST-derived microsatellites in the wheat genome",* Euphytica, 153, 73-85, 2007.

[178] Newcomb R.D., Crowhurst R.N., Gleave A.P., Rikkerink E.H.A., Allan A.C. et al**, "***Analyses of expressed sequence tags from apple",* Plant Physiology, 141, 147-166, 2006.

[179] Kumpatla S.P., Mukhopadhyay S., *"Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species",* Genome, 48, 985-998, 2005.

[180] Cardle L., Ramsay L., Milbourne D., Macaulay M., Marshall D., *"Computational and experimental characterization of physically clustered simple sequence repeats in plants"*, Genetics, 156, 847-854, 2000.

[181] Zeng S., Xiao G., Guo J., Fei Z., Xu Y., *et al*, *"Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, Epimedium sagittatum (Sieb. Et Zucc.) Maxim"*, BMC Genomics, 11, 1-11, 2010.

[182] Luo M., Dang P., Guo B., He G., Holbrook C.C., *et al.*, *"Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut",* Crop Science, 45, 346-353, 2005.

[183] Proite K., Leal-Bertioli S.C.M., Bertioli D,J., Moretzsohn M.C., Silva F.R. *et al*, *"ESTs from a wild Arachis species for gene discovery and marker development",* BMC Plant Biology, **7,** 1-10, 2007.

[184] Peakall R., Gilmore S., Keys W., Morgante M., Rafalski A., *"Cross-species amplification of soybean (Glycine max) simple sequence repeats (SSRs) within the genus and other legume genera: implications for the transferability of SSRs in plants"*, Molecular Biology and Evolution, 15, 1275-1287, 1998.

[185] Thiel T., Michalek W., Varshney R., Graner A., *"Exploiting EST databases for the development and characterization of genederived SSR-markers in barley (Hordeum vulgare L.)"*, Theoritical and Applied Genetics, 106, 411-422, 2003.

[186] Woerdenbag HJ, Lugt CB, Pras N. *"Artemisia annua L.: a source of novel antimalarial drugs"*. Pharmacy World Science, 19, 169-81, 1990.

[187] Van Nieuwerburgh F.C., Vande casteele S.R., Maes L., Goossens A., Inze D., Van Bocxlaer J., Van Bocxlaer J., Deforce D.L., *"Quantification of artemisinin and its biosynthetic precursors in Artemisia annua L. by high performance Liquid Chromatography- electrospray quadrupole time-offlight tandem mass spectrometry"*. Journal of Chromatography A, 1118, 180-187, 2006.

[188] Woerdenbag H.J., Pras N., Bos R., Visser J.F., Hendricks H., Malinger T.M.*" Analysis of artemisinin and related sesquiterpenoids from Artemisia annua L. by combined Gas Chromatography/ mass spectrometry"*. Phytochemical Analysis, 2, 215, 1991.

[189] Liu S., Tian N., Li J., Huang J., Liu Z., *"Simple and rapid micro-scale quantification of artemisinin in living Artemisia annua L. by improved gas chromatography with electron-capture detection",* Biomedical chromatography, 23, 1101-1107, 2009.

[190] Liersch R., Soicke, H., Stehr, C., Tullner, H.V., *"Formation of artemisinin in Artemisia annua during one vegetation period",* Planta Medica, 52, 387–390, 1986.

[191] Chan K.L., Chris K.H., Jinadasa S., Yuen, J.H., *"Selection of high artemisinin yielding Artemisia annua"*. Planta Medica. 61, 285–287, 1995.

[192] Woerdenbag H.J., Pras N., Chan N.G., Bang B.T., Bos R., van Uden W. et al. *"Artemisinin, related sesquiterpenes, and essential oil in Artemisia annua during a vegetation period in Vietnam,"* Planta Medica, 60,272-275 ,1994.

[193] Krishna S., Woodrow C.J., Staines H.M., Haynes R.K., Mercereau-Puijalon O., *"Re-evaluation of how artemisinins work in light of emerging evidence of in vitro resistance",* Trends Mol. Med., 12, 200–205, 2006.

[194] Chen D., Ye H., Li G., *"Expression of a chimeric farnesyl diphosphate synthase gene in Artemisia annua. L. transgenic plants via Agrobacterium tumefaciens-mediated transformation"* Plant Science, 155, 179-185, 2000.

[195] Jung M., ElSohly H.N., McChesney J.D., *"Artemisinic acid: a versatile chiral*

synthon and bioprecursor to natural products", Planta Medica, 56, 624, 1990.

[196] Sen N., Banerjee B., Das B. B., Ganguly A., Sen T., Pramanik S., Mukhopadhyay S., Majumder H. K., *"Apoptosis is induced in leishmanial cells by a novel protein kinase inhibitor withaferin A and is facilitated by apoptotic topoisomerase IDNA complex",* Cell Death and Differentiation, 14, 358-367, 2007.

[197] Borrmann S., Szlezak N., Faucher J.F., Matsiegui P.B., Neubauer R., Biner RK., Lell B., Kremsner P.G., *"Artesunate and praziquantel for the treatment of Schistosoma haematobium infections: a doubleblind, randomized, placebo-controlled study*" Journal of Infectious Disease, 184, 1363–1366, 2001.

[198] 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.

[199] Dong N.H., Thuang N.V., *"Breeding of high leaf and artemisinin yielding Artemisia annua variety"*, In: Paper Presented in National conference on malaria: current Status and Future Trends, Chulabhorn Research Institute. Bangkok, Thailand, Feb., 16-19, 2003.

[200] Ro D.K., Paradise E.M., Ouellet M., Fisher K.J., Newman K.L., Ndungu J.M., Ho K.A., Eachus R.A., Ham T.S., Kirby J., Chang M.C., Withers S.T., Shiba Y., Sarpong R., Keasling J.D., *"Production of the antimaralial drug precursor artemisinic acid in engineered yeast"*. Nature 440, 940-943, 2006.

[201] Zeng Q.P., Qiu F., Yuan L., *"Production of artemisinin by genetically modified microbes",* Biotechnology Letter, doi:10.1007/s10529-007-9596-y, 2007.

[202] Newman J.D., Marshall J., chang M.C.Y., NowrooZi F., Paradise E., Pitera D., Newman K.L., Keasling J.D., *"High-level production of amorpha-4, 11-diene in a two phase partionong bioreactor of metabolically engineered Escherichia coli",* Biotechology and Bioengineering, 5, 684-691, 2006.

[203] Zhang L., Jing F., Li M., Wang Y., Wang G., Sun Tang K., *"Development of transgenic Artemisia annua plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA mediated gene silencing",* Biotechnology and Applied Biochemistry, 52, 109-207, 2009.

[204] Aquil S., Husaini A.M., Abdin M.Z., Rather G.N., *"Overexpression of HMG-CoA reductase gene leads to enhanced artemisinin biosynthesis in transgenic A. annua L. plants"*, Planta Medica, 75, 1-6, 2009.

[205] Weathers P.J., Bunk G., McCoy M.C., *"The effect of phytohormones on growth and artemisinin production in Artemisia annua hairy roots"*, In Vitro Cellular and Devlopmental Biology Plant, 41:47–53; 2005.

[206] Pachepsky Y.A., Timlin D., Varallyay G., *“Artificial neural networks to estimate soil water retention from easily measurable data”,* Soil Science Society of American Journal, 60, 727-733, 1996.

[207] Elizondo D.A., McClendon R.W., Hoogenboom G.,*“Neural network models for predicting flowering and physiological maturity of soybean“*, Transactions of the ASAE, 37, 981-988, 1994.

[208] Mehlich A., *“New buffer pH method for rapid estimation of exchangeable acidity and lime requirement of soils”,* Communication in Soil Science and Plant Analysis, 7, 637-652, 1976.

[209] Mebius L.J., *“A rapid method for the determination of organic carbon in soil”,* Analytica Chemica Acta, 22, 120-124, 1960.

[210] Rump H.H., Krist H., “*Laboratory manual for the examination of water”,* Waste Water and Soil. VCH, Weinheim. second ed. pp. 11-113, 1992.

[211] Kimble J.M., Knox E.G., Holzhey C.S., *“Soil survey laboratory methods for characterizing physical and chemical properties and mineralogy of soils”, In Application of Agriculture Analysis in Environmental Studies”,* American Society for Testing and Materials, Philadelphia.23-31, 2005.

[212] Olsen S.R., Sommers L.E., “Phosphorus. In *Methods of soil analysis,”* eds. A. L. Page, R. H. Millerand and D. R. Keeney, 9, 403-429, 1982.

[213] Howard D., Mark B., “*Neural Network Toolbox. Version 4*”. The MathWorks, Inc. Natick, MA, USA. 2000.

[214] Clifford G., Lau Y., “*Neural Networks”,* Theoretical Foundations and Analysis. IEEE, New York, NY. 1992.

[215] Marchese J.A., Casiraghi V., Lira R., Tedesco A.C., Rehder V.L.G., *“Flowering of Artemisia annua L. plants submitted to different photoperiod and temperature conditions”*, Proceedings of the 1st Latin American Symposium on MAP, Acta Horticulturae, 569, 275–280, 2002.

[216] Singh A., Vishwakarma R.A., Husain, A., “*Evaluation of Artemisia annua strains for higher artemisinin production”,* Planta Medica, 64, 475–476, 1988.

[217] Knobloch K.H., Berlin J., *“Influence of phosphate on the formation of Indole alkloids and phenolic compounds in cell suspension cultures of Catharanthus roseus. I. Comparision of enzyme activities and product accumulation”,* Plant Cell Tissue Culture, 2, 333-341, 1983.

[218] Sachs J., Malaney P., *“The economic and social burden of malaria”*, Nature,

415, 680-685, 2002.

[219] Surolia N and Surolia A *"Triclosan offers protection against blood stages of malaria by inhibiting enoylACP reductase of Plasmodium falciparum"* Nature Medicine, 7, 167–17, 2001.

[220] Padmanaban, G., Rangarajan, P.N., *"Emerging targets for antimalarial drugs"*, Expert Opinion in Therapeutic Targets 5, 423-441, 2001.

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

[222] Zia M., Abdul M., Chaudhary M.F., *"Effect of growth regulators and amino acids on artemisinin production in the callus of Artemisia absinthium",* Pakistan Journal of Botany, 39,799-805, 2007.

[223] Chaurasia O.P., Ahmed Z., Ballabh B., (2007). Satish Serial Publishing House, Delhi. India.

[224] Ewing B., Green P., *"Base-calling of automated sequencer traces using phred. II. Error probabilities",* Genome Research. 8, 186–194, 1998.

[225] Delabays N., Collet G., Benakis A., *"Selection and breeding for high artemisinin (Qinghaosu) yielding strains of Artemisia annua",* Acta Horticultute, 330, 203-207, 1993.

[226] Heywood V.H., Humphries C.J., *"Anthemideae-systematic review"*. In V.H. Heywood, J.B. Harborne and B.L. Turner, (eds.), The Biology and Chemistry of the Compositae, Academic Press, London, 11, 852-888, 1977.

LIST OF 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).