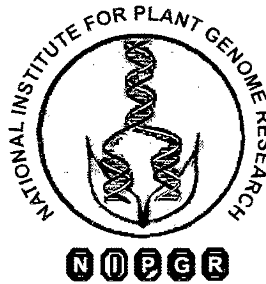
Marker development and molecular mapping in chickpea (*Cicer arietinum* L.)

Thesis submitted to
Jawaharlal Nehru University
for
The award of the degree
of

Doctor of Philosophy

by

Shalu Choudhary



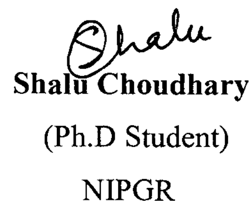
NIPGR
New Delhi-110067, INDIA

2009

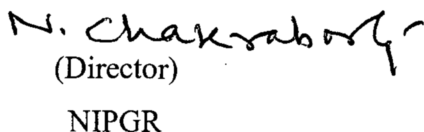
NATIONAL INSTITUTE FOR PLANT GENOME RESEARCH
(Formerly known as National Centre for Plant Genome Research)
(An Autonomous Institution of the Department of Biotechnology, Ministry of Science and Technology, Government of India)
Aruna Asaf Ali Marg, Post Box Number 10531, New Delhi-110067

CERTIFICATE

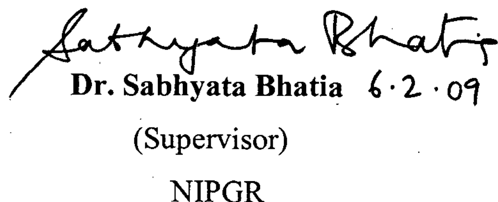
The research work embodied in this thesis entitled **'Marker development and molecular mapping in chickpea (*Cicer arietinum* L.)'** has been carried out at the National Institute for Plant Genome Research, New Delhi. This work is original and has not been submitted so far in part or in full, for the award by any degree or diploma by any university.



Shalu Choudhary
(Ph.D Student)
NIPGR

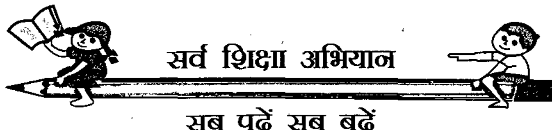


(Director)
NIPGR



Dr. Sabhyata Bhatia 6·2·09
(Supervisor)
NIPGR

Ph. : 26741612, 26741614, 26741617; Fax : 91-11-26741658
E-mail : nipgr@nipgr.res.in; Website : http://www.ncpgr.nic.in, http://www.nipgr.res.in



Dedicated to

'My Family'

Acknowledgements

Everything in this world is achieved with the blessings of the almighty and our faith in him and with his name I commenced my thesis.

It gives me immense pleasure to express my sound perception of gratitude and humble regards to my supervisor ***Dr. Sabhyata Bhatia*** *for his excellent guidance, unflinching support, suggestions and encouragements throughout the course of this study. I am grateful to her for inculcating and infusing in me the confidence to take up challenges, ability to think big and to be an independent researcher. Her friendliness nature, enthusiasm, always finding a better way of writing something and practical approach towards life has always motivated me during the course of my Ph.D.*

I render my sincere thanks to ***Prof. Asis Datta****, Ex- Director, National Institute for Plant Genome Reseach, New Delhi, for providing the research facilities to complete this research. I thank* ***Dr. Sudip Chattopadhyay, Prof. Sushil Kumar, Dr. Debasis Chattopadhyay, Dr. Niranjan Chakraborty, Dr. Subhra Chakraborty, Dr. Alok Krishna Sinha, Dr. Manoj Prasad, Dr. Praveen Verma*** *and Late* ***Dr. Jayanti Sen*** *for their timely advice and cooperation.*

I would be failing in my duty if I don't express my thanks to my teachers Prof. ***T.C. Bhalla*** *and* ***Dr. Arvind Bhatt*** *(H.P. University) who always encouraged me to pursue science.*

I am thankful to ***University Grants Commission*** *(UGC), India, for providing me the finical support for carrying out this investigation and NIPGR for in-kind support.*

During my years at NIPGR, I have met a lot of people that filled my achievements with loads of fun, joy, happiness and everlasting memories. From the first day I stepped in the lab, I got constant help and guidance from my seniors ***Dr. Bhumika Shokeen*** *and* ***Dr. Niroj Sethy*** *which was unforgettable. My achievements during my research work reflect the support and help provided by them. Bhumika Shokeen with an endless helping altitude solves my difficulties and confusion at every step and done extensive efforts in critically evaluating and giving valuable comments for my thesis. I would also like to give a big thanks to* ***Rashmi*** *for being helpful and trustful friend. I also acknowledge my labmates-* ***Santosh, Priyanka, Shefali, Sumit*** *and* ***Dr. Divya Srivastava*** *for their cooperation and support.*

I give my special thanks to ***Drs. Varsha di, Rakesh shukla, Jyoti di, Kundu da****, for their concern and valuable suggestions at times. I shall always cherish the love, encouragement and care endowed by them. I am highly thankful to all my juniors* ***Parul, Deepti, Monica, Prabhkar, Kundan, Sushil****, for their support and timely help.*

My note of acknowledgements would be incomplete if I forget the most hard working and happy people who all ease my burden- the trainees associated with me- ***Rohit, Saurabh, Gaurav, Ragvendra, Jayasmita*** *and* ***Chetan****.*

I am equally thankful to my batchmates ***Vinay, Kamal, Ramayya, Manikant, Santosh, Aarti****, and* ***Nasheeman*** *all of of whom have greatly assisted me during this study. A special thanks goes to* ***Vineeta*** *and* ***Arachna*** *who is always with me at the time of need.*

The technical support and assistance of ***Mr. Rajendra kumar, Arun, Ashok ji, Shobharam ji, Negi ji****, and* ***Surendra*** *was important and I acknowledge all of them with sincere thanks. I am deeply indebted to* ***Bahadhur ji, Manmohan ji*** *and* ***Mishra ji*** *for generous help at various times in the present study.*

Life is just not professional; it has its root under the roof of loved ones. I greatly feel that the patience and faith by ***My Family*** *in me has made me strong enough to face any situation.*

The love of ***Bhabhi-Bhaiya*** *and cheerful* ***Manas*** *has always smoothened my path and has helped me in achieving my goals. My younger brother* ***Deepanker*** *deserves a very special thanks as he has always boosted my morale by having faith in me and by just loving me. But still it is* ***My Mummy Papa,*** *who is always present in person to hold me whenever I am in need. I know that words can never express the gratitude I have for them but still "****My Sincere Thanks****' to them for believing in me, and I will always be in debt for what they all have done for me.*

Finally, but not least, I wish to thank ***Pritesh Gupta*** *for his love and always been source of positive thoughts and strong determination that has given me the inspiration and impetus to finally complete this thesis. I look forward to now relaxing and enjoying a happy future together!*

Shalu Choudhary

Abbreviations

accn	Accession
AFLP	Amplified Fragment Length Polymorphism
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	base pair(s)
cDNA	Complementary DNA
cM	centiMorgan
cv	cultivar
CTAB	Cetyltrimethylammonium bromide
DNA	deoxyribonucliec acid
dNTPs	Deoxy nucleotide triphosphates
EDTA	Ethylene Diamine Tetra Acetic acid
EST	Expressed Sequence Tag
ESTPs	Expressed Sequence Tag Polymorphisms
EtBr	Ethidium Bromide
FMs	Functional Markers
g	Gram
ha	Hectare
hrs	Hours
IARI	Indian Agriculture Research Institute
ICRISAT	International Crop Research Institute for Semi-arid Tropics
IPTG	Isopropyl -D-thiogalactopyranoside
ISSR	Inter Simple Sequence Repeat
ITPs	Intron-targeted primers
Kb	kilobase
LB	Luria broth
M	molar
min	minutes
ml	millilitre
mM	millimolar
MAS	Marker-assisted selection
MFR	Microsatellite-flanking region
mg	Milligram
NBPGR	National Bureau of Plant Genetic Resources
NCBI	National Centre for Biological Information
ng	nanogram
nmoles	Nanomoles
OD	Optical Density
PAGE	Polyacrylamide gel electrophoresis
PEG	Poly ethylene glycol
PCR	Polymerase chain reaction
PIP	Potential intron Polymorphisms
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RGA	Resistance Gene Analogue
RIL	Recombinant Inbreed Line

RNAse A	Ribonuclease A
SAMPL	Selective Amplification of Microsatellite Polymorphic Loci
SSC	Sodium chloride-Sodium citrate Buffer
SSR	Simple Sequence Repeats
STMS	Sequence Tagged Microsatellite Sites
rpm	Revolutions per minute
TBE	Tris Borate EDTA buffer
TE	Tris- EDTA
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tris	Tris hydroxymethyl amino methane
U	Units
X-gal	5-Bromo-4-chloro-3-indolyl-D-galactopyranoside
v/v	Volume/volume
w/v	Weight/volume
μCi	Microcurie
μg	Microgram
μl	Microliter
μM	Micromolar

CONTENTS

Certificate
Acknowledgement
Abbreviations

Chapter1 Introduction 1-5

Chapter 2 Review of Literature 6-34

2.1 Chickpea: an edible pulse crop of the Fabaceace family, subfamily Faboideae
2.1.1 Classification
2.1.2 Cultivar types
2.1.3 Origin
2.1.4 Area, production and productivity
2.1.5 Uses
2.1.6 Biotic and abiotic stress: major constraints on chickpea productivity

2.2 Molecular markers in plant genome analysis
2.2.1 Microsatellites
2.2.1.1 Evolution of microsatellites
2.2.1.1.1 Structural factors influencing microsatellite variability
2.2.1.1.2 Directional evolution of microsatellites
2.2.1.2 Isolation of microsatellites
2.2.2 Transcriptome-based/ Functional molecular markers
2.2.2.1 Development of functional molecular markers
2.2.2.2 ESTs a source of PCR-based functional markers
2.2.3 EST-SSR markers
2.2.3.1 Frequency and distribution of SSRs in ESTs
2.2.4 Applications of EST-SSR markers
2.2.4.1. Functional diversity
2.2.4.2 Mapping
2.2.4.3 Transferability and comparative mapping
2.2.5 Expressed sequence tag polymorphic (ESTP) markers
2.2.6 Potential intron-polymorphic (PIP) markers

2.3 Genome analysis in chickpea
2.3.1 cDNA libraries available in chickpea
2.3.2 Molecular breeding
2.3.2.1 Microsatellite markers in chickpea
2.3.2.1.1 Isolation of SSRs and development of STMS markers
2.3.2.2 Utilization of developed chickpea STMS markers
2.3.2.2.1 Genetic diversity studies
2.3.2.2.2 Cross-species amplification of chickpea microsatellite markers
2.3.2.3 Genetic linkage map of *Cicer* genome
2.3.2.4 Mapping of agronomic traits and important genes in chickpea.

Chapter 3 Materials and Methods 35-53

3.1 Materials
3.1.1 Plant materials
3.1.2 The inter-specific Recombinant Inbreed Lines (RIL) mapping population of chickpea
3.1.3 Tissue collection
3.1.4 Primers used for mapping in this study
3.1.5 Vectors and bacterial strains used
3.1.6 Chemicals and Materials used

3.2 Methods
3.2.1 General sterilization procedures
3.2.2 Genomic DNA isolation
3.2.3 Isolation of genomic DNA by GenElute^{TM} Plant Genomic DNA Miniprep kit (SIGMA)
3.2.4 Quantification of genomic DNA
3.2.5 RNA isolation
3.2.5.1 Isolation of RNA from Chickpea
3.2.5.2 RNA quantification
3.2.5.3 Denaturing formaldehyde gel for RNA electrophoresis
3.2.6. Purification of DNA fragments
3.2.6.1 Elution of DNA from gels
3.2.6.2 Purification of PCR amplified products
3.2.7. Cloning of purified DNA fragments
3.2.7.1 Ligation
3.2.7.2 Preparation of Competent Bacterial Cells by $CaCl_2$ method
3.2.7.3 Transformation
3.2.7.4 Confirmation of inserts by colony PCR
3.2.8 Isolation of plasmid DNA
3.2.8.1 Alkaline lysis miniprep method (Sambrook et al. 1989)
3.2.8.2 Alkaline lysis midiprep method
3.2.8.3 Purification of Plasmid by PEG Precipitation for Sequencing
3.2.9 Sequencing
3.2.10 Construction of cDNA library
3.2.10.1 First strand cDNA synthesis
3.2.10.2 Second-strand synthesis
3.2.10.3 Double-strand cDNA polishing
3.2.10.4 Ligation and transformation
3.2.10.5 Selection of positive clones and EST sequencing
3.2.11 EST processing
3.2.11.1 EST preprocessing and contig assembly
3.2.11.2 Functional annotation
3.2.12 Mining of microsatellite motifs from chickpea ESTs and designing of STMS primers
3.2.13 PCR amplification
3.2.14 Preparation of agarose gel and running conditions
3.2.14.1 Metaphor agarose gel
3.2.14.2 Polyacrylamide gel
3.2.15 Development of EST based markers

3.2.15.1 Expressed sequence tag polymorphic (ESTP) markers
3.2.15.2 Potential intron polymorphic (PIP) markers
3.2.16 Data analysis for genetic diversity studies
3.2.17 Sequence alignments
3.2.18 Mapping
3.2.18.1 Genotyping using various markers
3.2.18.2 Linkage analysis and map construction
3.2.19 Northern Hybridization
3.2.19.1 Transfer of total RNA on Nylon Membrane
3.2.19.2 Restriction digestion of Plasmid DNA
3.2.19.3 Radioactive probe preparation and purification
3.2.19.4 Hybridization, washing and Autoradiography

Chapter 4 Generation and analysis of chickpea ESTs **54-70**
4.1 Introduction
4.2 Results
4.2.1 Construction of a cDNA library from developing seeds of chickpea
4.2.2 Annotation and functional classification of unigenes
4.2.3 Northern analysis
4.3 Discussion

Chapter 5 Development of EST-SSR markers in chickpea **71-82**
5.1 Introduction
5.2 Results
5.2.1 Identification of microsatellites from chickpea ESTs
5.2.2 Development of functional chickpea EST-SSR markers
5.3 Discussion

Chapter 6 Development of ESTP and PIP markers **83-97**
6.1 Introduction
6.2 Results
6.2.1 Development of ESTP markers
6.2.2 Development of PIP markers
6.2.3 Screening for polymorphism between parents of mapping population
6.2.4 Sequence analysis
6.3 Discussion

Chapter 7 Exploiting chickpea EST-SSR markers for genetic diversity and cross transferability studies across genus *Cicer* and related genera **98-113**
7.1 Introduction
7.2 Results
7.2.1 Analysis of intra-specific genetic diversity in chickpea using genic markers
7.2.2 Cross-species transferability across *Cicer*
7.2.2.1 Amplification pattern
7.2.2.2 Sequence analysis of size variant alleles from *Cicer* species
7.2.2.3 Phylogenetic analysis
7.2.3 Cross-genera transferability and sequence variation of chickpea EST-SSRs across legumes
7.3 Discussion

Chapter 8 Construction of an inter-specific linkage map of chickpea **114-132**
8.1 Introduction
8.2 Results
    8.2.1 Screening for parental polymorphism and genotyping of polymorphic markers
    8.2.2 Segregation distortion
    8.2.3 Construction of the inter-specific linkage map of chickpea
8.3 Discussion

Chapter 9 Summary and Conclusions **133-139**

Chapter 10 References **140-160**

Appendix

Chapter 1: Introduction

Sequence information has brought out new perspectives and approaches in the field of biological research. Large-scale sequencing of cDNAs to produce Expressed Sequence Tags (ESTs), an alternative to the whole genome sequencing and comparing the resulting sequences with public databases are currently being used as a fast and efficient method for gene discovery and gene profiling studies. Consequently, ESTs have become the fastest growing segment of the public DNA databases and now has also attracted researchers for molecular marker development since they represent part of the transcribed portion of the genome. Molecular markers, defined as 'constant landmarks in the genome' are powerful tools in modern agriculture and have been deployed in various aspects of plant genetics and breeding including genome and comparative mapping, phylogeny and population genetics, parental selection and species identification, association studies and QTL analysis. Of the myriad molecular markers available, PCR-based markers especially SSR-derived STMS are the work-horse of gene mapping projects by virtue of their codominant inheritance, multiallelic nature, reproducibility, good genome coverage and hyper variability representing the defined positions in the genome. Consequently, during recent years the high resolution genetic maps are being developed at an unprecedented speed in several economically important crops, a prerequisite step for further tagging of agronomically important traits or map-based cloning required for marker-assisted selection.

However, most of the markers developed and used in the past belong to either the transcribed or the non-transcribed region of the genome, often described as random or anonymous markers. In recent years, advances in genetics and genomics have led to generation of new tools such as the functional molecular markers (FMs) and bio-informatics that could more efficiently and precisely assist in the crop improvement programs. In this regard, the swelling EST databases has provided attractive alternative source for development of EST-derived markers that are pronounced to be more useful for breeding, as genes can be tested directly for their roles in various crop traits, thus being more informative and applicative than anonymous markers. Subsequently, several classes of transcriptome-based molecular markers have been developed such as EST-SSRs, EST-SNPs, COS (Conserved Orthologous Sites) and ESTPs (Expressed Sequence Tag Polymorphism) in several economically important crops such as cereals, pines, grapes, *Medicago* etc. Recently, EST datasets in conjunction with bioinformatic tools have allowed the design of intron targeted primers (ITPs) even in species whose genome is yet to be sequenced or characterized.

Although functional markers are reported to be less polymorphic compared with random markers in crop plants, they are anticipated to be more relevant to the goals of marker-assisted breeding. For example, FMs used for diversity studies will perhaps give a better estimate of genetic diversity by capturing variations in transcribed and known-function genes, thus having implications in plant breeding programs where parents need to be selected on the basis of their genetic divergence. Moreover, by virtue of sequence conservation of the expressed regions of the genome, FMs are the preferred marker system for cross-transferability studies across related species aiding in gene introgression programs, comparative genome analysis (Varshney et al. 2005a; 2005b), depiction of gene evolution and phylogenetic studies.

Chickpea (*Cicer arietinum* L.) is the third most important grain legume crop, valued for its nutritive seeds that fulfills the protein and starch requirements in the diet of poor and vegetarian population and is therefore often referred to as 'poor man's meat'. The plant in association with *Rhizobia* fixes atmospheric nitrogen and is a suitable rotation crop for agricultural practices. Despite the growing demand and continual efforts by chickpea breeders, chickpea yield is still unstable and productivity is stagnant at unacceptably low levels. The low genome variability and susceptibility of crop to several biotic and abiotic stresses are the major constraints that hampered chickpea improvement programs (Millan et al. 2006). Research has concentrated on the development of improved germplasm for resistance/tolerance to biotic and abiotic stresses and more recently has focused on the use of genetics and biotechnological tools to enhance the chickpea productivity. With the development of modern genetic tools such as DNA-based markers, linkage maps and genomic resources like BAC libraries, ESTs, chickpea molecular genetic studies have significantly progressed. However availability of genomic resources in chickpea are still limited and lag far behind those available in other economically important crops and therefore urgent efforts need to be made in this direction that will enable us to carry out function/genomics -assisted breeding.

Till date microsatellite based markers especially STMS have emerged as the most efficient and reliable source for detecting genetic variation in chickpea. Subsequently, microsatellites have been isolated from the chickpea genome through different approaches utilizing conventional genomic libraries (Hüttel et al. 1999; Winter et al. 1999), BAC library (Lichtenzveig et al. 2005) and microsatellite enriched library (Sethy et al. 2003; Sethy et al. 2006a) leading to the availability of 694 total STMS markers in chickpea. The genomic

derived SSRs along with other molecular markers such as RAPD and ISSR have already been implicated in chickpea for diversity estimation, germplasm characterization, elucidating *Cicer* phylogeny, construction of molecular maps and transferability studies. However most of them neither have genic functions nor have they shown close linkage to coding regions. Moreover, the high developmental costs, species-specificity and their association mostly with non-coding regions have limited the applicability of aforementioned markers for direct tagging of genes, offsetting the gene introgression programs and comparative genomic studies. Hitherto, in chickpea, little emphasis has been made towards the development of gene-based molecular markers except for a report by Buhariwalla et al. 2005. Thus attempts need to be expedited to capitalize the new avenues for accelerating the chickpea breeding programs.

Presently, the central goal of chickpea geneticists is to generate an integrated genetic map of the crop, comprising loci of both economic and scientific importance (Millan et al. 2006). The advent of STMS markers has facilitated construction of inter- and intra-specific linkage maps in chickpea and also provided the possibility of unifying different genetic maps and to develop a consensus map. However chickpea demonstrates only 20%-40% DNA polymorphism, the currently available markers are insufficient for construction of a dense genetic map and for use in marker-assisted selection and map-based cloning of agronomically important genes. The most extensive genetic map of chickpea (Millan et al. 2006) till date encompasses 2483.3 cM of the genome and have only a few functional markers mapped (Pfaff and Kahl 2003). Moreover, the limited amount of genomic resources, especially ESTs, in the databases (only 1311 chickpea ESTs till 2005) has further impeded molecular genetic analysis in chickpea. High density genetic maps of gene based markers represent a powerful resource to enhance genome analysis and for identification of candidate genes for agronomically important loci. Thus there is an immediate need to generate a high density DNA marker map of chickpea anchored with gene based markers that will foster marker assisted selection for crop improvement. However in chickpea, due to the limited availability of the gene-based markers, it is imperative to enrich the EST database and utilize this resource in the breeding programs through generation of EST based functional markers. Simultaneously, EST generation will also boost the chickpea functional genomics which is still in its infancy. Therefore, in the present study following objectives were proposed:

1) Generation and analysis of expressed sequenced tags (ESTs) from chickpea seeds
2) Development of EST based microsatellite markers (EST-SSRs) from chickpea seed ESTs and NCBI EST database
3) Development of chickpea Expressed Sequence Tag Polymorphism (ESTP) and Potential Intron Polymorphism (PIP) markers from chickpea ESTs
4) Utilization of the chickpea EST-SSR markers for analysis of genetic diversity and cross-species transferability across genus *Cicer* and related legumes
5) Construction of an inter-specific linkage map of chickpea using genomic and genic SSRs and other gene based markers

Organization of thesis:

The thesis is organized into nine chapters including this introductory chapter and the contents of each are as follows:

Chapter 2. Review of literature

This chapter reviews the available literature on the different types of EST-based functional molecular markers and their applications in plant breeding. Further, the status of the available genomic resources and their utilization in molecular breeding in chickpea were reviewed.

Chapter 3. Material and Methods

This chapter lists the material used and the protocols of techniques used in the present study

Chapter 4. Generation and analysis of chickpea ESTs

This chapter describes the way the chickpea ESTs were generated, assembled and functionally annotated using the bioinformatics tools

Chapter 5. Development of EST-SSR markers in chickpea

This chapter describes how the generated ESTs as well as the available chickpea EST database was utilized for the development of large number of novel EST-SSR markers

Chapter 6. Development of ESTP and PIP markers

This chapter describes how the other kinds of chickpea EST-based markers such as ESTP and PIP were developed from chickpea ESTs

Chapter 7. Exploiting chickpea EST-SSR markers for genetic diversity and cross transferability studies across genus *Cicer* and related genera

This chapter demonstrates the potential of the chickpea genic SSR markers for analysis of genetic diversity and cross species transferability

Chapter 8. Construction of an inter-specific linkage map of chickpea

This chapter describes how all the EST based molecular markers developed in the present study as well as some of the reported but still unmapped markers were utilized for construction of a detailed genetic linkage map of chickpea

This is followed by summary and list of references.

Chapter 2: Review of Literature

2.1 Chickpea: an edible pulse crop of the Fabaceace family, subfamily Faboideae

Chickpea (*Cicer arietinum* L.) is an annual, self-fertilizing diploid (2n=2x=16) cool-season grain legume (Fig. 2.1) that ranks third in world legume production (FAOSTAT 2006) after dry beans and peas. The crop is broadly grown in arid and semi-arid tropics, mainly on the Indian subcontinent but also in Mediterranean basin, West Asia and North Africa (WANA) regions, United States and recently expanded to Australia and Canada. Being a staple diet component, chickpea provides protein and starch to the predominantly vegetarian population of the developing countries or to those who cannot afford expensive animal protein and is considered to be a health food in Western countries. Among the dry edible legumes, nutritional compositions and protein digestibility of chickpea is highest as anti-nutritive components are negligible (Williams and Singh 1987). Thus, chickpea is considered a functional food or nutraceutical (Agharkar 1991; McIntosh and Topping 2000). The plant has the ability to fix atmospheric nitrogen (upto 140 kg N/ha) through symbiotic relationship with *Rhizobia* (Fig. 2.1B), thus maintains soil-fertility which contributes to the sustainability of cropping systems in cereal-legumes rotation. Additionally the crop also helps in preventing the build-up of diseases, insect pests and weeds (Singh et al. 1999). Contrary to soybean, chickpea contains insignificant amount of isoflavones (USDA-ARS, 2004) however provide more beneficial dietary-carotenoids such as β-carotene, cryptoxanthin, lutein and zeaxanthin than genetically engineered "Golden rice" (Abbo et al. 2005).

2.1.1 Classification

Chickpea with a relatively small genome size of 750 Mb (Arumuganathan and Earle 1991) belongs to monogeneric tribe *Cicereae* Alef. The genus *Cicer* consists of 43 species of which 9 are annuals including cultivated chickpea, 33 perennial and 1 is unclassified. All these species are grouped into four sections based on morphological characteristics, life cycle and geographic distribution: *Monocicer*, *Chamaecicer*, *Polycicer* and *Acanthocicer* (van der Maesen, 1987). All annual species except *Cicer chorassanicum* (*Chamaecicer*) belong to the section *Monocicer*. Chickpea's wild relatives have attracted much attention as they are potential sources of a wide range of agronomically favourable traits for the improvement of the genepool of the cultigen (Muehlbauer et al. 1994; Singh and Ocampo 1997; Singh et al. 1998).

Crossability and fertility of hybrids in interspecific crosses have been used as a basis to elucidate the genetic relatedness of wild annuals with the cultivated species. van der

A

B

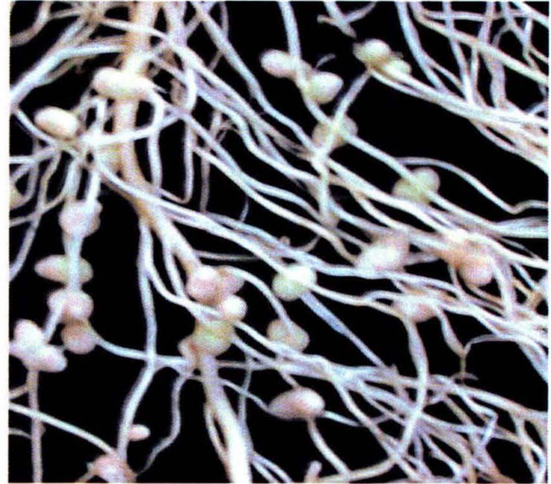


Figure 2.1: Overview of chickpea **A)** Chickpea plant **B)** Nodulated roots of chickpea **C)** Kabuli types with white flower and seeds and **D)** Desi cultivar with pink flower and seeds

C

D

Maesen (1972) and Ladizinsky and Adler (1976a, 1976b) classified the annuals into three genepools i.e. primary, secondary and tertiary. Based on this classification and results of karyotype (Ocampo et al. 1992; Ahmad 2000), specific enzyme activity (Tuwafe et al. 1988; Gaur and Slinkard 1990a,b; Ahmad et al. 1992; Kazan and Muehlbauer 1991; Labdi et al. 1996; Tayyar and Waines 1996), seed storage protein characteristics studies (Ahmad and Slinkard 1992) and DNA-based techniques like synthetic oligonucleotide fingerprinting (Weising et al. 1992; Sharma et al. 1995a), RFLPs (Banerjee et al. 1999), microsatellite-tagged RFLPs (Serret et al. 1997), satellite DNA (Staginnus et al. 1999), RAPD (Ahmad 1999; Iruela et al. 2002; Sudupak et al. 2002), ISSR (Rajesh et al. 2002) and sequence-tagged microsatellite site markers (Choumane et al. 2000), a revised classification has been proposed by Croser et al. 2003 (Table 2.1). These classifications have been supported by recent studies based on DNA-based molecular markers like AFLP (Sudupak et al. 2004; Nguyen et al. 2004; Shan et al. 2005), ISSR (Sudupak et al. 2004), STMS (Sethy et al. 2006b), chloroplast sequences (Javadi and Yamaguchi 2004) and EST-based markers (Buhariwalla et al. 2005).

Table 2.1: Proposed genepools of wild annual *Cicer* species (adapted from Croser et al. 2003).

Proposed genepool	**Ladizinsky and Adler 1976a,b**	**Verma et al. 1990**	**Muehlbauer et al. 1994**	**Singh et al. 1999a,b**
Primary	*C. arietinum*	*C. arietinum*	*C. arietinum*	*C. arietinum*
	C. reticulatum	*C. bijugum*	*C. echinospermum*	*C. judaicum*
		C. judaicum	*C. reticulatum*	*C. reticulatum*
		C. pinnatifidum		
		C. reticulatum		
Secondary	*C. echinospermum*	*C. echinospermum*		*C. bijugum*
				C. cuneatum
				C. echinospermum
				C. pinnatifidum
				C. yamashitae
				C. chorassanicum
Tertiary	*C. bijugum*		*C. bijugum*	
	C. cuneatum		*C. judaicum*	
	C. judaicum		*C. judaicum*	
	C. pinnatifidum		*C. pinnatifidum*	
Omitted	*C. chorassanicum*	*C. chorassanicum*	*C. cuneatum*	
	C. yamashitae	*C. cuneatum*	*C. yamashitae*	
		C. yamashitae		

2.1.2 Cultivar types

The two most common types of chickpea recognized are, "*Kabuli*" (from kabul, Afghanistan) morphologically characterized by large seeds with smooth round white or pale color (more than 26 g per 100 seeds) (Fig. 2.1C) and the small "*Desi*" type (Hindi: local) angular seeds which are dark-brownish to yellow in color (Fig. 2.1D). 'Kabuli' also referred as "macrosperma/macrocarpa" normally has non-pigmented flowers and is primarily grown in WANA, the Americas, and Europe. Whereas the desi type also known as "microsperma/microcarpa" accounts for about 80 percent of the world's chickpea production and predominates in Asia, parts of Africa, and Australia. 'Kabuli' plants compared to 'desi' type are generally taller, late maturing and lack anthocyanin pigmentation on the pods and stems. Studies revealed that dark colored 'desi' type has more protein content (235 $g\,kg^{-1}$) and thicker seed coat compared to the 'kabuli' type that has comparatively lower protein content (226 $g\,kg^{-1}$). On the contrary, the 'Kabuli' type possess more dietary fiber content, particularly cellulose and hemicellulose. Regarding the origin of 'Kabuli' and 'Desi' types, it is believed that Kabuli chickpea originated from the Desi type in the Mediterranean region through natural mutation and selection and spread to India in the east and Chile in the west (Moreno and Cubero 1978, Gil and Cubero 1993, Jana and Singh 1993). Differences in agro-climatic conditions and farmers' preferences for the species have produced geographical differentiation of interest both in agriculture and crop evolution. Though separated for centuries, no hybridization barrier exists between the two groups (Muehlbauer and Singh, 1987) but the slow transfer of genes between the two types has been reported (Hawting and Singh 1980; Maynez et al. 1993).

2.1.3 Origin

Chickpea is one of the pulse crops domesticated in the Old World ca 7000 years ago and is associated with the Neolithic origin of Near Eastern agriculture (Lee-Yadun et al. 2000). In a report by Vavilov (1926), southwest Asia and the Mediterranean were identified as the two primary centers of origin of chickpea, while Ethiopia was identified as a secondary center of origin. Later, Ladizinsky 1975; van der Maesen 1987; Singh (1997) reported that chickpea most probably originated in southeastern Turkey and adjoining areas of Syria. Wild *C. reticulatum* interfertile with cultivated cultigen, is regarded as an annual progenitor of chickpea (Ladizinsky and Adler 1976a) and perennial progenitor is proposed as *C. anatolicum* (Tayyar and Waines 1996).

2.1.4 Area, production and productivity

During 2002-2004, the global chickpea production was 8.0 million tons from an area of 10.1 million ha, giving an average productivity of 786 kg ha -1. Of the total world production 91% is produced in Asia, and in Asia, India accounts for 74.8% production. Currently, Australia followed by Turkey and Mexico are the three topmost exporters of chickpea and India shares 30% of total world imports despite largest area under cultivation (www.crnindia.com).

2.1.5 Uses

Chickpea is valued for its nutritive seeds with high protein content, 25.3-28.9 %, after dehulling (Hulse, 1991). On an average, chickpea seed contains 23% protein, 64% total carbohydrates, 47% starch, 5% fat, 6% crude fiber, 6% soluble sugar and 3% ash. Digestibility of protein varies from 76-78% and its carbohydrate from 57-60%. (Hulse 1991). Chickpea seeds are eaten fresh as green vegetables, parched, fried, roasted, and boiled; as snack food, sweet and condiments; seeds are ground and the flour can be used as soup, dhal, and to make bread; prepared with pepper, salt and lemon it is served as a side dish (Saxena 1990). Sprouted seeds are eaten as a vegetable or added to salads. Animal feed is another use of chickpea in many developing countries. Acid exudates from the leaves can be applied medicinally or used as vinegar. Chickpeas yield 21% starch suitable for textile sizing, giving a light finish to silk, wool, and cotton cloth" (Duke, 1981).

2.1.6 Biotic and abiotic stress: major constraints on chickpea productivity

Despite a proposed yield potential of 6 metric tonnes ha−1 (Singh 1987), actual yields have remained low compared to other pulse crops (world average 0.8 metric tonnes ha−1; FAOSTAT 2005), mainly because of biotic and abiotic stresses that reduce yield and yield stability. The necrotrophic foliar fungal disease Ascochyta blight caused by *Ascochyta rabiei* (Pass.) Labrousse and the soil-borne necrotrophic fungal disease Fusarium wilt caused by *Fusarium oxysporum* f. sp. *ciceris* are the two most serious diseases in North India, Pakistan, USA and the Middle East (sometimes causing 100% losses, Smithson et al. 1985). About 10-100% losses in yield due to *Fusarium* wilt is a regular feature in chickpea growing states of India. Other diseases of chickpea are more geographically localised and include pod borer (*Helicoverpa armigera*) in Australia and India (Nene and Reddy 1987), root rots *(Rhizoctonia bataticola)* in the tropics and sub-tropics (Kraft *et al.*. 2000) and rust in high-altitude regions (Nene and Reddy 1987). Pod borer *(Helicoverpa armigera)* is the most important pest, feeds

on leaves and developing seeds (Smithson et al.1985) among storage insects, specifically *Bruchid* sp. are a serious pest of stored chickpea.

In order of importance, drought, cold and salinity are the three main abiotic stresses that affect chickpea growth and productivity worldwide (Croser et al. 2003). As 90% of chickpea crops are cultivated under rain-fed conditions, drought is of major concern (Kumar and Abbo 2001). Throughout the chickpea production areas, the crop is additionally subjected to extremes of temperature and moisture supply, and to deficiency or toxicity of minerals in the soil (Saxena 1993). In West Asian and North African countries, low temperature causing freezing injury or death or delayed onset of podding reduces yield tremendously (Singh, 1987). Generally drought and stress lead to onset of other stresses particularly heat and salinity (Singh et al. 1994).

2.2 Molecular markers in plant genome analysis

There has been a progressive evolution in methods used to determine the genetic structure of individuals. Hitherto the complete genome of individuals can be sequenced for identifying the variations, however the geneticists still have to rely on genetic markers as a means of characterizing the genotypic variation of individuals. The development and use of molecular markers for the detection and exploitation of DNA polymorphism is one of the most significant developments in the field of molecular genetics. As compared to morphological /biochemical markers, molecular markers have several advantages as they are phenotypically neutral and are not influenced by pleiotropic and epistatic interactions, and their expression is not dependent on age/part of the plant (Jones et al. 1997). During the last two decades, DNA-based markers have led to the construction of whole genome linkage maps in many plant and animal genomes, a crucial step for several downstream applications such as gene cloning, genome analysis, and marker-assisted selection of agricultural crops (Cullis 2002; Dodgson et al. 1997; Paterson 1996a).

Ever since the development of molecular markers, their techniques are constantly being modified to enhance their sensitivity, resolution and utility to detect genetic discontinuity and distinctiveness. These DNA based markers are generally classified as hybridization-based markers and polymerase chain reaction (PCR)-based markers. Whereas the first generation DNA markers, like restriction fragment length polymorphism (RFLP) markers, have proved to be very useful, their development and use is laborious, time-consuming, expensive, and unsuitable for high throughput automation (Rafalski and Tingey 1993; Paterson 1996b). For these reasons, PCR-based markers, such as random amplified

polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSRs), have become popular for molecular genetic studies (Paterson 1996b). Out of the PCR-based markers, SSR markers became the markers of choice for plant and animal genomes during the last decade because of their hypervariability (ability to detect high levels of polymorphism) coupled with the requirement of small sample size (genomic DNA) for their analysis and their suitability for automation and high-throughput assays (Hearne et al. 1992; Morgante and Olivieri 1993; Powell et al. 1995). However the use of SSRs does necessitate their development from the species under study which involves cloning, sequencing and bioinformatics analysis. Nevertheless, the choice of the marker system is largely dependent on the intended application, expense involved and the ease of use.

2.2.1 Microsatellites

Microsatellites (Litt and Luty 1989) also known as simple sequence repeats (SSRs), short tandem repeats (STRs) or simple sequence length polymorphism (SSLP) are iterations of 1–6bp nucleotide motifs, found ubiquitously in all eukaryotic genomes (Tautz & Renz 1984; Hancock 1996; Gupta et al. 1996; Powell et al. 1996). The uniqueness and the importance of microsatellites arose from their multiallelic nature, codominant inheritance, relative abundance, hypervariable nature, extensive genome coverage and simple detection by PCR using the locus-specific markers that flank the microsatellite motifs and are termed as STMS (Sequence Tagged Microsatellite Site) markers as shown in Fig. 2.2. SSR markers find wide applicability in various fields ranging from genetic mapping (Roder et al. 1998; Brondani et al. 2006), genome fingerprinting (Sharma et al. 1995a) to genetic diversity analysis (Jarne and Lagoda 1996; Alghanim and Almirall 2003).

Microsatellites can be classified as perfect, compound or imperfect according to the form of microsatellite present in them and termed as Di-, Tri-, Tetra- etc with respect to the arrangement or organization of nucleotide string (Alghanim and Almirall 2003). Perfect SSRs consist of uninterrupted repeats of a single motif, e.g. (AT)n; compound SSRs are made up of adjacent tandem arrays of different motifs, e.g. (GT)n(AT)n; and imperfect SSRs are present as an interruption with in the repeat, e.g. (GT)nGG(GT)n.

2.2.1.1 Evolution of Microsatellites

Microsatellites are among the fastest evolving nucleotide sequences with length mutation rates varying from 10^{-2} and 10^{-5} mutations per locus per generation (Weber and

STMS
Forward primer
binding sites

Genomic DNA

STMS
Reverse primer
binding sites

GAGAGAGAGAGAGAGAGAGA Plant 1

GAGAGAGAGAGAGAGAGAGAGAGAGAGAGA Plant 2

GAGAGAGAGAGAGAGAGA Plant 3

Unique flanking
region

PCR-Amplification

Gel Analysis

P1 P2 P3

Figure 2.2: Molecular basis for using microsatellites as STMS (Sequenced Tagged Microsatellite Sites) markers

Wong 1993; Levinson and Gutman 1987). The most likely mechanism associated with length variation is the slipped strand mispairing (slippage) during DNA replication, the main source of mutation (Levinson and Gutman 1987) and recombination between DNA molecules resulting in loss or gain of one or more repeats. Based on these phenomena, classically two models have been proposed (Deka et al. 1991): the *Infinite Allele Model* (**IAM**, Kimura and Crow 1964) and the *Stepwise Mutation Model* (**SMM**, Kimura and Ohta 1978; Chakraborty and Nei 1982). In short, SMM describes mutation of microsatellite alleles by the loss or gain of a single tandem repeat, resulting in alleles already in the population. In contrast, IAM describes mutations involving the loss or gain of any number of tandem repeats but always resulting in an allele state not previously encountered in the population. However, despite several studies, the precise nature of mutation of microsatellite sequences and how they evolve is still controversial in literature. Recent studies have indicated the involvement of complex mutational processes including the point mutations (substitution, insertions or deletions) either in the flanking regions or within repeats (Primmer et al. 1998; Makova et al. 2000) and several other structural factors like allele size and purity, the upwardly biased mutation process and constraints on allele length that need to be considered while predicting the evolution of microsatellites.

2.2.1.1.1 Structural Factors influencing microsatellite variability:

Several studies have been attempted to determine the structural factors that would influence the variability at a microsatellite locus. Among them, some are discussed below:

A. Number of repeat units

Of the myriad factors influencing the mutation rates of tandem repeats, the number of identical repeat motifs is the most important one, with the mutation rates being highest in longer repeats (Levinson and Gutman 1987; Bryan et al. 1997). Several population studies have further substantiated the fact that a positive correlation exists between the diversity and repeat count (Beckmann and Weber 1992; Ostrander et al. 1993) as slippage increases with increase in motif number (Levinson and Gutman 1987; Weber 1990).

B. Composition of the repeat unit

Mutation rates (and hence the variability) of different types of microsatellites-di, tri or tetra is dependent on their nucleotide composition. Most of the studies indicate that dinucleotides have the highest mutation rate, on average followed by tri- and tetranucleotide repeats (Chakraborty et al., 1997; Kruglyak et al., 1998; Lee et al., 1999). The biological

basis for differences in stability of tracts of repeat units with similar size and different nucleotide composition remains unclear, although GC content appears to be atleast one factor, as the most polymorphic tri- and tetranucleotide motifs are AT- rich (Gastier et al. 1995).

C. Repeat type/length purity

This parameter substantially influences the microsatellite stability independent of the composition of the repeat unit (di, tri or tetra). Weber (1990) found (CA)n perfect repeats as the best predictors of polymorphism compared to compound and interrupted microsatellites. The interrupting bases generally stabilize the repeat tracts, reduce the slippage, thereby are responsible for monomorphism and finally their accumulation may lead to death of microsatellite loci (Taylor et al. 1999; Zhu et al. 2000). The most polymorphic and hence the most useful microsatellite motifs are the uninterrupted arrays (Weber 1990) as the imperfections induce microsatellite stability (Richards and Sutherland 1994; Pe‘pin et al. 1995).

2.2.1.1.2 Directional Evolution of microsatellites

Earlier thought to be a symmetrical process of expansion and contraction of repeats (Di Renzo et al. 1994; Kruglyak et al. 1998), microsatellite evolution has turned out to be a complex mechanism involving several hypothesis and contradictions. Researchers have suggested that the microsatellite evolution is directional involving two kinds of directionality. Firstly, directionality is at the individual level in which addition of repeats is more frequent than deletions (Weber and Wong 1993; Amos et al. 1996; Primmer et al. 1996). Wierdl et al. (1997) suggested that this could arise due to the tendency of the mismatched repeats to occur on the elongating strand rather than on the template strand.

Secondly, directionality is reported at the species level wherein microsatellites in some species are consistently longer than their homologues in other's related species. Rubinsztein et al. (1995) suggested that this process operates because of the differences in mismatch repair enzyme or difference in population size. Since long microsatellites are selected as markers, Ellegren et al. (1997) referred to the effect as an ascertainment bias and argued that the cloning process was the cause of observed length difference. But later, additional evidence demonstrating longer microsatellites in humans even when the loci were cloned from chimpanzees (Amos and Rubinsztein 1996) supported the directionality. Similarly, longer alleles were observed in sheep regardless of microsatellite identification procedures used in case of sheep or cattle (Crawford et al. 1998).

2.2.1.2 Isolation of microsatellites

From the first description of microsatellite markers in plants by Condit and Hubbell (1991), various protocols for isolation of microsatellites have been developed (Bruford et al. 1996; McDonald and Potts 1997; Hammond et al. 1998; Schlotterer 1998) and the details of the methodologies are reviewed by different authors (e.g., Chambers and MacAvoy 2000; Zane et al. 2002; Squirrell et al. 2003). A review by Zane et al. (2002) describes some of the technical advances that have been made in recent years to facilitate development of microsatellites from plants since they have been reported to be relatively less abundant in plant genomes compared to animals. In contrast to conventional methods, although SSR enrichment methods with selection either before or after genomic library construction improve the output and reduce development costs (Edwards et al. 1996), they still require considerable efforts and resources for isolation of SSRs thereby restricting their use to only a few of the important crops. Therefore recently, the paradigm has shifted towards an alternative source of SSR development i.e. from the transcribed region of the genome. Over the course of time, both wet lab and in silico approaches have been tremendously applied for the development of genic-SSR markers.

2.2.2 Transcriptome-based molecular markers/Functional molecular markers

Most of the molecular markers, developed and used in the past were related to genomic DNA (gDNA), and therefore could belong to either the transcribed region or the non transcribed region of the genome. These DNA-based markers derived from any region of the genome (mentioned as anonymous (non-coding) region) have been described as random DNA markers (RDMs) by Andersen and Lübberstedt 2003. During the last few years, molecular marker technology in higher plants has witnessed a shift from the so-called random DNA markers (RDMs) to the molecular markers representing the transcriptome/genes, commonly known as functional markers (FMs). Now days, FMs are preferred over RMs because they are completely linked to the desired trait allele, which is relevant to the goals of marker-assisted breeding. Since such markers are derived from the gene responsible for the trait of interest and target the functional polymorphism in the gene, they allow selection in different genetic backgrounds without revalidating the marker–quantitative-trait-locus (QTL) allele relationship. Thus, they have also been referred to as 'perfect markers', even though different alleles with the same polymorphism (resulting from intragenic recombination, insertion, deletion or mutation) might produce different phenotypes.

2.2.2.1 Development of functional molecular markers

Current genome sequencing studies and accumulation of sequence information have dramatically changed experimental plant biology by providing researchers with overwhelming avenues for gene discovery. Also the time required to localize and fully describe a gene is rapidly decreasing due to the development of Expressed Sequence Tags (ESTs). EST, a biological technique developed in the 1980s (Putney et al. 1983, Adams et al. 1991) is a rapid and cost-effective method for generating data on the coding capacity of genomes and has become the fastest growing segment of the public DNA databases (Wolfsberg and Landsman, 1997). EST provides researchers with a quick and inexpensive route for new gene discovery, genome annotation and comparative genomics.

ESTs are unedited, automatically, processed, single-read sequences produced from cDNAs (small DNA molecules reverse-transcribed from the cellular mRNA population) (Fig. 2.3). In plants, the EST approach was initially used for the model species *A. thaliana* (Höfte et al. 1993) and rice (Yamamoto and Sasaki 1997) and subsequently a large variety of EST sequences from other species have been deposited in the dbEST database of NCBI. Currently, from the number of publicly available ESTs, nearly 986,000 nucleotide sequences representing the Fabaceae family are known and among them 92% of the ESTs are derived from the model legumes *Medicago truncatula*, *Lotus japonicus* and the crop legume soybean (*Glycine max*) (Ramirez et al. 2005; Jayashree et al. 2006). EST data have been directly applied for gene discovery (Somerville and Somerville 1999; Ohlrogge and Benning 2000), evaluation of the genome-wide gene content and structure (Van der Hoeven et al. 2002), as well as in transcript-profiling studies both through wet lab and in silico approaches (Schena et al. 1995; Sreenivasulu et al. 2002, 2004). Besides the above mentioned utility, ESTs emerged as an efficient source for developing molecular markers which has had a revolutionary impact on gene mapping and more generally, on orphan crops whose genome sequence is not available.

With the availability of large EST collections in the database, different types of functional markers have been developed from many plant species (Gupta and Rustgi 2004). Based on methodologies similar to those used for development i.e. both "hybridization based" (as RFLP probes) and "PCR based" (as PCR primers), functional markers have been developed from the expressed region of the genome and are listed in Table 2.2. The table describes the different marker sources (e.g. mRNA, cDNA, ESTs, gene/genome sequences, etc.), the classes of markers derived from each of these sources and key features of each of these marker classes. Major classes of these markers belonging to the expressed region of the

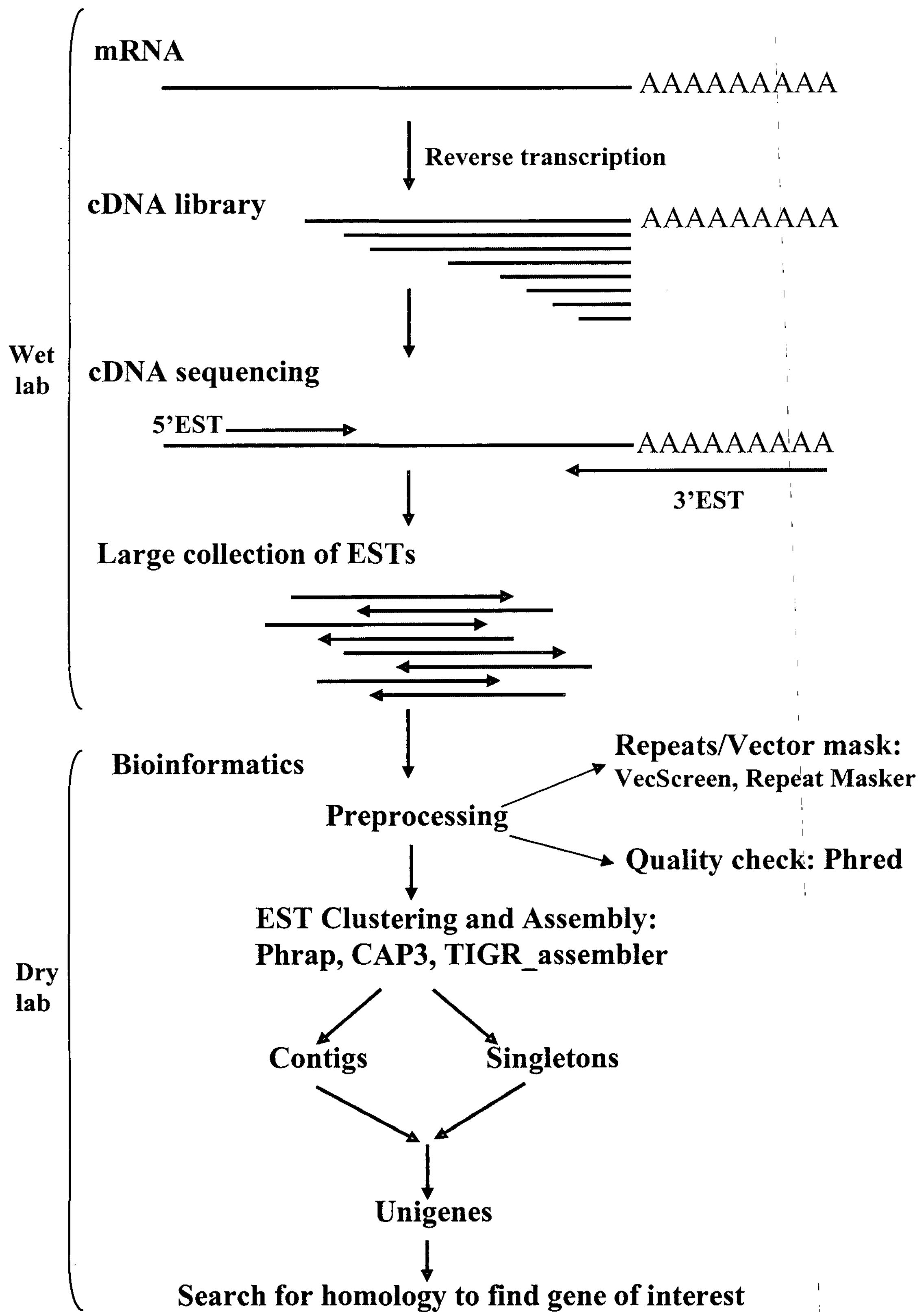


Fig. 2.3 Summarized steps of cDNA cloning, EST generation and *insilico* analysis as obtained from Rudd et al. 2003

genome were also developed earlier from the genomic sequences (including coding and non-coding regions) and were discussed in several reviews (Gupta et al. 2002; Semagn et al. 2006). However, a newer area is the development of the so-called Gene targeted markers (GTMs) and the FMs from sequences representing genes/ ESTs.

Table 2.2: Different type of functional markers (FMs) that can be developed from the expressed region of the genome along with their key features and references (obtained from Gupta and Rustgi 2004).

Marker type	Key Features	References
(I) cDNA/mRNA as a source of molecular markers		
cDNA-RFLPs	cDNA clones directly used as probes. Can be anonymous/gene-specific	Gupta et al. 2002
cDNA-AFLPs	cDNA instead of gDNA utilized for development of AFLP markers	Habu et al. 1997
RNA-fingerprinting differential display (RNAF-DD)	mRNAs of differential expressed genes are selectively amplified and used as markers	Sandhu et al. 2002
cDNA-SSRs	Library of locus-specific tags for SSR sequences is generated	Scott et al. 2000; Decroocq et al. 2003
(II) EST databases as a source of PCR-based molecular markers		
insilico AFLP	cDNA-AFLP banding patterns are correlated with the virtual fingerprints predicted from the reference database of model genome	Qin et al. 2001; Peters et al. 2001
EST-SSRs	In silico mining of SSRs from EST sequences available in the databases	Kantety et al. 2002
EST-SNPs	SNPs can be identified in the ESTs either from different accessions of the same species or through assembly of contigs	Kota et al. 2003
Insertion and indels (InDels)	Identification of insertion/indels through resequencing or data mining	Bhattramakki et al. 2002
Expressed Sequence Tag Polymorphisms (ESTPs)	PCR-based markers designed from the expressed regions of the plant genomes	Brown et al. 2001
Conserved orthologue set markers (COS)	Identifying conserved sequences through comparing available whole genome sequences with the EST databases.	Fulton et al. 2002
PCR-based markers targeting introns, exons and promoter regions (ITPs, PIPs)	Genomic DNA sequences indicating the positions of introns, exons and promoter regions can be used to develop PCR primers flanking these regions with high specificity	Holland et al. 2001; Wang et al. 2005
(III) Known gene sequences as a source of molecular markers (GTM)		
Amplified consensus genetic markers (ACGMs)	Based on the sequences of known genes, homologous genes are amplified in other genomes and assayed for polymorphism	Brunel et al. 1999
SSRs, SNPs, SCARs, CAPSs and RFLPs for known genes	PCR-based markers for known genes are developed. These could be in the form of SSRs, SNPs, SCARs, CAPs and RFLPs	Tartarini et al. 2003
Gene specific tags	GSTs are gene-specific primer pairs, which can amplify a specific gene from the genome	Samson et al. 2003
rDNA as markers	rRNA is used as a probe for hybridization of Southern-blots	Rogers and Bendich 1987
Resistance gene analogues (RGAs)	Degenerate primers are designed against the conserved domains of resistant genes	Hüttel et al. 2002
Exon-retrotransposon amplification polymorphism (ERAP)	A gene-specific primer combine with LTR retrotransposon primer is used	Pfaff and Kahl 2003

2.2.2.2 ESTs - a source of PCR-based functional markers

With the availability of expressed sequence tag (EST) sequencing projects for gene discovery programs in several plant species, a wealth of DNA sequence information has been generated which allows the possibility of detecting and genotyping these expressed sequences. ESTs provide valuable, although incomplete sequence information. However, they represent expressed genomic regions and are thought to identify the parts of the genome with the most biological significance (Rudd 2003). In the absence of complete genome sequences, the desire to generate high-density genetic maps of the different plant genomes remains a priority for the direct identification of specific genes. Thus, the large-scale EST sequence database appears to be a promising alternative for the rapid and inexpensive development of PCR based markers such as SSRs, ESTPs (Expressed Sequenced Tags Polymorphism) and SNPs (Single Nucleotide Polymorphisms) (including indels) etc. (Fig. 2.4) to the development of traditional ‘‘anonymous’’ markers following standard methods (Gupta et al. 2003 and 2004; Pashley et al. 2006). Moreover, to exploit the EST resources efficiently for genetic analysis and breeding, researchers are utilizing more competent methodology like intron-targeted amplified primers (ITAPs) for detecting sequence polymorphisms that could be effectively applied for DNA fingerprinting and mapping studies (Fig 2.4).

2.2.3 EST-SSR markers

2.2.3.1 Frequency and distribution of SSRs in ESTs

Identification of SSRs from gene sequences of plant species was started as early as 1993 by Morgante and Oliveri. Computational analysis has permited the rapid discovery of ever-increasing microsatellites from the swelling database from many genomes, demonstrating the presence of SSRs both in protein-coding genes and expressed sequence tags (ESTs) (Morgante et al. 2002; Li et al. 2002). In protein-coding regions of all known proteins, 14% proved to contain repeated sequences, which was 3 times higher in eukaryotes than in prokaryotes (Marcotte et al. 1999).The frequency of microsatellites in ESTs was observed to be significantly higher than in genomic DNA of several plant species (Toth et al. 2000; Morgante et al. 2002; Kumpatia and Mukhopadhyay 2005) even though repeat numbers and total length of SSRs observed was small (Kantety et al. 2002; Thiel et al. 2003). The frequency of SSRs in ESTs has been widely studied in both monocots (Kantety et al. 2002; Varshney et al. 2002; Thiel et al. 2003) and dicots (Scott et al. 2000; Eujayl et al. 2004;

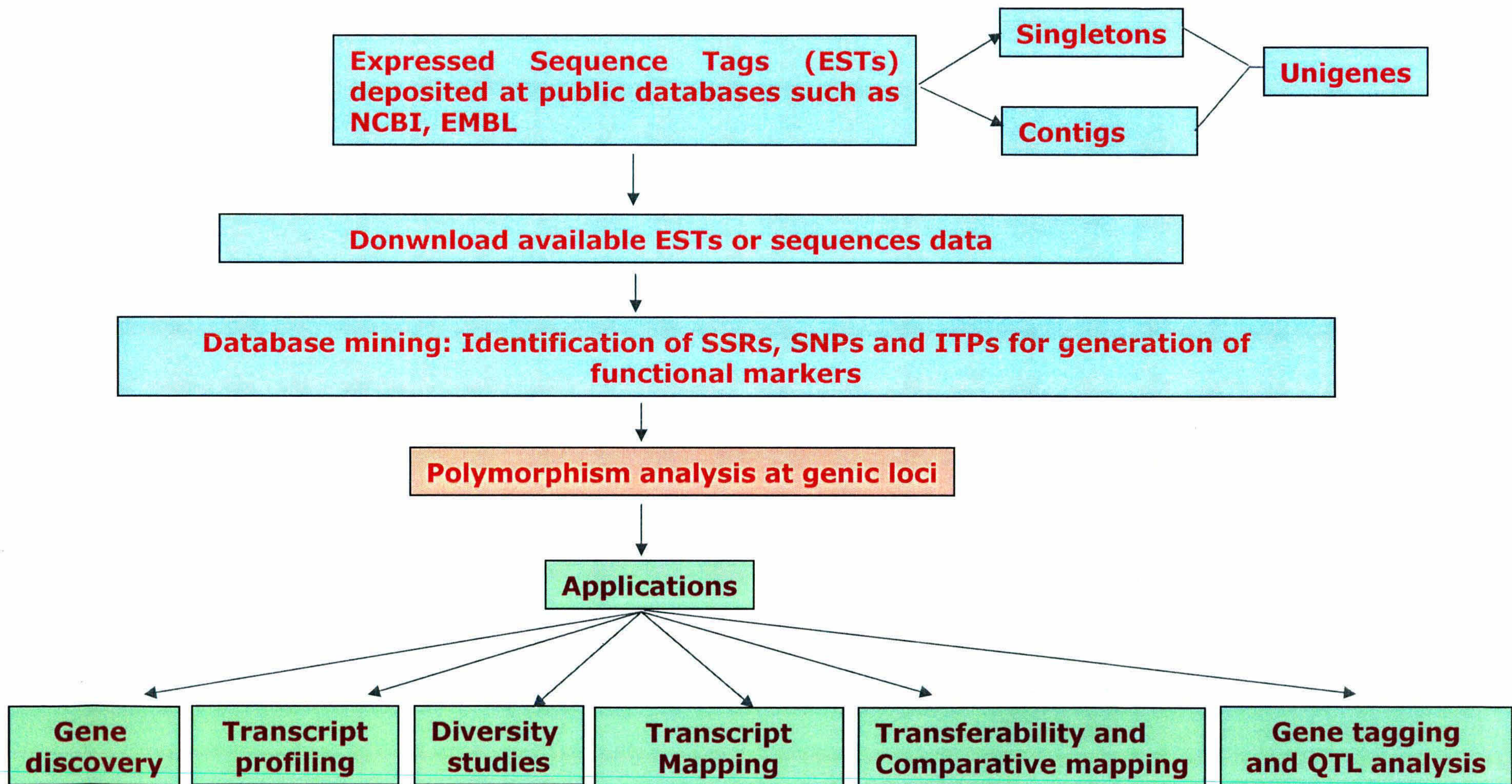


Figure 2.4: Development and applications of EST-based molecular markers (as adapted from Varshney et al. 2005).

Kumpatia and Mukhopadhyay 2005). In general, the higher SSR-EST frequency was observed for dicot species (in the range of 2.6-10.6%, Kumpatia and Mukhopadhyay 2005) than to monocots (in the range of 1.5-4.7%, Kantety et al. 2002). Furthermore, the abundance of SSRs mined from a sequence database depends on the SSR search criteria, the size of the dataset and the database mining tools (Varshney et al. 2005a).

With regard to motif types, trinucleotides are the most abundant repeat types followed by di and tetra nucleotides in the EST sequences as studied in several crop plants (Cardle et al. 2000; Varshney et al. 2002; Jayashree et al. 2007). This is attributed to an absence of frameshift mutations due to variation in the number of trinucleotide repeats (Varshney et al. 2002). However, this is quite contrary to genomic DNA sequences where dinucleotides are the most abundant motif (Akagi et al. 1996; Chin et al. 1996). Among the trinucleotides, AAG/ TTC, AGG/TCC and AGA/ CTT were found to be most abundant motifs in dicots (Eujayl et al. 2004; Kumpatia and Mukhopadhyay 2005; Poncet et al. 2006) whereas in monocots majorly CCG motif was abundant (Varshney et al. 2002) owing to high G+C content and consequent codon usage bias. Contrary to high AT repeats observed in genomic DNA (Wang et al. 1994; Cardle et al. 2000), AG/CT is the most frequently observed dinucleotde SSRs in ESTs of plants (Morgante et al. 2002; Kumpatia and Mukhopadhyay 2005).

Over the years, the presence of SSRs in gene transcripts has been shown to be substantially involved in regulating gene expression and function (Kashi et al. 1997; Li et al. 2004; Varshney et al. 2005a). In transcribed regions, microsatellites are present more frequently in 5' UTRs than in coding regions or 3'-UTRs (Wren et al. 2000; Morgante et al. 2002; Fujimori et al. 2003) suggesting that they can potentially act as factors in regulating gene expression. The 5' UTRs contains more triplets than 3' UTRs as reported in humans, *Arabidopsis*, barley (Wren et al. 2000; Morgante et al. 2002; Theil et al. 2003) indicating that repeats of these kinds are under strong selection pressure (Richard and Dujon 1997). Moreover, recent studies have substantially demonstrated that length of SSRs in coding regions might be associated with phenotypic variations. For example, variation in the number of GA repeats in the 5' UTR of waxy genes was correlated with amylase content (Ayer et al. 1997).

2.2.4 Applications of EST-SSR markers

2.2.4.1 Functional Diversity

Characterization of genetic variation within natural populations and among breeding lines is crucial for effective conservation and exploitation of genetic resources for crop improvement programs. DNA-based molecular markers especially SSR markers has been suggested extremely useful for precise and reliable characterization and discrimination of genotypes (Gupta et al. 1996; Mohammadi et al. 2003). Recently the emphasis has been shifted from genomic SSRs to EST-SSRs which belong to the ranscribed region of the genome and may have a role in gene expression or function. Evaluation of the germplasm with SSRs derived from ESTs might enhance the role of genetic markers by assaying variations in transcribed and known-function genes (Eujayl et al.. 2002; Wang et al. 2007) hence reflecting the better relationships between species or varieties.

Assessment of genetic diversity among cultivars has been done for a number of plants using EST-SSR markers (Table 2.3) and has been found useful for elucidating genetic relationships. Although in comparison to genomic SSRs, EST-SSRs displayed less polymorphism in germplasm characterization and genetic diversity studies (Cho et al. 2000; Eujayl et al. 2001; Chabane et al. 2005), attention has been focused on them as they detect the "true genetic diversity" available within or adjacent to the genes (Eujayl et al. 2002; Maestri et al. 2002; Thiel et al. 2003).

Table 2.3: Assessment of genetic diversity using EST-SSR markers in plants (adapted from Varshney et al. 2005a).

S.No.	Plant	No. of genotypes used	No. of EST-SSR markers used	Average PIC	References
1	Wheat	64	137	4.1	Eujayl et al. 2002
		52	20	0.1-0.7	Eujayl et al. 2001
		52	20	0.1-0.7	Gupta et al. 2003
		60	25	0.46	Wang et al. 2006
		75	37	0.41	Fu et al. 2006
2	Barley	38	54	-	Thiel et al. 2003
		7	75	-	Kota et al. 2001
3	Sugarcane	5	21	0.23	Cordeiro et al. 2001
4	Spruce	23	44	-	Rungis et al. 2004
5	Coffee	15 *C. arabica* 8 *C. canephora*	18	0-0.7 0-0.82	Aggarwal et al. 2007
6	Grape		10	-	Scott et al. 2000
7	*Crotelaria*	17 C. *juncea* and 9 other species	58	-	Wang et al. 2006

2.2.4.2 Mapping

Genetic maps are fundamental for understanding the genetic control of plant characters and have been successfully applied in many applications in plant genetics and breeding including gene tagging, map-based cloning, QTL mapping and marker-assisted selection. During the past two decades, the step from the quite limited polymorphism in morphological traits/mutants and isozymes to the high pace of development of molecular markers resulted in availability of molecular maps for many plant species. The first large scale efforts to produce genetic maps was performed mainly using RFLP markers in maize, rice, tomato, Arabidopsis (Helentjaris et al. 1986; Beavis and Grant 1991; Shoemaker et al. 1992). Among the innumerable number of marker technologies, currently codomiant or STMS markers remain a standard for linkage map construction in several crops like rice, wheat, barley, cotton etc. However, due to the species specific nature of genomic derived STMS markers, the swelling EST database has attracted geneticists as an alternative source for marker development as they represent coding regions of the genome. High density genetic maps of gene-based markers provide a resource for trait/gene identification, candidate gene identification, marker saturation at independent trait loci and represent high density loci. In recent years, several EST based/SSR markers have been integrated into framework genetic linkage maps of several plants which are summarized in Table 2.4.

Table 2.4: High density linkage map constructed in plants using STMS markers developed from genomic and/or genic regions (adapted from Varshney et al. 2005a).

S.No.	Plant	No. of gSSR loci mapped	No. of eSSR loci mapped	References
1	Barley	-	339	Pillen et al 2000
		-	76	Thiel et al. 2003
			185	Varshney et al. 2007a
2	Wheat	-	90	Yu et al. 2004
		-	126	Nicot et al 2004
		-	-	Gao et al 2004
		185	65	Torada et al. 2006
3	Cotton	511	95	Han et al. 2004
		-	193	Park et al. 2005
		495	1122	Guo et al. 2007
4	Kiwifruit	-	138	Fraser et al. 2004
5	Strawberry (*Fragaria*)	35	-	Sargent et al. 2007
6	Rice	-	91	Temnykh et al. 2000
7	Rye	-	39	Khlestkina et al. 2004
8	*Medicago*	-		Sledge et al. 2005
9	Soybean	991	24	Song et al. 2004
10	Tall fescue	-	91	Saha et al. 2004
11	White clover	30	335	Barrett et al. 2004
12	Pepper	41	139	Yi et al. 2006
13	Melon	46	16	Gonzalo et al. 2005

2.2.4.3 Transferability and comparative mapping

The field of comparative genomics has evolved as a powerful tool for transferring genetic information from model species to genetically more complex species (Gale and Devos 1998; Paterson et al. 2005). Comparative genetic mapping has been well studied in several plant families, and is best addressed in cereals (Bennetzen et al. 2000; Devos and Gale 2000), crucifers (Lagercrantz et al. 1996), species of nightshade family (e.g. tomato, pepper and potato) (Tanksley et al. 1992; Livingstone et al. 1999) and also recently initiated in legumes (Choi et al. 2004a). Comparative genome analysis has revealed conservation of gene order and gene content among the genomes of closely related species (Gale and Devos et al. 1998) and can greatly facilitate gene discovery and map-based cloning of genes of agronomic importance (Sorrells et al. 2003; Jaiswal et al. 2006). Comparisons of genome sequence of model plants *Arabidopsis*, *Oryza* and *Medicago* to major crop genomes revealed high genome conservation. Recently, Rajesh et al. 2008 established the macrosynteny between chickpea and *M. truncatula* based on 500 Kb genomic sequences.

Initial comparative mapping studies predominantly utilized morphological, isoenzyme and RFLP probes (Weeden et al. 1992) that was superseded by PCR-based codominant marker systems such as trans-specific and ITAP markers recently (Intron targeted amplified polymorphic sequence) (Choi et al. 2004b; Aubert et al. 2006). Sequence analysis of SSR loci indicated high conservation of microsatellite flanking regions across related taxa (Westman and Kresovich 1998; Decroocq et al. 2003; Kuleung et al. 2004). In general, the extent of STMS marker transferability between species depends on the evolutionary rates of microsatellite-flanking sequences as well as on the nature of microsatellites themselves. Several reports of SSR marker transferability across species within a genus (Peakall et al. 1998; Choumane et al. 2000; Eujayl et al. 2004) or across genera (Pandian et al. 2000; Saha et al. 2004) are well illustrated. However, the transferability of genomic SSR markers across the genus borders is quite limited (Peakall et al. 1998). In this context, EST-SSR markers gained popularity over genomic SSRs owing to their association with coding regions, and high conservation between genomes, consequently facilitate their use in comparative mapping (Yu et al. 2004; Varshney et al. 2005b; Sargent et al. 2007; Zhang et al. 2007). The transferability of genic SSRs across species and genera has been reported in several plants which are summarized in Table 2.5.

TH-16433 631.523 Cu578 Ma

Table 2.5: Interspecific and generic transferability of genic SSR markers in plants (adapted from Varshney et al. 2005a).

S.No.	Plant	EST-SSR markers used	Species/ genera used	Percent transferability	References
1	Wheat	64 98	18 wild species and 5 genera 8 genera	83.59 40.68	Gupta et al. 2003 Zhang et al. 2005
2	Barley	165 47	Wheat, rye and rice 5 acc. of *H. bulbosum* and two of wheat and rice	 80 to 40	Varshney et al. 2002 Thiel et al. 2003
3	Grape	10 3 6 8	- 8 species of 4 *Vitaceae* genera 25 species from 5 *Vitaceae* genera 62 genotypes of 46 species	- - 55.0 -	Scott et al. 2000 Rossetto et al. 2002 Arnold et al. 2002 Decrooq et al. 2003
4	Sugarcane		2 species *Erianthus* and *Sorghum*	-	Cordeiro et al. 2001
5	*Medicago*	209	24 genotypes from six species and subspecies pea, fababean and chickpea	- 38	Eujayl et al. 2004 Gutierrez et al. 2005
6	Pine	8 53	3 *Pinus* species 6 Pine species	100 64.6 to 94.2	Liewlaksaneeyanawin et al. 2004 Chagne et al. 2004
7	Tall fescue	145	5 genotypes from *Fescue* species; ryegrass, wheat, rice	82.5 and 71.0	Saha et al. 2004
8	Coffee	25	7 *Coffea* species	75 to 86	Poncet et al. 2006
9	*Actinidia*	20	120 genotypes from 21 species, 5 other races	-	Fraser et al. 2005

2.2.5 Expressed sequence tag polymorphic (ESTP) markers

ESTPs are PCR-based markers, which can detect length and sequence polymorphisms prevalent in the expressed regions of plant genomes. The method involves the designing of primers separated by an amplifiable EST segment, selected randomly, and using these primers for PCR amplification of genomic DNA. DNA polymorphism is generally detected by simple gel electrophoresis or restriction sites, or by examining mobility differences using DGGE or SSCP. Since the ESTP primers target the expressed genes, they are particularly useful for QTL mapping, direct tagging of genes that affect agronomic traits and for comparative mapping studies. ESTP technique was successfully implicated for the first time in loblolly pine. Inherited in a codominant manner, ESTP markers have been exploited mostly in coniferous species such as Pines (Temesgen et al. 2000; 2001), spruce (Schubert et al. 2001) and *Cryptomeria* (Tsumura et al. 1997) for enriching linkage maps and as anchored loci for comparative mapping studies. Later, to increase its efficiency and to reduce members of gene families amplified, one primer was placed within or near the 3′ UTR, and the melting

profile of the expected product was optimized by adding GC clamp to the end of one of the primers.

2.2.6 Potential intron polymorphic (PIP) markers

Introns are noncoding sequences interspersed in genes and are abundant in eukaryotic genomes. While predicting intron–exon structure of the eukaryotic genes, Detusch and Long 1999 concluded that on an average, genes have 3.7 introns of 40–150 nt each. In fruit fly and human genomes, introns constitute 11.0% and 24 % respectively (Venter et al. 2001). In general, the selection pressure in intronic regions is much less than exonic regions attributing to more variations in introns compared to exons (Small et al. 2004). For detecting polymorphisms, length is the most perceptive variation in introns, popular as intron length polymorphism (ILP) and gained momentum as a valuable source for genetic marker development. Therefore, geneticists emphasized towards intron-targeted strategy for primer designing that was anticipated to yield more polymorphism (and therefore more efficient) than conventional EST-PCR based primer strategies.

Lessa (1992) for the first time introduced intron-targeted PCR, in which a non-coding intron was amplified using primers designed from highly conserved exon sequences. This approach, termed as 'Exon-Primed Intron-Crossing (EPIC)-PCR' has been shown to yield substantial variability, mainly from intron length polymorphism, and demonstrated more extensive applications than those generated from non-coding sequences. Subsequently a large number of ILP markers have been generated and successfully used in several population genetic surveys and mapping studies (Côrte-Real et al. 1994; Daguin & Borsa 1999, Bierne te al. 2000). ILP markers sharing the properties of SSR markers including specificity, co-dominance, neutral nature and feasibility of use are more advantageous as they directly reflect variation within genes and thus are more useful for marker-assisted selection. Moreover such markers may shed light on intron evolution and are thus valuable for gene structure prediction. However till date, studies exploiting this newly developed molecular marker are very limited in plants. Only few studies have employed ILP markers in plants (Holland et al. 2001 in *Avena*; Choi et al. 2004b in *Medicago*; Wei et al. 2005 in *Rhododendron*; Wang et al. 2005 in *Oryza*; Panjabi et al. 2008 in *Brassica*).

Obviously, for developing ILP markers, suitable introns are required which is largely restricted to model organisms where whole genome and cDNA/EST sequences are available. Recently the field of comparative genomics has facilitated the design of genome analysis

tools in orphan crops lacking sequence information. The assumption that the exon-intron structures are mostly conserved among homologous genes from different species (Batzoglou et al. 2000) has opened new doors for designing ILP primers even in non-model species where EST sequence are available but genome sequence is not. Based on the genome sequences of *Arabidopsis*, Choi et al. 2004a, Wei et al. 2005 and Panjabi et al. 2008 designed intron-flanking primers in the target species that anneal to the conserved exonic regions and amplify the introns. In addition, both studies have supported that intron vs exon regions are more efficient in polymorphism detection than conventional EST-PCR methods and thereby are a reliable source for mapping of transcribed genes. Similarly Feltus et al. 2006 employed this strategy in Sorghum and *Pennisetum* using *Oryza* genome. Yang et al. 2007 created a web based database platform for developing such markers termed as 'PIP' (Potential intron polymorphic) markers in target plants and also designed PIP markers in 59 plant species (Fig. 2.5).

Furthermore, ILP markers because of conserved nature of exons also facilitate cross-taxon amplification that foster the comparative genomic analysis/transfer of genetic information to and fro i.e. from model to orphan crops. The cross species ILP markers have been applied for understanding the genome conservation among legume crops (Choi et al. 2004a) and recently a GeMprospector program has been developed (Fredslund et al. 2006) for designing cross-species genetic markers.

2.3 Genome analysis in chickpea

2.3.1 cDNA libraries available in chickpea

Over the last 30 years, much research has been focused on improving either Ascochyta blight or Fusarium wilt resistance through traditional and molecular breeding approaches, with little emphasis on genes and pathways of gene regulation controlling these traits. To overcome this gap, recent studies have focused on the use of functional genomics tools to uncover important genes involved in resistance/tolerance to both biotic and abiotic stresses. Subsequently, several cDNA libraries have been generated from different tissues and resistant genotypes of chickpea both against biotic and abiotic stresses for example: Coram and Pang et al. 2005 generated ESTs from stem and leaf of Ascochyta-resistant genotype, Boominathan et al. 2004 reported from drought-stress SSH library and Buhariwalla et al. 2005 generated ESTs from SSH library of root–tissues prepared from two drought tolerant genotypes. However currently there are only 1311 chickpea ESTs in the NCBI EST database,

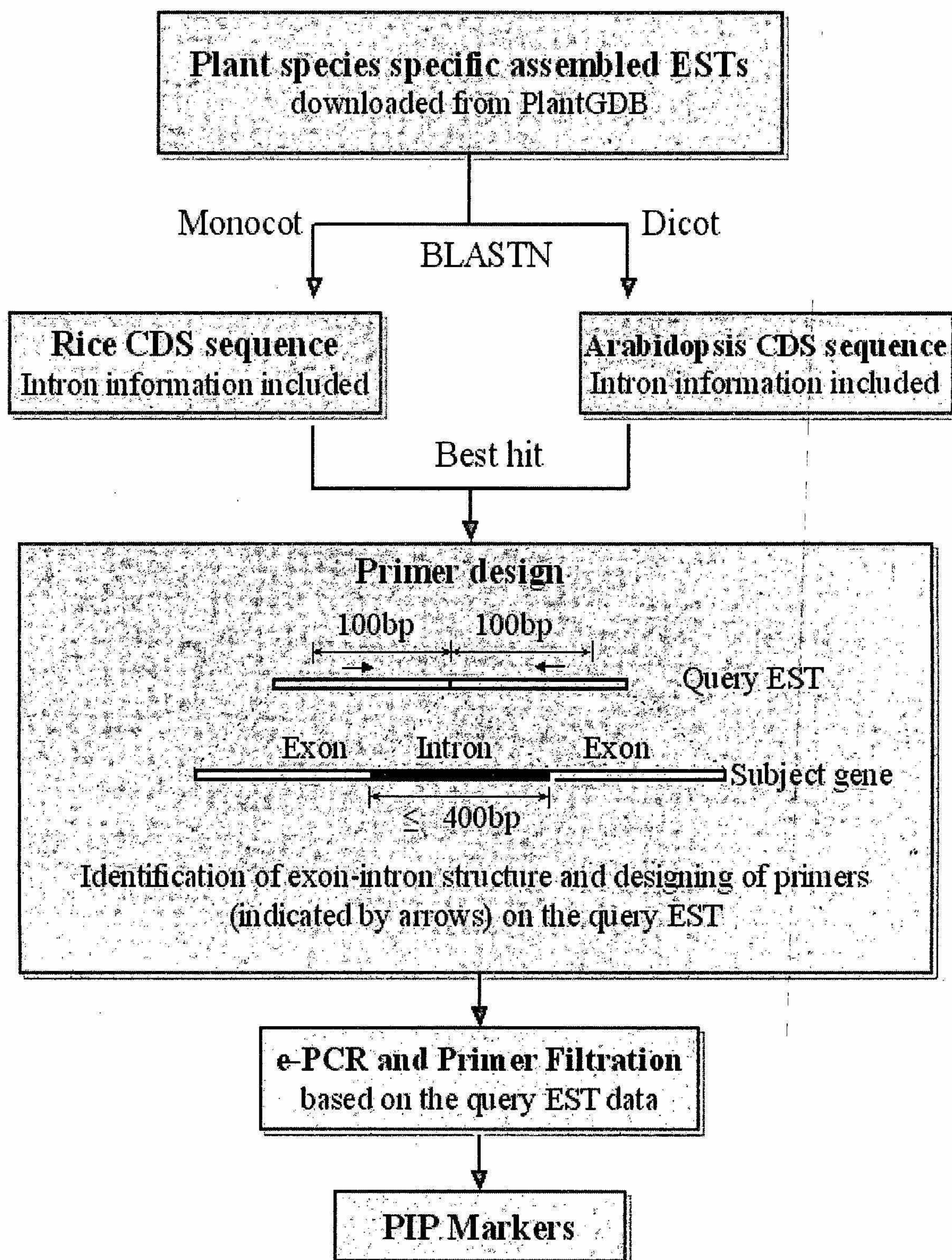


Figure 2.5: Methodology of developing PIP markers from query species (adapted from Yang et al. 2007)

which is quite insignificant when compared with the number of ESTs available from model legumes. Expression profiling of the available chickpea ESTs has been carried out using microarray and SuperSAGE to identify suites of genes responding to particular stresses (Coram and Pang 2005, 2007; Winter et al. 2006; Nimbalkar et al. 2006 and Mantri et al. 2007; Molina et al. 2008), however the application of these resources is still in its infancy.

2.3. 2 Molecular breeding

Genetic bottlenecks imposed during domestication and breeding practices narrowed the gentic base of crops as compared to wild founder species. This holds true especially for obligatory self-pollinating species as chickpea, resulting in highly invariable genomes (Tanksley and McCouch 1997). Preliminary investigations based on morphological, seed protein profile and isoenzyme studies (Muehlbauer and Singh 1987; Gaur and Slinkard 1990 a, b; Kazan et al. 1993), commonly used DNA markers such as restriction fragment length polymorphisms (RFLPs) (Udupa et al. 1993), randomly amplified polymorphic DNAs (RAPDs) (Sharma et al. 1995a) and more recently amplified fragment length polymorphisms (AFLPs) revealed that the degree of polymorphism within *C. arietinum* is very low and there is a need for more efficient and reliable methods for chickpea breeding programs. Oligonucleotides fingerprinting studies i.e. in-gel hybridization of restriction-digested genomic DNA with microsatellite-specific probes conducted by Weising et al. 1992 and Sharma et al. 1995a revealed ample genetic variations at both intra and inter-specific levels in chickpea. In continuation, Sharma et al. (1995a) employed MP-PCR technique using 28 different microsatellites that generated highly reproducible banding patterns and produced inter-specific polymorphisms in *Cicer*. Further, supported by ISSR markers (Chowdhury et al. 2002), chickpea biologists concluded that the tandemly repeated microsatellites are a potent source of molecular markers for chickpea genome analysis.

2.3.2.1 Microsatellite markers in chickpea

2.3.2.1.1 Isolation of SSRs and development of STMS markers

Initially an approach of conventional genomic library construction was utilized for the isolation of microsatellites in chickpea. For the first time Hüttel et al. 1999 screened 13,000 colonies from two small insert libraries (250bp to 400bp, and 400bp to 600bp respectively and covering 0.7% of the chickpea genome) with a set of two di- and eight trinucleotide repeat motifs in an attempt to identify the most abundant microsatellite motifs for the generation of STMS markers in chickpea. Among the probes used by them, (TAA), (GA) and

(GAA) motifs were found to be the most abundant (average spacing 60 Kb) and for the first time Hüttel et al. 1999 generated 22 STMS markers in chickpea. In another major study, Winter et al. 1999 screened a much larger genomic library (280,000 colonies), representing about 18% of the chickpea genome, for the most abundant (GA)n, (GAA)n and (TAA)n microsatellite motifs in the chickpea genome. From the screened clones, a total of 389 positive colonies were sequenced, of which 75% contained perfect repeats and a total of 174 primer pairs produced clear amplification patterns. Another study emanating from our laboratory (Sethy et al. 2003) 5,000 recombinant clones obtained from nuclear genomic library were screened with $(CA)_{10}$ and $(CT)_{10}$ motifs that resulted in development of 10 functional STMS markers.

More recently, Lichtenzveig et al. 2005 screened a large-insert Bacterial Artificial Chromosome (BAC) library from chickpea cv. Hadas for identification of microsatellite sequences. The BAC library consisted of 14,976 clones (average insert size 121kb) and were screened with eight synthetic SSR oligos $(GA)_{10}$, $(GAA)_7$, $(AT)_{10}$, $(TAA)_7$, $(TGA)_7$, $(CA)_{10}$, $(CAA)_7$ and $(CCA)_7$. Out of the 444 independent loci obtained after two phases of SSR isolation, 325 clones contained more than four repeat units from which 233 functional STMS markers were developed. The most abundant motifs were (TAA)n and (GA)n while the (TGA)n motif was the rarest one.

To generate the sufficient number of STMS markers for chickpea and to overcome the time, cost and labor-intensive generation of microsatellites, Sethy et al. 2006a constructed for the first time (in our laboratory), the microsatellite enriched genomic library from the chickpea nuclear genome for the identification of (CA/GT)n and (CT/GA)n motifs. A total of 255 (74 published; Sethy et al. 2006a and remaining under pers. communication) functional STMS markers were developed. With this study, a total of 694 functional STMS markers were available for chickpea genome analysis till date.

In an attempt to develop EST-based markers, recently, Buhariwalla et al. 2005 developed 106 EST-based markers of which only 14 are EST-SSR markers. Thus in chickpea, more research is necessary for developing a large number of genomic as well as EST-based markers to identify genomic regions and genes underlying important plant responses.

2.3.2.2 Utilization of developed chickpea STMS markers

2.3.2.2.1 Genetic diversity studies

Chickpea STMS markers developed through different approaches as mentioned above were found to be polymorphic in chickpea cultivars and applied for various aspects like germplasm evaluation and map construction. For testing pattern of polymorphism of STMS markers, Hüttel et al. 1999 analyzed the developed 22 STMS markers in a test set of one *C. reticulatum* and four *C. arietinum* accessions. For the first time they reported that, all the 16 CaSTMS markers could detect intraspecific polymorphism in chickpea (2-4 alleles). Further they tested two primer pairs CaSTMS10 and CaSTMS15 in 63 *C. arietinum* accessions from different geographic locations and obtained gene diversity values of 0.937 and 0.922, indicating the discriminating power of these markers. Correspondingly, Winter et al.1999 analyzed the 174 functional markers in six chickpea breeding cultivars and its wild relatives *C. reticulatum* and *C. echinospermum* and observed that 137 loci produced polymorphism between the species with at atleast two alleles on polyacrylamide gels.

In another study Udupa et al. 1999 examined the allelic variation of 12 $(TAA)_n$ microsatellite loci in a worldwide collection comprising of 72 landraces, four improved cultivars and two wild varieties (*C. reticulatum* and *C. echinospermum*). They obtained a high number of alleles per locus (14.1) and a high average genetic diversity (0.86) at this locus. Also in terms of microsatellite evolution, they suggested that amount of variation increased with increase in average number of repeats or vice versa. To further gain an insight into the dynamics of microsatellite evolution, Udupa et al. 2004 studied allelic variation at a closely linked (TA)n and (TAA)n microsatellite loci in 114 land races of chickpea sampled worldwide and suggested that the two loci were separated by 27 bp and were under linkage disequilibrium.

Moreover, ICRISAT has developed a genotyping kit for chickpea, which could be used as a reference kit to compare with other genetic diversity studies. ICRISAT has genotyped 9 chickpea accessions (representing the largest genetic diversity based on phenotyping data) with 35 SSR markers (http://www.icrisat.org/gt-bt/Marker_Kits.htm) and is proposed that it compare the allelic composition of these controls (genotype pools) with other genetic diversity studies. Recently using 48 SSR markers, Upadhyaya et al. 2008 formed an ideal set of chickpea germplasm referred as 'genotype-based reference set' of 300 accessions that is useful resource for allele mining, association genetics, mapping and cloning

gene(s), and in applied breeding for the development of broad-based elite breeding lines/cultivars.

2.3.2.2.2 Cross-species amplification of chickpea microsatellite markers

The need for marker-assisted exploitation of primary and secondary gene pool of crops for quality improvement has triggered research aiming at transfer of STMS markers from one species to another. In this context, Choumane et al. 2000 illustrated the potential of chickpea STMS primer pairs for genome analysis of wild *Cicer* species including one accession of a perennial species *C. anatolicum*. The transferability ranged from 92.2% in *C. reticulatum* to 50% for *C. cuneatum*. Based on the amplification patterns, they arrived at two conclusions (i) STMS from chickpea were of limited use as syntenic markers beyond the first crossability group and (ii) Sequence of amplicons derived from species of other crossability groups differed so much from chickpea sequence that they most probably represented different loci.

In another attempt, Pandian et al. 2000 examined the transferability of STMS primers from field pea and chickpea to four other major pulses. The potential transferability of the STMS primer pairs among the species, expressed as the total mean percentage of positive amplification, was 53% for the field pea STMS primers and 9% for chickpea STMS primer pairs. The individual mean percentage of successful amplification of chickpea STMS primer pairs across lentil, vetch and field pea accessions were 5%, 3% and 18% respectively. The transferability of these STMS primer pairs indicates a high level of sequence conservation in these loci across the genera.

Choumane et al. 2004 studied the conservation of chickpea STMS markers in dry pea and lentil. Amplifications of 58.5% and 70.7% were observed in lentil and pea respectively though the size and number of amplicons differed between the three genera. To determine the nature of amplified fragments, 63 selected amplified loci were hybridized to the labelled probe $(TAA)_5$ showing that only 69.8% loci in lentil and 66.6% loci in dry pea hybridized to the probe indicating conservation of homologous genomic sequences. The sequence analysis at loci (Tr7 amplifying chickpea and lentil, Ts35 amplifying chickpea and pea and Ta176 amplifying two alleles in chickpea) revealed low level of conservation of repeat motifs indicating that care should be taken while using the cross-genera STMS markers or confirmed with sequence data.

2.3.2.3 Genetic linkage map of *Cicer* genome

Attention has been focused on mapping the chickpea genome to mark genomic regions related to disease resistance genes (Winter et al. 2000) and other yield related traits (Cho et al. 2002; Rajesh et al. 2002). The former linkage maps in chickpea were based on morphological and isoenzyme markers (Gaur and Slinkard 1990 a, b; Kazan et al. 1993) spanning about 250 cM of the chickpea genome. However these markers bear the disadvantage of ever-changing environmental influences and exhibit extremely low polymorphism within the chickpea cultivars (Kazan and Muehlbauer 1991; Udupa et al. 1993), thus compelling researchers to work on inter-species crosses (*C. arietinum* x *C. reticulatum; C. arietinum* x *C. echinospermum*) rather than intra-specific.

Incorporation of molecular markers like RFLP and RAPD was expanded with the work of Simon and Muehlbauer (1997) that developed the integrated genetic linkage maps of chickpea resulting from three inter-specific mapping populations that covered 550cM. Subsequently, development of SSR markers, an efficient marker system for genetic analysis in plants has accelerated the molecular breeding efforts in chickpea. The characterization of 174 STMS loci from the chickpea genome proved to be informative at an intra-specific level in *Cicer arietinum* (Hüttel et al. 1999; Winter et al. 2000). The suitability of these markers for construction of a genetic map in an interspecific population was well demonstrated (Winter et al. 2000). They mapped a total of 120 STMS markers in a population of 90 RILs from a inter-species cross between the chickpea cultivar ICC4958 and *C. reticulatum* accession PI 489777. The resulting first co-dominant DNA marker map contained 112 markers in 11 linkage groups covering 613 cM. Clustering as well as random distribution of loci were observed. Segregation of 46 markers (39%) that deviated significantly ($P > 0.05$) from the expected 1:1 ratio, and the majority of these loci (73%) were located in 3 distinct genomic regions. This STMS marker map represented the "first landmark map" of the chickpea genome (Winter et al. 1999).

Building on this skeleton map, Winter et al. 2000 further established the integrated linkage map of the chickpea using 130 RILs from the above cross and tagged three loci that confer resistance against *Fusarium* wilt as the parental lines *C. reticulatum* accession PI 489777 is susceptible and ICC4958 is resistant to races 0, 4 and 5 of *Fusarium oxysporum*. At a LOD score of 4.0, 303 markers covered 2077.9 cM in 8 large and 8 small linkage groups at an average distance of 6.8 cM between the markers (Winter et al. 2000). Clustering of markers in central regions of linkage groups was observed.

Since an intra-specific linkage map will be more reliable as compared to inter-specific as former targets traits of breeding interest, chickpea geneticists are focused on generation of an integrated genetic map of chickpea, comprising loci of both economic and scientific importance. In recent years, several intra-specific linkage maps are available for chickpea with various mapping populations (Cho et al. 2002, 2004; Flandez-Galvez et al. 2003a; Udupa and Baum 2003; Cobos et al. 2005; Radhika et al. 2007; Taran et al. 2007) employing the only set of common markers reported by Winter et al. 2003 and lately the markers developed by Lichtenzveig et al. 2005, illustrating that STMS markers are indeed elite anchor markers for merging genetic maps in chickpea rather than RAPD and ISSR markers (Millan et al. 2006). The genetic linkage maps developed to date with DNA based molecular markers in chickpea are summarized in Table 2.6. However, chickpea still necessitates a large number of molecular markers preferentially developed from different sources in order to construct a high coverage genome map for marker assisted selection.

Table 2.6: Genetic linkage maps constructed till date for chickpea (adapted from Varshney et al. 2007b).

Reference	Populations	Markers	Linkage groups	Map Size (cM)
Gaur and Slinkard (1990 a,b)	F_2 populations from inter- and intra- species crosses between five accessions of *C. arietinum*, six accessions of *C. reticulatum* and one accession of *C. echinospermum*	3 morphological 26 isozymes Total: 29	7	~200
Kazan et al. (1993)	Eight F_2 populations from inter-species crosses between two *C. arietinum* varieties (desi and kabuli) and *C. reticulatum* and *C. echinospermum*	5 morphological 23 isozymes Total: 28	8	257
Simon and Muehlbauer (1997)	Three F_2 and one F_3 populations from inter-species crosses between *C. arietinum* x *C. reticulatum*	9 morphological 27 isozymes 10 RFLP 45 RAPD Total: 91	10	550
Winter et al. (1999)	Inter-specific cross of *C. arietinum* (ICC4958) x *C. reticulatum* (P.I.489777) with 131 RILs	120 STMS	11	613
Winter et al. (2000)	Inter-specific cross of *C. arietinum* (ICC4958) x *C. reticulatum* (P.I.489777) with 131 RILs	118 STMS 96 DAF 70 AFLP 37 ISSR 17 RAPD 8 isozymes 3 cDNAs 3 *Fusarium* resistance loci Total: 354	16	2078
Santra et al. (2000)	Inter-species cross of *C. arietinum* (FLIP 84-92C) x *C. reticulatum* (P.I.599072) with 142 RILs	111 RAPD 21 ISSR 11 isozymes 1 morphological 3 QTLs for *Ascochyta* blight resistance Total: 146	9	981.6

Hüttel et al. (2002)	Inter-species cross of *C. arietinum* (ICC4958) x *C. reticulatum* (P.I.489777) with 131 RILs	6 RGA	8	-
Tekeoglu et al. (2002)	Inter-species cross of *C. arietinum* (ICC4958) x *C. reticulatum* (P.I.489777) with 142 RILs	Integration of 55 STMS 1 RGA	8	1174.4
Cho et al. (2002)	Intra-species (*C. arietinum*) cross of ICCV2 x JG62 with 76 RILs	55 STMS 20 RAPD 3 ISSR 2 morphological Total: 80	14	297.5
Udupa and Baum (2003)	Intra-specific RIL arising from ILC1272 x ILC3279	52 STMS 3 *Ascochyta* blight resistance loci Total: 55	8	419
Flandez-Galvez et al. (2003a)	F_2 population arising from intra-specific (*C. arietinum*) cross between ICC12004 x Lasseter	54 STMS 3 ISSR 12 RGA Total: 69	8	534.4
Pfaff and Kahl (2003)	Inter-species cross of *C. arietinum* (ICC4958) x *C. reticulatum* (P.I.489777) with 131 RILs	Integration of 47 gene-specific markers Total: 296	12	2483.3
Collard et al. (2003)	Inter-species cross of *C. arietinum* (Lasseter) x *C. echinospermum* (P.I. 527930) with F_2 population	54 RAPD 14 STMS 9 ISSR 6 RGA 2 QTLs for *Aschochyta* blight resistance Total: 83		570
Cho et al (2004)	Intra-specific cross arising from P.I.359075 x FLIP 84-92C	53 STMS 4 QTLs for *Aschochyta* blight resistance Total: 53	11	318.2
Cobos et al. (2005)	Two intra-specific (*C. arietinum*) crosses arising from CA2139 x JG62 (80 RILs) and CA2156 x JG62 (79 RILs)	118 RAPD 13 STMS 3 ISSR 4 morphological Total: 138	10	427.9
Abbo et al. (2005)	Inter-species cross of *C. arietinum* (Hadas) x *C. reticulatum* (Cr205) with 120 F_2 population	91 STMS 2 CytP450 4 QTLs for beta-carotene conc. 1 QTL for lutein conc. 3 QTLs for seed weight Total: 93	9	344.6
Iruela et al. (2006)	A population of 106 $F_{6:7}$ RILs derived from intra-species (*C. arietinum*) cross of ILC3279 x WR315	17RAPD 1 ISSR 3 STMS 1 morphological 2 QTLs for *Aschochyta* blight resistance Total: 22	-	-
Cobos et al. (2006)	A population of 97 $F_{6:7}$ RILs derived from inter-species cross of *C. arietinum* (ILC72) x *C. reticulatum* (Cr5-10)	16 RAPD 3 ISSR 14 STMS 1 Isozyme 1 Morphological 1 QTL for *Aschochyta* blight resistance Total: 35	10	601.2
Lichtenzveig et al. (2006)	A population of 120 RILs derived from intra-species (*C. arietinum*) cross of Hadas x ICC5810 with F_5 population	233 SSRs 3 QTLs for *D. rabei* resistance 2 QTLs for time to flowering Total: 233	-	-
Radhika et al.	A composite intra-species map of *C. arietinum*	44 RAPD	8	739.6

(2007)	JG62 x Vijay and Vijay x ICC4958 (Combined population size of 186 $F_{8:9}$ RILs)	16 ISSR 165 SSR 2 RGA 1 ASAP 2 morphological 8 QTLs for seed weight Total: 230		
Taran et al. (2007)	A population of 186 F_2 plants derived from intra-species cross ICCV96029 x CDC Frontier	144 SSR 1 Morphological 3 QTLs for *Ascochyta* blight resistance Total: 145	8	1258

2.3.2.4 Mapping of agronomic traits and important genes in chickpea

A) Mapping of resistance genes and resistance gene analogue

Genetic mapping for disease resistance genes mostly focused on tagging agronomically relevant genes such as ascochyta (Tekeoglu et al. 2002; Udupa and Baum, 2003; Collard et al. 2003; Flandez-Galvez et al. 2003b; Cho et al., 2004) and fusarium resistance genes (Benko-Iseppon et al. 2003; Sharma et al. 2004). The genetics of resistance to ascochyta blight has been extensively analyzed but confounding results have aroused because of varied nature of the fungal isolate and cultivar. Depending on the cultivars tested, the isolates of the fungus and the methods of disease screening either qualitative or quantitative resistance have been reported. Till, date, a number of QTL for aschochyta resistance was reported by different research groups (Santra et al. 2000; Millan et al. 2003; Udupa and Baum 2003; Flandez-Galvez et al. 2003b; Collard et al. 2003; Cho et al. 2004; Iruela et al. 2006; Taran et al. 2007). Majorly the QTLs against *A. rabei* are located on LG2 (*Ar1a* Cho et al. 2004) and LG 4 (Udupa and Baum 2003; Cho et al. 2004; Iruela et al. 2006, Taran et al. 2007) but additional secondary QTLs in LG3 and LG6 have been reported (Cho et al. 2004; Taran et al. 2007).

Eight physiological races of *Fusarium oxysporum* f. sp. *ciceris* (0, 1A, 1B/C, 2, 3, 4, 5 and 6) have been reported so far (Haware and Nene, 1982; Jim´enez-D´ıaz et al. 1993) whereas additional races are suspected from India. Mapping efforts to identify QTLs and markers that linked to different races have identified markers linked to six genes governing resistance to six races (0, 1A, 2, 3, 4 and 5) of the pathogen and their position on chickpea linkage maps have been elucidated (Benko-Issepon et al. 2003; Sharma et al. 2004; Cobos et al. 2005). These genes lie in two separate clusters on two different chickpea linkage groups. While the gene for resistance to race 0 (*$foc0_1$* and *$foc0_2$*) is situated on LG 5 flanked by RAPD marker $OPJ20_{600}$ (3 cM apart) and STMS marker TR59 (2 cM apart) of the genetic map of Winter et al. 2000 (Rubio et al 2003; Cobos et al. 2005) those governing resistance to races

1A, 2, 3, 4 and 5 (*foc1* and *foc3*, *foc4* and *foc5*) spanned a region of 8.2 cM on LG 2 flanked by STMS markers GA16 and TA96 (*foc1–foc4* cluster) and TA96 and TA27 (*foc3–foc5* cluster) respectively. Moreover resisatnce genes of both Ascohyta (*ar1* and *ar2a* loci) and Fusarium (*foc* gene clusters) as well as pathogenesis related genes are identified on LG2 and are thus considered as hotspot for pathogen defense (Udupa and Baum 2003; Hüttel et al. 2002; Pfaff and Kahl 2003).The cluster of five resistance genes was further subdivided into two sub clusters of 2.8 cM and 2.0 cM, respectively (Sharma and Muehlbauer 2007).

In an attempt to isolate resistance gene from NBS-LRR regions, Hüttel et al. 2002 first time isolated the 13 resistance gene analogos (RGAs) from the chickpea genome (8 from *C. arietinum* and 5 from *C. reticulatum*) using the degenerate primers approach and mapped them on their reported maps. They obtained clustering of RGAs, their association with a *Fusarium* R-gene cluster and their distribution on four of eight already established linkage groups indicating its usefulness for marker-assisted selection and as a pool for resistance genes of *Cicer*. In another report, Pfaff and Kahl 2003, mapped forty-seven gene-specific primers majorly coding against defense responses on previously constructed core map. They exploited the sequence data of chickpea for designing the gene-specific primers. Recently, Palomino et al. 2008 constructed composite linkage map and mapped six RGAs in different chickpea linkage groups where major QTLs conferring resistance to ascochyta blight and fusarium wilt have been reported.

B) Mapping of yield related traits in chickpea

Besides resistance genes, geneticists also focused on other yield related traits of chickpea like seed mass, seed weight and seed size (Rao et al 1994), double podding (Rajesh et al. 2002), chilling tolerance at flowering (Clarke & Siddique, 2003), flowering time (Or et al. 1999) etc. The positive effect of the gene controlling the double podding trait (s) on yield stability was evaluated on near isogenic lines (NILs) derived from a cross of CA-2156 (single-podded) with JG-62 (double-podded) using STMS, RAPD, and ISSR markers (Rajesh et al. 2002) They linked STMS marker TA80 to double podding gene (recognized by symbol '*s*') at a distance of 4.8cM apart and suggested possible use of this marker for marker-assisted selection.

Moreover, the agronomic traits like flower color (*B/b*), seed coat thickness (*Tt/tt*) and double podding (*Sfl/sfl*) were also studied on the mapping populations derived from two crosses (Kabuli x Desi), with JG62 as a common parent (Cobos et al. 2005). Flower colour

and seed coat thickness were mapped on LG1 and were flanked by STMS marker GAA47. This particular marker is also reported to be linked to anthocyanin pigmentation locus (*p*) (Santra et al. 2000) and it might represent the same locus. The single-/double-podding locus (*Sfl/sfl*) was located on LG9 jointly with the STMS marker TA80 and two RAPD markers $OPS04_{1078}$ and $OPAB03_{1196}$. Lichtenzveig et al. 2006 through single trait analysis detected two QTLs for time to flowering (One on LG1 between the markers H1F022 and GAA40 and other on LG2 between markers H4B09 and H1B06) in an intra-species (*C. arietinum*) cross of kabuli variety Hadas x desi variety ICC5810. In addition, to study the genetics governing seed carotenoids in chickpea, Abbo et al. 2005 identified four QTLs for beta-carotene concentration, one QTL for lutein concentration on LG8B and three QTLs for seed weight in an inter-specific population of chickpea cv. Hadas x *C. reticulatum* (Cr205). These QTLs may assist in improving the nutritional quality of chickpea.

Chapter 3: Materials and Methods

3.1 Materials

3.1.1 Plant Materials

Seeds of chickpea genotypes and the annual wild *Cicer* species were procured from the germplasm unit of International Crops Research Institute for Semi-arid Tropics (ICRISAT) and Indian Agriculture Research Institute (IARI). Accessions of legumes namely blackgram, mungbean, *Trifolium*, pigeonpea and lentil were obtained from National Bureau of Plant Genetic Resources (NBPGR), New Delhi, pea and soybean from Maharana Pratap Agriculture University (MPAU), Udaipur, India and *Medicago* from Australian Medicago Genetic Resource Centre, SARDI, Australia, The germplasm used is listed below along with their source (Tables 3.1, 3.2 and 3.3). All the plant materials were grown at the NIPGR field sites.

Table 3.1: List of *C. arietinum* genotypes used in the present study alongwith the germplasm accession number and source.

S No.	Acc. no./name	Source	S.No.	Acc.no./name	Source
1	ICC15802	Syria	16	ICC15518	Morocco
2	ICC156947	-do-	17	ICC15407	-do-
3	ICC16976	Portugal	**Indian accessions**		
4	ICC7676	-do-	18	ICC12947	Rajasthan
5	ICC16800	-do-	19	ICC10945	-do-
6	ICC16761	-do-	20	ICC15406	Madhya Pradesh
7	ICC12866	Ethiopia	21	ICC13124	-do-
8	ICC12726	-do-	22	ICC506	Andhra Pradesh
9	ICC3485	Jordan	23	ICC283	Bihar
9	ICC6293	Italy	24	ICC5383	-do-
10	ICC3631	Iran	25	ICC791	Punjab
11	ICC16487	Pakistan	26	ICC5477	Uttar Pradesh
12	ICC8195	-do-	27	ICC11121	-do-
13	ICC7272	Algeria	28	JG62	ICRISAT
14	ICC13780	Spain	29	ICCV2	-do-
15	ICC8444	Tunisia	30	Puas362	IARI

Table 3.2: List of annual wild *Cicer* accessions analyzed in this study. Species names and source country are mentioned.

Sl. No.	Accession No.	Species	Source country
01	ICC17121	*C. reticulatum*	Turkey
02	ICC17164	*C. reticulatum*	Turkey
03	ICC17159	*C. echinospermum*	Israel
04	ICC17122	*C. bijugum*	Turkey
05	ICC17125	*C. bijugum*	Turkey
06	ICC17126	*C. pinnatifidum*	Turkey
07	ICC17200	*C. pinnatifidum*	Syria
08	ICC17209	*C. pinnatifidum*	Syria
09	ICC17148	*C. judaicum*	Lebanon
10	ICC17150	*C. judaicum*	Israel

Table 3.3: The list of legume accessions used in this study along with accession number, species and pulse name.

S No.	Acc. no.	Species	Pulse	S No.	Acc. no.	Species	Pulse
1	NRC37[c]	*Glycine max*	Soybean	15	IC337447[d]	*Cajanus cajan*	Pigeonpea
2	MAUS47[c]	*Glycine max*	Soybean	16	IC396014[d]	*Cajanus cajan*	*Pigeonpea*
3	PRATAP[c]	*Glycine max*	Soybean	17	IC342955[d]	*Phaseolus mungo*	Blackgram
4	BRAGG[c]	*Glycine max*	Soybean	18	IC328538[d]	*Phaseolus mungo*	Blackgram
5	IC381277[d]	*Lens esculenta*	Lentil	19	IC397612[d]	*Phaseolus mungo*	Blackgram
6	IC334282[d]	*Lens esculenta*	Lentil	20	IC362567[d]	*Phaseolus mungo*	Blackgram
7	IC384444[d]	*Lens esculenta*	Lentil	21	IC279013[d]	*Pisum sativum*	Field pea
8	IC383609[d]	*Lens esculenta*	Lentil	22	IC356344[d]	*Pisum sativum*	Field pea
9	IC411188[d]	*Trifolium alexandrinum*	Berseem Clover	23	RFP-19[c]	*Pisum sativum*	Field pea
10	IC411189[d]	*Trifolium alexandrinum*	Berseem Clover	24	RFP-18[c]	*Pisum sativum*	Field pea
11	IC508311[d]	*Trifolium alexandrinum*	Berseem Clover	25	SA27783[e]	*Medicago truncatula*	Barrel Medic
12	IC411183[d]	*Trifolium alexandrinum*	Berseem Clover	26	SA11959[e]	*Medicago truncatula*	Barrel Medic
13	IC347150[d]	*Cajanus cajan*	Pigeonpea	27	SA3235[e]	*Medicago truncatula*	Barrel Medic
14	IC339040[d]	*Cajanus cajan*	Pigeonpea	28	SA3780[e]	*Medicago truncatula*	Barrel Medic

3.1.2 The inter-specific Recombinant Inbred Lines (RIL) mapping population of chickpea

The inter-specific mapping population arose from a cross between *C. arietinum* ICC4958 x *C. reticulatum* PI489777 were kindly gifted by Dr. Fred Muehlbauer, Washington State University, USA. Briefly, *C. arietinum* ICC4958, a fusarium wilt resistant chickpea cultivar was crossed with closely related wild annual species *C. reticulatum* PI489777 (fusarium susceptible). For RIL development, F2 plants were propagated to the F7 or F8 using the single-seed descent method. The segregating material consisting of 129 recombinant inbred lines was selected for linkage analysis and map construction.

3.1.3 Tissue Collection

To construct the chickpea cDNA library, developing seeds of cultivar *C. arietinum* ICCV2 that was grown under field conditions were collected at 20 DAA (days after anthesis/flower opening). For northern analysis, pods of chickpea at different stages of development were harvested and stored at -80°C for use when required.

3.1.4 Primers used for mapping in this study

S. No.	Primers/clones used	Source	Designated names
1	272 chickpea genomic STMS markers	Developed in our laboratory (10 by Sethy et al. 2003; 74 by Sethy et al. 2006; 7 by Choudhary et al. 2006 and 181 are available in the thesis of Niroj Sethy, 2007 pers. comm.)	NCPGR series
2	97 chickpea EST-SSR markers	Developed as part of this study (listed in chapter 5; Table 5.1)	CESSRDB from chickpea EST database and CESSR from inhouse generated ESTs
3	15 *Medicago* eSSRs	Gutierrez et al. 2005	MTEST
4	80 ESTPs	Developed as part of this study (listed in chapter 6; Table 6.1)	CEST
5	110 intron based primers	Developed as part of this study (listed in chapter 6; Table 6.2)	PIPs

3.1.5 Vectors and bacterial strains used

Vector/Strains	Material	Source
Plasmids	pGEMT-Easy, pUC19, pTZ57R/T, TopoT/A	Promega, MBI-Fermentas Invitrogen,
Bacterial strains	*Escherichia coli* DH5α	Invitrogen

3.1.6 Chemicals and Materials used

Molecular weight Markers	1Kb ladder, 100bp ladder, 50bp ladder, Low Molecular Weight ladder, pUC19-*Msp*I digest, λ DNA	New England Biolab, Fermentas
Membrane(s)	Hybond N^+, Whatman 3MM paper	Amersham BioSciences, Whatman
X-ray film	Hyperfilm™ MP	Amersham , Kodak
Enzymes	Commonly used restriction enzymes	NEB
	Taq DNA Polymerase	Clontech, NEB
	T4 DNA Ligase	Fermentas, NEB
	RNase	BioBasic, Amersham
Antibiotics	Ampicllin	SIGMA
Dyes	Ethidium Bromide, Xylene cyanol Methylene Blue	Amersham, Sigma
Radioisotope	[α^{32}P] dCTP,	Amersham Biosciences, Perkin Elmer
Disposable filters	PVDF 0.45 μm filter unit	Millipore
Kits used	GenElute DNA extraction kit Qiagen gel extraction kit PCR product purification kit TOPO TA cloning kit SMART-cDNA library lit Montage™ PCR centrifugal filter	Sigma Qiagen Millipore Invitrogen Clontech Millipore
Culture media components	Tryptone Yeast Extract Agar	Difco/ Pronadisa
Locally available	Chloroform, Isopropanol, iso-amyl alcohol, $CaCl_2$, NaCl, NaOH, Glucose, $MgCl_2$, Potassium acetate, Glycerol, Acetic acid, NaH_2PO_4, Na_2HPO_4, $MgSO_4$, HCl, H_2SO_4,	Qualigens, HiMedia and

chemicals	Glycine, LiCl, Sucrose, Pot. Dichromate, Sodium hypochlorite, tri-Sodium citrate, Formaldehyde, Saturated phenol, H_2O_2	Merck
Fine chemicals	CTAB, Acrylamide, Bis-Acrylamide, TEMED, IPTG, Sephadex G-50, EDTA, X-gal, MOPs, Agarose, Metaphor agarose	Amersham, Sigma, Cambrex
Plasticware	Microcentrifuge tubes, micro tips, PCR tubes, reagent bottles, 96 well PCR plates, oak-ridge and falcon tubes	Tarsons, Polylab and Axygen
Glassware	Reagent bottles, flasks, measuring cylinders, trays, beakers, culture tubes	Borosil, Schott Duran

3.2 Methods

3.2.1 General sterilization procedures

The glassware, culture media and autoclavable plastiwares were sterilized by autoclaving at 121^{0}C under 15lb psi pressure for 20 min. The antibiotics and other heat labile components were filter sterilized with dispensable syringe driven PVDF filter unit of 0.22µm pore size (Millex^{TM}, Millipore, USA).

3.2.2 Genomic DNA isolation

For genomic DNA isolation (Doyle and Doyle, 1987), 3gm fresh young leaves were ground to a fine powder in a mortar and pestle using liquid nitrogen. The powdered material was transferred to a 50 ml oakridge tube, suspended in 15 ml of pre-warmed DNA isolation buffer (2% w/v CTAB, 1.4M NaCl, 20mM EDTA, 100mM Tris-Cl, 0.2% β-mercaptoethanol) and incubated at 65^{0}C for 1hr with occasional stirring. The tube was allowed to cool down to room temperature. Following chloroform: isoamyl alcohol (24:1) extraction for 10 min, the aqueous phase was separated by spinning at 8,000 rpm for 20 min at 20^{0}C and transferred to a fresh oakridge tube. For precipitation of DNA, 0.6 volume of ice-cold isopropanol was added and kept overnight at 4^{0}C for enhancing the yield. The DNA was then pelleted by centrifugation at 10,000 rpm for 15 min at 4^{0}C, washed with 70% ethanol, air dried and finally dissolved in 1.0 ml of autoclaved MQ water. To get rid of the RNA content of the extracted genomic DNA, the samples were treated with DNase-free RNaseA (final concentration of 10µg/ml) by incubating at 37^{0}C for 1 hour. The enzyme was removed using an equal amount of chloroform: iso amylalcohol (24:1) and the DNA was precipitated by adding 2 volumes of absolute ethanol, washed with 70% ethanol, dried and dissolved in 200 µl of 0.1 TE buffer (pH 8.0) and stored at -20^{0}C.

3.2.3 Isolation of genomic DNA by GenElute™ Plant Genomic DNA Miniprep kit (SIGMA)

The genomic DNA of mainly wild *Cicer* species or accessions whose leaf materials were limited was chosen for isolation through this method. Fresh green leaves (100mg) were grounded to fine powder in a 1.5 ml microcentrifuge tube with liquid nitrogen. Three hundred and fifty microlitres of lysis solution (part A) and 50 µl of lysis solution (part B) were added to the tube and vortexed to mix contents. The mixture was then incubated at 65^0C for 20 min. To the lysate, 130µl of precipitation solution was added, mixed by inversion and incubated on ice for 5 min. After centrifugation at 12,000rpm for 5 min the supernatant was transferred to the blue filtration column for further separation of debris. Alongside a binding column was prepared by adding 500 µl of column preparation solution and centrifuge at 13000 rpm for 1 min. To the filtrate collected from the blue filtration column, 700 µl of binding solution was added and mixed thoroughly by inversion. From the mix, 700 µl was transferred to the binding column, spinned for 1 min and the flow-through was discarded. The above step was again repeated for the remaining mixture. The column was then transferred to a fresh collection tube pursued by washing twice with 500 µl of wash solution. After a brief spin of 3 min with the wash solution the column was transferred to a new collection tube. To the column 100 µl of pre-warmed (65^0C) elution buffer was added and centrifuged for 1 min to elute DNA.

3.2.4 Quantification of genomic DNA

DNA concentration was checked by agarose gel electrophoresis. All the DNA samples were electrophoresed on 0.8% agarose gels in 1X TBE buffer (pH 8.0) with known concentration of uncut λ DNA. The gel was stained in ethidium bromide solution in a final concentration of 10µg/ml and scanned in a gel documentation system. The concentrations of DNA samples were compared with the uncut λ DNA (25ng/µl) and diluted accordingly.

3.2.5 RNA isolation

Before starting RNA work, mortar, pestle, glassware, spatula, cylinders, flasks and other required materials were baked at 250°C for 5-6 hrs. Gel electrophoresis assembly and other plasticwares were treated with 3% H_2O_2 overnight.

3.2.5.1 Isolation of RNA from Chickpea

Before start grinding, mortar and pestle were chilled with liquid nitrogen and then about 0.8 g of plant tissue (2-4 seeds) was crushed to fine powder without letting it to thaw.

The powdered material was transferred to a 1.5 ml eppendorf tube containing 500 μl each of extraction buffer (200mM Sodium acetate pH - 5.2, 1% SDS, 10mM EDTA pH 8.0) and phenol (saturated with DEPC water pH -7.0). This was vigorously shaken in order to homogenize the sample. The homogenized samples were centrifuged at 14,000xg for 10 min at room temperature and the aqueous layer was collected in a fresh tube. This was extracted twice with phenol: chloroform (1:1) by centrifuging at 14,000g for 10 min at 4°C. The upper phase was transferred to a new tube and 0.3 volume of 10M LiCl was added, mixed and kept at 4°C for 1-4 hrs or overnight for RNA precipitation. Nucleic acids were pelleted by centifugation at 10,000 rpm for 10 min at 4°C. Finally the pellet was washed twice with 2.5M LiCl and once with 70% ethanol by dislodging the pellet from the surface of tube with vigorous shaking and centrifuging at 10,000x g for 10 min at 4°C. The pellet was air-dried for 10 min and dissolved in adequate volume of DEPC-treated autoclaved water or for long term storage, the ethanol washed pellet was suspended in 75% ethanol and kept at -80°C.

3.2.5.2 RNA quantification

The water dissolved RNA was incubated at 55°C for 10 min and quickly chilled on ice. After brief centrifugation, it was collected at the bottom of tube and tapped gently to mix. Two microlitre of the RNA was diluted 500 times by adding 1 ml of DEPC-treated water and mixed thoroughly. The O.D of this diluted RNA was taken at 260 nm spectrophotometer (BIORAD) against DEPC-treated water as blank. Concentration of the RNA was calculated according to the following formula-

$$\text{RNA conc. } (\mu g/\mu l): \frac{O.D_{260} \times 40 \times \text{Dilution factor}}{1000}$$

Purity of the RNA was checked by taking O.D at 260 and 280 nm wavelengths. The RNA was considered as pure if the ratio of O.D (260/280) falls between 1.7-2.0 (<1.7 is typically protein contamination).

3.2.5.3 Denaturing formaldehyde gel for RNA electrophoresis

Total RNA was run in 1.5 % denaturing formaldehyde gel. For preparation of gel, 1.5g agarose was added to 72 ml DEPC treated water and boiled for 2 min. Once the temperature comes down to 55°C, 18 ml formaldehyde (2.2 M or 6%) and 10 ml of 10X MOPS buffer was added. 10X MOPS had the following constituents (0.2M MOPS pH 7, 20mM sodium acetate and 10mM EDTA (pH 8.0). The contents were mixed by swirling taking adequate precautions as formaldehyde is harmful for eyes. The molten gel was poured

in casting tray with combs already fitted into it. Meanwhile, RNA samples to be loaded were prepared by mixing 2.0 µl of RNA (10–20 µg) with 2 µl of 10X MOPS buffer, 4.0 µl of formaldehyde, 10.0 µl of formamide and 1.0 µl of ethidium bromide (10µg/ml). The samples were incubated at 65^{0}C for 15 minutes and then chilled on ice. Further 2 µl of DEPC treated formaldehyde gel loading buffer (50% glycerol, 1mM EDTA pH 8.0, 0.25% bromo phenol blue and 0.25% xylene cyanol) was added. The samples were run at 30-40 Volts for 5-6 hours in 1X MOPS buffer.

3.2.6. Purification of DNA fragments

3.2.6.1 Elution of DNA from gels

a) From agarose gels

The digested product was electrophorsed on 1.2% agarose/EtBr gel and the desired DNA fragment was excised out by using sterile blade and collected in a 1.5 ml micro-centrifuge tube. The gel elution was performed by using MinElute gel extraction kit (Qiagen, Germany) according to the manufacturer's instructions. To the excised gels, three volumes (one volume of gel, 100 mg ~ 100 µl) of buffer QG was added and incubated at 50°C for 10-15 min with occasional vortexing until the gel slice dissolve completely. One gel volume of isopropanol was added to solubilize DNA and mixed properly by inversion. The solution was transferred to the MinElute column which was kept on a 2 ml collection tube and centrifuged at 13,000 rpm for 1 min. After discarding the flow-through and keeping the column in the same collection tube, 500 µl of QG buffer was added and centrifuged for 1min. Again the flow-through was discarded and 750 µl buffer PE was added and centrifuged. The column was again centrifuged for an additional 1 min in order to remove the residual ethanol. For elution of DNA, the column was transferred to a clean 1.5 ml micro-centrifuge tube and 10 µl of elution buffer (100mM TrisCl, pH 8.0) was added and kept for 5 min before spinning at 13,000 rpm for 2 min. The eluted DNA was checked on 1.2% agarose gel alongwith appropriate DNA ladder.

b) From PAGE gels

The PCR amplified products were separated on 6% polyacrylamide gels (acrylamide: bis-acrylamide, 19:1) in 1X TBE for 3-4 hrs at 20 mA and stained with ethidium bromide. The desired portion of the gel was cut with a sharp blade, chopped into small pieces and was transferred to a new 1.5 ml micro-centrifuge tube containing 500 µl of 1X TE buffer (pH 8.0). Then it was kept at 37°C for 4 hr/overnight for elution. After a brief spinned, the

supernatant was transferred to a fresh micro-centrifuge tube. The DNA was precipitated by adding two volumes of absolute ethanol and one-tenth volume of 3M sodium acetate (pH 5.2) followed by spinning at 13,000 rpm for 20 min. Lastly, the DNA pellet was washed twice with 70% ethanol, dried and dissolved in 10 µl of TE (pH 8.0) buffer. The eluted DNA was checked on 1.2% agarose gel alongwith the size marker.

3.2.6.2 Purification of PCR amplified products

Montage™ PCR centrifugal filter devices (Millipore Corp.): The PCR amplified products were purified using this kit according to the manufacturer's instructions. PCR reaction volume was diluted up to 400 µl with autoclaved MQ and transferred to a sample reservoir placed on a fresh 1.5 ml microcentrifuge tube. The assembly was centrifuged at 1,000g for 15 min and after discarding the supernatant, the sample reservoir was placed upright into a new 1.5 ml microcentrifuge tube. Twenty microlitre of distilled water was added to the purple end of the reservoir. The reservoir was inverted and centrifuged at 1,000g for 2 min to obtain purified DNA.

3.2.7. Cloning of purified DNA fragments

3.2.7.1 Ligation

The DNA fragments used in this study were ligated to either pGEM-T Easy or pTZ57R/T vectors according to the manufacturer's instructions. For example the ligation with pGEM-T Easy vectors was carried out in a 10 µl reaction volume containing 2X ligation buffer, vector to insert molar ratio of 1:3 and 1 µl T4 DNA ligase (3 Weiss Units/µl). The reaction mix was incubated at 4^0C for overnight. Similarly the ligation with pTZR57R/T vectors was carried out in 30 µl reaction mix containing 10X reaction buffer, 3 µl of PEG 4000 solution, vector to insert molar ratio of 1:3 and 1 µl of T4 DNA ligase (5U/µl). The ligation reaction was carried out at 22^0C overnight for maximum recombinants.

3.2.7.2 Preparation of Competent Bacterial Cells by $CaCl_2$ method

Competent *E. coli* DH5α bacterial cells were prepared by the $CaCl_2$ method (Sambrook et al. 1989) with minor modifications. For pre-culturing, single colony of bacterial cell was inoculated in 5 ml of LB media and kept for overnight growth at 37^0C with vigorous shaking (200-250rpm). From the O/N grown culture, one ml of inoculum was used to inoculate 100ml LB medium and was grown under similar conditions until the O.D600 reached to a 0.3-0.4. The culture was chilled on ice for 20min, transferred to a 50ml Oak-

ridge tube and bacterial cells were harvested by centrifugation at 5000rpm for 5min at 4^0C in a Sorvall® RC5C plus centrifuge with SA-600 rotor. The cells were suspended by gently swirling the tube in 0.5 volumes (of original culture) of prechilled 100mM $CaCl_2$ followed by incubation on ice for 30min. Again the cells were collected by centrifugation and resuspended in 0.1 volumes ice-cold 100mM $CaCl_2$ as above and placed for O/N incubation at 4^0C-8^0C. The competent cells were either used directly, or for storage, glycerol was added to achieve the final concentration of 15%. One hundred microlitres of the cell suspension was dispensed in 1.5ml micro-centrifuge tubes and snap-frozen in liquid nitrogen. The frozen competent cells were stored at - 80^0C for future use.

3.2.7.3 Transformation

Competent *E. coli* cells were transformed according to the standard protocol of Hanahan (1983). A vial of competent cells, stored at -80^0C, was cautiously thawed on ice avoiding any temperature shock. The ligated product or plasmid was directly added to 100μl competent cell suspension, mixed by gentle tapping and subsequently kept on ice for 30min. All the steps of transformation were carried out in laminar hood under sterile conditions. The cells were then given a heat shock at 42^0C for 90s and quick chilled on ice for 5min. This was followed by addition of 0.9ml of LB and the cells were allowed to grow at 37^0C for 45min with gentle shaking. The transformed competent cells were grown on LB plates containing ampicillin (50mg/ml), X-gal (20mg/ml) and IPTG (200mg/ml) for blue-white selection. The plates were then incubated at 37^0C overnight.

3.2.7.4 Confirmation of inserts by colony PCR

The colony PCR was performed by carrying out the lysis of randomly selected white bacterial colonies at 95^0C for 10min in 10μl of autoclaved MQ water. PCR reaction was carried out using 10 μl of cell lysate, vector specific primers (M13F 5'-AACAGCTATGACCATG-3'/ M13R 5'- TGACCGGCAGCAAATG-3', 1μM each) alongwith 10X PCR buffer, 1.5mM Magnesium chloride, dNTPs mix (2.5mM) and 0.5U Taq DNA polymerase. The amplification profile used was: initial denaturation of 2min at 94^0C followed by 30 cycles of denaturation at 94^0C for 30s, annealing at 58^0C for 30s and extension at 72^0C for 1min and additional extension of 7min at 72^0C. The amplified product was electrophoresed on 1.2% agarose/EtBr gel in 1X TBE with standard size marker to confirm the presence and size of the inserts.

3.2.8 Isolation of plasmid DNA

3.2.8.1 Alkaline lysis miniprep method (Sambrook et al. 1989)

A single colony of bacterial recombinant clone was used to inoculate 5ml of LB medium containing ampicillin (50mg/ml) and allowed to grow overnight at 37^{0}C. The bacterial cells were harvested by centrifugation at 10,000rpm for 1min at RT. The pellet was suspended in 200µl of ice-cold solution I (50mM glucose, 10mM EDTA, 25mM TrisCl) by vortexing. Then, 300µl of freshly prepared solution II (0.2N NaOH,1%SDS) was added, mixed gently by inversion, and incubated for 5min at room temperature. This was followed by addition of 300µl of ice cold solution III (3M potassium acetate, pH 4.8) and was again incubated on ice for 5min. This mixture was then centrifuged at 13,000rpm for 15min at RT and the clear supernatant was transferred to a fresh 1.5ml microcentrifuge tube taking care to avoid any white precipitate. The supernatant was subjected to RNase treatment (20µg/ml) at 37^{0}C for 60min. The supernatant was extracted twice with 400µl of chloroform, followed by separation of upper aqueous phase containing the plasmid. Equal volume of isopropanol was added to precipitate the plasmid DNA by centrifugation at 10,000rpm for 15min at room temperature. This was followed by washing with 70% ethanol. The pellet was dried at 37^{0}C and dissolved in 32µl of sterile water.

3.2.8.2 Alkaline lysis midiprep method

A single colony of bacterial cell containing the desired clone was inoculated to the 100ml of LB medium containing ampicillin (50mg/ml) and allowed to grow overnight at 37^{0}C. The bacterial cells were harvested by centrifugation at 5000rpm for 10min at 4^{0}C. The pellet was resuspended in 5ml of ice-cold solution I (50mM glucose, 10mM EDTA, 25mM TrisCl). Then 5ml of freshly prepared solution II (0.2N NaOH,1%SDS) was added and mixed gently by inversion, and incubated for 5min at room temperature followed by addition of 5ml of ice cold solution III (3M potassium acetate, pH 4.8) and the mixture was incubated on ice for 15min. This mixture was then centrifuged at 14,000 rpm for 30min at 4^{0}C and the supernatant was transferred to a fresh oakridge tube. The supernatant was subjected to RNase treatment (20µg/ml) at 37^{0}C for 45min. The supernatant was extracted twice with phenol: chloroform: isoamyl alcohol (25:24:1) followed by separation of upper aqueous phase containing the plasmid in a fresh Oakridge tube. Equal volume of isopropanol was added to precipitate the DNA by centrifugation at 10,000rpm for 30min at room temperature. This was followed by washing with 70% alcohol. The pellet was dried at 37^{0}C and dissolved in 100µl of sterile water.

3.2.8.3 Purification of Plasmid by PEG Precipitation for Sequencing

For sequencing purpose, plasmid DNA dissolved in 32 µl was purified by adding eight microlitre of 4M NaCl and 40µl of 13% polyethylene glycol (PEG 8000) and the mixture was incubated on ice for 30min. DNA was pelleted by centrifuging at 12,000rpm for 20min at 4^0C. The pellet obtained was washed twice with 70% alcohol, dried and dissolved in 20µl of sterile water. Visual quantification of DNA was done.

3.2.9 Sequencing

One hundred and fifty nanogram of plasmid DNA in 2µl, isolated and purified as mentioned in section 3.2.8.1 and 3.2.8.3 was used for automated sequencing using the Big Dye Terminator kit (Applied Biosystems) on the ABI PRISM 3700 automated DNA sequencer.

For direct sequencing of PCR amplified fragments, 75-150ng of PCR product in 2µl was used with the specific primers at a concentration of 1µM.

3.2.10 Construction of cDNA library

The Clontech's BD SMART^{TM} cDNA synthesis kit used in the present study is an efficient and reliable kit for producing high-quality cDNA even from nanogram quantities of total or poly A+ RNA and is useful in cases where the starting tissue is limited. Figure 3.1 presents a brief methodology of constructing cDNA library as described below:

3.2.10.1 First strand cDNA synthesis

First strand cDNA synthesis was performed using 1.0 µg of total RNA isolated from chickpea 20 DAA developing seeds excluding pod wall. One microlitre each of 3' CDS Primer II A (12 µM) and oligonuclotide (12 µM) was added to 1.0 µg of RNA in microcentrifuge tube and total volume was made to 5.0 µl. The contents were mixed and spinned briefly. After incubation at 72°C for 2 min the tube was cooled on ice and briefly centrifuged. To this, following components were added: 2.0 µl 5X first-strand buffer, 1 µl dNTP mix (10mM each), 1 µl DTT (20mM) and 1 µl Powerscript Reverse Transcriptase (20 units/µl). After mixing and brief spin, the tube was incubated at 42°C in thermal cycler for 1.0 hrs. The reaction was terminated at 72°C for 15 min and then kept on ice. A 2.0 µl aliquot of first strand cDNA synthesis was transferred to a 0.5 ml prechilled microcentrifuge tube for PCR amplification or stored at -20°C for future use.

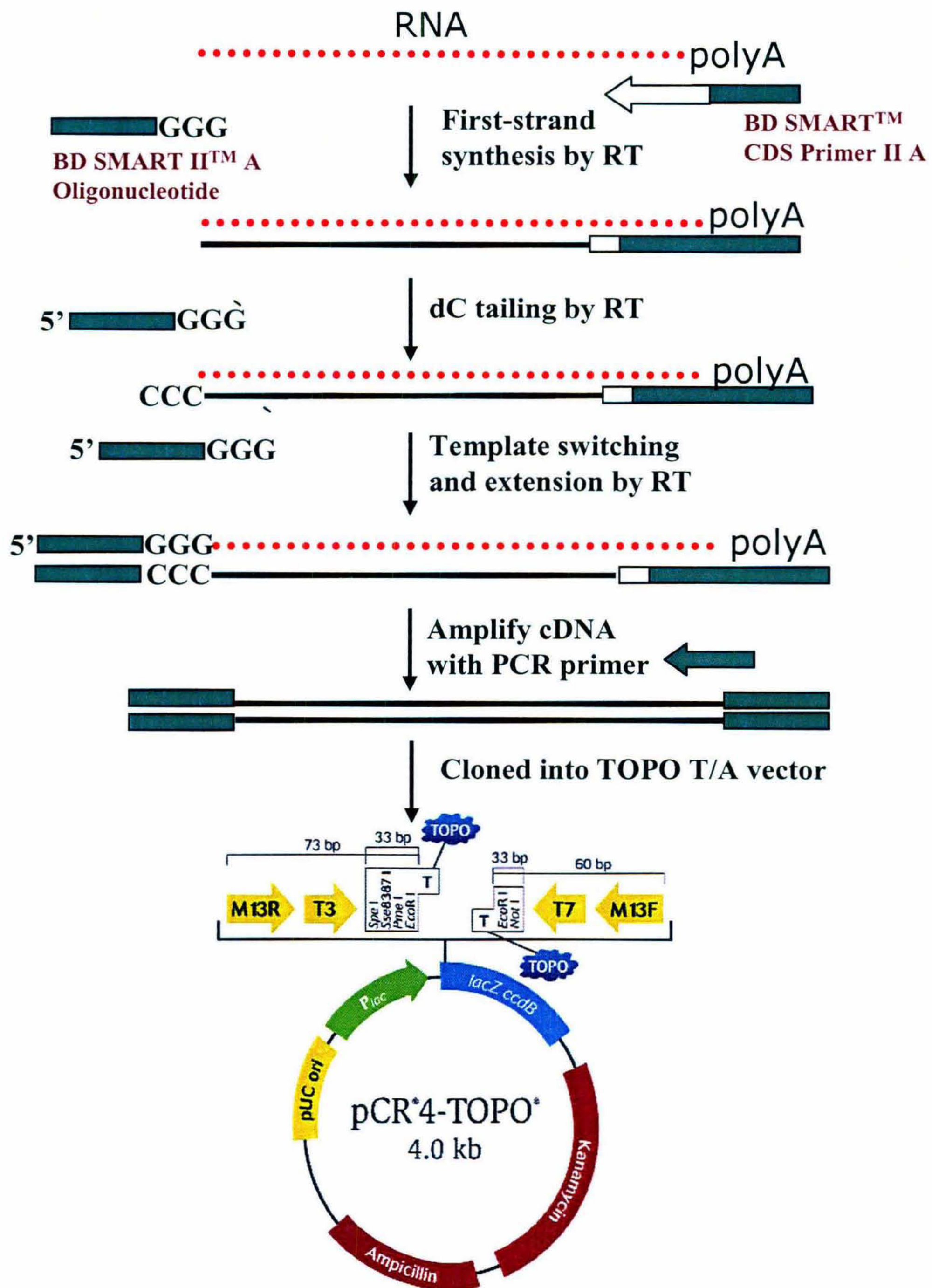


Figure 3.1: Flow chart of the cDNA cloning (adapted from BD SMART cDNA synthesis kit).

3.2.10.2 Second-strand synthesis

Before proceeding to cDNA amplification, the thermal cycler was preheated to 95°C. The master mix of 98.0 μl volume containing the following components was prepared, mixed and briefly centrifuged: 80.0 μl sterile water, 10.0 μl 10X second-strand buffer, 2.0 μl 50X dNTP mix (10mM each) and 4.0 μl 5' PCR Primer II A and 2.0 μl 50X Polymerase mix. The aliquot of master mix was added to all the first-strand synthesis reaction tubes (2.0 μl in each) and mixed properly by gentle flicking and short spin. The tubes were placed in a preheated thermal cycler and following program was used for amplification: 95°C for 1 min, denaturation at 95°C for 50 sec, annealing at 65°C for 30 sec and extension at 68°C for 6 min. Initially the 12 cycle was set for amplification and 5.0 μl of amplified product was electrophoresed on 1% agarose/EtBr gel in 1X TAE buffer along with 1.0 Kb ladder (0.1 μg) as a control. The size range of the ds cDNA was compared to control and depending on the visual intensity of the smear, 2-3 additional cycles could be performed.

3.2.10.3 Double-strand cDNA polishing

To make the double-stranded cDNA blunt-ended, the following steps were performed. Two microlitre of Proteinase K was added to 50 μl (2-5 μg) amplified ds cDNA and the contents was mixed and spin briefly. The tube was then incubated at 45°C for 1 hr in a thermal cycler. After short spin, the tube was heated at 90°C for 10 min to inactivate the enzyme. The tube was placed on ice for 2 min and kept for incubation at 16°C (water bath) for 30 min after adding 3 μl (15 units) of T4 DNA Polymerse (NEB). Then the tube was heated at 72°C for 10 min to terminate the reaction. For DNA precipitation, twenty-eight microlitres of 4M NH_4OAc and 210 μl of 95% ethanol was added, vortexed thoroughly and centrifuged at 14,000 rpm for 20 min at room temperature. The supernatant was removed carefully and 500 μl of 80% ethanol was added. The tube was centrifuged at 14,000 rpm for 10 min and the supernatant was removed. The pellet was air dried for about 10 min to evaporate residual ethanol and dissolved in 7.0 μl of sterile H_2O.

3.2.10.4 Ligation and transformation

Since the cohesive vector was selected for cloning, it was necessary to make polished ds cDNA having overhang A/T sequence. For this, the following components were added to the final polished product to make the reaction volume of 10.0 μl: 1X PCR buffer, 1.0 μl dATP (1mM) and 1.0 μl Taq Polymerase (5U) and placed at 72°C for 25 min in a thermal cycler. Using the TOPO T/A cloning kit (Invitrogen), a 2.0 μl aliquot of the tailed product

was ligated into pCR 4- TOPO vector according to manufacturer's instructions. Briefly, 2.0 μl of the ds cDNA was mixed with 1.0 μl of salt solution and 1.0 μl of TOPO T/A vector and the total volume was made to 6.0 μl. The reaction mix was incubated at RT for 30 min and then transferred on ice. The ligated product was used to transform the chemically competent *E.coli* DH5α cells supplied by the above kit. In this, 6.0 μl of the ligated product was added to 50 μl of competent cells suspension and was incubated on ice for 30 min followed by a heat shock at 42^0C for 30 s. After chilling on ice, 250 μl of the SOC media was added. The cells were grown at 37^0C for one hr, after which 125 μl of the transformed product was plated onto IXA (IPTG, X-gal and Ampicillin) selective plates.

3.2.10.5 Selection of positive clones and EST sequencing

Of the blue and white colonies obtained on the IXA plates, white colonies were picked up and gridded on fresh plates. The putative recombinants were identified through colony PCR (sections 3.2.7.4) and the 2 μl amplified product of each positive clone was sequenced using universal primers either in 5' or 3' direction (section 3.2.9).

All sequence data were submitted to GenBank dbEST database to obtain the accession numbers (ES544474-ES544489, EX151623-EX152143, EX567533-EX567971 and EY457878-EY457905).

3.2.11 EST processing

3.2.11.1 EST preprocessing and contig assembly

The adapter sequences were removed manually and vector sequences were trimmed using the VecScreen program of NCBI. The sequences were maintained in a fasta format word files and sequences <150bp were discarded. To reduce the redundancy, the EST sequences were clustered using the CAP3 program (Huang and Madan 1999) and assembled into contigs and singletons collectively known as unigenes.

3.2.11.2 Functional annotation

The generated ESTs were compared against the sequences in the non-redundant protein database (nr) at the NCBI using the BLASTX program. Default parameters of the program were used and the expectation value (e-value) cutoff was set at 1e-10 for sequence similarity searches. Based on putative functions, ESTs were classified according to the Clusters of Orthologous Groups of Proteins – KOG available at NCBI (http://www.ncbi.nih.gov/cog/new/shokog.cgi). For further annotation and classification of the unigenes, Gene Ontology (GO) system available at the TAIR site (Ashburner et al. 2000)

was employed. For this, the chickpea unigenes were blasted against the proteome of *Arabidopsis* from TAIR (Release 7). All *Arabidopsis* hits with an expectation value of 1e-5 or better were submitted to the GO annotation search tool (Berardini et al. 2004) and relative frequencies of gene counts assigned to the different GO functional classes were displayed (http://www.arabidopsis.org/) as pie charts using Microsoft Excel.

3.2.12 Mining of microsatellite motifs from chickpea ESTs and designing of STMS primers

The chickpea EST sequences were searched for repeat motifs using the softwares like TROLL (Tandem Repeat Occurrence Locator; Castelo et al. 2002) or TRF (Tandem Repeat Finder) considering dinucleotides of $\geq$5 and trinucleotides of $\geq$4 repeats. The identified microstellites were classified on the basis of type, nature and number of motifs. Primers were designed against the microsatellite flanking regions using the Primer 3.0 program (Rozen and Skaletsky 1997) with the parameter set as: primer length of 20-24 bp with optima of 22bp, primer T_M = 50-65^{0}C with an optimal of 55^{0}C, 35-50% GC content with optima of 40% and 200-500bp amplified product size.

3.2.13 PCR amplification

PCR amplifications were carried out in 10µl reaction volume containing 1X PCR buffer, 0.2mM of each dNTPs, 0.5µM of each primer, 25ng of genomic DNA and 0.5 units of Taq DNA polymerase (Titanium, Clontech). Reactions were carried out in an Icycler (BIORAD Laboratories, USA) thermal cycler using the touch-down amplification profile: an initial denaturation at 94^{0}C for 2 min followed by 18 cycles of 94^{0}C for 20s, 65^{0}C for 50s with 0.5^{0}C decrease in each subsequent cycle, 72^{0}C for 50s. This was followed by 20 cycles of 94^{0}C for 20s, 55^{0}C for 50s, 72^{0}C for 50s and final extension of 7 min at 72^{0}C. The primers that do not amplify in this PCR conditions were again re-amplified with a touchdown profile of 60^{0}C to 50^{0}C with all the other conditions keeping constant as above. PCR products were electrophoresed on either 3% Metaphor agarose gels or 6% or 8% polyacrylamide agarose gels depending on the resolution pattern along with size markers and stained with ethidium bromide.

3.2.14 Preparation of agarose gel and running conditions

3.2.14.1 Metaphor agarose gel

Three percent Metaphor (Cambrex, East Rutherford, N.J) agarose gels containing 0.15µg ethidium bromide/ml were used to separate PCR amplification products. The gel was

prepared according to manufacturer's instructions with slight modifications. Briefly for 3% metaphor agarose gel, metaphor agarose and agarose were taken in 3:1 ratio and added to prechilled IX TBE buffer. Care was taken to avoid the formation of agarose clumps in the buffer and mixed well. After the addition, the metaphor agarose was allowed to swell by incubating the mixture at 4^0C for 1-1.5 hr. The resulting solution was weighed and boiled in a microwave for 2min. The conical flask was swirled in order to dissolve the agarose properly. After complete dissolution the flask was weighed again and the distilled water was added to make up the weight loss. The solution was cooled down to 55^0C, and gel was casted after adding the EtBr (0.15μg/ml). The PCR products were mixed with the tracking dye, loaded on gel and electrophoresed at 6V/cm for 2 hrs in 1X TBE.

3.2.14.2 Polyacrylamide gel

Six or eight percent polyacrylamide gels were prepared with 30% Acrylamide: bisacyrlamide (29:1) dissolved in autoclaved MQ water. To make 8% PAGE gels, following mix of 150 ml was prepared: 40 ml of acrylamide:bisacrylamide (30%; 29:1) solution, 30 ml of 5X TBE, 150μl of TEMED, 300μl of 10% (w/v) ammonium persulfate and 70 ml of RO water was added to make up the final volume. The resulting solution was mixed well and poured onto assembled glass plates. After insertion of comb, the gel was allowed to polymerize for 30-60 min. and fitted onto the electrophoresis tank. Both the lower and upper tank was filled with 1X TBE buffer but in the lower tank EtBr was added to a final concentration of 10μg/ml. The amplified products were loaded on the gel, electrophoresed for 3-4 hrs at 150 mA and visualized under UV transilluminator.

3.2.15 Development of EST based markers

3.2.15.1 Expressed sequence tag polymorphic (ESTP) markers

Primers were designed from the randomly selected (any non-specific) regions of the EST sequences using the Primer 3 Program (mentioned in section 3.2.12). Amplification of primers was carried out using the same touch down protocol described in section 3.2.13 and was resolved on 8% PAGE gels as mentioned above.

3.2.15.2 Potential intron polymorphic (PIP) markers

The unigenes from chickpea were searched for designing of intron-targetted primers using the PIP program (Yang et al. 2007). In brief, the program identifies the possible intron positions in the query species (chickpea in the present case) by aligning the query EST sequences with the the subject CDS using BLASTN (note: monocot and dicot plants were

compared with rice and *Arabidopsis*, respectively) at a threshold of 80% similarity and atleast of 200bp overlapping. In the next step, the program designed intron flanking exon-exon primers on a 200 bp sequence cut from the query EST with 100 bp on each side of the target intron. The output window will provide the following information: query EST sequences; possible intron positions in the query sequences; intron lengths in the subject species; primer pairs bracketing single introns; primer positions in query sequences and sizes of PCR products without introns in query species. For amplification of PIP markers, the chickpea genomic DNA was amplified using the same touch down protocol described in section 3.2.13 and were resolved on 1.2% agarose gels for validation and on 8% PAGE gels for polymorphism analysis (described in section 3.2.14.2).

3.2.16 Data analysis for genetic diversity studies

For each microsatellite locus, alleles were scored across all the accessions both manually and with the help of the gel documentation system (Alpha Innotech Corp., USA) in a binary matrix where '1' represented the presence of a band and '0' the absence of a band. Genetic parameters like alleles per locus, effective alleles per locus (Ne), observed heterozygosity (H_O), expected heterozygosity ($He = 1 - \Sigma p_i^2$, where p_i is the frequency of the *i*th allele), Fixation index (Fis) and Shannon's information index (I) were calculated using POPGENE Version 1.32 (Yeh and Boyle 1997). Pairwise genetic similarity was calculated among the 30 accessions and six annual *Cicer* species. The similarity matrix (D = 1-S) was used for constructing the dendrogram using the UPGMA (Unweighted Pair Group Method with Arithmetic Averages) algorithm on NTSYS-pc (version 2.1, Rohlf 1998) software.

3.2.17 Sequence alignments

For each individual fragment six random clones were sequenced using both M13 forward and reverse primer and the consensus sequences were retained. The multiple sequence alignments along with the originally cloned allele were performed either using CLUSTALW (1.83) or MAFFT (version 5.667).

3.2.18 Mapping

3.2.18.1 Genotyping using various markers

All the primers including genomic as well as genic-derived STMS and EST-based (developed in the present study as well as reported primers) that exhibited polymorphism between the parental lines of mapping population *C. arietinum* ICC4958 and *C. reticulatum* PI489777 were selected for genotyping in the 129 RILs. The PCR amplification was carried

out using the protocol and conditions mentioned in 3.2.13. The amplified products were analyzed on 3% Metaphor agarose gels or 6% or 8% PAGE gels along with standard size markers and visualized using EtBr staining (described in section 3.2.14).

Moreover, EST-SSR/ESTP primers amplifying larger size products and showing polymorphism were sequenced directly (as mentioned in section 3.2.9) to allow new primers to be designed to amplify smaller products. The obtained genomic sequence was aligned with its mRNA sequence using the Splign program of NCBI and intronic sequences were identified. New primer pairs were then designed using the primer 3.0 program based on either intron- intron or intron-exon sequences to amplify smaller sized fragments and polymorphism was detected and scored as described above.

3.2.18.2 Linkage analysis and map construction

The amplified banding patterns were scored manually as 'A' for *C. arietinum* ICC4958 type banding pattern, 'B' for *C. reticulatum* PI489777 type banding pattern and heterozygous loci were scored as missing data and we did not consider them for mapping. The data matrix was used as an input file for map construction using JOINMAP® 4.0 program (van Ooijen 2006).

The chi square test was performed for identification of markers with aberrant segregation using the locus genotyping frequencies of JOINMAP. To identify linkage groups, grouping of markers were performed using the minimum independence LOD threshold of 2 and a maximum of 5.0 with a step up of 0.5. The groups showing maximum number of markers and highest linkage at the variable LODs were selected (maximum at LOD3). The groups were converted to maps with the help of the regression algorithm with the following settings: used linkages with REC smaller than: 0.49, LOD larger than: 0.01, threshold for removal of loci with respect to jumps in goodness-of-fit: 5.000, number of added loci after which to perform a ripple: 2 and Kosambi's mapping function.

3.2.19 Northern Hybridization

3.2.19.1 Transfer of total RNA on Nylon Membrane

A 20µg total RNA sample from different stages of seed development was loaded on denaturing gel (as mentioned in section 3.2.5.3) and electrophoresed at 40V for 4-6 hrs till one-third of the gel. After taking the photograph in Geldoc system, the gel was rinsed with DEPC treated water for 30 min to remove formaldehyde and it was equilibrated with 20X SSC for 10 min. The RNA was transferred to Hybond–N^+ Nylon membrane (Amersham,

UK) by vertical capillary action using 20X SSC for 16 h. After that the RNA was cross-linked to the nylon membrane in UV crosslinker (Stratagene, USA) at 1200kJ/cm^2 and this RNA cross-linked membrane was treated with 5% glacial acetic acid for 15 min. To check the RNA transfer on the membrane, it was stained with 0.04% methylene blue (Solution prepared in 0.5 M Na-acetate, pH 5.2. Excess of the stain on the membrane was removed by washing with sterile MQ water. Image of ribosomal RNA was captured on Fluor-S™ MultiImager (Bio-Rad, USA) at highest resolution available to show equal loading of RNA. The hybridized nylon membrane was wrapped in a saran wrap to avoid it from drying.

3.2.19.2 Restriction digestion of Plasmid DNA

Digestion reaction was carried out in 25μl reaction volume containing plasmid DNA (5-10μg), 1X reaction buffer and 20U restriction enzyme (NEB). The reaction mix was incubated for 4-6 hrs at 37^0C. The digested products were electrophoresed on 1.2% agarose/EtBr gel in 1X TBE along with standard size marker to confirm the size of insert. The DNA fragments were eluted out from the agarose gel as mentioned in section 3.2.6.1.

3.2.19.3 Radioactive probe preparation and purification

The probe was prepared using random primers labeling NEBlot® kit (Amersham Biosciences). For probe preparation, in 1.5 ml micro-centrifuge tube 30 ng of eluted DNA (fragment to be used as probe) was taken in final volume of 10 μl. The dsDNA was denatured for 5 min in boiling water bath and quickly chilled on ice. For 50 μl reaction, the following components were added in the order- 27 μl of MQ H_20, 5.0 μl of 10X labeling Buffer, 2.0 μl of dATP, 2.0 μl of dGTP, 2.0 μl of dTTP, 1.0 μl of radioactive α^{32}P-dCTP (3000 Ci/mmole, Perkin Elmer) and 5 units of Klenow polymerase enzyme. The final mixture was incubated at 37°C for one hour in water bath.

For purification of free radioactive dNTPs from the mixture, Sephadex G-50 column was prepared in a 0.5ml microcentrifuge tube by creating a hole at the bottom and that was plugged with the glasswool. These tube was placed in a 1.5ml microcentrifuge tube and packed with Sephadex G-50 beads equilibrated in 1x TAE (pH 8.0) by centrifugation repeatedly at 1000g for 1min 3-4 times. The packed column was then transferred to a fresh eppendorf tube and the 50μl reaction mix (labeled oligonucleotide) was loaded on the packed column and centrifuged at 3000g for 1min to elute the purified probe.

3.2.19.4 Hybridization, Washing and Autoradiography

The spotted nylon membranes were pre-hybridized for 4 hrs at 60° C in a pre-hybridiztion buffer (0.1M Sodium phosphate buffer pH 7.2, 10% SDS and 0.5M EDTA). The purified probe was denatured for 10 min in boiling water bath and quick chilled for 5 min. After a brief spin, the probe was added directly to the pre-hybridization solution kept in hybridization bottle. The probe was left for hybridization for 14-16 hr at 60°C in hybridization incubator. Hybridized nylon membranes were washed twice with 2X SSC and 1% (w/v) SDS for 15 min at hybridization temperature followed by low stringent condition with 1X SSC and 1% (w/v) SDS at RT for 10 min. The filters were then wrapped in saran wrap to avoid drying and the X-ray film was exposed to the membrane in the Hypercassette^{TM} (Amersham Pharmacia biotech, U.K) and autoradiographed at -80^{0}C for 24-48 hrs depending upon the signal intensity. Subsequently, the X-ray film was developed using Developer and Fixer solutions (Kodak Affiliate Products, India).

Chapter 4: Generation and analysis of chickpea ESTs

4.1 Introduction

Large-scale sequencing of cDNAs to produce Expressed Sequence Tags (ESTs) has become the method of choice for the rapid and cost-effective generation of data revealing the coding capacity of genomes. Consequently, several EST sequencing projects are underway or are being initiated for numerous organisms, generating millions of short, single-pass nucleotide sequence reads which are fast accumulating in public databases (http://www.ncbi.nlm.nih.gov/dbEST). These large-scale EST data provides an extensive reservoir for discovery of new genes, transcript profiling, and generation of molecular markers for map-based cloning. Referred as 'poor man's genome' this resource forms the core foundation for various genome-scale experiments within the 'as yet unsequenceable' genomes. In a variety of plants, a large number of ESTs have been generated from different developmental stages and in response to a variety of environmental conditions to identify expressed genes. Concomitantly, extensive computational strategies have been developed to organize and analyze these data for understanding the complex plant genetic mechanisms and interactions. Besides being a powerful tool for functional genomic studies, ESTs have served as expansive resources for generation of functional molecular markers that are useful for construction and saturation of linkage maps that define transcribed regions of genomes which will prove directly useful in marker assisted selection and map-based cloning of economically important traits.

In chickpea, functional genomics studies have been restricted due to the lack of sufficient public-domain genomic resources such as ESTs. Recently efforts in this direction have started with the major aim of understanding the molecular mechanism of resistance/tolerance to major biotic and abiotic stresses. The first endeavor of large scale development of ESTs from a chickpea Ascochyta-resistant genotype was undertaken by Coram and Pang 2005 and simultaneously various other cDNA libraries were also made (Boominathan et al. 2004; Romo et al. 2004; Buhariwalla et al. 2005) under different conditions. In spite of this, currently there are only 1311 chickpea ESTs available in the public database, which is reasonably inadequate when compared to the other legume crops since of the 986,000 Fabaceae ESTs, 92% are derived only from *M. truncatula*, *L. japonicus* and *G. max* (Ramírez et al. 2005). Thus there is a pressing need in chickpea to expand the EST database for understanding the complex agronomic traits.

In the present chapter, special emphasis was given to sequencing of cDNAs from developing seeds of chickpea since the long term aim of the laboratory was to capture the

transcriptome associated with the key stages of seed development through EST resources. Therefore, for the first time, seed related ESTs were generated in chickpea not only for the functional dissection of gene expression during seed development but also to generate EST based molecular markers for mapping seed related traits. Further, the generated ESTs were assembled and functionally annotated using the bioinformatics tools.

4.2 Results

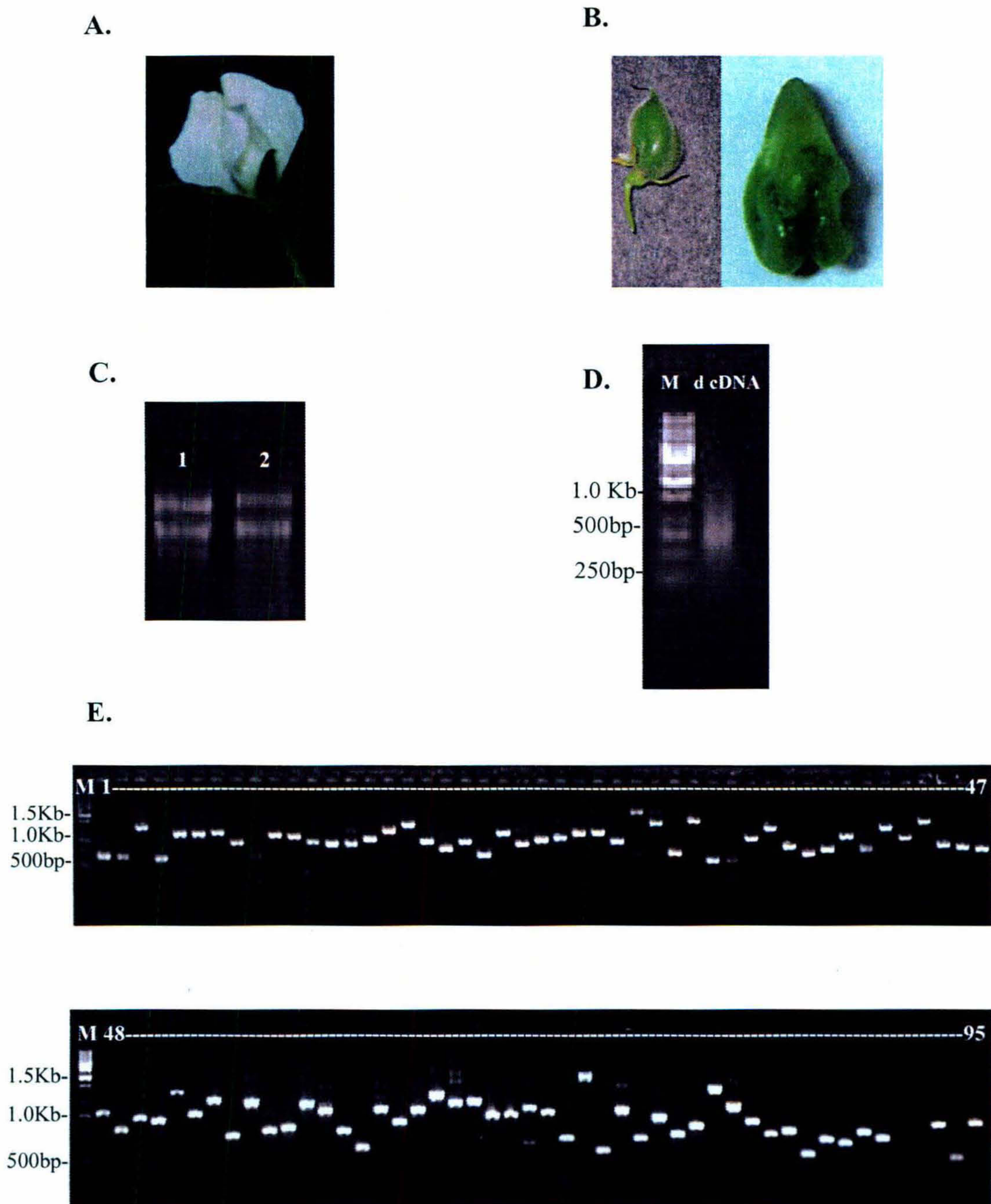
4.2.1 Construction of a cDNA library from developing seeds of chickpea

To generate chickpea ESTs, a cDNA library from chickpea seeds (20 DAA) was constructed as described in section 3.2.10. In brief, total RNA was isolated from 20 DAA chickpea seeds excluding pod wall (Fig. 4.1) and was used to synthesize double-stranded cDNA (Fig. 4.1). The DNA fragments were polished, tailed with 'A' nucleotide and cloned in TOPO T/A vector (Invitrogen). A total of 2760 white recombinant colonies were obtained after blue white selection. Insert amplification was obtained in a total of 2123 recombinant clones (inserts ≥ 200 bp were selected) with an insert size ranging between 300 bp to 3.0 kb, with an average of 700 bp (Fig. 4.1 E). All the recombinant clones containing inserts were sequenced using M13 Forward/Reverse primers. Finally 1897 EST sequences were obtained after trimming off vector sequences and ignoring insert sequences shorter than 150 bp. Further, the 1897 ESTs were assembled into contigs using the CAP3 program that gave a total of 1037 unigenes, consisting of 307 contigs (encompassing 1167 ESTs) and 730 singletons (70.3%). Nine contigs had 10-20 ESTs, five contigs had ≥ 20 ESTs while the largest contigs contained 49 ESTs. The average read-length of these unigenes was 350 bp (Table 4.1). Overall, the redundancy level of EST collection was 61.5% (1167/1897) and the frequency of EST distribution after clustering is shown in Fig. 4.2. A summary of the EST library is shown in Table 4.1 All the unigenes (1037) were submitted to the dbEST and their accession numbers were obtained (ES544474-ES544489, EX151623-EX152143, EX567533-EX567971, EY457878-EY457905).

4.2.2 Annotation and functional classification of unigenes

To assign the putative function, all the 1037 unique chickpea sequences were annotated using the BLASTX program of NCBI. 819 ESTs showed significant homology above the cut off value E-10 and were analyzed after manual screening. Of the 819 ESTs listed in Table 4.2, 58.6% (479) showed significant homology to previously identified genes, 21.80% to unknown/hypothetical proteins and 19.60% showed no homology. Unknown

Figure 4.1: Construction of cDNA library of chickpea cv. ICCV2. **A)** Open flower **B)** 20 DAA (Days after Anthesis) developing seed **C)** Total RNA preps of chickpea seed without pod wall **D)** Double-stranded cDNA between 300bp to 1.0 kb, M: 1.0 Kb ladder and **E)** Lanes 1-95: Colony PCR products showing the size of inserts, M: 1.0 kb ladder



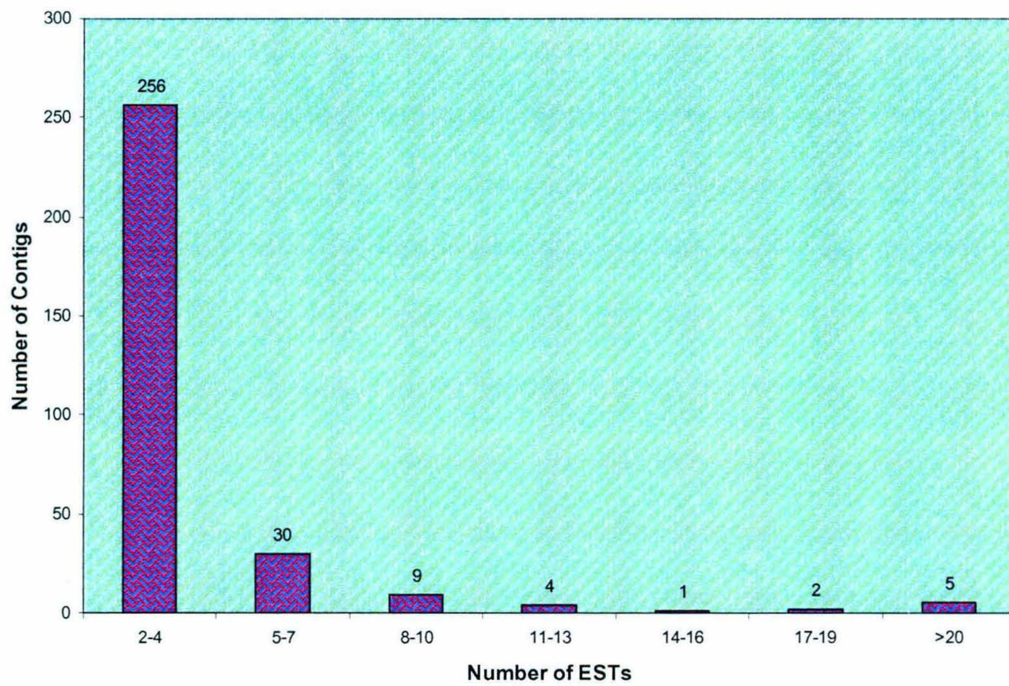


Figure 4.2: Distribution and number of EST members in a contig

proteins refer to that subset of the ESTs that show a significant similarity (high e-value) to genes reported in the public database but do not have an assigned function till date. A total of 0.7% (73/1037) of the ESTs had significant amino acid sequence match to *Medicago* whereas only 0.2% were found to be similar to previously identified chickpea genes. It was observed that the highly abundant ESTs assembled in the contigs comprising of >10 ESTs are those of putative lipid transfer proteins, proteinase inhibitors, seed-specific protein, Chlorophyll-a/b binding proteins, MAPK, serine carboxypeptidase, photosystem II reaction centre and broadly represented the degree of expression of the respective genes in developing seeds (Table 4.3). On the basis of the KOG (Clusters of Eukaryotic Orthologous groups of proteins), the 656 chickpea unigenes excluding ' no homology' sequences, were further functionally classified by sorting into 22 putative functional groups (Fig. 4.3). Of these, 195 did not `match with the KOG database, hence remain unclassified. The remaining 461 sequences showed homology to many classes of proteins where the main groups of proteins were related to information storage and processing (30.4%), especially those related to translation, ribosomal structure and biogenesis (21.0%); cellular processes and signaling including post-translational modifications, protein turnover and chaperones (16.8 %) and metabolism (25.3%) including energy production and conversion (6.0%) and carbohydrate transport and metabolism (4.4%).

To get a better overview of the chickpea unigenes, they were further annotated on the basis of the existing annotations for the proteome of *Arabidopsis* assigned by the Gene Ontology (GO) Consortium. Of the 1037 chickpea unigenes, 622 (60.0%) of the chickpea unigenes matched with the categorized proteins of *Arabidopsis* with the E-value threshold set to $\leq 10^{-7}$. Relative frequencies of GO hits for *C. arietinum* unigenes assigned them to the functional categories: Biological Process, Molecular Function, and Cellular Component (as defined for the *Arabidopsis* proteome) and are presented in Fig. 4.4. However, 40.0% chickpea transcripts could not be functionally classified since they had no significant match with the *Arabidopsis* proteins.

4.2.3 Northern analysis

Based on the annotated sequences, five clones reported to be specifically involved in seed development in other plant systems were selected for expression analysis. These included, two clones having putative function for oleosin and conglutin-delta and were previously identified in our laboratory (unpublished results) while the remaining three namely pectinesterase, heat-shock protein binding and a seed-specific clone were selected

Figure 4.3: Functional classifications of the *C. arietinum* ESTs according to the KOG database. Blue bars represent frequency of sequences with homology to genes involved in cellular processes and signaling; Maroon color bars, information storage and processing; Yellow bars, metabolism and pink bars, poorly characterized ESTs.

CELLULAR PROCESSES AND SIGNALING

M Cell wall/membrane/envelope biogenesis
O Posttranslational modification, protein turnover, chaperones
T Signal transduction mechanisms
U Intracellular trafficking, secretion, and vesicular transport
V Defense mechanisms
Z Cytoskeleton

INFORMATION STORAGE AND PROCESSING

A RNA processing and modification
B Chromatin structure and dynamics
J Translation, ribosomal structure and biogenesis
K Transcription
L Replication, recombination and repair

METABOLISM

C Energy production and conversion
D Cell cycle control, cell division, chromosome partitioning
E Amino acid transport and metabolism
F Nucleotide transport and metabolism
G Carbohydrate transport and metabolism
H Coenzyme transport and metabolism
I Lipid transport and metabolism
P Inorganic ion transport and metabolism
Q Secondary metabolites biosynthesis, transport and catabolism

POORLY CHARACTERIZED

R General function prediction only
S Function unknown

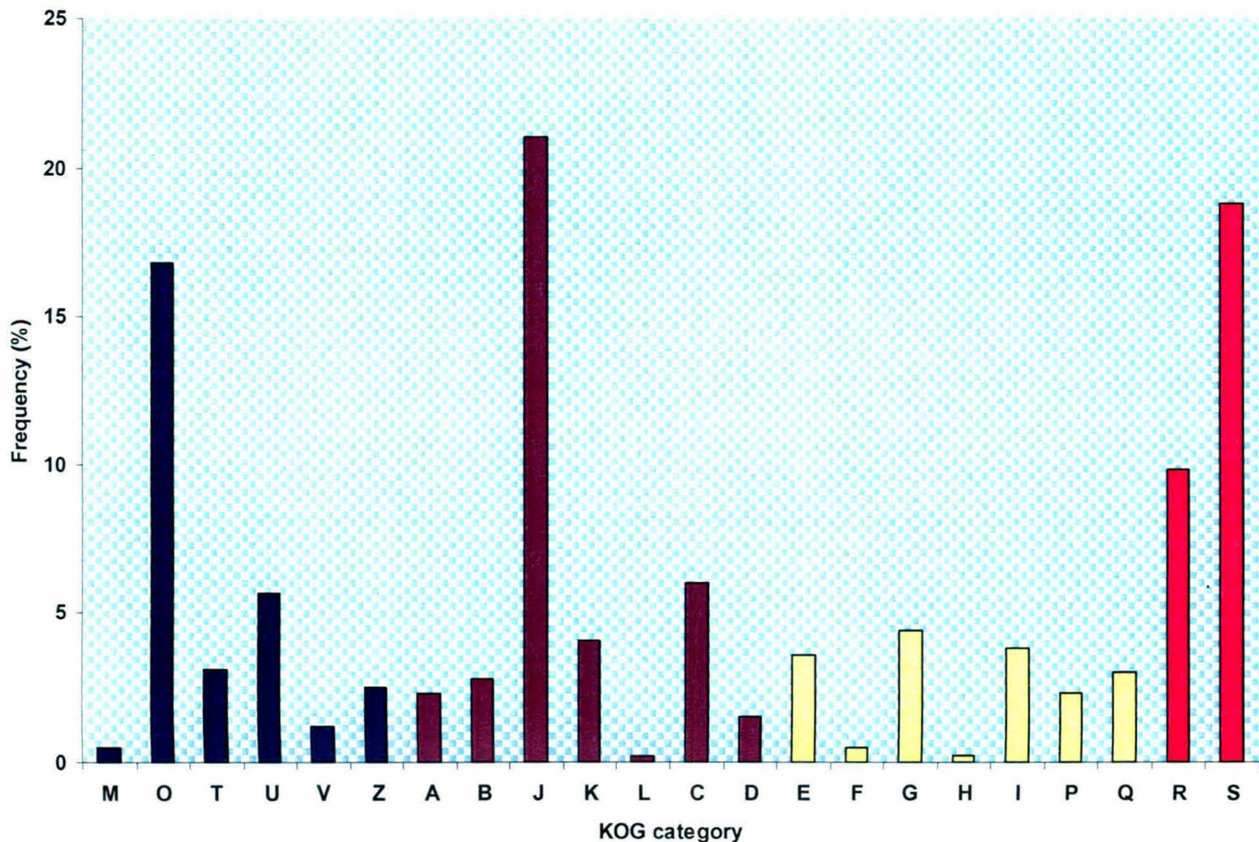
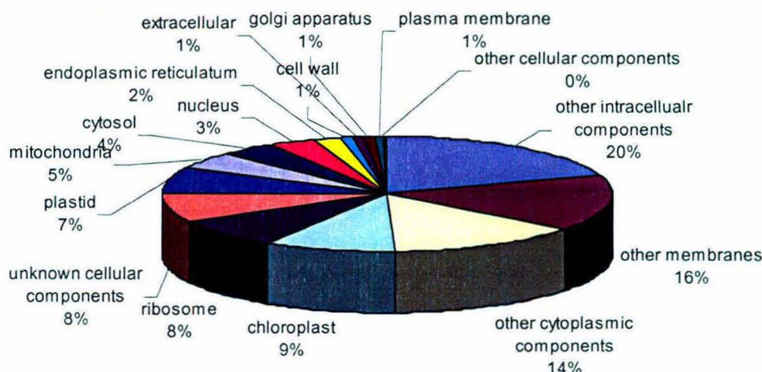
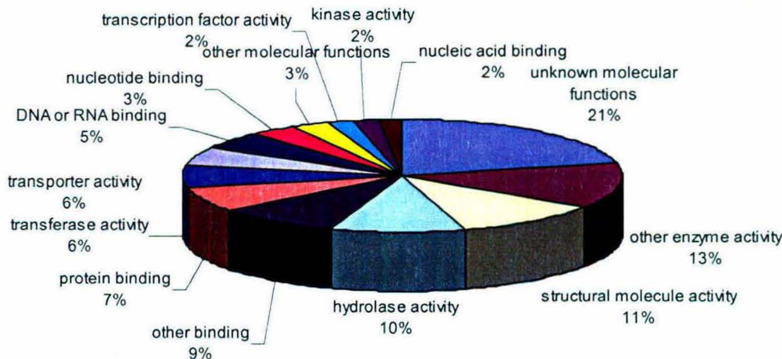
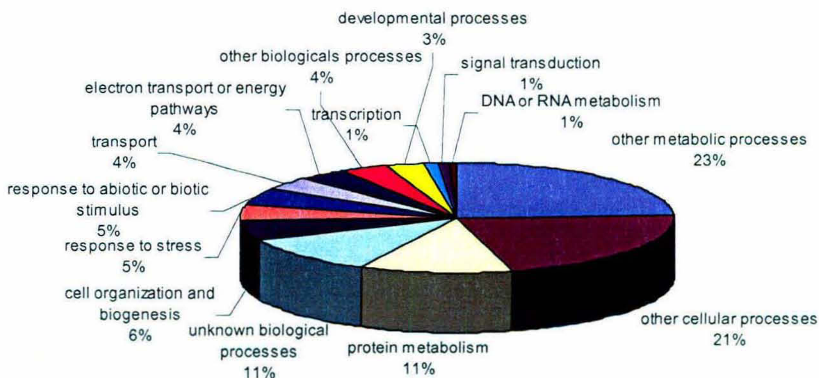


Figure 4.4: Gene Ontology (GO) classification of the *C. arietinum* EST library. The relative frequencies of GO hits for chickpea unigenes assigned to the GO functional categories Cellular Component, Molecular Function and Biological process as defined for the *Arabidopsis* proteome.







from the library described in this study. Northern analysis indicated that oleosin and conglutin-delta as expected, were expressed at later stages of seed development i.e. at 35-40 DAA in chickpea (Fig. 4.5). Pectinesterase and seed-specific protein were expressed throughout the developmental stages of chickpea seed as compared to control (opened flower) (Fig. 4.5). Putative heat shock protein binding also showed a similar pattern of expression.

4.3 Discussion

ESTs have proven to be an effective method for gene discovery and evaluation of mRNA expression patterns associated with specific plant tissues or growth conditions (Ramírez et al. 2005). Seeds of legumes have high nutritional value both for human consumption and animal feed and determining the molecular processes involved in legume seed development, a phenomenon that takes place within the time course of flower and pod development, is a major thrust area for biologists. With respect to grain legumes, earlier studies mainly focused on specific subjects such as sugar metabolism, storage protein synthesis etc. (Weber et al. 1997). In the past years, due to the wide complexity and varied genome size of legume crops, most of the proteome and transcriptome studies of legume seed biology were carried out in the annual barrel medic *M. truncatula* (Gallardo et al. 2003 and 2007; Firnhaber et al. 2005), a model legume characterized by a process of seed development very similar to that of other legumes. Recently, several seed related cDNA libraries were also made from other economically important legumes such as common bean, guar etc. (Ramírez et al. 2005; Naouminka et al. 2007). Chickpea is valued for its high nutritive seed (protein content ranges from 25.3-28.9 %). Yet, limited molecular analysis or EST generation from seed tissues has been done so far. The current research is focused at generating the EST database for chickpea seeds that display a wide range of morphological characters like seed shape, size, color, texture etc thus affecting the productivity.

Seed development in legumes is generally divided into three stages: early embryogenesis, seed filling followed by maturation. The storage compounds found in most mature seeds accumulate mostly during seed filling stage that usually ranges from 10-30 DAP (Days after Pollination) in most crops (Gallardo et al. 2003). Thus we targeted towards the generation of ESTs from seed filling stage of chickpea that determines the seed composition and factors that might affect the yield. The cDNA library generated from 20 DAA developing chickpea seed yielded 2123 recombinant clones with 76.9% efficiency. The assembly process resulted in a collection of 1037 unigenes showing 61.5% redundancy that

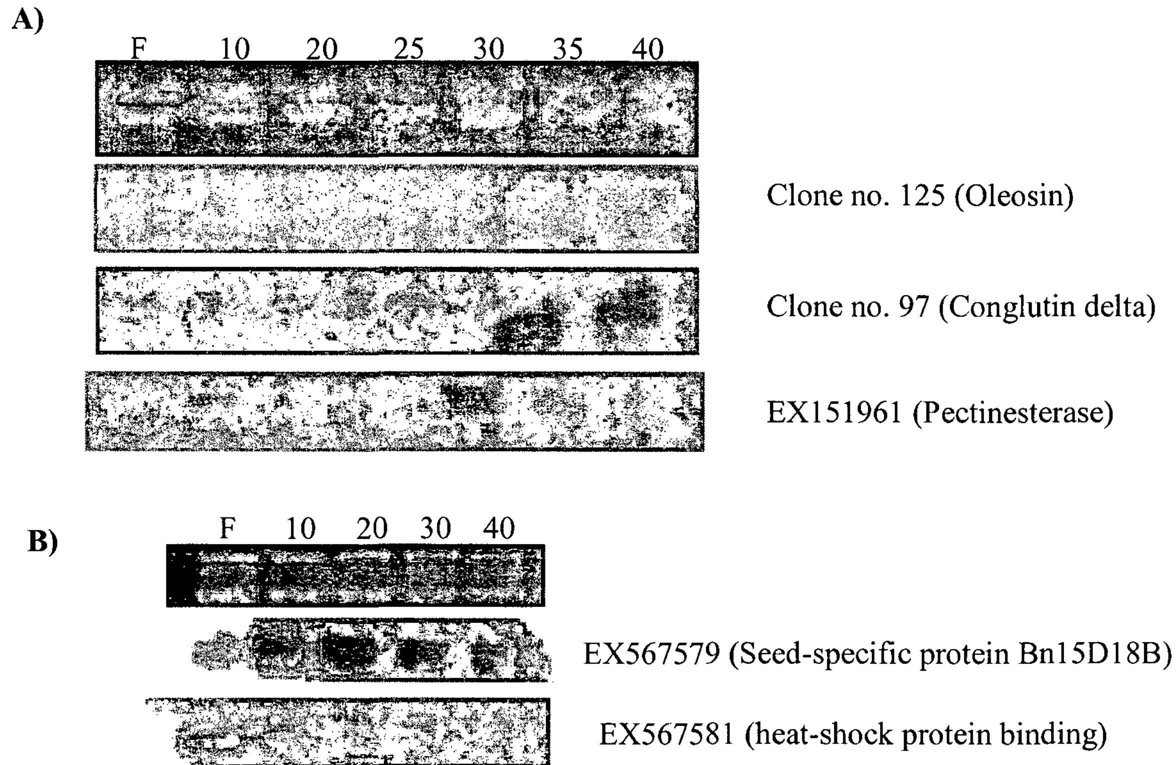


Figure 4.5: Northern blot analysis. Transcript accumulation of *C. arietinum* cDNAs that were expressed in developing seeds (10-40 DAA). Flower (F) was used as internal control and 25s rRNA was shown as loading control at the top panel. The corresponding ESTs along with annotations assigned by BLASTX homology search are mentioned on the right.

was quite consistent with other studies (Low et al. 2007). BLASTX analysis indicated that only 58.6% ESTs could be assigned putative functions illustrating that relatively few number of plant ESTs have been derived specifically from developing seeds till date (White et al. 2000). 'No hit' of approx. 20% chickpea unigenes in GenBank imparts an interesting pool of novel proteins with a function that may be of special relevance for developing seeds or may represent the chickpea specific transcriptome.

Further, expression patterns of chickpea seed-specific proteins obtained in the present study such as oleosin, seed specific unigene (showing homology to Bn15D18B of *Brassica*) and conglutin-delta was quite congruent to others reports (Dong et al. 2004; Naoumkina et al. 2007). Oleosins are proteins associated with lipid bodies mainly synthesized during seed development. In oil yielding plants, like *Brassica*, oleosins are mainly reported to be expressed at high level during the latter stages of seed development. Conglutin-delta is related to the 2S super family of storage proteins (protein family characterized by high levels of cysteine and glutamine) and widely reported in dicot seeds including economically important genera like *Brassica*, *Pisum* and *Arabidopsis*. In *Brassica*, accumulation of transcript Bn15D18B was reported both in embryo and seed coat illustrating the common function of this gene during seed development. Up-regulation of chickpea transcripts i.e. pectinesterase involved in the cell wall metabolism and heat-shock protein unveiled their role during seed development. Transcript accumulation of pectinesterase could either indicate the high rate of synthesis of new cell wall during seed development or the role of cell wall components in co-ordinating gene expression programmes during seed development, as observed in maize (Jose-Estanyol et al. 1992). Thus the present study enriches the collection of ESTs in this important crop and provides an opportunity in future to analyze a large number of chickpea seed related unigenes for in-depth understanding of molecular processes or mechanisms involved during seed development.

The present study herein first time reports the ESTs from chickpea developing seeds that in future would aid in evaluating the molecular mechanisms of seed development. The study underpinned the chickpea EST database by providing 1037 novel unigenes that would aid in expediting the functional genomic studies. Broad GO functional categorization of these unigenes was illustrated. Further these ESTs have been systematically explored for development of different types of functional molecular markers such as SSRs, ESTPs and ITPs which was applied in different fields ranging from diversity studies, transcript mapping to comparative mapping (see chapter 5, 6 and 8) studies.

Table 4.1: Summary of *Cicer arietinum* developing seed EST library

Total ESTs	2123
Total high quality ESTs	1897
Average insert size (bp)	700
Avg sequence length (bp)	350
Number of contigs	307 (encompassing 1167 ESTs)
Number of singlets	730
Number of unigenes	1037
Observed redundancy	61.5%
Homolgy (%) to known sequences	80.4%
No homologs	19.6%

Table 4.2: List of chickpea unigenes obtained from 20 DAA cDNA library. The EST_id of unigenes showing significant homology above the cut-off value (E-10) along with their homology match against the NCBI protein database, score and E-value are mentioned.

S.No	EST_id	Match, Score, E-value	S No	EST_id	Match, Score, E-value
1	EX567533	ADP-ribosylation factor *(M. sativa)*, 317, 4e-85	406	EX567523	No significant similarity
2	EX567534	ubiquitin fusion protein (*A. hypogaea*), 247	407	EX151761	putative 21kD protein (*Medicago)*, 178, 2e-43
3	EX567535	proline-rich protein (*G. max*), 155, 2e-40	408	EX151762	hypothetical protein (*Arabidopsis*), 148, 4e-39
4	EX567536	photosystem I psaH protein (*Trifolium*), 197, 9e-49	409	EX151763	Like-Sm ribonucleoprotein-related (*M. .truncatula*), 157, 2e-37
5	EX567537	Protein disulfide-isomerase precursor (*M. sativa)*, 196, 9e-49	410	EX151764	disulfide oxidoreductase (*Arabidopsis*), 169, 5e-48
6	EX567538	No significant similarity	411	EX151765	SEP1 (Stress Enchance Protein), 84, 2e-18
7	EX567539	Gibberellin regulated protein (*Medicago*), 136, 6e-31	412	EX151766	RNA polymerase II transcription (*Arabidopsis*), 105, 1e-21
8	EX567540	No significant similarity	413	EX151767	No significant similarity
9	EX56754	No significant similarity	414	EX151768	NFD2 RNA binding / ribonuclease III (*Arabidopsis*), 138, 2e-36
10	EX567542	No significant similarity	415	EX567524	No significant similarity
11	EX567543	unknown protein (*Oryza*), 80, 2e-40	416	EX151769	lipid binding (*Arabidopsis*), 140, 3e-32
12	EX567544	Alpha-helical ferredoxin (*Medicago*), 250, 2e-65	417	EX151770	unknown protein (*Arabidopsis*), 201, 2e-50
13	EX567545	alpha-mannosidase (*Arabidopsis*), 330, 9e-97	418	EX151771	No significant similarity
14	EX567546	cyclophilin A (Beauveria bassiana), 84.7, 6e-47	419	EX151772	No significant similarity
15	EX567547	hypothetical protein (*O. sativa)*, 94,4, 6e-18	420	EX151773	2OG-Fe(II) oxygenase (*M. truncatula*) , 137, 6e-31
16	EX567548	60S ribosomal protein L39 (*Oryza*), 78.6, 1e-13	421	EX151774	No significant similarity
17	EX567549	putative PSII-P protein (*Trifolium*), 238, 4e-75	422	EX151775	conserved hypothetical protein (*M. truncatula*), 204, 1e-51
18	EX567550	protease inhibitor (*G. max*), 125, 5e-27	423	EX151776	DnaJ-like protein heat shock protein(*Arabidopsis*), 123, 7e-28
19	EX567551	RNA binding (*Arabidopsis*), 89, 9e-25	424	EX151777	No significant similarity
20	EX567552	chlorophyll a/b binding protein (*Cicer*), 214, 2e-126	425	EX567525	No significant similarity
21	EY457888	unknown protein (*A. thaliana*), 77, e-13	426	EX151778	actin [*Gossypium hirsutum*), 267, 2e-71
22	EX567553	hypothetical protein (Medicago) 174, 1e-42	427	EX151779	Nicotiana lesion-inducing like (*Arabidopsis*), 87.4, 6e-24
23	EX567555	SAH7 (*A. thaliana*), 60.8, 5e-19	428	EX151780	Cu/Zn superoxide dismutase II (*Pisum sativum*), 193, 2e-75
24	EX567556	ribosomal protein S3 (*M. truncatula*), 360, 1e-120	429	EX151781	Chloroplast 50S ribosomal protein (*P. sativm*), 69.3, 5e-11
25	EX567557	40S ribosomal protein (*Solanum*), 105, 6e-45	430	EX151782	DNA binding (*Arabidopsis*), 122, 3e-46
26	EX567558	Photosystem II reaction center (*Spinacia oleracea*), 110, 4e-23	431	EX151783	No significant similarity
27	EX567559	10 kDa photosystem II polypeptide (*T. pretense*), 239, 7e-62	432	EX151784	No significant similarity
28	EX567560	Nonspecific lipid-transfer protein (*Cicer arietinum*), 229, 6e-59	433	EX151785	No significant similarity
29	EX567561	Ribosomal protein (*A.thaliana)*, 100, 6e-20	434	EX151786	unknown [*Arabidopsis thaliana*], 184, 3e-45
30	EX567562	No significant similarity	435	EX151787	acyl carrier protein 1 (*Cicer arietinum*), 77, 3e-24
31	EX567563	hypothetical protein 91 (garden pea), 72, 7e-12	436	EX151788	ADP-ribosylation factor [*Hyacinthus)*, 142, 5e-33
32	EX567564	Defender against cell death 1 (DAD-1) (*Pisum sativum*), 211, 2e-53	437	EX151789	no significant similarity
33	EX567565	No significant similarity	438	EY457880	*14-3-3 brain protein homolog [Vicia)*
34	EX567566	AT3g22430 (*Arabidopsis*), 86.7 4e-16	439	EX151790	unknown protein (*Arabidopsis*), 120, 5e-43
35	EX567568	CCT (*M. truncatula*), 115, 2e-24	440	EX151791	No significant similarity
36	EX567569	chlorophyll a/b-binding (*A. thaliana*), 69.7, 5e-11	441	EX151792	unknown protein (*Arabidopsis*), 110, 6e-23
37	EX567570	desaturase-like protein (*T. repens)*, 255, 7e-67	442	EX151793	No significant similarity
38	EX567571	Os08g0191600 (*Oryza sativa*), 198, 1e-49	443	EX151794	unknown [*Euphorbia esula*) , 58.2, 5e-08
39	EX567572	Ribosomal protein (*A. thaliana*), 92, 2e-17	444	EX567526	NOI nitrate-induced protein (*Arabidopsis*), 96.3, 5e-19
40	EX567574	Vacuolar (H+)-ATPase G subunit (*M. truncatula*), 267, 5e-22	445	EX151795	unknown protein (*Arabidopsis*), 107, 3e-22
41	EY457893	hypothetical protein (*Nicotiana)*, 93.2, 3e-27	446	EX151796	NADH dehydrogenase subunit 3 (*Daucus carota*), 180, 2e-44
42	EX567575	serine carboxypeptidase (*Pisum*), 267, 5e-70	447	EX151797	MADS box protein AP3-like (Lotus), 60.5,

43	EX567576	No significant similarity
44	EX567577	lipid transfer protein (*Tamarix*), 115, 2e-24
45	EX567578	No significant similarity
46	EX567579	seed specific protein Bn15D18B (*Brassica*), 77.4 5e-13
47	EY457894	ATP sulfurylase (*G. max*), 156, 3e-37
48	EX567580	ribosomal protein L17 (*Castanea sativa*), 232, 1e-59
49	EX567581	70 kDa heat shock cognate protein 3 (*Vigna*), 176, 3e-43
50	EX567582	temperature-induced lipocalin (*Medicago*), 67.4, 2e-10
51	EX567583	copper chaperone (*Populus alba*), 117, 2e-26
52	EX567584	hypothetical protein (*A. thaliana)*, 139, 2e-31
53	EX567585	unknown protein (*A. thaliana*), 52.4, 3e-18
54	EX567586	nucleotide-sugar transporter (*A. thaliana*), 165, 2e-39
55	EX567587	Os02g0704900 (*Oryza sativa*), 124, 5e-41
56	EX567588	cytochrome-c oxidase (*Arabidopsis*), 96.3, 6e-19
57	EX567589	Annexin (*M. truncatula)*, 142, 4e-33
58	EX567590	ATP binding protein (*A. thaliana)*, 74.7 , 2e-17
59	EX567591	No significant similarity
60	EX567592	Os06g0653900 (*O. sativa)*, 106, 5e-22
61	EX567593	Ribosomal protein S27a (*Medicago*), 155, 4e-63
62	EX567594	ethylene-responsive tran (*Retama*), 182, 8e-45
63	ES544479	OSH1 (*A. thaliana)*, 171, 2e-41
64	EX567595	glutaredoxin (*Tilia platyphyllos)*, 161, 2e-38
65	EX567596	HAD-superfamily subfamily (*Medicago)*, 79, 8e-14
66	EX567597	SKP1 component (*M. truncatula)*, 198, 3e-59
67	EX567598	3-ketoacyl-CoA thiolase (*Mangifera indica*, 190, 6e-47
68	EX567599	light-harvesting chlorophyll a/b binding (*Trifolium repens*), 251, 2e-70
69	EX567600	hypothetical protein (*Medicago)*, 62.4, 2e-08
70	EX567601	cytochrome b6 (*Citrus sinensis*), 131, 2e-35
71	EX567602	No significant similarity
72	EX567603	NADH dehydrogenase subunit (*Beta vulgaris*), 129, 7e-29
73	EX567604	annexin (*M. sativa*), 256, 7e-67
74	EX567605	ribosomal protein S14 (*Pisum sativum*), 125, 3e-27
75	EX567606	Thioredoxin-related (*Medicago*), 159, 3e-55
76	EX567607	proteasome subunit (*Medicago*), 107, 2e-22
77	EX567608	hypothetical protein (*Oryza*), 155, 8e-37
78	EX567609	ribosomal protein L31 (*Lactuca sativa*), 138, 2e-31
79	EX567610	unknown protein (*Arabidopsis*), 149, 1e-34
80	EX567611	40S ribosomal protein (*Zea mays*), 169, 6e-43
81	EX567612	putative epsilon subunit of mitochondria (*Cicer*), 145, 1e-33
82	EX567613	unknown protein (*Arabidopsis*), 103, 3e-21
83	EX567614	structural constituent of ribosome (*Arabidopsis*), 88.23e-16
84	EX567615	ATP binding proteein (*A. thaliana)*, 116, 4e-25
85	EX567616	hypothetical protein (*Oryza*), 81.3, 2e-14

		4e-13
448	EX151798	Helix-turn-helix (*M. truncatula*), 145, 2e-49
449	EX151799	No significant similarity
450	EX151800	Ribosomal protein S23, (*M. truncatula*), 280, 4e-74
451	EY457899	No significant similarity
452	EX151801	zinc finger (*M. truncatula*), 140, 3e-35
453	EX151802	No significant similarity
454	EX151803	ribosomal protein [*Petunia x hybrida*), 283, 6e-75
455	EX151804	cyanate hydratase (*Arabidopsis*), 226, 6e-58
456	EX151805	BiP [*Glycine max*), 102, 1e-26
457	EX151806	putative histone deacetylase [*Trifolium pretense*), 97.8, 1e-19
458	EX151807	At5g26850 (*Arabidopsis*), 115, 8e-30
459	ES544480	RNA-binding protein (*M. truncatula*), 114, 3e-31
460	EX151808	ADP-ribosylation factor (*M. truncatula*), 323, 5e-87
461	EX151809	2OG-Fe(II) oxygenase (*M. truncatula*), 99.4, 4e-20
462	EX151810	Glutathione peroxidase (*M. truncatula*), 205, 1e-51
463	EX151811	unknown protein (*Arabidopsis*), 68.6, 8e-11
464	EX151812	putative stress-responsive protein [*Fragaria x ananassa*), 90.5, 4e-22
465	EX151813	peptidyl-prolyl cis-trans isomerase (*Arabidopsis*), 163, 1e-52
467	EX151814	thioredoxin H (*Nicotiana alata*), 79, 2e-23
468	ES544481	No significant similarity
469	EX567527	No significant similarity
470	EX151815	40S ribosomal protein S30-like [*Oryza*), 76.3, 2e-17
471	EX151816	CP12 [*Pisum sativum*), 62, 1e-11
472	EX151817	eIF4-gamma/eIF5/eIF2-epsilon (*M. truncatula*), 86.3, 4e-16
473	EX151818	Scorpion short chain toxin (*M. truncatula*), 73.2, 3e 12
474	EX15181	putative aminopeptidase (*A. thaliana*), 223, 4e-57
475	EX151820	At4g32110 (*Arabidopsis*), 83.6, 5e-15
476	EX151821	cytochrome P450 (*Arabidopsis*),72.4, 6e-12
478	EX151822	unknown protein [*Arabidopsis thaliana*), 50.4, 7e14
479	EX151823	No significant similarity
480	EX151824	vf14-3-3c protein (*Vicia faba*), 250, 2e-77
490	EX151825	ribosomal protein (*Oryza sativa*), 149, 4e-55
491	EX151826	putative developmental protein [*Nicotiania benthamiana*), 143, 7e-33
492	EX151827	rbcL ribulose 1,5-bisphosphate (*V. radiata*), 102, 2e-20
493	EX151828	serine carboxypeptidase [*Vigna radiata*), 71.6, 1e-11
494	EX151829	cytochrome P450 monooxygenase CYP75A (*M. truncatula*), 223, 6e-73
495	EX151830	Ribosomal protein (*Arabidopsis*), 76.3, 1e-12
496	EX151831	unknown protein (*Arabidopsis*), 120, 9e-28
497	EX151832	SYP121; t-SNARE (*Arabidopsis*),100, 6e-29
498	EX151833	No significant similarity
499	EX151834	No significant similarity
500	EX151835	ribulose 1,5-bisphosphate carboxylase (*C. arietinum*),103, 3e-21
501	EX151836	Zinc finger, RING-type (*M. truncatula*), 180, 3e-84

86	EX567617	Os08g0190800 (*Oryza sativa*) , 112, 1e-23
87	EX567618	unknown protein (*Arabidopsis*), 175, 1e-42
88	EX567619	No significant similarity
89	EX567620	unknown protein (*Arabidopsis*), 252, 1e-65
90	EX567621	unknown protein (*Arabidopsis*), 149, 4e-35
91	EX567622	hypothetical protein (*Oryza*) , 84.7, 2e-15
92	EX567623	Plastocyanin (*Pisum*) , 120, 5e-26
93	EX567624	unknown protein (*A. thaliana*) , 135, 6e-33
94	EX567625	ferredoxin I (*Trifolium pratense*), 131, 2e-29
95	EX567626	Chlorophyll a/b-binding (*A. thaliana*), 111, 8e-30
96	EX567627	Haem peroxidase (*Medicago*) , 317, 1e-85
97	EX567628	NAD(P)H-quinone oxidoreductase (*Vicia*) , 89, 8e-17
98	EX567629	Eukaryotic translation initiation factor (*Onobrychis viciifolia*)
99	EX567630	Homeodomain-related (*M. truncatula*), 135, 5e-35
100	EX567631	ribosomal protein S3a (*Cicer*) , 174, 2e-50
101	EX567632	isoflavone reductase-like protein (*Vitis*) , 74.7, 2e-12
102	EX567633	SAM (*M. Truncatula*), 183, 7e-61
103	EX567634	hypothetical protein (*Oryza*) ,64.7, 2e-09
104	EX567635	unknown protein (*Arabidopsis*) , 111, 3e-23
105	EX567636	Os03g0606200 (*Oryza sativa*) , 102, 5e-21
106	EX567637	Luminal-binding protein 3 (*N tabcum*), 170, 4e-41
107	EX567638	hydroxyproline-rich glycoprotein (*Pisum*) , 3e-58
108	EX567639	Wound-induced protein 109(*P. vulgaris*), 88.6, 1e-16
109	EX567640	No significant similarity
110	EX567641	Ribosomal protein S9 (*Medicago*) , 196, 2e-67
111	EX567642	hypothetical protein (*Medicago*), 101, 1e-20
112	EX567643	germin-like protein (*Phaseolus vulgaris*), 110, 2e-45
113	EX567644	No significant similarity
114	EX567645	Rubredoxin-type Fe(Cys)4 (*Medicago*), 92.4, 4e-30
115	EX567646	Photosystem I protein (*Medicago*), 366, 1e-99
116	EX567647	unknown protein (*Arabidopsis*), 186, 3e-49
117	EX567648	seed specific protein Bn15D18B (*Brassica*), 77.4, 6e-13
118	EX567649	No significant similarity
119	EX567650	S25 ribosomal protein (*Medicago* , 135, 1e-30
120	EX567651	unknown protein, (*Nicotiana tabacum*), 53.1, 3e-07
121	EX567652	Ribosomal protein L27 (*Medicago*), 179, 1e-43
122	EX567653	Ubiquitin (*Medicago truncatula*), 152, 6e-36
123	EX567654	hypothetical protein (*Trifolium*), 234, 9e-66
124	EX567655	TOM22-V (*Arabidopsis thaliana*), 90.9, 5e-17
125	EX567656	hypothetical protein, 51.6, 3e-15
126	EX567657	Ribosomal protein S11 (*Medicago*) , 214, 2e-54
127	EX567658	cyclin D (*Pisum sativum*), 61.6, 1e-08
128	EX567659	SHI (SHORT INTERNODES); (*Arabidopsis*), 90.1, 7e-17

502	EX151837	unknown protein (*Arabidopsis*), 132, 1e-29
503	EX151838	No significant similarity
504	EX151839	Pex19 protein (*M. truncatula*),198, 2e-68
505	EX151840	unknown [*Arabidopsis thaliana*],184, 1e-45
506	EX151841	ribosomal protein L16 (*Lotus corniatus*),157, 3e-37
507	EX151842	NADH dehydrogenase (*Jasminum abyssinicum*), 106, 4e-22
508	EX151843	DNA-directed RNA polymerase (*Pisum sativum*), 167, 1e-59
509	EX151844	No significant similarity
510	EX151845	Domain of unknown function (*Medicago*), 147, 1e-37
511	EX151846	cationic amino acid transporter (*Arabidopsis*), 79, 1e-13
512	EX151847	unnamed protein product [*Agrobacterium rhizogenes*), 148, 4e-34
513	EX151848	vacuolar membrane ATPase subunit c" [*Citrus limon*), 92.8, 1e-17
514	EX151849	Glycosyl hydrolases (*Oryza*), 311, 2e-83
515	EX151850	alpha-mannosidase [*Arabidopsis thaliana*], 384, 3e-105
516	EX151851	repressor protein Drl/DrAp (*G. max*), 151, 4e-37
517	EX151852	Glycosyl hydrolases .(*Oryza sativa*), 256, 4e-69
518	EX151853	Lipoxygenase, LH2 (*M. truncatula*), 146, 9e-34
519	EX151854	unnamed protein product [Agrobacterium) , 147, 6e-34
520	EX151855	NAD-binding site (*M. truncatula*), 72.8, 6e-27
521	EX151856	MADS box protein AP3-like [Lotus), 76.6, 9e-13
522	EX151857	ORF18 [*Agrobacterium rhizogenes*, 76.6, 1e-12
523	EX151858	amidohydrolase (*M. truncatula*), 171, 7e-42
524	EX151859	Os04g0683100 [*Oryza sativa*) , 102, 2e-25
525	EX151860	Proteinase inhibitor (*M. truncatula*), 80.9, 4e-17
526	EX151861	No significant similarity
527	EX151862	carbohydrate transporter/ sugar (*Arabidopsis*), 52.4, 7e-06
528	EX151863	unknown (*Lobochlamys segnis*)
529	EX151864	Mitochondrial import receptor (*S . .tubeosumr*), 134, 5e-33
530	EX151865	Transcriptional factor B3 (*M. truncatula*), 120, 1e-26
531	EX151866	squamosa promoter binding protein (*Antirrhinum*),83.2, 6e-25
532	EX151867	Ribosomal protein (*M. truncatula*), 63.2, 3e-10
533	EX151868	Type II chlorophyll a/b binding (*Pisum*) , 226, 5e-58
534	EX151869	unknown protein (*Arabidopsis*), 79.3, 7e-15
535	EX151870	No significant similarity
536	EX151871	unknown protein (*Arabidopsis*), 99.8, 3e-30
537	EX151872	PsRT17-1 (*Pisum sativum*),111, 5e-38
538	EX151873	fiber protein Fb15 (*M. truncatula*), 104, 2e-21
539	EX151874	unnamed protein [*Pisum satvium*), 170, 4e 41
540	EX151875	Serine carboxypeptidase (*Oryza*), 101, 9e-21
541	EX151876	No significant similarity
542	EX151877	unknown protein (*Arabidopsis*), 68.6, 3e-13
543	EX151878	Photosystem I reaction centre (*M. truncatula*), 122, 1e-26
544	EX151879	TPR repeat (*M. truncatula*), 106, 7e-22

129	EX567660	Ribosomal L18ae protein (*Medicago*) , 179, 4e-79
130	EX567661	unknown protein (*Arabidopsis*) , 195, 1e-48
131	EX567662	hypothetical protein (*Nicotiana*) , 96.7, 5e-39
132	EX567663	60S ribosomal protein L30 (*Lupinus*) , 177, 3e-45
133	EX567664	putative senescence-associated (Pea), 84.3, 2e-15
134	EX567665	No significant similarity
135	ES544483	No significant similarity
136	EX567666	hypothetical protein (*Oryza*) , 183, 5e-45
137	EX567667	
138	EX567668	ribulose 1,5-bisphosphate carboxyl. (*Cicer*) , 340, 4e-93
139	EX567669	No significant similarity
140	EX567670	unknown protein (*Arabidopsis*) , 169, 8e-41
141	EX567671	hypothetical protein (*Oryza sativa*) , 70.5, 5e-11
142	EX567672	14-3-3 Protein (*Glycine max*), 125, 8e-28
143	EX567672	14-3-3 Protein (*Glycine max*), 125, 8e-28
144	EX567673	No significant similarity
145	EX567674	Tubulin binding cofactor(*Medicago*), 205, 6e-52
146	EX567675	60S ribosomal protein L6 (*Cicer*), 213, 1e-62
147	EX567676	early light inducible protein (*Trifolium*), 127, 3e-28
148	EY457895	Photosystem I reaction centre (*Medicago*), 102, 8e-22
149	EX567677	xyloglucan endotransglycosylase (*Malu*) , 125, 7e-28
150	EX567678	unknown protein (*Oryza sativa*), 128, 5e-38
151	EX567679	photosystem ii core complex (*Medicago*), 70.5, 6e-17
152	EX567680	Cystinosin/ERS1p repeat (*Medicago*), 175, 2e-42
153	EX567681	hypothetical protein (*Cicer*), 71.6, 7e-11
154	EX567682	protease inhibitor (*Glycine*), 87.4, 3e-16
155	EX567683	Pectinesterase inhibitor (*Medicago*), 89.7, 6e-17
156	EX567684	ribosomal protein L33 (*Castanea sativa*), 228, 2e-58
157	EX567685	hypothetical protein (*Medicago*), 170, 2e-52
158	EX567686	40S ribosomal protein (*Fragaria* x *ananassa*), 284, 3e-75
159	EX567687	mitogen-activated protein kinase (*Citrus sinensis*), 672, 0
160	EY457896	No significant similarity
161	EX567688	unnamed protein (*Agrobacterium*), 149, 3e-34
162	EX567689	unknown protein (*Arabidopsis*), 81.6, 3e-27
163	EX567690	ZIM (*Medicago truncatula*), 140, 1e-38
164	EX567691	No significant similarity
165	EX567692	hypothetical protein (*Oryza*), 68.6, 1e-10
166	EX567693	Brassinosteroid-regulated protein (*Glycine max*), 163, 6e-39
167	EX567694	C/VIF2 (*Arabidopsis*), 100, 2e-20
168	EX567695	Complex 1 LYR protein (*Medicago*), 128, 1e-28
169	EX567696	putative phospholipid hydroperoxide (*Cicer*), 167, 3e-40
170	EX567697	subtilisin-type protease precursor (*Glycine*), 94.7, 1e-18
171	EX567698	cationic peroxidase (*Cicer*), 368, 9e-117
172	EX567699	MtN4 (*Medicago truncatula*), 113, 1e-23
173	EX567700	Glyceraldehyde-3-phosphate dehydrogense (*Pisum*), 197, 3e-49

545	EX151880	multicatalytic endopeptidase (*Cicer*) , 171, 2e-41
546	EX151881	histone H2A protein [*Oryza*), 138, 2e-31
547	EX151882	No significant similarity
548	EX151883	No significant similarity
549	EX151884	structural constituent of ribosome (*Arabidopsis*), 72.4, 6e-12
550	EX151885	putative protein (*Arabidopsis*), 122, 7e-27
551	EX151886	proteosome component (*Arabidopsis*), 85.5, 2e-21
552	EX151887	hypothetical protein (*Arabidopsis*), 71.2, 2e-11
553	EX151888	unknown protein (*Arabidopsis thaliana*)
554	EX151889	Oleoyl-acyl carrier protein (*Coriandrum sativum*) , 51.6, 5e-10
555	EX151890	Sterol desaturase (*M. truncatula*), 125, 2e-33
556	EX151891	receptor-like GPI-anchored protein (*T. pretense*), 56.2, 2e-12
557	EX151892	Zinc finger (*M. truncatula*), 77, 1e-17
558	EX151893	Chlorophyll A-B binding protein (*M. truncatula*), 85.5, 7e-16
559	EX151894	GTP-binding signal recognition (*M. truncatula*), 171, 2e-41
560	EX151895	No significant similarity
561	EX151896	No significant similarity
562	EX151897	peroxidase [*Dimocarpus longan*), 228, 6e-65
563	EX151898	conserved hypothetical protein (*M. truncatula*),79.7, 9e-14
564	EX151899	Ribosomal L29e protein (*M. truncatula*), 126, 4e-28
565	EX151900	No significant similarity
566	EX151901	KH, type 1 (*M. truncatula*), 107, 4e-24
567	EX151902	hypothetical protein (*M. truncatula*), 120, 2e-26
568	ES544485	unknown [*Arabidopsis thaliana*), 154, 2e-3
569	EX151903	putative protein (*Arabidopsis*), 119, 3e-26
570	EX151904	Ycf1 [*Medicago truncatula*), 95.9, 6e-19
571	EX151905	No significant similarity
572	EX151906	40S ribosomal protein (S1 *Fragaria x ananassa*, 137, 5e-31
573	EX151907	Inorganic pyrophosphatase (*M. truncatula*), 97.4, 2e-19
574	EX151908	putative protein [*Arabidopsis thaliana*), 290, 3e-77
575	EX151909	putative heat shock protein 82 (*Oryza*), 109, 4e-23
576	EX15191	unknown protein [*Oryza sativa*) , 122, 1e-52
577	EX151911	Plant self-incompatibility S1 (*M. truncatula*), 84.3, 4e-15
578	EX567528	U2 snRNP auxilliary factor (*M. truncatula*), 165, 6e-40
579	EX151912	No significant similarity
580	EX151913	xylogen protein 1 (*Zinnia elegans*), 68.9, 1e-10
581	EX151914	unknown protein (*Arabidopsis*), 92.8, 5e-18
582	EX151915	No significant similarity
583	EX151916	No significant similarity
584	EX151917	Concanavalin A-like lectin/glucana (*M. truncatula*), 140, 1e-32
585	EX151918	No significant similarity
586	EX151919	No significant similarity found
587	EX151920	hypothetical protein (*Cicer arietinum*), 107, 2e-22
588	EX151921	No significant similarity
589	EX151922	No significant similarity

174	EX567701	No significant similarity
175	EX567702	60S ribosomal protein L23a (*Daucus*), 135, 1e-30
176	EX567703	hypothetical protein (*Phaseolus vulga*), 91.3, 2e-17
177	EX567704	hypothetical protein (*Medicago*), 76.6, 4e-13
178	EX567705	CP12 [(*Pisum sativum*), 170, 8e-41
179	EX567706	Os11g0634500 (*Oryza sativa*), 125, 6e-28
180	EX567707	Ribosomal protein L19e (*Medicago*), 171, 2e-80
181	EX567708	profilin-like (*Solanum tuberosum*), 193, 9e-48
182	EX567709	ribosomal protein L17 (*Castanea*), 248, 1e-64
183	EX567710	hypothetical protein (*Cleome spinosa*), 152, 6e-36
184	EX56771	hypothetical protein (Oryza), 149, 1e-34
185	EX567712	Photosystem I reaction center (*Medicago*), 224, 3e-57
186	EX567713	No significant similarity
187	EX567714	LHC II Type III chlorophyll a /b bi Brassica napus, 64.7, 9e-20
188	EX567715	40S ribosomal protein S5 (*Cicer*), 371, 2e-101
189	EX567716	glycine-rich RNA-binding (*Pisum*), 179, 2e51
190	EX567717	Harpin-induced 1 (*Medicago*), 181, 2e-56
191	EX567718	No significant similarity
192	EX567719	No significant similarity
193	EX567720	chloroplast pigment-binding protein (*Nicotiana*), 240, 4e-62
194	EX567721	2OG-Fe(II) oxygenase (*Medicago*), 231, 3e-61
195	EX567722	ribosomal protein L37 (*Glycine max*), 170, 4e-41
196	EX567723	60S ribosomal protein L34 (Pea), 175, 1e-42
197	EX567724	ultraviolet-B-repressible protein (*Gossypium hirsutum*), 122, 2e-26
198	EX567725	actin depolymerizing factor-like (*Arachis hypogaea*), 268, 2e-70
199	EX567726	structural constituent of ribosome (*Arabidopsis thaliana*), 80.9, 5e-14
200	ES544475	unknown protein (*Arabidopsis*), 79, 1e-13
201	EX567727	transcription factor homolog BTF3 (*Lotus japonicus*) , 214, 3e-54
202	EX567728	H+-transporting two-sector ATPase (*Medicago*) , 98.6, 2e-19
203	EX567729	calmodulin cam-203 [Daucus carota, 258, 1e-67
204	EX567730	DNA-directed RNA polymerase (*Medicago*), 238, 3e-61
205	EX567731	Prefoldin (*Medicago*), 82.4, 7e-15
206	EX567732	No significant similarity
207	EX567733	Protein of unknown function (*Medicago*), 250, 6e-65
208	EX567734	putative invertase inhibitor (*Cicer*), 119, 1e-25
209	EX567735	No significant similarity
210	EX567736	unknown protein (*Arabidopsis*), 91.7, 1e-17
211	EX567737	60S Ribosomal Protein L44 (*Neurospora crassa*) , 72.8, 1e-21
212	EX567738	putative senescence-associated (Pea), 98.6, 1e-19
213	EX567739	60S ribosomal protein L21 (*Oryza*), 306, 6e-82
214	EY457897	No significant similarity
215	EX567740	heat-shock protein 80 (*Euphorbia*), 367, 4e-100
216	EX567741	unknown protein (*Oryza sativa*), 93.2, 6e-18

590	EX151923	Os07g0483400 [*Oryza sativa*), 127, 6e-28
591	EX151924	DNA-directed RNA polymerase II 8 (*Musa balbisiana*), 140, 2e-32
592	EX151925	GroES-like [*Medicago truncatula*), 97.1, 2e-19
593	EX151926	unknown protein (*Arabidopsis*), 139, 1e-31
594	EX151927	No significant similarity
595	EX151928	hypothetical protein [*Homo sapiens*), 62, 2e-08
596	EX151929	unknown [*Arabidopsis thaliana*] , 70.5, 2e-11
597	EX151930	Os03g0125000 [*Oryza sativa*), 73.9, 7e-23
598	EX567529	No significant similarity
599	EX151931	Lhca5 protein [*Arabidopsis thaliana*), 85, 5e-21
600	EX151932	No significant similarity
601	EX151933	6b-interacting protein 1 (*Nicotiana*) , 64.3, 2e-09
602	EX151934	chalcone isomerase 3 [*Glycine max*), 74.3, 2e-12
603	EX151935	No significant similarity
604	EX151936	orf174 [*Beta vulgaris*), 81.6, 1e-14
605	EX151937	Ribosomal protein L6E (*M. truncatula*), 270, 5e-71
606	EX151938	Allergen V5/Tpx-1 related (*M. truncatula*), 335, 5e-91
607	EX151939	thiosulfate sulfurtransferase [Datisca) , 174, 1e-42
608	EX151940	Ribosomal protein S19e (*M. truncatula*), 266, 6e-70
609	EX151941	No significant similarity
610	EX567530	No significant similarity
611	EX151942	Peptidase (*M. truncatula*), 328, 1e-88
612	EX151943	No significant similarity
613	EX151944	unknown protein (*Arabidopsis*), 84, 5e-15
614	EX151945	hypothetical protein (*Arabidopsis*), 121, 1e-26
615	EX151946	Allergen V5/Tpx-1 related (*M. truncatula*), 72, 8e-12
616	EX151947	unknown protein (Arabidopsis) , 135, 1e-30
617	ES544489	hsr203J homolog [*Pisum sativum*), 103, 1e-37
618	EX151948	No significant similarity
619	EX151949	No significant similarity
620	EX151950	Globin-like (*M. truncatula*), 168, 2e-40
621	EX151951	C2 (*M. truncatula*), 289, 8e-77
622	EX151952	No significant similarity found
623	EX151953	OSJNBa0018M05.20 [*Oryza sativa*) , 162, 6e-39
624	EX151954	putative glutathione transporter (Zea) , 165, 2e-39
625	EX151955	Nascent polypeptide-associated compex (*M. truncatula*), 328, 2e-88
626	EX151956	unknown protein (Arabidopsis), 73.2, 8e-12
627	EX151957	34 kDa outer mitochondrial (*Solanum*) , 205, 5e-67
628	EX151958	Glyceraldehyde-3-phosphate dehydrogenase (*Magnolia quinquepeta*), 183, 4e-45
629	EX151959	hypothetical protein (*M. truncatula*), 114, 1e-24
630	EX151960	glycosyltransferase [*Plantago major*), 182, 5e-57
631	EX151961	enzyme inhibitor/ pectinesterase (*Arabidopsis*), 98.6, 1e-19
632	EX151962	40S ribosomal protein S5 [*Cicer arietinum*), 266, 2e-70

217	EX567742	Generic methyltransferase (*Medicago*), 82.4, 5e-16
218	EX567743	Ribosomal protein L7 (*Medicago)*, 244, 4e-63
219	EX567744	Ribosomal protein 60S (*Medicago*) , 101, 2e-20
220	EX567745	calmodulin binding (*Arabidopsis*), 95.1, 6e-22
221	EX567746	hypothetical protein OsJ_022416 , 70.9, 2e-11
222	EX567747	hypothetical protein OsI_015409 , 105, 3e-21
223	EX567748	Annexin, putative (*Medicago*) , 2 33 , 3e-75
224	EX567749	Acyl-CoA-binding (*Panax)*, 149, 4e-35
225	EX567750	hypothetical protein OsI_020638 , 99.4, 1e-19
226	EX567751	mtACP-1 (*Arabidopsis)*, 178, 2e-43
227	EX567752	cytochrome P450 (*Nicotiana*) , 93.6, 3e-18
228	EX567753	No significant similarity
229	EX567754	Os07g0662500 (*Oryza sativa*) , 134, 2e-30
230	EX567755	Xyloglucan endotransglycosylase (*Populus*), 73.6, 3e-12
231	EX567756	hypothetical protein (*Medicago)*, 79.7, 8e-14
232	EX567757	dUTP diphosphatase/ hydrolase (*Arabidopsis)*, 219, 4e-56
233	EX567758	cytochrome-c oxidase (*Arabidopsis)*, 59.3, 5e-12
234	EX567759	Glycosyl hydrolases (*Oryza*) , 282, 7e-79
235	EX567760	ETC complex I subunit (*Medicago)*, 97.8, 4e-42
236	EX567761	No significant similarity
237	EX567762	No significant similarity
238	EX567763	No significant similarity
239	EX567764	alpha-L-arabinofuranosidase (*Malus domestica*), '103, 3e-21
240	EX567765	Polygalacturonase, 302, 6e-81
241	EX567766	PGR5 (*Portulaca grandiflora*), 97.8, 2e-28
242 •	EX567767	Inorganic pyrophosphatase (*Medicago)*, 97.4, 2e-19
243	EX567768	Tetratricopeptide-like helical (*Medicago)*, 106, 5e-22
244	EX567769	putative protein (*Arabidopsis*), 72.4, 2e-11
245	EX567770	No significant similarity
246	EX567771	HCF106 [*Pisum sativum*], 87, 3e-16
247	EX567772	hypothetical protein NitaMp, 93.2, 9e-25
248	EX567773	unknown protein (*Oryza sativa*) , 131, 9e-31
249	EX567774	putative protein (*Arabidopsis thaliana*), 50, 3e-10
250		
251	EX151623	CCT (*Medicago truncatula)*, 157, 7e-46
252	EX151624	Ribosomal protein S4 (*Medicago*), 97.4, 2e-19
253	EX151625	No significant similarity
254	EX151626	putative DNA-damage-repair. (*T. Pratense*), 382, 7e-105
255	EX151627	auxin efflux carrier (*Oryza sativa*) , 95.1, 8e-19
256	EX151628	No significant similarity
257	EX151629	beta-D-galactosidase (Pyrus), 244, 3e-63
633	EX151963	hypothetical protein (*M. truncatula*), 118, 7e-27
634	EX151964	No significant similarity
635	EX151965	No significant similarity
636	EX151966	hypothetical protein AC124956g13v2 (*M. truncatula*), 162, 1e-38
637	EX151967	unknown protein [*Arabidopsis thaliana*), 167, 3e-40
638	EX151968	glutathione transferase (Arabidopsis), 157, 7e-40
639	EX151969	unknown protein [*Oryza sativa*) , 138, 1e-31
640	EX151970	hypothetical protein [*Cucumis melo*), 64.7, 9e-13
641	EX151971	MADS box protein AGL11 [*Lotus corniculatus var. japonicus*), 252, 7e-66
642	EX151972	putative casein kinase [*Oryza sativa*) , 124, 2e-27
643	EX151973	unknown [*Prunus armeniaca*), 70.1, 3e-16
644	EX151974	Chlorophyll A-B binding protein [*Medicago truncatula*], 111, 3e-23
645	EX151975	No significant similarity
646	EX151976	Ribosomal L22e protein (*M. truncatula*), 1 30, 3e-29
647	EX151977	2OG-Fe(II) oxygenase (*M. truncatula*), 218, 1e-55
648	EX151978	photosystem II protein I (*Ranunculus*) , 78.2, 1e-13
649	EX151979	unnamed protein product [*Oryza sativa*), 115, 5e-25
650	EX151980	glutathione S-transferase GST 25 [*Glycine max*], 79.7, 4e-14
651	ES544488	No significant similarity
652	EX151981	No significant similarity
653	EX151982	cytochrome c oxidase (*Solanum tuberosum*), 176, 9e-43
654	EX151983	Auxin responsive SAUR protein (*M. truncatula*), 153, 5e-36
655	EX151984	ribosomal protein S13 [*Glycine max*), 266, 7e-70
656	EX151985	No significant similarity
657	EX151986	conserved hypothetical protein (*M. truncatula*), 96.7, 3e-19
658	EX151987	ribosomal protein L37 [*Glycine max*), 182, 1e-44
659	EX151988	NmrA-like family protein [*Solanum demissum)*, 149, 5e-35
660	EX151989	Oligosaccharyl transferase, STT3 subunit (*Medicago)* , 87.4, 2e-16
661	EX151990	Peptidyl-prolyl cis-trans isomerase (*M. truncatula*), 117, 5e-25
662	EX151991	notchless-like protein [*Solanum chacoense]*, 65, 9e-11
663	EX151992	OSH1 (Oas High Accumulation) , 200, 4e-50
664	EX151993	translocon-associated protein [*Solanum tuberosum*), 124, 7e-38
665	EX151994	unknown protein [*Arabidopsis thaliana*), 61.6, 2e-20
666	EX151995	El protein and Def2/Der2 allergen [*Medicago truncatula*], 177, 2e-43
667	EX151996	unknown protein [*Arabidopsis thaliana*), 115, 8e-25
668	EX151997	ribosomal protein S7-like protein [*Solanum tuberosum*), 178, 9e-67
669	EX151998	ribosome-associated protein [*Cicer arietinum*], 276, 6e-73
670	EX151999	iron-superoxide dismutase [*Pisum sativum*), 172, 1e-41
671	EX152000	RNA binding protein-like (*Arabidopsis thaliana*), 91.7, 1e-17
672	EX152001	Longin-like [*Medicago truncatula*), 179, 9e-44
673	EX152002	hypothetical protein OsI_024435 [*Oryza sativa*), 86.3, 5e-16

258	EX151630	S-adenosylmethionine-dep methyltranferase (*A. thaliana*), 202, 9e-51	674	EX152003	AT3g45010/F14D17_80 [*Arabidopsis thaliana*], 110, 3e-23
259	EX151631	unknown [*Arabidopsis thaliana*), 161, 2e-38	675	EX152004	Nucleic acid-binding, OB-fold (*Medicago truncatula*), 223, 1e-69
260	EX151632	No significant similarity	676	EX152005	No significant similarity
261	EX151633	Bet v I allergen [*Medicago truncatula*), 92, 2e-17	677	EX152006	SAM (and some other nucleotide) [*Medicago truncatula*), 340, 2e-92
262	EX151634	NADH-ubiquinone oxidoreductase (*Medicago*), 205, 4e-52	678	EX152007	unknown protein (*Arabidopsis thaliana*), 85.5, 4e-17
263		unknown (*Arabidopsis thaliana*), 109, 9e-23	679	EX152008	unknown [*Arabidopsis thaliana*),116, 7e-25
264	EX151635	unknown protein (*Arabidopsis*) , 117, 1e-25	680	EX152009	Ribosomal protein L10E (*M. truncatula*), 355, 1e-102
265	EX151636	lipid trans/seed-alpha amylase (*Medicago*), 104, 2e-21	681	EX152010	Ribosomal protein S16 (*M. truncatula*), 219, 1e-58
266	EX151637	HCNGP-like (*Medicago*) , 66.2, 9e-24	682	EX152011	xyloglucan endotransglycosylase hydro (*Apium graveolens*), 150, 5e-35
267	EX151638	No significant similarity	683	EX152012	Ribosomal protein L29 (*M. truncatula*), 218, 7e-56
268	EX151639	unknown protein (*Arabidopsis*), 92, 7e-18	684	EX152013	hypothetical protein MtrDRAFT_AC15 (*M. truncatula*), 73.9, 2e-12
269	EX151640	putative protein (*Arabidopsis*), 76.6, 1e-12	685	EX152014	No significant similarity
270	EX151641	ribosomal protein s3 (Lotus), 165, 7e-40	686	EX152015	No significant similarity
271	EX151642	Sugar transporter superfamily (*Medicago*), 293, 9e-80	687	EX152016	Zinc fingerRanBP2-type (*M. truncatula*), 164, 2e-39
272	EX151644	putative BY-2 cell cycle-related (*N. tabcum*), 180, 4e-44	688	EX152017	No significant similarity
273	EX151645	Remorin, C-terminal region (*Medicago*), 99, 6e-20	689	EX152018	Protein of unknown function DUF640 (*M. truncatula*), 249, 5e-65
274	EX151646	vacuolar H+-ATPase (*Hyacinthus orientalis*), 70.5, 2e-11	690	EX152019	MFP1 attachment factor 1 [*Glycine max*), 73.6, 7e-12
275	EX151647	Glycoside hydrolase(*Medicago*), 231, 1e-59	691	EX152020	No significant similarity
276	EX151648	unknown protein (*Arabidopsis*), 115, 4e-25	692	EX152021	ZIM [*Medicago truncatula*), 131, 3e-29
278	EX151649	Protein kinase (*Medicago*), 186, 6e-46	693	EX152022	cytochrome c oxidase subunit I [*Contracaecum osculatum*)
279	EX151650	cytochrome-c oxidase [*Arabidopsis*), 90.5, 6e-17	694	EX152023	ATVAMP714 (*Arabidopsis*), 201, 3e-59
280	EX151651	No significant similarity	695	EX152024	unknown protein (*Arabidopsis*), 134, 3e-30
281	EX151652	Protease-associated (*Medicago*) , 178, 7e-44	696	EX152025	No significant similarity
282	EX151653	2OG-Fe(II) oxygenase (*Medicago*), 103, 4e-21	697	EX152026	No significant similarity
283	EX151654	lipid transfer protein (*Cicer*)	698	EX152027	No significant similarity
284	EX151655	No significant similarity	699	EX152028	conserved hypothetical protein (*M. truncatula*), 120, 2e-26
285	EX151656	photosystem I reaction center (*Medicago*), 122, 4e-27	700	EX152029	hypothetical protein [*C. arietinum*), 117, 2e-48
286	EX151657	No significant similarity	701	EX152030	anaphase promoting complex [*Arabidopsis thaliana*], 194, 2e-48
287	EX151658	unknown protein [*Arabidopsis*), 100, 2e-20	702	EX152031	Myo-inositol-1-phosphate synthase [*Medicago truncatula*), 412, 5e-114
288	EX151659	histone H2B (*Cicer arietinum*), 190, 4e-47	703	EX152032	Myo-inositol-1-phosphate synthase [*Medicago truncatula*), 412, 5e-114
289	EX151660	Armadillo-like helical (*Medicago*), 89.4, 4e-17	704	EX152033	No significant similarity
290	EX151661	cytochrome-c oxidase (Arabidopsis), 57, 9e-11	705	EX152034	Ferredoxin [2Fe-2S][*Medicago truncatula*], 199, 1e-49
291	EX151662	hypothetical protein MtrDRAFT, 72.8, 2e-24	706	EX152035	structural constituent of ribosome (*Arabidopsis thaliana*], 124, 1e-27
292	EX151663	No significant similarity	707	EX152036	carrier/ steroid binding [Arabidopsis thaliana] , 61.6, 3e-08
293	EX151664	unknown protein [*Oryza sativa*) , 137, 2e-31	708	EX152037	putative beta-glucosidase [Oryza sativa), 108, 2e-22
294	EX567512	receptor-like kinase RHG1 [*Glycine*), 51.6, 1e-05	709	EX152038	No significant similarity
295	EX151665	heat shock protein binding (*A. thaliana*),132, 1e-31	710	EX152039	No significant similarity
296	EX151666	Cytochrome b5 (*Medicago*), 196, 9e-49	711	EX152040	unknown protein (*Arabidopsis thaliana*), 105, 1e-21
297	EX151667	ribosomal protein S18 (*O.sativa*), 87, 3e-17	712	EX152041	No significant similarity
298	EX151668	Glutamine synthetase(*M.sativa*), 350, 9e-96	713	EX152042	No significant similarity
299	EX151669	Ribosomal protein L15 (*Medicago*), 80.1, 5e-23	714	EX152043	RING-box protein (*Arachis hypogaea*), 215, 6e-55
300	EY457881	40S ribosomal protein S8 (*Medicago*), 169, 4.00E-43	715	EX152044	No significant similarity
301	EX151670	unnamed protein product (*Arabidopsis*), 75.5, 2e-15	716	EX567531	ATP sulfurylase [*Glycine max*), 156, 3e-37

302	EX151671	Lissencephaly type-1-like (*Medicago*), 225, 8e-58
303	EX151672	No significant similarity found
304	EX151673	Peptidase M20 (*Medicago truncatula*), 121, 2e-49
305	EX151674	cytochrome-c oxidase (*Arabidopsis*), 153, 3e-36
306	EX151675	lysosomal Pro-X carboxypeptidase (*Arabidopsis*), 270, 3e-71
307	EX151676	Tyrosyl-tRNA synthetase (*Nicotiana*) , 85.1, 8e-16
308	EX151677	ribosomal protein S4 (Lotus), 173, 6e-42
309	EX151678	60S ribosomal protein L21 (*O. sativa*, 124, 1e-27
310	EX151679	pyrimidine 5'-nucleotidase –related (*Oryza*) , 154, 4e-36
311	EX151680	HMG-I and HMG-Y, DNA-binding (*Medicago*), 150, 6e-35
312	EX151681	ribulose-phosphate 3-epimerase (*Medicago*), 154, 1e-36
313	EX151682	nucleoside diphosphate kinase (*Oryza sativa*) , 106, 6e-39
314	EX151683	No significant similarity
315	EX151684	SMC3 putative chromosome associated (*Arabidopsis*), 98.2, 2e-19
316	EX151685	cationic peroxidase (*Cicer arietinum*), 127, 2e-28
317	EX151686	acetyl-CoA carboxylase (*Glycine*) , 57.8, 2e-14
318	EX151687	hypothetical protein (*Cleome spinosa*), 89.4, 9e-17
319	EX151688	unknown protein (*Arabidopsis*),79, 1e-13
320	EX151689	No significant similarity
321	EX151690	flowering locus T like protein [(*Populus nigra*) , 273, 6e-72
322	EX151691	No significant similarity
323	EX151692	ferredoxin-thioredox (*Solanum*), 115, 2e-24
324	EX151693	Ribosomal protein S21e (*Medicago*), 137, 1e-31
325	EY457878	starch synthase-related protein (*Vicia*)
326	EY457879	alcohol dehydrogenase-lik (*Arabidopsis*),
327	EX151694	Peptidase M20 (*Medicago truncatula*), 112, 7e-24
328	EX151695	Annexin, putative (*Medicago truncatula*, 165, 1e-39
329	EX151696	core protein [*Pisum sativum*), 176, 6e-43
330	EX151697	Sterol Methyltransferase (*A. thaliana*), 243, 4e-66
331	EX151698	auxin-amidohydrolase (*Populus*) , 218, 2e-55
332	EX567513	60S ribosomal protein L21, (*O. sativa*), 119, 2e-35
333	EX151699	Regulator of chromosome condensation (*Medicago*), 181, 2e-5
334	EX151700	Nonaspanin (*Medicago truncatula)*, 70.9, 2e-11
335	EX151701	bHLH transcription factor (*A. thaliana*), 99, 6e-20
336	EX151702	unknown protein (*Arabidopsis thaliana*), 107, 2e-22
337	EX151703	Plastid Ribosomal Protein , 233, 4e-60
338	EX151704	Transcriptional factor B3 (*Medicago truncatula*), 95.9, 7e-19
339	EX151705	No significant similarity
340	EX151706	unknown protein (*Phytophthora sojae*), 127, 1e-47
341	EX151707	unknown protein (*Oryza sativa*) , 67.8, 2e-22
342	EX151708	ripening-related protein (*Arabidopsis*), 143, 7e-33
343	EX151709	ripening-related protein (*Arabidopsis*), 143, 7e-33
717	EX152045	No significant similarity
718	EX152046	putative VASA [Oryza sativa), 74.7, 3e-12
719	EX152047	No significant similarity
720	EX152048	No significant similarity
721	EX152049	No significant similarity
722	EX152050	Helicase, C-terminal, CCHC-type (*M. truncatula*), 316, 6e-85
723	EX152051	Histone H2A (*M. truncatula*), 110, 4e-23
724	EX152052	protein phosphatase type 2C [*Lotus japonicus*], 246, 3e-84
725	EX152053	At1g01540/F22L4_6 [*Arabidopsis thaliana*], 174, 3e-42
726	EX152054	No significant similarity
727	EX152055	developing seed L-asparaginase [*Lupinus angustifolius*], 283, 6e-82
728	EX152056	unknown protein [*Arabidopsis thaliana*], 58.5, 1e-10
729	EX152057	acetolactate synthase [*Solanum ptychanthum*], 278, 7e-84
730	EX152058	putative photosystem II protein (*Gossypium hirsutum*), 86.7, 4e-16
731	EX152059	ribonuclease T2 [*Cicer arietinum*), 369, 7e-103
732	EX152060	aspartate aminotransferase, 127, 2e-28
734	EX152061	Ribosomal L18ae protein (*M. truncatula*), 338, 8e-92
735	EX152062	UBC19; ubiquitin conjugating enzyme [*Arabidopsis thaliana*], 197, 2e-57
736	EX152063	ribosomal protein S12 [*Morus indica*], 127, 2e-28
737	EX152064	SAR1 (RAS) [*Arabidopsis thaliana*], 265, 2e-69
738	EX152065	No significant similarity
739	EX152066	putative protein (*Arabidopsis*), 65.5, 9e-10
740	EX152067	No significant similarity
741	EX152068	No significant similarity
742	EX152069	hypothetical protein MtrDRAFT_AC12 , 172, 5e-42
743	EX152069	No significant similarity
744	EX152071	No significant similarity
745	EX152072	AP2/EREBP transcription factor ERF-2 [*Gossypium hirsutum*], 152, 9e-36
746	EX152073	ribosomal protein L2 [*Glycine max*), 372, 5e-102
747	EX152074	S25 ribosomal protein (*M. truncatula*), 131, 1e-29
748	EX152075	No significant similarity
749	EX152076	unknown [*Arabidopsis thaliana*), 192, 8e-48
750	EX152077	No significant similarity
751	EX152078	molybdenum cofactor sulfurase-like (*O. sativa*) , 99.8, 6e-20
752	EX152079	No significant similarity
753	EX152080	No significant similarity
754	EX152081	conserved hypothetical protein (*M. truncatula*), 118, 1e-25
755	EX152082	conserved hypothetical protein (*M. truncatula*), 118, 1e-25
756	EX152083	copper chaperone homolog CCH (*G. max*), 152, 8e-36
757	EX152084	unknown protein (*Oryza sativa*), 202, 9e-51
758	EX152085	Ribosomal protein S13 [*M. truncatula*], 288, 1e-76
759	EX152086	GPX12Hv, glutathione peroxidase-like protein (*Hordeum vulgare*), 77.4, 2e-13

344	EX151710	unknown protein (*Arabidopsis*), 84.3, 4e-15
345	EX151711	60S ribosomal protein L1 (*Prunus*), 237, 4e-61
346	EX151712	Homeodomain-related (*Medicago truncatula*), 251, 2e-65
347	EX151713	T1N15.24 (*Arabidopsis*), 55.8, 2e-20
348	EX151714	Zinc finger, RING-type (*Medicago truncatula*), 88.2, 1e-16
349	EX151715	FK506 binding / peptidyl-prolyl (*Arabidopsis*), 161, 2e-38
350	EX151716	mitochondrial uncoupling (*Saccharum officinarum*), 79.3, 4e-28
351	EX151717	At3g24570 (*Arabidopsis*), 108, 1e-22
352	EX151718	Phenazine biosynthesis (*Medicago truncatula*), 223, 2e-63
353	EX151719	nucleic acid binding(*Arabidopsis*), 80.1, 3e-14
354	EX151720	Putative CGI-126 protein (*Arabidopsis*), 353, 3e-96
355	EX151721	DNA binding protein S1FA (*Medicago truncatula*), 72, 8e-12
356	EX151722	Glutathione peroxidase (*Medicago truncatula*), 126, 8e-43
357	EX151723	ribosomal protein S26 [*Pisum sativum*], 169, 1e-40
358	EX151724	unknown [*Glycine max*), 124, 1e-27
359	EX151725	Os01g0835900 (*Oryza sativa*) , 157, 4e-37
360	EX151726	UBC9 (*A.. thaliana*) , 246, 5e-64
361	EX151727	No significant similarity found
362	EX151728	CDC25 (*Arabidopsis*), 210, 5e-53
363	EX151729	Nucleic acid-binding (*Medicago truncatula*), 230, 3e-59
364	EX151730	Ribosomal protein (*Arabidopsis*), 149, 6e-35
365	EX151731	Plastid-specific 30S ribosomal (*Spinacia*), 84, 2e-15
366	EX151732	40S ribosomal (*Capsicum annum*), 147, 1e-46
367	EX151733	26S proteasome subunit (*Medicago truncatula*), 75.1, 1e-16
368	EX151734	No significant similarity
369	EX567514	Rhodopsin-like GPCR (*Medicago truncatula*), 110, 3e-23
370	EX567515	At3g17600/MKP6_15(*Arabidopsis*), 90.5, 5e-17
371	EX567516	No significant similarity
372	EX151735	hypothetical protein (*Arabidopsis*), 84, 4e-15
373	EX151736	No significant similarity
374	EX151737	No significant similarity
375	EX151738	PSI light-harvesting antenna chlorophyll (Pea), 71.2, 1e-11
376	EX151739	microtubule associated protein (*Cicer*), 204, 2e-51
377	EX151740	RPB5d [Brassica napus OR, 154, 2e-36
378	EX567517	Ribosomal L22e protein (Arabidopsis), 132, 3e-22
379	EX567518	Ribosomal L22e protein (*M. truncatula*), 164, 3e-39
380	EX151741	2OG-Fe(II) oxygenase (*Medicago truncatula*), 160, 6e-38
381	EX151742	AGG2 (G-PROTEIN GAMMA SUBUNIT 2) (*Arabidopsis*), 78.2, 3e-15
382	EX151743	No significant similarity found
760	EX152087	ubiquinol-cytochrome-c reductase [*Arabidopsis thaliana*], 137, 2e-31
761	EX152088	No significant similarity
762	EX152089	NADH dehydrogenase [*Arabidopsis thaliana*], 235, 1e-60
763	EX152090	disease-resistent-related protein (*Oryza sativa)*, 171, 1e-41
764	EX152091	cAMP response element binding (CRE (*Medicago truncatula*), 229, 1e-58
765	EX152092	H+-transporting two-sector ATPase, (*Medicago truncatula*), 172, 1e-41
766	EX152093	Dual specificity protein phosphatase [*Medicago truncatula*], 189, 8e-47
767	EX152094	Metallothionein-like protein 2 ((*M. truncatula*), 126, 5e-28
768	EX152095	Aldehyde dehydrogenase (*M. truncatula*), 227, 3e-58
769	EX152096	No significant similarity
770	EX152097	hypothetical protein MtrDRAFT_AC14 (*M. truncatula*), 71.2, 1e-11
771	EX152098	developing seed L-asparaginase [*Lupinus angustifolius*], 128, 2e-28
772	EX152099	No significant similarity
773	EX152100	conserved hypothetical protein (*M. truncatula*), 61.2, 1e-16
774	EX152101	Os05g0524300 (*Oryza sativa*), 158, 1e-39
775	EX152102	light harvesting protein, 188, 3e-74
776	EX152103	AT5g08050/F13G24_250 [*Arabidopsis thaliana*], 125, 1e-27
777	EX152104	hypothetical protein [*Cucumis melo*), 105, 9e-22
778	EX152105	unknown protein (*Arabidopsis thaliana*), 117, 5e-42
779	EX152106	conserved hypothetical protein [*M. truncatula*], 270, 5e-71
780	EX152107	nutrient reservoir [*Arabidopsis*], 337, 2e-91
781	EX152108	No significant similarity
782	EX152109	No significant similarity
783	EX152110	unknown protein [*Arabidopsis thaliana*), 112, 1e-23
784	EX152111	No significant similarity
785	EX152112	hypothetical protein [*Vitis vinifera*), 47.4, 6e-04
786	EX152113	ADP-ribosylation factor (*Oryza sativa*), 175, 2e-42
787	EX152114	IS1 ORF2 [*Shigella*), 100, 1e-22
788	EX152115	hypothetical protein (*Arabidopsis*), 143, 6e-33
789	EX152116	DSBA oxidoreductase (*M. truncatula*), 96.3, 5e-19
790	EX152117	Nonspecific lipid-transfer protein (*Cicer arietinum*) , 169, 2e-40
791	EX152118	No significant similarity
792	EX152119	elongation factor 2 (*Triticum*), 209, 4e-85
793	EX152120	No significant similarity
794	EX152121	No significant similarity
795	EX152122	No significant similarity
796	EX152123	GTP-binding signal recognition (*M. truncatula*),255, 8e-71
797	EX152124	Annexin, putative (*M. truncatula*), 150, 3e-35
798	EX152125	unknown protein [*Arabidopsis thaliana*], 133, 8e-30

383	EX151744	No significant similarity found	799	EX152126	hypothetical protein (*Cicer arietinum*), 129, 6e-29
384	EX567519	Cyclin-like F-box (*Medicago truncatula*), 140, 3e-32	800	EX152127	No significant similarity
385	EX567520	Superoxide dismutase (*Medicago truncatula*), 149, 9e-72	801	EX152128	No significant similarity
386	EX567521	calcium binding protein (*Fagus sylva*) , 94.7, 2e-33	802	EX152129	No significant similarity
387	EX567522	Rhomboid-like protein (*Medicago*) , 102, 6e-40	803	EX152130	Proteasome subunit , 71.6, 1e-11
388	EX151745	No significant similarity	804	EX152131	ankyrin-like protein [*A. thaliana*], 182, 1e-44
390	EX151746	26S proteasome ATPase (*A. thaliana*), 103, 3e-21	805	EX152132	aspartyl aminopeptidase-like [*Arabidopsis thaliana*], 116, 3e-25
391	EX151747	No significant similarity	806	EX152133	No significant similarity
392	EX151748	2Fe-2S ferredoxin (*Medicago truncatula*), 122, 2e-49	807	EX152134	40S ribosomal protein (*Solanum*) , 192, 2e-47
393	EX151749	Enoyl-CoA hydratase/isomerase (*Medicago truncatula*), 184, 2e-45	808	EX152135	No significant similarity
394	EX151750	No significant similarity	809	EX152136	hypothetical protein (*Oryza*) , 109, 2e-66
395	EX151751	Enod8-like protein (*Medicago truncatula*), 88.6, 3e-21	810	EX152137	No significant similarity
396	EX151752	TO92a-2rc (*Taraxacum officinale*), 74.7, 2e-12	811	EX152138	CAT7 (Cationic Amino acid Transporter [*Arabidopsis thaliana*], 130, 2e-29
397	EX151753	unknown protein [(*Arabidopsis*), 73.9, 3e-12	812	EX152139	ribosomal protein S19 [*Glycine max*}, 120, 3e-26
398	EX151754	starch phosphorylase (*V. faba)*, 120, 4e-26	813	EX152140	unknown protein [*A. thaliana*], 69.7, 5e-11
399	EX151755	unknown protein [*Arabidopsis thaliana*), 104, 1e-26	814	EX152141	No significant similarity
400	EX151756	unknown protein (*Arabidopsis*), 220, 3e-56	815	EX152142	No significant similarity
401	EX567862	putative transcription factor EREBP , 90.9, 5e-35	816	EX152143	No significant similarity
402	EX151757	Calreticulin/calnexin [*M. truncatula*), 114, 2e-31	817	EX567532	catalytic/ hydrolase (*Arabidopsis*) , 69.3, 6e-11
403	EX151758	Ras GTPase [*Medicago truncatula*), 323, 3e-87	818	EX567840	sts14 pistil-specific (*Solanum Tuberosum*), 60, 7e10
404	EX151759	hypothetical protein [*Citrus x paradi*), 219, 4e-66	819	EX567957	unnamed protein product [Arabidopsis thaliana] 50, 3.00E-10
405	EX151760	Ferritin-2, (*Vigna*) , 187, 5e-46			

Table 4.3: List of the most abundant reads obtained from the 20 DAA cDNA library and their Blast homologies.

No. of reads	Blast homology	GenBank No.	Best e- value
49	No significant similarity		
38	hypothetical protein *Oryza sativa*	BAD07869	6.00E-18
27	Nonspecific lipid-transfer protein *Cicer arietinum*	O23758	6.00E-59
27	hypothetical protein NitaMp027 *Nicotiana*	YP_173374	5.00E-39
20	seed specific protein Bn15D18B *Brassica napus*	AAP37971	6.00E-13
19	protease inhibitor *Glycine*	AAC97524	3.00E-16
14	No significant similarity	-	-
12	mitogen-activated protein kinase *Citrus sinensis*	ABM67698	0.0
12	Photosystem II reaction center *Spinacia oleracea*	Q41387	4.00E-23
11	chlorophyll a/b binding protein *Cicer arietinum*	CAA10284	2.00E-126
10	putative serine carboxypeptidase *Pisum sativum*	CAC19488	5.00E-70
10	hypothetical protein OsJ_009198 *Oryza sativa*	EAZ25715	2.00E-15

Chapter 5: Development of EST-SSR markers in chickpea

5.1 Introduction

Among the molecular markers, STMS (Sequenced Tagged Microsatellite Sites) have emerged as a promising source for marker assisted selection and map-based cloning especially in highly inbreeding species such as chickpea where genetic polymorphism is very low (Udupa et al. 1993; Labdi et al. 1996). In chickpea, microsatellite repeats are known to be abundant as well as efficient in detecting genetic variability (Sharma et al. 1995b; Hüttel et al. 1999; Winter et al. 1999; Udupa et al. 1999; Lichtenzveig et al. 2005). For isolation of microsatellites in chickpea, different strategies have been applied worldwide including conventional genomic library screening procedure (Hüttel et al.1999; Winter et al. 1999; Sethy et al. 2003), screening BAC libraries with synthetic oligonucleotides (Lichtenzveig et al. 2005) and construction of microsatellite enriched-genomic library (Sethy et al. 2006a). These approaches have led to the development of a total of 694 chickpea STMS markers, of which 265 emanated from our laboratory (unpublished results). Nonetheless, the available microsatellite markers for chickpea are still insufficient for constructing dense linkage map and to tag genes related to traits of interest. Moreover, whilst the SSRs developed till date are excellent markers for providing a linkage framework, they are most likely developed from the non-coding regions of the genome and thereby cannot directly pinpoint genes of known function or traits of economic importance. In this regard, the swelling EST databases have emerged as a potential source of microsatellite markers more so, since they are derived from coding regions that are relatively well conserved among taxa (Morgante et al 2002) Moreover, the SSRs have been reported to be more abundant in transcribed regions than in the non-transcribed DNA (Morgante et al. 2002; Fujimori et al. 2003). Thus, ESTs are an excellent, inexpensive, alternative resource for mining SSRs in species where a plethora of EST sequences are available.

Today, EST-SSR markers have been reported for a number of plant species such as rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001), wheat (Gupta et al. 2003), barley (Thiel et al. 2003), *Medicago* (Eujayl et al. 2004), coffee (Poncet et al. 2006; Aggarwal et al. 2007), pepper (Yi et al. 2006) and citrus (Chen et al. 2006). Similar to genomic SSRs, EST-SSRs have proved to be profoundly useful for many applications in plant genetics and breeding such as molecular mapping, genetic diversity analysis and cross transferability across related species and genera (Varshney et al. 2005a). Moreover, as a result of their association with coding sequences, they provide the possibility of direct gene tagging for QTL mapping of agronomically important traits. In chickpea, limited EST sequences are

available in the public database and these resources have remained unutilized till the initiation of the present study (except Buhariwalla et al. 2005 who reported only 14 EST-SSR markers). Thus there was an urgent need to expand the genomic resources of chickpea especially ESTs (undertaken in the previous chapter 4) and utilize them as a source of markers in order to achieve a significant progress towards function-associated chickpea breeding.

Hence, the present chapter reports the development of STMS markers from the genic regions of chickpea. For the large-scale development of EST-SSR markers from chickpea, both the available NCBI chickpea EST database as well as the in-house developed seed-specific ESTs were mined for SSRs. Further, the pattern of distribution of SSRs in the chickpea transcribed regions was also analyzed.

5.2 Results

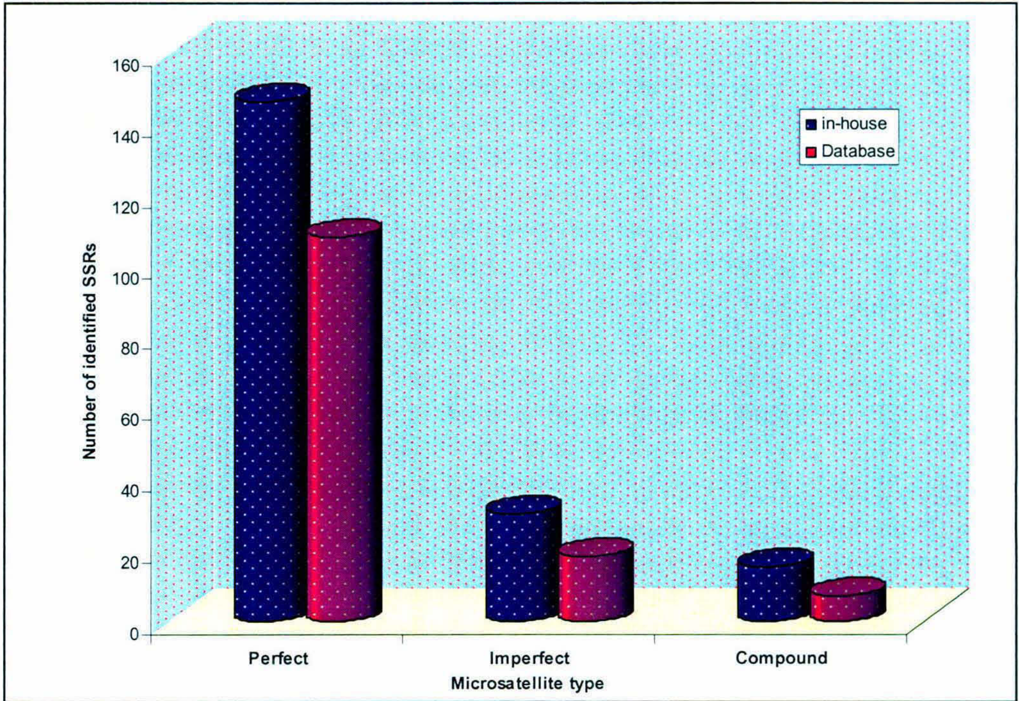
5.2.1 Identification of microsatellites from chickpea ESTs

ESTs from the two sources namely the publicly available chickpea ESTs in the NCBI database and the seed-specific ESTs generated in this study were used for development of the EST-SSR markers. Firstly, using the 1309 chickpea EST sequences reported in the NCBI database till January 2007, representing approximately 0.76Mb, 133 microsatellite motifs were identified (based on the criteria described in Material and Methods section 3.2.12) in 117 non-redundant sequences. Secondly, from the 1037 unigenes generated in-house (reported in the previous chapter), a total of 191 SSRs were identified in 167 ESTs (SSR-ESTs) comprising 16.0% of total EST sequences. Twenty SSR-EST sequences contained ≥ 2 microsatellite motifs. Hence from these two EST resources, a total of 324 SSR motifs (133 +191) were identified in 284 (117 + 167) ESTs.

Sequence analysis of these 324 SSRs revealed that the maximum 254 (78.3%) were perfect repeats whereas 48 (14.8%) were imperfect and 22 (6.7%) were compound (Fig. 5.1 A). The copy number of the dinucleotide repeat motifs at the perfect loci varied from 5 to 20 and the trinucleotide motifs from 4 to 14. A diverse range of SSR motifs was present which varied widely with trinucleotide repeats (51.5%) being the most abundant followed by di- (38.8%), tetra- (4.9%) and pentanucleotide (4.6%) motifs (Fig. 5.1 B). The most frequently occurring dinucleotide motifs were GA (75.5%) followed by AT (15.3%) and GT (9.0%) whereas no single GC motif was identified from both sources (Fig. 5.2 A). Among trinucleotide motifs, AAG (36.0%) was predominant followed by AAT (14.0%), ACC

Figure 5.1: Distribution of microsatellite motifs in *C. arietinum* ESTs **A)** on the basis of organization **B)** on the basis type of repeat motif

A)



B)

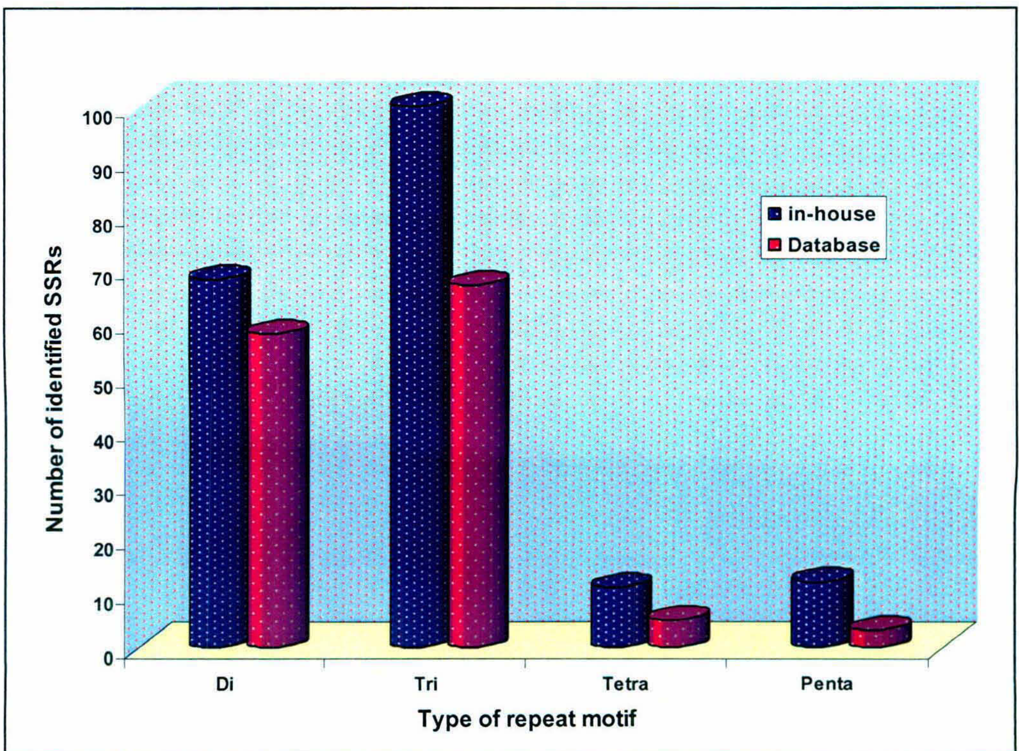
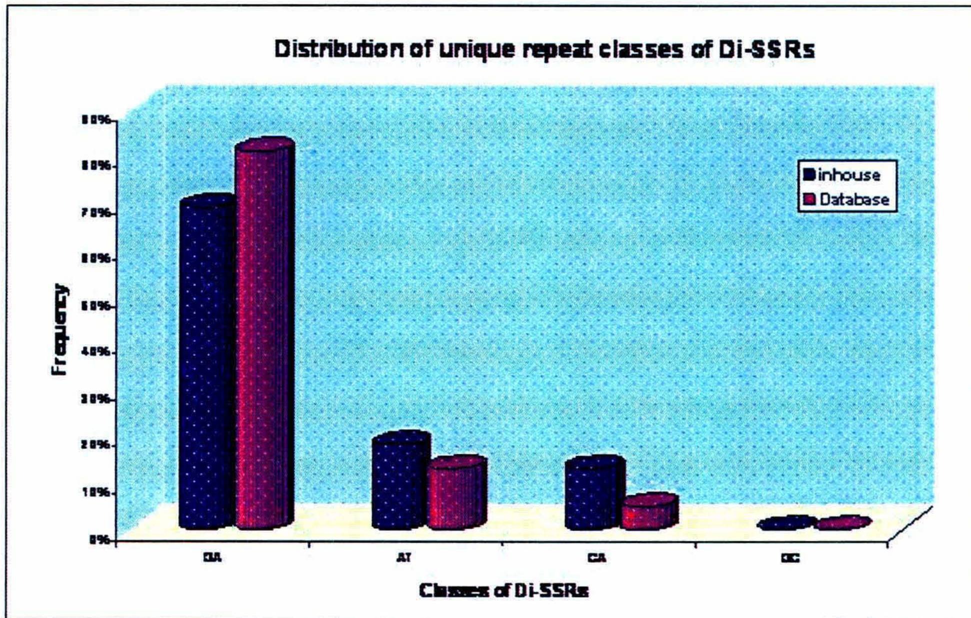
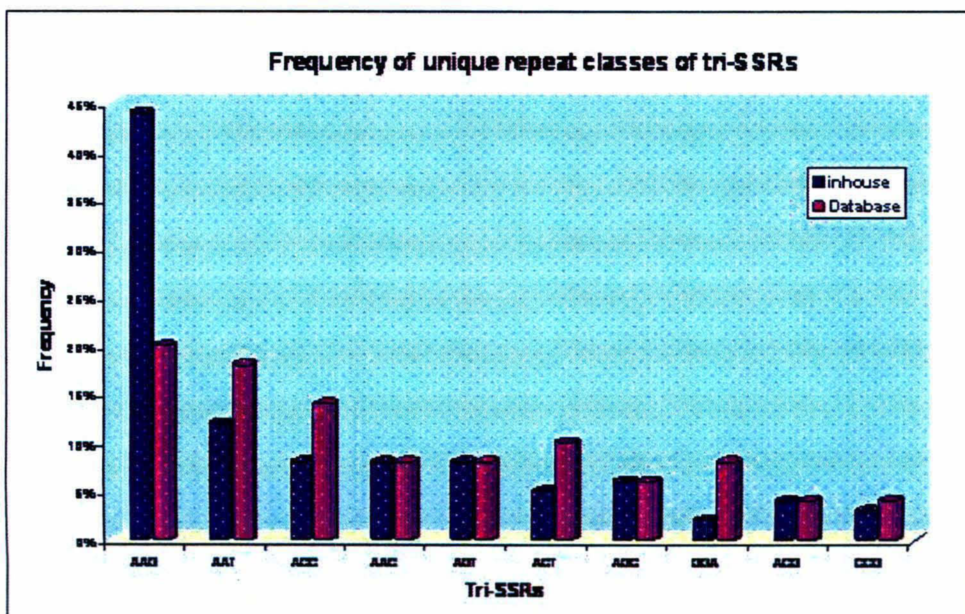


Figure 5.2: Frequency of unique repeat classes of SSRs in *C. arietinum* ESTs obtained from database and inhouse generated cDNA library **A)** of dinuclteotide motifs and **B)** of trinucleotide motifs

A)



B)



(10.0%), AAC (9.3%), AGT (8.0%), ACT (6.6%), AGC (6.0%), GGA and ACG (4.0%) with a minimal abundance of CCG motif (3.3%) (Fig. 5.2 B).

5.2.2 Development of functional chickpea EST-SSR markers

Of the identified 284 microsatellite containing EST sequences, 217 (83 + 134) primer pairs could be designed complimentary to the conserved sequences flanking microsatellite repeats (Fig. 5.3). For the remaining sequences, attrition was due to one of the following reasons: a) the small length of microsatellite containing EST sequences, b) the motifs were too close to the cloning sites of ESTs c) the flanking sequences were inappropriate for designing high-quality primer pairs (e.g. low GC content). Of the 217 primer pairs, 135 (48 +87) primers were finally synthesized and utilized in the present study. The 82 loci for which primers were not synthesized were mostly those EST sequences harboring dinucleotide motifs of only 5 repeat units, since the chances of polymorphism exhibited by them were very less especially in a self-pollinated crop like chickpea. All these 135 primer pairs were validated in *C. arietinum* cv ICCV2 and Pusa 362 and their sequences along with GenBank no. are listed in Table 5.1. This analysis yielded 97 functional primer pairs that amplified expected size fragments whereas of the remaining 38 primer pairs, 27 primers did not yield any amplification product and 11 gave higher size products (marked in Table 5.1) hence these were excluded from further analysis. Further of the 97 functional primer pairs, 86 primers amplified a single allele, 11 primers amplified 2 to 4 alleles and 9 primers produced a fragment somewhat larger than expected (Table 5.1), suggesting the presence of introns in the corresponding genomic DNA. Representative patterns of amplification obtained using the EST-SSR primers are shown in Fig. 5.4. Of the The BLASTX search for the ESTs (from which 135 primer pairs were designed) against the NCBI database revealed that 62.2% ESTs had significant homology to reported proteins, whereas 26.0% of them represented unknown/hypothetical proteins and 11.8% to novel sequences (Table 5.1).

5.3 Discussion

Nowadays, microsatellite markers have assumed great significance in biological research. They are becoming the preferred molecular markers for plant geneticists and are being widely applied for varietal identification, linkage mapping and marker-assisted selection. Since the abundance of microsatellites in plants is low in comparison to animals (Wang et al. 1994), the isolation of microsatellites from plant genomes is always technically demanding (Squirrell et al. 2003). However, to overcome these key obstacles for

Figure 5.3: Identification of repeat motifs and designing of STMS primers from chickpea ESTs **A)** CESSRDB16 **B)** CESSRDB27 and C) CESSR26. The microsatellite motifs are shown in bold red color and arrows indicate forward and reverse sequences selected for primer designing.

A) CESSRDB16:Genebank No. AJ487472

AGAGTGGGTTTTATGGAAAATCACCACTCACCAAACAGAACAATGCTATGCATGAT
GTTTCATCGGTTGCAGCAGCTATTGTTAGAGAATTACTCTAGATACTATACTTTCTC
AAACTTGTGTGCTCTCAAATAAAATATAAGAAAAAAATATATTGAAAAATTAT**TAT**
ATATATATATATATATATATGTAAAGTAACTTTTCAAATATATTTTTATTTATATTTTA
TTTCGAAATAAATATAT**TTATTATTATTAGTTAATTATTA**TATCAACTTAGTAGTACT
ATTAATTATGTATATATTTACTAATCTAATGCAATTGTTGTGTTTGTTTGGAACTAA
AAAAAAAAAAAAAA AAA

B) CESSRDB27:Genebank No. AJ271660

TCTAGAGCTATTACTTTGCTATCTGAAAACAATCCTCTTACTGGCACAAAGGGTGAG
ATTAGGAAGCAATGCAGTGTTGCCAACAAGCAGCACTTTGATGAACAACCTTGAAT
GGAGTTTCATATTCTAATGTCTCCAGACTATGTAAGA**TATTATTATTATTATTATTA**
TGCAAAAATAAAAGCTTTGGTCTTCATGGGTTGGTTTGTAAGGTCAATGAATTGAG
CCTTTGGAGGTTGTGGAGTTCATCTTATGGGGATTGGATAAATTATATGCTTTTAGT
TTTTAAT TAATCTTTTGAGATGTAACT

C) CESSR26:Genebank No. ES544483

AATAGTGGCAAAATCGAAATTCAACCCAACATGAAGGGCTATCATACTGAAGTT
GTCATATTCTCTC**CTTCTTCTTCTT**CCTTTATTAGCTATACCCAACCTTAAAACCA
AGCTTTTAATTTTGTCTCTCTTTTTTCTCTGTGTTTTTTTTTTTCTTTTAATTTTTTTT
ATATTGTTTATTTCTTCATTGGGTTTTGACCGTTGCTTTTTTTGTTAACTTCTTTCTC
TATTCACA**CTCTCTCTCTCTCTCTCTCT**GATTCCTTCTACACTGTCATTGATCAAA
AGCAANGTTTTGTTCTTGTTCTCTATTTACACAAAAAAAAAAAAAAAA

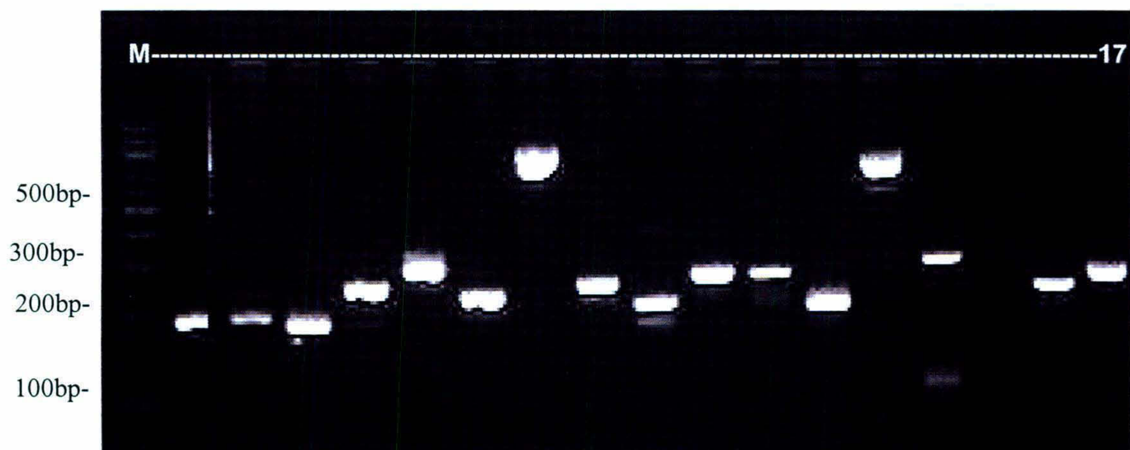


Figure 5.4: Validation of developed chickpea EST-SSR primers in chickpea acc. ICCV2. The amplified products were resolved on 3% Metaphor agarose gels. M indicates 100bp ladder and Lanes 1-17 represents amplified products obtained with chickpea EST-SSR markers.

microsatellite development, several enrichment methods have been developed (Zane et al. 2002) that increase the efficiency of microsatellite identification by upto 90% as compared to conventional microsatellite isolation techniques (Billotte et al. 1999; Zane et al.2002).

Initially the conventional approach of screening small insert plasmid genomic library was utilized for the isolation of microsatellites in chickpea that yielded 22 STMS markers by Hüttel et al. 1999; 174 by Winter et al. 1999 and another 10 by Sethy et al. 2003. Recently Lichtenzveig et al. 2005 also screened BAC libraries with synthetic oligonucleotides and developed a set of 233 chickpea STMS markers. With the advent of microsatellite enriched libraries (Zane et al. 2002), the task of developing more number of STMS markers was initiated in our laboratory that resulted in the development of 265 new chickpea STMS markers (74 were published and the remaining under personal communication are available for use). However despite the construction of genomic DNA libraries in chickpea, only a limited number of SSR markers (694) have been developed so far. Moreover, up till now, majority of the de novo developed STMS markers in chickpea belong to anonymous DNA fragments of unknown function that usually failed to reflect differences in genetic traits during germplasm evaluation.

With the recent escalating emphasis on functional genomics studies in several organisms, key focus has been put on the generation of functional markers. At present over three million sequences from approximately 200 plant species have been deposited in publicly available plant EST databases. These ESTs provide an attractive alternative source for mining SSRs to complement existing genomic SSR collections. Nowadays, development of SSR markers through data mining has become a fast, efficient and low-cost choice for plant species where large numbers of EST sequences are available, thus eliminating time-consuming and expensive steps of genomic library construction. Further, being associated with the transcribed regions of the genome, these markers are precious tools for breeding applications. Using this strategy, EST-SSRs have been identified and used for a variety of applications in a number of plant species like, *Arabidopsis* sp., (Areshchenkova and Ganal 2002), cotton (Han et al. 2004; Guo et al. 2007), grapes (Scott et al. 2000; Decroocq et al. 2003), *Medicago* (Eujayl et al. 2004), coffee (Poncet et al. 2006; Aggarwal et al. 2007), pepper (Yi et al. 2006), citrus (Chen et al. 2006) and cereals such as rice, barley, wheat, rye, tall fescue (Varshney et al. 2005b). However in chickpea, only about 1300 ESTs were publicly available (upto January 2007). Hence, our study utilized this resource for developing EST-SSR markers which finally yielded only 35 markers. As a result of the limited number

of publicly available chickpea ESTs, we were compelled to undertake generation of more new EST sequences (as described in chapter 1) and use them for the development of a novel set of EST–SSR functional markers. Our efforts resulted in the large scale development of EST-SSR markers in chickpea (first report in chickpea) which will not only be a significant addition to the limited set of SSR markers available in chickpea, but will have the added advantage of marker-trait associations.

For development of EST-SSR markers, both the publicly available NCBI database and in-house ESTs were mined for SSRs. A total of 324 SSRs were identified in 284 ESTs (since some ESTs had more than one repeat motif), which represented 13.8% of the screened ESTs. This SSR frequency was comparable with those obtained in citrus (10.6%, Chen et al. 2006), pepper (10.7%, Yi et al. 2006) and in other dicot species (Kumpatia and Mukhopadhyay 2005). However, Kantety et al. 2002 obtained comparatively lower frequency of EST-SSRs ranging from 1.5-4.7% in monocots. The abundance of SSRs mined from a sequence database depends on the SSR search criteria, the size of the dataset and the database mining tools (Varshney et al. 2005a). On applying stringent SSR criteria with a minimum of 20bp, about 5% of ESTs have been shown to contain SSRs in plants (Varshney et al. 2005a), whereas the same when applied to this set of chickpea EST sequences, only 3.1% sequences contained SSRs. Collectively, these observations clearly illustrate that EST sequences are a promising source of SSR discovery and concur with the observation of Morgante et al. 2002 that SSRs occur at a high frequency in the expressed regions of the plant genome. Further, the abundance of microsatellites in the expressed sequences of many species makes these markers very interesting because of a possible role in gene expression or function. It has been widely observed that microsatellites are present more frequently in 5' UTRs than in coding regions or 3'-UTRs (Wren et al. 2000; Morgante et al. 2002; Fujimori et al. 2003) and could have effect on the gene transcription and/ or regulation. For example, CAG repeat in a 5'UTR of human calmodulin-1 (*hCALM*1) gene when deleted causes a decrease in expression by 45% (Toutenhoofd et al.1998). Moreover, recent studies have substantially demonstrated that length of SSRs in coding regions might be associated with phenotypic variations. In an experimental study in rice, the variation in the number of GA or CT repeats in the 5' UTR of the waxy gene was found to be correlated with amylase content (Ayers et al. 1997; Bao et al. 2002).

The abundance of trinucleotide motifs in the chickpea coding sequences (51.5%) was in close agreement with observations in monocot and dicot plants (Kantety et al. 2002; Tian

et al. 2004 and Yi et al. 2006) establishing the need of the coding regions to maintain the reading frame (Varshney et al. 2002; Li et al. 2004). In contrast, the non coding regions of the eukaryotic genomes have been found to contain mainly dinucleotide repeats (Toth et al. 2000). The predominance of GA motifs among dinucleotides in the chickpea ESTs was similar to reports in cereals (Varshney et al. 2002) and dicots like *Medicago*, soybean and *Arabidopsis* (Tian et al. 2004). Similarly among trinucleotides, the abundance of AAG motifs in chickpea was quite consistent with the findings of Kumpatia and Mukhopadhyay 2005 who surveyed the abundance and distribution of various types of SSRs in dicot species. However, the earlier studies on chickpea microsatellites have reported the $(TAA)_n$ motif to be most abundant (Udupa et al. 1999). Regarding the monocot genomes, CCG motifs were reported to be most frequent whereas AAT motifs were least (<1%) (Varshney et al. 2002). Moreover it was also observed that even though the EST-microsatellites contained lesser number of repeat motifs than the genomic microsatellites (gSSRs) reported earlier (Sethy et al. 2006a), they proved to be highly informative in the genetic diversity and cross-species transferability studies as demonstrated later in our study (chapter 7) as well as in other crops (Scott et al. 2000; Thiel et al. 2003).

Of the 284 (117 + 167) microsatellite containing chickpea EST sequences, STMS primers could be designed only for 217 (76.4%) sequences. Such reduction in the number of EST-SSR markers have also been seen in other crops (Thiel et al. 2003; Aggarwal et al. 2007). Further only 97 (71.8%) of the designed 135 primer pairs could be validated. The excluded primers (28.2%) either did not amplify or produced anomalous sized fragments in *C. arietinum*. Such findings have also been encountered in other plant species (Cordeiro et al. 2001; Thiel et al. 2003) and in general, SSR amplification rates usually range from 60 to 90% in plants (Varshney et al. 2005a). Among the 86 primer pairs amplifying a single allele, 9 gave an amplification product larger than expected suggesting the presence of an intron in the genomic sequence that was not accounted for in the EST sequence or a lack of specificity which may lead to amplification of another copy of the gene family. Generally, it has been reported that the incidence of markers amplifying product of higher size due to the presence of introns is higher in EST-SSRs as compared to anonymous SSRs (Cordeiro et al. 2001; Kota et al. 2001). However, such large sized bands may cause problems during analysis since fragments above 500 bp are difficult to score accurately for small differences in fragment size (Thiel et al. 2003).

In conclusion, in the present study a novel set of 97 chickpea EST-SSR markers were developed from two different sources of ESTs – the NCBI chickpea database and the inhouse generated ESTs from chickpea 20 DAA seed. The study was the first attempt at large-scale characterization of SSRs from the coding regions of the chickpea genome. Additionally, considerable efforts have been made to provide an insight into the distribution and composition of different types of SSR motifs in the chickpea transcribed regions. It is known that EST-SSR markers can be potentially used for genetic diversity studies, construction of transcript maps, across genera transferability and comparative mapping in legumes. Hence in the following chapters (7 and 8) attempts have been made to fully utilize the generated EST-SSR resources for various applications leading to enhancing the chickpea breeding programmes through MAS selection.

Table 5.1: List of chickpea EST-SSR markers developed in this study. The designed primer pair sequences, microsatellite repeat motif structure, expected allele size (bp) in *C. arietinum* cv. Pusa362, number of amplified alleles (N_a), GenBank accession numbers and their putative functions based on BLASTX results are mentioned. Serial nos. 1-34 represent markers designed using the chickpea database whereas serial nos. 35-121 represent markers designed using the in-house generated ESTs.

S.No.	Primer name	Primer sequence (5'→3')	Motif	Expected size (bp)	N_a	Putative function	GenBank No.
1	CESSRDB2	CGGGCAGGTATTGAATTGTAA/ GAAAGGTTTACAGCCGTTGG	$(CT)_{17}$	169	1	No homology	CD051322
2	CESSRDB3	TTATCACTTGTTATTGTCCTCTAAG/ AATTTATGGACCCCATGTAA	$(TAA)_6$	197	1	No homology	AJ609280
3	CESSRDB4	GAAGAGGTAGCGGAGGAG/ CAAGCAACAGTTTTCACTCA	$(GGT)_3N_3(GGT)_2N_6(GGT)_3$	274	1	RNA and export binding	AJ609279
4	CESSRDB5	CCGACATCTCTTCTCAATTC/ CTTTAGGTGGTGGTTGTTGT	$(TCA)_{14}$	177	1	SAT5 gene	AY370650
5	CESSRDB6	GACACTTGTTTCTCCTCGTC/ TTTGGTTCAATCATTTTCTTT	$(TG)_5N_2(TG)_2$	340	NA	Lectin	AJ006765
6	CESSRDB7	AAGTGGTGTCGGTAATGGT/ TAATACCAAAAGCATGCACA	$(GGT)_5N_6(GGT)_2$	196	1	Glycine- rich protein	AJ487469
7	CESSRDB8	CACATACAGAGCGAAACAAA/ AAACCAAACAACAAACCATC	$(TCA)_4$	196	NA	QOR gene	AJ487465
8	CESSRDB10	CCCTTAATCAATTCA CCTCA/ TTATCCAAACCAATGATTCC	$(TAA)N_3(TAA)_6$	192	1	No homology	AJ005947
9	CESSRDB11	AATCTAACAGCAACGACGAT/ ATCAAGCTTCTTCTGCACAT	$(CCA)N_3(CCA)_3$	298	1	Unknown protein	AJ006048
10	CESSRDB13	ATCTGGGAGCTTGTGAGTTA/ TTGTATCTCCTTCAGATGGC	$(AT)_5$	260	1	Hypothetical protein	AJ012683
11	CESSRDB15	CTTACGATTTCTCCTCCCTT/ TTTCTCATACCGAATCCTTG	$(GCT)N_6(GCT)_4$	276	2	Hypothetical protein	AJ012681
12	CESSRDB16	ATGCTATGCATGATGTTTCA/ GTTCCAAACAAACACAACAA	$(TA)_{10}$, $(TTA)_4N(TTA)N(TTA)_2$	295	2	Invertase inhibitor	AJ487472
13	CESSRDB17	CAGAGAACACACAGAGCGTA/ ATGATCGTCAGAACGAAGAG	$(TC)N_3(TC)_5$	298	NA	ATP-ase subunit	AJ487471
14	CESSRDB18	TGCAAATAAAGCCTTCAAGT/ GAAAGTGGGAAAATGCAATA	$(TA)_5N_2(TG)_2T(TG)_3$	242	1	No homology	AJ487042
15	CESSRDB20	CAATACAGAGCACCAAAA/ CCTCCAAGTATAATGCCAAG	$(CTT)_4(CAT)_4$	299	HS	ADP-glucose	AF356004
16	CESSRDB21	GTGTATCGGTCAGGAAAAGA/ GGTACACACCACAATTCACA	$(ATT)_2N_2$ $(AAT)_3$	259	3	Plantacyanin	AJ012693
17	CESSRDB23	GTGTGGACCTGAAATTGAGT/ GAATATGGGAACAAGTGCAT	$(TA)_5$	221	1	Ribonuclease T2	AJ012689
18	CESSRDB24	TGTGCTTGACTTGTTCACAT/ TATGCATCCTCATTTTCTCC	$(GGC)_4$	283	1	β-amylase	AJ006763
19	CESSRDB26	GGTGCATTCTCTTCCATAAG/ TGCAAATCT TTAACCAAACA	$(GT)_5$	273	1	Expansin	AJ004959
20	CESSRDB27	GGTGAGATTAGGAAGCAATG/ TATCCAATCCCCATAAGATG	$(TAT)_7$	215	1	Cationic peroxidase	AJ271660
21	CESSRDB28	TCATACTTCTCCCCCATACA/ ATTAACGCCTTTTCTTTCCT	$(CCA)_2N_2(CCA)_3$, $(ACT)_4$	267	HS	PM protein	AJ299396
22	CESSRDB29	TTTAGTTGCACAACAACAGC/ AAATCCACATCCAAAAAGGT	$(TGA)_5$	176	1	GTP-binding	AJ299064
23	CESSRDB31	CAGGTTTCTCGACCAGTTAC/ CAGCCTTTTCCTTCTCAATA	$(CAG)N(CAG)_4$	262	NA	Cyclin	AF287306
24	CESSRDB33	GCTGCACAAAAAGTACATGA/ ATCCATCGAAACACCAATAG	$(GA)N(GA)N_2(GA)_5$	234	1	Pal gene	AJ250836
25	CESSRDB34	AACCTAAAGCCGAAAAGAAG/ CTCCCGTGAAGTAATAGTCG	$(AAG)N_4(AAG)_4$	251	1	Histone H2B	AJ400863
26	CESSRDB35	TCTAGAGCTAGCCAAAGGAA/ GCATCGTAATCATCGGTACT	$(GAT)_7$	272	1	UDP-glycose	AJ400861

27	CESSRDB36	CAGATCCGTTTGCTATTGAT/ CCGCTTCGATTTACTACCTA	$(TTC)_4$	256	NA	Tubby-like	AJ400860
28	CESSRDB38	GAGTAAGATGGCACAGTGGT/ GTATCTATTAGCGAAGCGGA	$(CCG)_4$	197	2	Cysteine proteinase	X93220
29	CESSRDB39	CTGAGGTTAATGTGAAAGGC/ GTCAACATCACATGCTCAAC	$(GGT)_4$	257	1	Glycine- rich protein	AJ275314
30	CESSRDB40	GAAATTAGGAAGCATTGTGC/ AATTGATTGAACCCACTTGT	$(TTAT)_4$	188	1	Peroxidase	AJ275313
31	CESSRDB41	GAACCAATAAAGCCTTGAAA/ TGACCAATTGATACAATCCA	$(GCT)_4$, $(TTTA)_4$	247	3	PM intrinsic polypeptide	AJ275307
32	CESSRDB42	GAGACAAAGATAGTGGCTGG/ TATTAATCACTCGCACGACA	$(TAAAT)_4$ $(GTTT)_3$	235	2	ABA-responsive protein	AJ275304
33	CESSRDB44	ATCCTTTCCTTGTTGTGCTA/ TTTAGTGAAGCATTGTTGGA	$(CTTT)_3$, $(TTG)_4$	267	1	Cytochrome P450	AJ012581
34	CESSRDB45	AGATGGTTTGAATGTTGAGG/ CACTTGACCCTTTGATTGTT	$(AT)_7(AG)_5$	295	3	Cytochrome P450	AJ249802
35	CESSRDB47	ACGAAGAAAGTTCCTGTGAA/ ACCGAAAACCTGATTCATTA	$(TTA)_2N_4(TTA)_4N$ (TTA)	240	4	Histone H1	AJ006767
36	CESSRDB49	GGCACGAGAAATTAATGAAG/ CCTCACAACCACTATCGTCT	$(GGT)_4N_3(GGT)$	288	NA	Chitinase	AJ012821
37	CESSRDB51	ACTATTACAAGAGCCCACCC/ CATAATGGTAAGGAGGTGGA	$(CAA)_4$	297	1	Extensin	AJ006770
38	CESSRDB53	CCCTTAATCAATTCACCTCA/ GCTTCTTATCCAAACCAATG	$(TAA)_6$	197	1	No homology	AJ005947
39	CESSRDB54	AGTGTTGTGGGTTTCATTTC/ TTGATTTGCCAAAGTACACA	$(TTA)_5$	221	2	Trans-membrane protein	AJ005869
40	CESSRDB55	CGATTATCTCAACTTTTGGC/ ACATGCACACGACAAATAAA	$(TA)_5$, $(ACT)_5$	136	2	Transcriptional regulator	AJ005000
41	CESSRDB56	TGTCTGGAACAACAAGTGAG/ GCCAATCAGATTTCCTCTTA	$(ATG)_4$	247	1	Myb family TF	CK149113
42	CESSRDB57	AACTCCATACCCAAAGTGTG/ GAAGCGAGGGTATTAAGATG	$(AT)_{11}$	211	NA	Unknown protein	CK148697
43	CESSRDB58	TTGAGGGAAGAGGGAGATTG/ TGATTACGCCAAGCTCAGAA	$(AG)_{20}$	238	HS	Unknown protein	CK149122
44	CESSRDB59	TCCATTGAGATTTGGAGACT/ TTGGGAAAGGGCCTTAAA	$(AT)_{10}$	220	NA	LTP protein	CK149001
45	CESSRDB60	AGCTCGTTCCTTGTCAACT/ GAAGGTTGCGTCATCATCT	$(CAA)_5$	232	NA	TIF3	CK149112
46	CESSRDB61	GCAGAATGGGAGATAATGAA/ TGCTGATTCTGATGTCTACG	$(CTT)_7$	233	1	bZIP TF	CK149116
47	CESSRDB62	ACATCCCTTCAATGAACATC/ ATTGGATATCGGTGTGTTGT	$(TTC)_6$	240	NA	Hypothetical protein	CK149128
48	CESSRDB63	GTTGCAAAGCATCCTTCA/ CTTCCTCCCACTTCTCTTCT	$(CT)_{20}$	235	NA	No homology	CK149141
49	CESSR11	TCAATTCGTTCATACTGAGGTTTC/ GGATGCTATTTCAAGCGATTCT	$(CT)_8$	336	NA	Adomet	EX151630
50	CESSR12	TTAACAAGCATCCCCTTTCAC/ TGGAAGTAGTGCCTCCACAAT	$(AGA)_5$	400	NA	ADP ribosylation factor	EX567533
51	CESSR13*	AAGAAGCTGAAGCCGAAGATG/ CATCCTCCTGAGTCTTGGCTAT	$(ACA)_6$, $(GA)_5$	399	1	Unknown protein	EX151631
52	CESSR14	GGCACAAGGTATCTCCACAA/ ATGCTTGCCTCAACCTCAGA	$(TGC)_6$	300	1	Unknown protein	ES544474
53	CESSR15	CATGACATCCTCAATCCTTGG/ TAGCGACAAATCTTAGCCGTAG	$(TGC)_6$	300	1	Unknown protein	ES544475
54	CESSR16	GGGGGTCAGCTCTATTATCTTC/ GTTTCTAGATTCGTCGCAAGGT	$(TTC)_6$	371	NA	Unknown protein	EX151755
55	CESSR17	GCCGAATTTTGAGGAAACAA/ TGCAGCTCTTCAGCAGTGTT	$(GAT)_6$	391	NA	TF EREBP	
56	CESSR18	CTGTTTCCATCCAAAGCCATA/ AACAGCGAACAATCACGATAAC	$(TC)_8T$, $(TTTA)_5TT$, $(TTCAA)_3$	297	1	Hypothetical protein	ES544476
57	CESSR19	AAAGAAGAGAGATTACCCTTCTCA/ CGAGAATCCTAGCACGCTCATA	$(TCAAC)_5$	299	1 (780bp)	DNA-binding	ES544477
58	CESSR20	CGAAACTCGAACGTGCAAT/ TTTGGCGAATTTGAAAGGAG	$(GATTC)_6$, $(ATTTA)_5$	386	1	Unknown protein	ES544478

59	CESSR21	CCTCAACGCTCATTCTTCTTCT/ CCCCAAGGAACCATTCTAAGAT	$(CTT)_6$	233	1	OSH1	ES544479
60	CESSR22	CCCTTATCCACAACGTTACCA/ AACATCTGCATTTCCCCAAG	$(CTT)_5CT$, $(AGA)_5$	371	1	RNA binding	ES544480
61	CESSR23	CGCGTAAACGTTATTCTCTTCA/ CATCATTTCCCTTAGCATCCTT	$(TTC)_7$, $(CTT)_3$	399	1	Glutathione peroxidase	EX151810
62	CESSR24	TGAGAGGAAAATAACGGTAGCA/ TGGATGGAGAGAAAGGAGAAGA	$(TCT)_9TC$	198	1	No homology	ES544481
63	CESSR25	CTATGGCAAAAAGCATCACAAG/ ATAGCCATGGCCACATTAAACT	$(CCG)_6$	363	1	No homology	ES544482
64	CESSR26	GGCAAAATCGAAATTCAACC/ TGATCAATGACAGTGTAGAAGG	$(CTT)_4$, $(CT)_{10}$	275	1	No homology	ES544483
65	CESSR27*	AAAGGGCGTTGAAAGAAAGA/ AAGTCCTCGCACACTTCGAT	$(AG)_{11}$	293	1	Unknown protein	EX567680
66	CESSR28	ACTTTTCTCCGACCAAATGC/ AATGGAGAACCCCAAAAACC	$(TTC)_7$, $(TCT)_7$, $(GAA)GA(GAA)_8$	364	1	Hypothetical portein	ES544484
67	CESSR29	CAATGGTGACTGGTGGTTTG/ ACCAATTGACGCAGAGATCC	$(TCT)_8$	278	1	Unknown protein	ES544485
68	CESSR30	TCGGACCACAAGAGCATCTA/ CGTGGAAGAAAGGAATGTTG	$(CT)_9TT(CT)_4$, $(CAC)_6$	388	1	Actin binding	ES544486
69	CESSR31	ACGTAGGTTAAGGTTGCTGGTC/ TTCAACGTGTTCGAAAGCTC	$(AAG)_8$	113	1	Unknown protein	ES544487
70	CESSR32	GCAATTTCGAGGTCAACAGAA/ GGGACTTGCGTCTTTTTCCT	$(CCT)_5CCA$	238	NA	Unknown protein	EX567918
71	CESSR33	TTAGCAGCATCCCTTCTGTGTA/ CAGCAACAAAACCAATCATAGC	$(CCG)_6$	297	1 (487 bp)	No homology	EX151964
72	CESSR34	CATTGCTCAAAGCCAATTCA/ TCGATGAATCGGAACAAACA	$(TAT)_4T$, $(TGT)_6TG$	294	1	No homology	ES544488
73	CESSR37	CCCTTCATCAACCTCTTCTCTG/ GCATGTGCAGATACCAACCTAA	$(AGA)_4AG$	344	1 (545 bp)	Ribosomal protein	EX567650
74	CESSR38	GGGGTCAGCTCTATTATCTTCC/ TGAACTTGTGACCTCTTGAAGG	$(TTC)_6$	333	NA	PsRT17 Pea	EX151872
75	CESSR40*	GAAGAGCTTCTTGCTACCCATC/ CTTGAAAACGTCAGGTTTGGAG	$(GAT)_5GA$	297	1	Nascent polypeptide	EX151955
76	CESSR41*	AGTTTTGGAGAAGTGGTTGAGG/ :GGGCAGTAACCTTGGTGAATTT	$(GTG)_6$	361	1	Glycine rich RNA	EX567716
77	CESSR42	TGGTTCAAGAAAAGAAGGTAGTG/ CGGTTCACTAATGCAAAAACCT	$(ACC)_5$	298	1	Hsr2035 homolog	ES544489
78	CESSR43	CATTAAAGCTAGGAGTTTGTGCTG/ ACGGTACCATACCCGACTACAT	$(CTA)_4$	386	1	Bimodular protein	EX567535
79	CESSR44	TAAACATCGTTGTTGGCGTGAG/ CAAAAAGGACGTCAACTATGTCT	$(CCT)_5$	657	NA	No homology	EX151981
80	CESSR45	CAGAGACACACACATTAGAAAAAC/ CTGAACTCTTCAACATGACAAGG	$(TTC)_4TT$, $(TTG)_4$	298	1 (720 bp)	Auxin responsive	EX151983
81	CESSR46	GGCCACGAAGAAAGGAATAAA/ GCTTCGAACCTTGTGCTATCAT	$(AGGAG)_3$, $(CGA)_4$	300	1	No homology	EX152041
82	CESSR47	GAGTTCCACATTGTCACAGGAA/ AATGCAACAGTCCTTGTGGATA	$(TTC)_5$	541	1	Germin like protein	EX567643
83	CESSR48	AAGTTGATGAAGATGGCTGGTT/ CAGCAGCTGTCTCAAAAGTCTC	$(ATG)_4$	451	NA	TF BTF3	EX567727
84	CESSR49	GGATCATCAAGATCTTCCACTTC/ CCCAAGCTTGTTTTAGAGGACA	$(CAA)_4CA$	325	1	Unknown protein	EX152018
85	CESSR51	CACATGAACAGAAAAAGGGACA/ GCATGTTGAGCCAAAGCTAAAT	$(TTTG)_5$	205	1	Pleckstrin domain	EX567864
86	CESSR52	GCAACAACAATGGCACTTAGAC/ AGGTTTTACAAAGAGCGCTACG	$(AAC)_4AA$	394	1 (745 bp)	Unknown protein	EX151786
87	CESSR53	GACAGTGATTCATTTCACCTCAA/ CACTGGTCCTCATTCCTTCTTT	$(AGAAG)_4$	316	NA	Unkown. protein	EX151859
88	CESSR55	CTATGGCAAAAATCATCACTGC/ AGGCCCCACCTATGAAGATAAT	$(GTT)_4GT$	376	1	Hypothetical protein	EX567812
89	CESSR57*	GTCGCAAGCTTGTCTAACACAC/ GCAAGATGACAAGAGTCGATTT	$(AG)_7$	381	1	Mit. Acyl carrier prot.	EX567751
90	CESSR58	ATGCGGAGAGAAGAAGATGAAA/ ACTCCTCTTGCAATCGCTTATT	$(TTA)_3TT$, $(TAT)_4$	280	HS	Heat shock protein	EX151665
91	CESSR59	CAGAGACCACCTCAAGAACCTT/ ACATAACCAACTGCTCCCAAAT	$(CCA)_4CC$	366	NA	Delirium	EX567776
92	CESSR60	AAGGTCTGCTTCTTGTTTGGAG/ TTAGGTCGTTCTCCCTGAATGT	$(AAC)_5$	403	1	ANAC075 TF	EX567830

93	CESSR61	CACTCTTCCCTCCCTTTCTTTA/ GAATCAGGGTAGGTTTGTTTGC	$(CT)_7T(CT)_2$	257	1	Armadillo like helical	EX151660
94	CESSR62	ACCAGCTGCTAGACCTGATGTT/ GCAATAAAACAAAATCCTCACACC	$(TGA)_5TG$, $(TAT)_3$	245	1	Receptor-like kinase	
95	CESSR63*	TCTTGTGTTGTGTGTGCCATAA/ GTTTTCGCGTAGCTTCAGTTCT	$(TTC)_6$	362	1	Unknown protein	EX151706
96	CESSR64*	CCCGTGACACTTTAATCTCTCC/ TAGCTGGGAAAGGAAAACAGAA	$(CT)_6N_2(CT)_8$	126	1 (250 bp)	No homology	
97	CESSR65	CTCCTCCACTCATCTTCATCTTC/ GAGAAGGTGTTTCCGGTAAAAGT	$(CTT)_2N_3(CTT)_2N_2$ (CTT)	352	NA	P0034E02.34	EX567847
98	CESSR66	TGATGAAGAGTGACATGGATGAA/ CCTCCACATGGTAGCATCAAT	$(ATA)_6$	333	1	Homodomain-related	EX567630
99	CESSR67	GAAGAGAATGGCATTGGGTTT/ ATTCACACGCGCAAAACAGTTA	$(TTC)_5$	356	NA	Unknown protein	EX567863
100	CESSR68	AATGGCCACCATTTTCTCATC/ AAACGTTCTTTCCATCCTTCTG	$(ACC)_6$	330	1	Hypothetical protein	EX151762
101	CESSR69	TCCAAGCCTGATATTGACTCAT/ GAAAGTCTGCCTCATCGTCAAC	$(GAT)_6$N3(GAT)	292	1 (784 bp)	Hypothetical protein	EX151775
102	CESSR70*	GCGAGTTGCTGTAAGGAAGAA// GTGGCACTACTCAAGTTTCACG	(AAG)6N5(GA)3	318	1	unknown Protein	EX567773
103	CESSR71	TTGTAGTTCTCCTCTCTCTCTCTC/ CATCAAAACCAAACCTATGGAG	(CT)N$(CT)_8$	295	1	Unknown protein	EX567905
104	CESSR72	ATTTCACTCCTCACTTCTCACC/ CACGAAAATCGGATGATTCAG	$(CT)_7$	345	1	Unknown protein	EX151914
105	CESSR73	TCTTCTCCCATTCGTTGTTGAT/ GATCTTCTGTTCCTCAGCCAAC	$(CTT)_3N_3(CTT)_6N_6$ $(AT)_4$, $(GGC)_3N_4(GGC)$	363	1	No homology	EX151922
106	CESSR74	TCTCTGAGCGAGTTTGTGTGAG/ ACCATAAATCCTCTGGAAAGCA	$(GA)_4N_2(GA)_7$	300	1	Ribosomal protein	EX151940
107	CESSR75	TGAAGGCAACAACAACAACA/ TCCTTCATATGCTGGTAATAGG	(CAA)4, (AAC)4, (CT)6	284	NA	SAM	EX567633
108	CESSR76	TGCACCACCAATATGCTCTTAC/ CTCTCTTTGCTTGGATTCCACT	$(GAT)_4$, $(GA)_4$	334	HS	Oxygen-evolving	EX567549
109	CESSR77	CCAACTTAAACTCATTTCGTCTCA/ CCAAGATGTGTTTTTGATGATG	$(GA)_2N_2(GA)_4$, $(CAT)_4$	173	1	DUF647	EX567970
110	CESSR78	ATTGCTGAGGCTGTGAATTGTA/ CCCAATACATCAAAGATAGATCG	$(TTAAA)_3$	373	1	Lipid transfer	EX567577
111	CESSR79	CCATTTCCAATTCCAATCAGAG/ TCACTGCACTGCTGATCTTCTT	$(GA)_{10}$	396	NA	Calcineurin B-like	EX567554
112	CESSR80	TCACCCTTTCTTCTTCAACTTC/ GAACGCATAAAATAGTCGCTGA	$(GA)_7$	260	1	No homoloy	EX151949
113	CESSR81	AGCTTGTTCCACTTGAGCTTTC/ AGGTTCCTTGATGCGTTTAGAG	(CA)2N2(CA)N_4(CA)	382	HS	Hypothetical protein	EX567883
114	CESSR82*	CTGTGGATCCCAACAAGCAT/ TACCAGAGGAAG GAGAAGAGG	$(CAG)_4$	286	1	Mitochondrial receptor	EX151864
115	CESSR83	CCTCTTCATCATTCAGCATCAC/ AGGGTAATCATTGGGAAAGGAA	$(TTA)_4$	399	HS		EX152062
116	CESSR84	AACCTCAAAAGAAACAACCCTTC/ TCCCATCTGCTCCTATTGATTT	$(AAAC)_3$AAAAAC	393	HS	PGR5	EX567766
117	CESSR85	ATGTACTTGGTCTGGTCCGTCT/ ACCTTTTCGGCGTTTCTTTTAC	$(GCT)_4$	279	1	Hypothetical protein	EX151920
118	CESSR86	ACCTTCCATCTTCAATCCCTAA/ CGTGATCCTCATCCACATACAT	$(ATTC)_3$ATCC	377	1		EX567555
119	CESSR87	ACTCATCACCCTCAACCT CAAC/ GCGACATTCACTAGCTCTTTGA	$(TCA)_3$TC	364	1	Ketoacyl-CoA	EX567598
120	CESSR88	CCTCAACCGTTACAAACCTCTC/ AATTCACCAGGAAGAACCAAAC	$(AAG)_4$	395	1	Histone H2B	EX151659
121	CESSR89	GGAAGAAGGTTCGTGTGAAGTG/ TGAAGC AAGGAACAACTCAGAT	$(TA)_6$	371	HS	Unknown protein	EX567621
122	CESSR90*	CGCTCTCTTCTCTTCGAAAAAC/ CTCAACCACTCGGTCTCTATCA	$(GAA)_4$	373	1	ADP-ribosylation factor	EX151808
123	CESSR91	TGGCTGTTTCATCTTACCAATC/ CTCCACCAGAAATCTGTTCAAA	$(CCA)_4$	397	1 (650 bp)	Early light inducible	EX567676
124	CESSR93	ACGAGAAGAGCATTGCATTTG / TAACGGCCTCTTTAGTCTGCTC	$(AAC)_3$AAG	353	1	Pex19	EX151839
125	CESSR94*	GAGAGCCTAACATAACCAATCACA/ AGGCTTAATGGCAATGAAAGTG	$(CATCA)_4$	351	1	Nucleoside diphosphate	EX151682
126	CESSR95	TGGAAAGATTGGTCTTGAACC/	(GGA)4	298	NA	Hypothetical	EX151735

		CGAAGTTGCTTCCTCTTGTGATA				protein	
127	CESSR96*	TTTCCAACAAACACACACTCTTC/ ATCAGCTCCAAGTGCAGAAAAT	$(TC)_8$	361	2	Acyl carrier protein 1	EX151787
128	CESSR97	TCTCGTTCCTTCAAGTGTGAAA/ CTCCTAATCAAACAAGCCGAAT	(TGC)4	352	NA	Ribosomal protein 60S	EX567744
129	CESSR98	TACACTACGAATCTGATCCATGC/ ACTGACTTCGTTGAAGAAGAGGT	$(AATCA)_3$	345	1	Hypothetical protein	EX567925
130	CESSR99	ATGGTACTTGCAAGAGAGCTTGT/ AAACCAGAAACAGAGATCCAGGT	$(AGA)_5$	380	1	Unknown protein	EX567955
131	CESSR100	ATAACTCGCGCCTCTGTCTAAG/ GTAGCCCTGGTGTAAATCTGCT	$(AGA)_4AG$	363	HS	Ribosomal protein	EX152074
132	CESSR101	TCCAAGGGGAAAAAGTATCACA/ CAATCACTTCAGTGTGACTGTCT	$(GTG)_4$	241	1	Copper domain	EX567915
133	CESSR102	CCAAATTCCTTTGAGACATCA/ AGTCGGTTTGAAGCTGAGGA	$(AAC)_5$	333	HS	cAMP binding	EX152091
134	CESSR103	CACGAAGCAAACCCAGTACA/ GACCCAATTCTTGCTTCTTCA	$(AAAAC)_3$	363	1	Seed-alpha amylase	EX151636
135	CESSR104	AGCGGCTTACTCGAAATCATAA/ TTTGATAGTTCTAGTCCAACAGCAC	$(AT)_6$	392	1	Hypothetical protein	EX567831

HS: Higher Size ; NA: No Amplification ; *: amplification at 60-50^{0}C touchdown

Chapter 6: Development of ESTP and PIP markers

6.1 Introduction

The recent development of applied technologies in biology is leading to an enormous production of information in the area of plant genomics, especially through the sequencing of different genomes. During the last two decades, Expressed Sequence Tags (ESTs) have been the most widely available, sequenced nucleotide commodity from plant genomes and are regarded as the valuable key resource for gene discovery and mapping programs (Somerville and Somerville 1999; Ramíirez et al. 2005). Concomitantly in the recent years, molecular markers that in the past were primarily captured from non-transcribed regions, have taken a step forward towards development from the expressed regions of the genome (Gupta et al. 2002). Popularly known as 'Functional Markers (FMs)', these EST-derived markers apart from sharing the desirable properties with random markers (RDMs) such as locus specificity, hyper variability and codominant inheritance, have some intrinsic advantages owing to their association with coding regions. In the recent past, these markers have been extensively applied for mapping gene rich regions of the genome in the backdrop of anonymous markers (Temesgen et al. 2001; Choi et al. 2004b; Guo et al. 2007) that would largely assist in direct tagging of traits if linked to mapped genes, comparative mapping and understanding the pattern of genome structure, function and evolution.

However, assigning positions to ESTs on linkage maps is still an area desiring attention of researchers. Being representatives of coding regions, EST sequences are highly conserved and thus known to exhibit low levels of polymorphism in comparison to random markers. Therefore, different types of PCR-based approaches have been developed in plants to exploit polymorphism associated with such sequences. These include the most common ones such as SSRs, (Varshney et al. 2005a), SNPs (Gupta et al. 2001; Rafalski et al. 2002) and the less widely used such as Conserved Orthologous Sites (COS; Fulton et al. 2002), Expressed Sequence Tag Polymorphism (ESTPs; Harry et al. 1998) and Intron Targeted primers (ITPs; Lessa 1992).

ESTPs or gene-based markers, designed from random sequences from the ESTs have been mostly reported in woody species such as pines (Brown et al. 2001; Temesgen et al. 2001), spruce (Schubert et al. 2001) etc. to generate a consensus genetic map.Recently another strategy termed as 'intron-targeted primers', where non-coding introns are amplified using DNA primers based on the conserved exon sequences has also gained momentum for finding DNA polymorphisms in eukaryotic genomes (Côrte-Real et al. 1994; Daguin & Borsa 1999, Bierne te al. 2000). The perception that intronic regions are more variable than exonic

regions due to less evolutionary constraints lead to a new source of potentially neutral genetic markers for use in linkage mapping, phylogeny, evolutionary and comparative genomic studies. Additionally, the intron targeted primers are reported to be more efficient in polymorphism detection than conventional EST-PCR primers (ESTPs), making the gene mapping programs more efficient (Wei et al. 2005; Ishikawa et al. 2007). However, this newly developed approach has been applied vastly in animals and is currently very limited in plants, primarily restricted to systems where fully characterized genes/whole genome sequences are available (Holland et al. 2001; Wang et al. 2005). Recently however, comparative genomics has hastened the strategy in non-model crops also where only EST sequences are available. Using ESTs of targeted species and genomic sequences of *Arabidopsis* homologs, Choi et al. 2004a; Wei et al. 2005 and Panjabi et al. 2008 successfully designed intron-targeted primers (ITPs) and demonstrated their utility for linkage mapping. Moreover, the intron-targeted primers are highly suitable for cross-transferability studies because of the conserved nature of exons and thus useful for identification of genetic determinants of a trait in less-studied taxa (Wang et al. 2005; Choi et al. 2004a). The cross species ITP markers have been applied for understanding the genome conservation among legume crops (Choi et al. 2004a) and recently several web based programs like GeMprospector (Fredslund et al. 2006) and PIP (Yang et al. 2007) have been developed that would facilitate the design of intron targeted primers even in orphan crops.

In chickpea, attempts to mine SSRs from EST sequences have already been reported in Chapter 5. However, other EST based markers also need to be developed in order to increase the array of available markers and also to fully exploit/utilize the generated EST resources of chickpea. The additional markers will be extensively used to construct a high density, gene rich linkage map of chickpea. In chickpea, no report of EST-PCR markers is available except that of Buhariwalla et al. 2005 who developed 38 functional EST-PCR primers. Therefore, the present study was undertaken with the aim of developing novel EST-based PCR markers in chickpea. Two different types of markers namely the Expressed Sequence Tag Polymorphisms (ESTPs) and the intron targeted primers termed as potential intron polymorphisms (PIP, being reported here for the first time in chickpea) were developed from chickpea unigenes. The primers were characterized, percent polymorphism was estimated between the 2 chickpea cultivars crossed to generate the mapping population and molecular basis of length variation was elucidated.

6.2 Results

6.2.1 Development of ESTP markers

For developing ESTP primers, the approach of conventional EST-PCR method was used which is based on targeting of random EST regions for primer design. For this, the chickpea seed related unigenes (described in chapter 4) were utilized. A total of 80 ESTP primers were designed using the primer 3 program as mentioned in section 3.2.12 to amplify a product of up to 500 bp from EST sequences. These ESTP primers were validated in chickpea cultivar ICCV2 which resulted in 39 primers producing expected size bands whereas 25 primers amplified genomic regions measurably larger than those from their corresponding ESTs indicating the presence of one or more introns and the remaining 16 primers did not amplify at all even under varying amplification conditions. Of the 25 primers that produced higher size alleles, 19 produced bands upto 800bp in size whereas 6 amplified products > 1.0 Kb or gave unspecific/complex products and were therefore not considered. Hence a total of 58 (39 + 19) ESTP primers were available for use. The primer sequences, product size (expected and observed), putative function and GenBank accession number of all the 80 ESTP primers are mentioned in Table 6.1.

6.2.2 Development of PIP markers

Intron-flanking exon-exon based chickpea primers were designed using the program PIP (Potential Intron Polymorphism) as mentioned in section 3.2.15.2. A total of 1337 chickpea ESTs (1037 inhouse ESTs + 270 other ESTs obtained from our institute, (Chattopdhyay et al., pers. comm.)) were uploaded into the PIP program. In brief, the program first screens the sequences (chickpea ESTs in the present study) that show best sequence alignment with the *Arabidopsis* CDS sequences (with intron) and then predicts the possible intron positions in the query species. In the next step, the program designs intron flanking exon-exon primers in the query EST sequences (chickpea in this case) alongwith providing the information on intron lengths in the subject species (*Arabidopsis* in the present case), primer positions in query sequences and sizes of PCR products if without introns in query species (Fig. 6.1) Thus from the 1307 (1037 + 270) chickpea unigenes, the program designed a total of 110 primers that were designated as 'PIP' and were mostly predicted to amplify 100-120 bp fragments if without intron lengths. Amplifications using these primers were carried out in chickpea cultivar ICCV2 that yielded 76 functional primers producing alleles larger than expected (>100-120bp) and predictably contained introns (Table 6.2). The sizes of amplified products obtained with these primers ranged from 100 to 820bp and thus

Figure 6.1: Designing of intron-flanking exon-exon primers of chickpea based on the PIP program (Yang et al. 2007). Blue vertical arrows indicate predicted intron positions in the chickpea unigenes and horizontal arrows with bold characters indicate primer binding sites.

Three PIP primers designed from EX151729 sequences as marked by different color arrows for different primers: PIP27 (Pink), PIP28 (Green) and PIP29 (Purple).

were expected to contain introns. Of 76 primer pairs, four namely PIP2, PIP3, PIP29 and PIP104 amplified 2 alleles. Of the remaining 34 primers, 20 either did not amplify or produced very large fragments or complex banding patterns whereas 14 amplified predicted size of 100-120bp and therefore did not contain introns (Table 6.2). Hence a total of 76 PIP markers were developed in this study.

6.2.3 Screening for polymorphism between parents of mapping population

The 58 ESTP and 76 PIP functional primers were screened for length polymorphism between the parents of the mapping population i.e. *C. arietinum* ICC4958 and *C. reticulatum* PI489777 as mentioned in section 3.1.2. Of the 58 chickpea ESTP primer pairs, 10 were found to be polymorphic between the mapping parents. Similarly, of the 76 PIP primers, 24 primers revealed intron polymorphisms between the parents. A representative gel of polymorphic and monomorphic primers is shown in Fig. 6.2.

6.2.4 Sequence analysis

To test the specificity of genomic fragments that amplified larger than expected sized bands and to unravel the molecular mechanisms of length polymorphisms within the exonic and intronic regions, direct sequencing of some of the amplified alleles produced by chickpea ESTP and PIP primers was carried out as described in section 3.2.9. Of the ten polymorphic ESTP primer pairs, two primers namely CEST44 (amplifying 415bp instead of expected 296bp; Fig. 6.3 A) and CEST86 (amplifying the same sized fragments as the expected size; Fig.6.3 B) were sequenced. Similarly size variant alleles amplified by two PIP markers namely PIP41 (Fig. 6.4 A) and PIP44 (Fig. 6.4 B) in the mapping parent's i.e. *C. arietinum* ICC4958 and *C. reticulatum* PI489777 were also analyzed at sequence level.

The primer pair CEST44 amplified variant sized alleles of 415 and 445bp in P1 (*C. arietinum*) and P2 (*C. reticulatum*) respectively. MAFFT (version 5.667) sequence alignment (Fig. 6.3 A) resulted in the following conclusions: 1) it confirms the PCR specificity that indeed corresponded to the ESTs from where the primers were designed 2) the presence of introns of 119 bp in P1 and 149 bp in P2 contributes to higher sized alleles 3) the length variation was due to presence of indels in the intronic regions and 4) point mutations at several nucleotide positions (exonic as well as intronic regions) shown in Fig. 6.3 A. At another locus CEST86, sequence comparisons revealed that indels was the cause for length differences in P1 (430bp) and P2 (427 bp) (Fig. 6.3 B). Furthermore several point mutations were also observed at CEST 86 at nucleotide positions 59, 328, 329, 331 and 332.

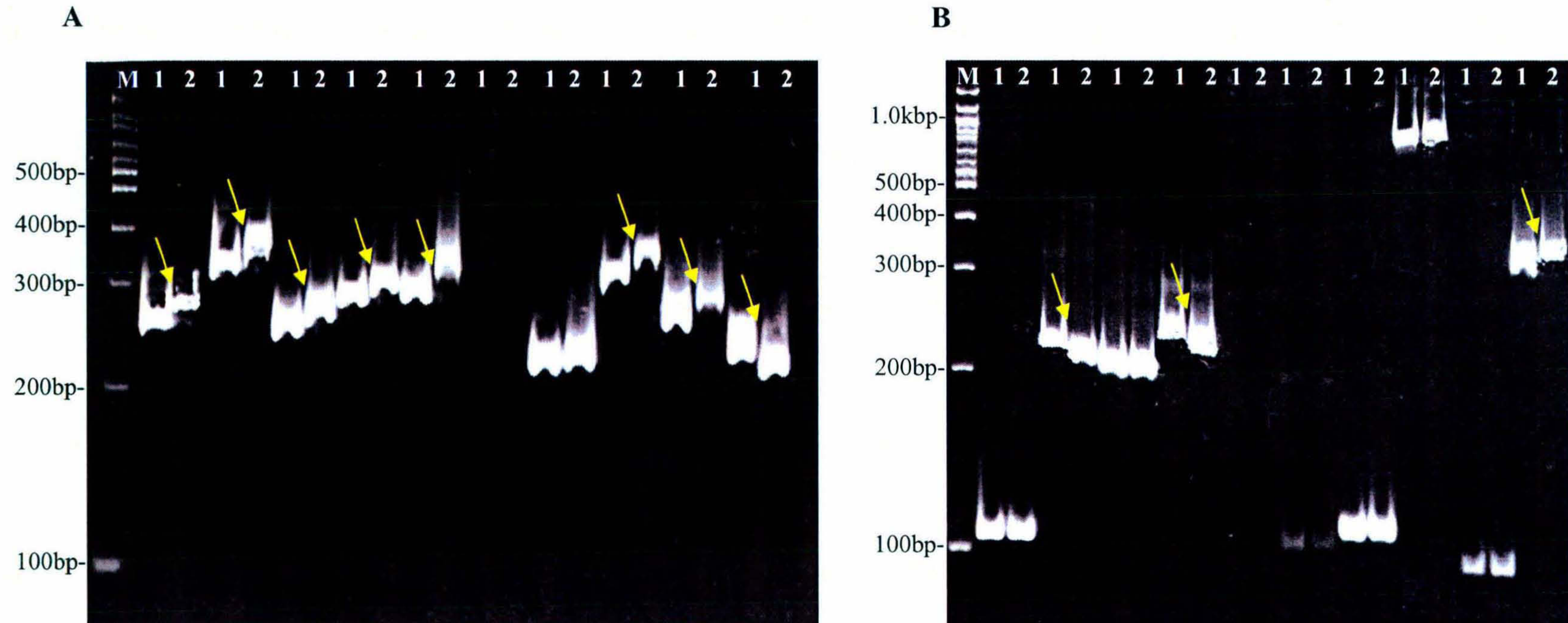


Figure 6.2: Screening of developed EST-based chickpea primers for polymorphism between parental DNA. The PCR amplification obtained with **CEST (A)** and **PIP (B)** primers were resolved on 6% PAGE gels. 1 indicates parent *C. arietinum* (ICC4958), 2 indicates parent *C. reticulatum* (PI489777). Arrows indicate polymorphic primers and M: 100 bp ladder.

Figures 6.3: Multiple sequence alignment of size variant alleles amplified using primer pairs (**A)** CEST44 and **(B)** CEST86 from ICCV2 and the mapping parents, ICC4958 and PI489777. The *asterisks* indicate similar sequences and - indicate alignment gaps. The primer-binding sites are represented by underlined bold letters and characters in bold shaded boxes indicate point mutations.

(A) CEST44

```
ICCV2        ATGGGGAACTGATGAAAACACTGTGATATCAATACTGGGTCACAGAACCTCACATCAGAT
ICC4958      ATGGGGAACTGATGAAAACACTGTGATATCAATACTGGGTCACAGAACCTCACATCAGAT
PI489777     ATGGGGAACTGATGAAAACACTGTGATATCAATATTGGGTCACAGAACCTCACATCAGAT
             ********************************** *************************

ICCV2        ACAATTAATTAGAAAATCTTATGAGGACATTTATCATGAGGATCTTGTCAAACGCTTGGA
ICC4958      ACAATTAATTAGAAAATCTTATGAGGACATTTATCATGAGGATCTTGTCAAACGCTTGGA
PI489777     ACAATTAATTAGAAAATCTTATGAGGACATTTATTTTGAGGATCTTGTCAAACGCTTGGA
             **********************************  ************************

ICCV2        GTCTGAGATCAAAGGAGATTTTGAGA----------------------------------
ICC4958      GTCTGAGATCAAAGGAGATTTTGAGGTGCATATT-----AGTAATTTAGTATCAAAGTCA
PI489777     ATCTGAGATCAAAGGAGAATTTGAGGTGCATATATATTAAGAAATTAAGTATCAAAGTCA
             ****************** *****

ICCV2        ------------------------------------------------------------
ICC4958      ATTTTGTCAATTTTATAAATAACATTCAATTAATTCATTTTCTTAAATTAAT--------
PI489777     AGTTTGTCAATTTTATAAATAACATTCAATTAATTCATTTTCTTAAATTAATTGTGTCTG

ICCV2        --------------------------------------------------------AAGC
ICC4958      ------------------ATATGTAAAATTTTGAAATTTTATTTTAATGAACAGAGAAGC
PI489777     ATATCTGTGTCAGTATCAGTGATTCAAATTTTAGAATTTTATTTTAATGAACAGA-AAGC
                                                                     ****

ICCV2        TGTGTACCGTTGGATATTGGAGCCTGCTGATCGTGACGCTGTTTTGGTCCATGTTGCTAT
ICC4958      TGTGTCCCTTTGGATATTGGAGCCTGCTGATCGTGACGCTGTTAGGTTCCATGTTGCTAT
PI489777     TGTGTACCGTTGGATATTGGAGCCTGCTGATCGTGACGCTGTTAAGGTCCATGTTGCTAT
             ***** ** **********************************  * *************

ICCV2        AAAGAGTGGAAAAAACTACAATGTGATTGTGGAAATTTCCTCTGTTCTTTCCCCTGAAGA
ICC4958      AAAGAGTGGAAAAAACTACAATGTGATTGTGGAAATTTCCTCTGTTCTTTCCCCTGAAGA
PI489777     CAAGAGTGGAAAAAACTACAATGTGATTGTGGAAATTTCCTCTGTTCTTTCCCCTGAAGA
             ************************************************************

ICCV2        GCTCTTTAATGTGAGGCGTGCCTATA          (296bp)
ICC4958      GCTCTTTAATGTGAGGCGTGCCTATA          (415bp)
PI489777     GCTCTTTAATGTGAGGCGTGCCTATA          (445bp)
             **************************
```

(B) CEST86

```
ICCV2        GAATTCAACTCGCCCTGAAATCGCCGCCGGAGTTCCTCGNTAGTCCGCCACACCCATAAT
ICC4958      GAATTCAACTCGCCCTGAAATCGCCGCCGGAGTTCCTCGNTAGTCCGCCACACCCATAAT
PI489777     GAATTCAACTCGCCCTGAAATCGCCGCCGGAGTTCCTCGNTAGTCCGCCACACCCATAAT
             ************************************************************

ICCV2        GGACTCTGAACCCCAATCCAAAAAGAAAACAAAAAACAAAAAGAAAAAGCAAACTGAACC
ICC4958      GGACTCTGAACCCCAATCCAAAAAGAAAACAAAAAACAAAAAGAAAAAGCAAACTGAACC
PI489777     GGACTCTGAACCCCAATCCAAAAAGAAAACAAAAAACAAAAAGAAAAAGCAAACTGAATC
             ********************************************************** *

ICCV2        TGAGCCTGAACCGGTTCAGGCAAAAGTCAAAGTTGTCCGTGAGAATCCAAACAAAATACC
ICC4958      TGAGCCTGAACCGGTTCAGGCAAAAGTCAAAGTTGTCCGTGAGAATCCAAACAAAATACC
PI489777     TGAGCCTGAACCGGTTCAGGCAAAAGTCAAAGTTGTCCGTGAGAATCCAAACAAAATACC
             ************************************************************

ICCV2        TCCATATATTGGTTATTTTCCTTCTGGTTTTGATCCGGTGAAGTCGACTTCTGTTTCGGC
ICC4958      TCCATATATTGGTTATTTTCCTTCTGGTTTTGATCCGGTGAAGTCGACTTCTGTTTCGGC
PI489777     TCCATATATTGGTTATTTTCCTTCTGGTTTTGATCCGGTGAAGTCGACTTCTGTTTCGGC
             ************************************************************

ICCV2        CGGTTTCCAAGTATATCGAAACAAGAACATGACTAAGAGGCTTGAGCTTGTTGTTAGTCC
ICC4958      CGGTTTCCAAGTATATCGAAACAAGAACATGACTAAGAGGCTTGAGCTTGTT-TTAGTCC
PI489777     CGGTTTCCAAGTATATCGAAACAAGAACATGACTAAGAGGCTTGAGCTTGTTGTTAGTCC
             **************************************************** *******

ICCV2        TGCTGGCTCTTCGGTTGATTTCGTCGGAACAAATTATATGGGTGAGGCTACGGGCTCACA
ICC4958      TGCTGGCTCTTCGGTTGATTT-GTCGGGGCCCAT-ATATGGGTGAGGCTACGGGCTCACA
PI489777     TGCTGGCTCTTCGGTTGATTTCGTCGGAACAAATTATATGGGTGAGGCTACGGGCTCACA
             ********************* *****  *  ** *************************

ICCV2        CCGATCAATGTATGCTCTTGGTGTGTTTGATAAGGAATCTCAGACACTTTAAGGTTGTGC
ICC4958      CCGATCAATGTATGCTCTTGGTGTGTTTGATAAGGAATCTCAGACACTTTAAGGTTGTGC
PI489777     CCGATCAATGTATGCTCTTGGTGTGTTTGATAAGGAATCTCAGACACTTTAAGGTTGTGC
             ************************************************************

ICCV2        CTATTGGTGC          (430bp)
ICC4958      CTATTGGTGC          (427bp)
PI489777     CTATTGGTGC          (430bp)
             **********
```

Figure 6.4: Multiple sequence alignment of variant sized alleles amplified using PIP primer pairs (**A)** PIP41 and **(B)** PIP44 in chickpea mapping parents' *C. arietinum* ICC4958 and *C. reticulatum* PI489777). The *asterisks* indicate similar sequences and - indicate alignment gaps. The primer-binding sites are represented by bold underlined letters and characters in bold shaded boxes indicate point mutations between parental lines.

(A) PIP41

```
ICCV2EST    TGTTCATGGTCAATGGAAGCNTTTTGACCTTAAGGTTAAGGACTCTAAAACCCTTCTCTT
ICC4958     TGTTCATGGTCAATGGAAGCATTTTGACCTTAAGGTTAAGGACTCTAAAACCCTTCTCTT
PI489777    TGTTCATGGTCAATGGAAGCATTTTGACCTTAAGGTTAAGGACTCTAAAACCCTTCTCTT
            ************************************************************

ICCV2EST    TGGTGAGAAGACCGTTACTGTTTTTGGAACTAGG--------------------------
ICC4958     TGGTGAGAAGACCGTTACTGTTTTTGGAACTAGGTAAAGTTTCTTTCTTTTTTCCTGTTT
PI489777    TGGTGAGAAGACCGTTACTGTTTTTGGAACTAGGCAAAGCAACTTTCTTTTTTCCTG---
            **********************************

ICCV2EST    ------------------------------------------------------------
ICC4958     TTATTTTTATTTTTATTTTTTATTTTAAATTTTAATATGTGATTGTGGATCTGAATGTTA
PI489777    ---TTTTTATTTTTATTTTTTATTTTAAATTTTAATATGTGATTGTGGATCTGAATGTTA

ICCV2EST    ------------------------------------------------------------
ICC4958     TATCCTTTTATATGTAATTAATGAATTAATTTTATGCTACTTTTATGTATAAGCATTTTT
PI489777    TATCCTTTTATATGTAATTAATAAACTAATTATATGCCACTTTTATGTATAACCAATTAG

ICCV2EST    -------------------------------------------AACCCTGAGGAGATTCC
ICC4958     TGAATTAATTTATTAATAAATCTGATTGGACATTAA--TTAGGAACCCTGAGGAGATTCC
PI489777    TGAATTAATTTATTAATAAATCTGATTGCATCTGAATGTTAGGAACCCTGAGGAGATTCC
                                                       *****************

ICCV2EST    ATGGGGTG        (119bp)
ICC4958     ATGGGGTG        (306bp)
PI489777    ATGGGGTG        (302bp)
            ********
```

(B) PIP44

```
ICCV2EST     TCTCAAAAGGTTGCAGTCTGAATTAATGTCTTTGAT------------------------
ICC4958      TCTCAAAAGGTTGCAGTCTGAATTGATGGCTTTGATGGTATGAATTTGGATCTCTTTTTC
PI489777     TCTCAAAAGGTTGCAGTCTGAATTGATGGCTTTGATGGTATGAATTTGGATCTCTTTTTC
             ************************************

ICCV2EST     ------------------------------------------------------------
ICC4958      ACTTTCTATGAATTTGGCCTCTTTTAGTTTATTGCTTTTTGTTTATGCAATGCAATGTAA
PI489777     ACTTTCTATGAATTTGGCCTCTTTTAGTTTATTGCTTTTTGTTTATGCAATGCAATGTAA

ICCV2EST     ------------------------------------------------------------
ICC4958      TTTGAGCACTACTTCAACAAGTAACTATTTTTCTTTTGACAAAAAAAAGTAATAACTTTG
PI489777     TTTGAGCACTACTTCAACAAGTAACTATTTTTCTTTTGACAAAAAAAAGTAATAACTTTG

ICCV2EST     ------------------------------------------------------------
ICC4958      TTGTAAGTATAAAGAGTTTTTTCACAACCAGGGTTTCTCAAATCGGTGCAGTCTGAATTA
PI489777     TTGTAAGTATAAAGAGTTTTTTCACAGCC-----TTCTCAAAT-GGTGCAGTCTGAATTA

ICCV2EST     ------------------------------------------------------------
ICC4958      ATGTCTTTGATGGTAAGAACCGTATATCCACTTTGTTTTGTGTTTCATGTCTTGGCAATA
PI489777     ATGTCTTTGATGGTAAGAACCGTATATCCACTTTGTTTTGCGTTTCATGTCTTGGCAATA

ICCV2EST     ------------------------------------------------------------
ICC4958      TCATGAATGTTTTAAGATAGTTTTTCTAATAGAAATTAATGGAATGGTTCAATTGAAAAG
PI489777     CCATTAATGTTTTAAGATAGTTTTTCTAATAGAAATTAATGGAATGGATCAATTGAAAAG

ICCV2EST     -------------------------------------------GATGAGTGGTGAGTCTG
ICC4958      AATATTATTAATGTTTTGCTTCTTAATTTAA---TGATTGTTAGATGAGTGGTGAGTCTG
PI489777     CATATTATTAATGTTTTGCTTCTAACTGCTAAGTGGATTGTTAGATGAGTGGTGAGTCTG
                                                         *****************

ICCV2EST     GTATATCTGCTTTTCCAGAGGAGGACAACATATTCTTATGGAAAGGAACAATAACAGGAA
ICC4958      GTATATCTGCTTTTCCAGA-GAGGACAACATATTCTTATGGAAAGGAACAGTAGCAGGAA
PI489777     GTATATCTGCTTTTCCAGA-GAGGACAACATATTCTTATGGAAAGGAACAGTAGCAAAGA
             ******************* ****************************************

ICCV2EST     GCAAAGA  (120bp)
ICC4958      GCAAAGA  (483bp)
PI489777     GCAAAGA  (480bp)
             *******
```

Additionally, some more ESTP loci were also sequenced to verify the amplified products (data not shown).

Similarly, MAFFT alignment at two PIP loci (PIP41; Fig. 6.4A and PIP44; Fig. 6.4 B) confirmed that PCR products amplified in chickpea were homologous to the sequences from which they were designed. Sequence analysis at the two loci revealed that indels in the sequences was the common cause responsible for polymorphism between the mapping parents. For example, at locus PIP41, 6bp deletion at positions 58-63 in PI489777 and 2bp (277 and 278) in ICC4958 was responsible for size variant alleles. Besides indels, both loci (PIP41 and PIP44) harbored several SNPs in their intronic regions (Fig. 6.4 A and B).

6.3 Discussion

The swelling EST databases strengthened by both structural and functional genomic projects have led to a range of contemporary genetic marker systems that have been exploited vastly in the recent past for enhancement of crop characteristics. Nowadays, molecular markers developed from EST sequences are preferred over anonymous markers due to inherent attributes of ESTs (represent the expressed regions of the genome). Since these EST-derived markers target specific genes, they provide an important opportunity to tag genes linked to desired trait/QTL and allow gene cloning and marker assisted breeding. The presence of SSRs in EST sequences provides a way to determine the genome positions of ESTs. However, the number of ESTs that possess polymorphic SSRs appears rather limited for mapping purpose especially in inbred crops like chickpea (discussed in Chapter 5). Therefore, for generating the high-density gene rich linkage map, sufficient numbers and different kinds of molecular markers are required. Thus in the present study, two new types of PCR based EST markers namely ESTPs and PIP were developed for the first time from chickpea ESTs and applied for map construction in order to localize their map positions (described in chapter 8). These mapped markers would provide a preliminary description of the organization of expressed genes (Jermstad et al. 1998) in the chickpea genome, tagging genes of interest and insights about genome evolution.

In the present study, ESTP primers were designed complementary to any randomly selected non-specific region of chickpea ESTs. The PCR amplification success rate of chickpea ESTP primers obtained was 72.5% (58/80) which was much higher than 45.2% reported by Buhariwalla et al. 2005. However, this observation alongwith the size polymorphism rate of 17.2% (10/58) obtained in the chickpea mapping parents were quite

consistent to those observed in pines (Temesgen et al. 2001) and *Rhododendron* (Wei et al. 2005). Generally it is observed that primers located close to 3'-end of the EST sequence tends to amplify genomic DNA more reliably than those designed from random coding regions (Temesgen et al. 2001). To detect polymorphisms, PCR amplifications using EST specific primers are generally followed by restriction enzyme digestion, heteroduplex analysis, denaturing gradient gel electrophoresis (DGGE), or single-stranded conformational polymorphism (SSCP) gels (Harry et al. 1998; Cato et al. 2001; Temesgen et al. 2001; Rowland and Dhanaraj 2003). However, in the present study, chickpea ESTP markers were resolved on 8% PAGE gels. Thus it could be anticipated that the polymorphism-generating efficiency of these markers may be further augmented by using the above mentioned analyses after PCR amplification.

To efficiently exploit the available chickpea EST resources for increased DNA marker development, another strategy i.e. primers targeting intron amplification was utilized in the present study. The burgeoning knowledge in model plants and well-studied crops have opened new doors in comparative biology, with great potential benefits for orphan species lacking sequence information. Comparative genomics has widely revealed the genome conservation among the closely related species (Choi et al. 2004a), thus providing a strategy to cross utilize the genomic tools from model to yet uncharacteized crops or vice versa to study genetic diversity, delineating gene coding regions and genomic rearrangements. In a need to identify efficient tools for genetic analysis, intronic sequences that are reported to be more variable than exons gained popularity as a valuable source for genetic marker development. Initially this study was restricted to only model plants. However, based on comparsions with the genome sequences of *Arabidopsis*, Choi et al. 2004a and b; Wei et al. 2005 designed intron-flanking primers in the target species that anneal to the conserved exonic regions and amplify across introns. With the development of web based program like PIP (Yang et al. 2007) and GeMProspector (Fredslund et al. 2006) recently, the designing of intron-targeted primers has accelerated rapidly even in species where genomic sequences are lacking.

The program 'PIP' employed in the present study predicts the intron positions in the EST sequences of query plant (chickpea) according to close homologs in model plants (*Arabidopsis*), with the intent of designing primers that anneal in conserved regions of exons and amplify across introns. Therefore, the efficiency and reliability of PIP marker exploitation is entirely dependent on the conservation of intron-exon junctions. The high PCR

amplification success rate of chickpea PIP markers observed herein (83.3%) and its capability to amplify intronic regions pointed to its efficiency in primer–template alignment which could be attributed to accuracy and reliability of the program. This result further indicates that intron positions are highly conserved in plants and therefore using model plants to predict intron positions in other plants is feasible (Yang et al. 2007). Besides the high PCR success rate, the high percent polymorphism 31.5% (24/76) of PIP markers compared to ESTP markers (17.2%) in the present study illustrated the presence of substantial levels of variation in intronic regions of chickpea thereby establishing PIPs as a reliable source of DNA markers. Panjabi et al. 2008 obtained similar percentage of polymorphism (32.0%) between the *Brassica* lines using PIP markers. Earlier studies have well documented that introns evolve more rapidly than exon sequences in both plants (Small and Wendel 2000) and animals (Hughes and Yeager 1997) suggesting that ITP markers could efficiently meet the need of genetic research and plant breeding. Moreover, the high percent transferability of ITP markers well documented in several studies has further enhanced their utility for comparative mapping, synteny studies and elucidation of phylogenetic relationships.

To confirm the PCR specificity and to elucidate the basis of length variations, two ESTP primers (CEST44 and CEST 86) and two PIP primers (PIP41 and PIP 44) were sequenced. At the CEST44 locus, the presence of an intron within the amplified genomic fragments resulted into higher sized band of 415bp instead of the expected 296bp (Fig. 6.3 A). The incidence of the amplified product being larger than the predicted size, possibly due to the presence of introns is higher in the EST based markers (Cordeiro et al. 2001; Temesgen et al. 2001). Similar observations were also obtained at the two PIP loci verifying PCR specificity and presence of introns (Fig. 6.3 A and B).

Amongst all the sequenced loci, except CEST86, indels in the intronic sequences emerged as the major factors for the polymorphism between the parents. Introns are known to possess bundles of neutral space where mutations can occur without any serious effect on gene function. The indels in these regions were probably not deleterious and hence the genotype possessing the deletions could sustain themselves. These deletions can be attributed to the unequal intra-strand recombination, segmental duplication or due to non-reciprocal translocation. The deletions in the intronic regions have been documented by several researchers in their findings (Chetelat et al. 1995; Hongtrakul et al. 1998). Li (1997) compared nucleotide substitution rates of different regions of the gene and found that the introns exhibit the highest levels of substitution followed by the UTRs and coding regions

thus exhibiting higher levels of polymorphism within species compared to the exonic sequences. Besides indels, point muations were also observed (Figs. 6.3 and 6.4) thus highlighting the prospects of SNP mapping in chickpea as these represent the most fundamental source of variation for molecular marker development.

Thus, the present study provides an arsenal of EST-based molecular markers for chickpea analysis. A total of 58 ESTPs and 76 functional PIP markers were developed for chickpea that could be potentially utilized for linkage mapping in the backdrop of anonymous markers (described in chapter 8). The high PCR success rate accompanied by the high percent polymorphism exhibited by chickpea intron targeted primers (developed and characterized for the first time in chickpea) indicated that ITPs would serve as excellent markers for the studies of genome evolution, comparative genomics and marker-assisted selection in chickpea. Sequence analysis illustrated that indels were responsible for size polymorphism and there is an abundance of point mutations in the chickpea genome that need to be exploited for crop improvement programs.

Table 6.1: Characteristics of ESTP markers designed from the chickpea EST sequences. The designed primer sequences, expected and observed allele sizes (bp) in *C. arietinum* cv. ICCV2, GenBank accession numbers and their putative functions based on BLASTX results are mentioned.

S. No	Primer	Primer Sequence	Expected Size (bp)	Observed size (bp)	Putative Function	GeneBank No.
1	CEST21	TTTTTGAGGAACTTGGAGGAGA/ CCAATCAGTCCCTTTAATTGGT	429	535	B-D galactosidase	EX151629
2	CEST22	TGAGCGGATAACAATTTCACAC/ TTGGATTGGATGATGTAAGCAC	366	680	L.arbinofuranosidase	
3	CEST23	CTGTGGAATTGGTGCTGATAAA/ AGACGTCTCTCATTGCCTTCTC	402	NA	Glutamine synthetase	EX151668
4	CEST24	CCATGGCAGTAGAACTGATGAC/ TGGAGATTGAGGTAAAGGTGCT	332	332	WRKY TF	EY457881
5	CEST25	TTTATTCCCTTTACCGGAACTG/ CTCCCCCATTGGAATATGAA	501	501	Legumin	
6	CEST26	GGGGAAGTTTGAAATTTGGTTC/ GAGTTGACGAGCTGTTTGACTG	354	498	A-D- mannosidase	EX567545
7	CEST27	TCATTGCTGCTAATCAAGGAGA/ GGAAATCCACCAACACAGATTT	411	530	Myoinositol phosphtase	EX152031
8	CEST28	AGATCTTCTCAAATGGCCGAAC/ CAACCTTCCCTTGTTCGTTTAC	443	515	Glutamine peroxidase	EX567696
9	CEST29	TTGGTGGACCAAATATGCAA/ GAGCTTGAGAATTTCTTGCTACA	357	357	Pectinesterase	EX151961
10	CEST30	CCATGCACTCTCTTTCTTGTCA/ CAGGGGTAAGATTTTTGCTT	386	NA	Glutathione transporter	EX151954
11	CEST31	ATGCCACCAAAGGTTCTGAC/ CCCAGCCTCTCTAAACACCTTA	393	790	Haem peroxidase	EX567627
12	CEST32	GAACGAGATGGTGGCGATTA/ TTGGCAGAATGTATGCACTTTAG	298	298	Wound-inuduced kinase	EX567687
13	CEST33	GTGGATTAAGTGTTTCCCTTGC/ TACCGGCGTAAGCTAATGAAAT	220	815	DNA-binding	EX151680
14	CEST34	GAGGTCACCGGAAAAGTCAA TGACCGCCATAAAACATCGT	300	HS	ZIM	EX152021
15	CEST35	CGCTCAAATTTCTCAACAACTC/ CCCTTTCTCTAGCACTGTCTTTC	337	825	Sugar transporter	EX151642
16	CEST36	CGACATTCCATATTCACTTCC A/ CTAAAG TCTCAACGCCTTGTTG	377	NA	Seed L-asparaginase	EX152055
17	CEST37	CATGACAAAGGATGGAAGGATT/ ATTTCCAAAGAGACTTTGAGG	243	243	Aspartate aminotransferase	EX152060
18	CEST38	TGATGAGGCCCAAGATCTGGTA/ AAGTCAAGGCGTTGCTCATCT	344	344	Ripening-related Arabid	EX151708
19	CEST39	CACTTAGAATCAATGGCAAGCA/AG TGAGATGTTCCATATTCCTG	399	720	Lipid-transfer protein	EX567560
20	CEST40	TTCCTTCAAGCTATTGTCACCA/ AGCATCATTTGCTCCATCTTCT	393	393	Invertase inhibitor chick	EX567734
21	CEST41	TGGTAGTGAACAAATGGTTTGC/ TAGTGCATTGGCACTTTAAGCA	398	398	Aminopeptidase	EX151819
22	CEST42	CACTTTGGCATTAGCAATCAAC/ AAATTAAGGACACTTGTGTCTGC	350	350	Gibberellin regulated Medi	EX567539
23	CEST43	AGGTGGTGCGTTCAATGTTATT/ GATTACGCATTCCAAGAAGGAA	348	680	Auxinamidohydrolase	EX151698
24	CEST44	ATGGGGAACTGATGAAAACACT/ TATAGGCACGCCTCACATTAAA	296	415	Annexin	EX567748
25	CEST45	AACAAGCCGAACGTTACGAAG/ GCCACTCTTAAACTCAGCAAGG	376	376	14-3-3 brain protein homolog	EY457880
26	CEST46	GCTTGAATTTTATGCTCCTTGG/ AGACCGCAGGTAAAAACTGTGT	385	780	Disulfide isomerase	EX567537
27	CEST47	TCGTGCTACACAAATCAAAACC/ AATTTTGCCCCTGAGCTTAAC	384	384	Ethylene-responsive TF	EX567594
28	CEST48	AGATTGCAGACATTCCATTCCT/ GTGACACTGGTGGTGATCTTGT	348	645	Calreticulin/calnexin	EX151757
29	CEST49	GACGCAACGGTCTGTGTATTTA/ CTTGTCTTTTTGGCTGATAATGG	340	340	Lissencephaly type-	EX151671
30	CEST50	GCTATCCATGGGTCTTGTCAAT/	318	318	Enoyl-CoA	EX151749

		GGTTGCAGCACTCTATCAACAT			hydratase/isomerase	
31	CEST51	ATAAAATCTGGGTTGGCACTGT/ CAAAGTTGTGGCATGTAAAACA	289	289	Starch synthase	EY457878
32	CEST52	ATCGTGATAACCATGGCAAGTA/ AATGTTGTCCTCACGCTTTTG	376	376	Seed-specific protein	EX567648
33	CEST53	TCTCCCAACAAAAACACACATC/ GATCAAACGTCAAACACTCTGC	388	NA	Acyl-carrier protein	EX567749
34	CEST54	GACTCGGATACCAGAAACGAAC/ GCAAAAGGAAAGATGAATCCAG	424	424	Hypothetical protein	EX567747
35	CEST55	GCCCTTTTCTTTTGCATATC/ CTAAAAGTGGGGGATCTCACTG	447	447	Flowering locus	EX151690
36	CEST56	TGGATTGGGTGCATAAGATAAG/ GAGGAATATCTTGCCAGCAGAC	448	HS	Unknown protein	EX151717
37	CEST57	CCCTATAGCCGGTCGTTATCTA/ TTCCAGGATCCATGAATT TACC	589	HS	Starch Phosphorylase	EX151754
38	CEST58	GTTCCGTCGAAAATTCCC TAAC/ CTCCCCTTTCCTCTTCACTAGA	437	550	Ferritin	EX151760
39	CEST59	CTTGCTCAATCTCTTTCCGTTT/ TGATCAGTTTTCTGACGCTGAC	424	424	Stress-enhanced protein	EX151765
40	CEST60	AGAGAGAGAGAG ATGCGGTGTT/ GCTCTTACTTCTTCTCTTCAGGT	508	NA	DnaJ protein	EX151776
41	CEST61	ATCAGGCAGATATTTGGAATGG/ TTAAGCTTCCTCAGACTGCAAC	338	HS	Zinc finger Protein	EX151801
42	CEST62	CCCTAGTTCGATCATCAA AACC/ TGTAGATTCCAACGATTTCAC G	349	349	Developmental protein	EX151826
43	CEST63	CGCTCTTCTCTTTCTCTTCTCG / AGTTACAGTACCAACCCTGGTG	338	NA	Lipooxygenase	EX151853
44	CEST64	AGGTGCAGCTCAAACTCTCTCT/ GCTTTTCTGCTTCTCTTCCAAA	447	447	Unknown protein	EX151967
45	CEST65	GGAATGCTGTTTCTCTCTGG/ GGCAGTGCAGACAGAGCTATAA	337	NA	Ring-box protein	EX152043
46	CEST66	GAAGGATTTCAACAAGGG ATTG/ TTTAGGCCCAGTTCAACAAGAT	383	383	Unknown Protein	EX152104
47	CEST67	TCACCAAGGATAAACTCAATGG/ CACTACACGTTTCCCAGTGTTC	325	720	Ankyrin related protein	EX152131
48	CEST68	TTGCATTAGCAGAGATGCAGAT/ GACAAACACCATCCTAAGCACA	491	HS	Ubiquitin fusion protein	EX567534
49	CEST69	CAAGTCCGCTTATCTCCAACTC/ AATTCAGCAGCAGTTCGAAAG	330	330	Hypothetical protein	EX567553
50	CEST70	CGGCCTAGTGAAAACAGATTG/ TTTTTCATTCAAGAGGGCATTC	272	272	Xyloglucan endotransglycosylase	EX567677
51	CEST71	CGTCAATTTAGGGATGTCCAAC/ AGGTTGAGGAGCATACTGTGGT	334	804	Acetyl-CoA carboxylase	EX151686
52	CEST72	CTTTGATAATACCGCTGCTGAA/ TGATTCATGTCTAGCAAAGTGAG	402	402	Beta-glucosidase	EX152037
53	CEST73	AAAACCTTCCTCCAACTTCACA/ CTGCGATTTCAATCTCATGG	502	502	Anaphase-promoting complex	EX152030
54	CEST74	GCCCTTAAGCAGTGGTATCAAC/ CCATGGTGTCACAATCTCATTC	397	397	No homolgy	EX567541
55	CEST75	GCAAGCATCAGTGGATCAAATA/ CTGATCTTGCACAGATTCTTCG	354	354	Peroxidase	EX151897
56	CEST76	TGTGTGTTGATGTTGCTGTTCA/ GGTCCTGGCTTGCACTATAAAA	301	301	Plant-self incompability	EX151911
57	CEST77	TTGTCTTGAATGTCACCAAAGC/ AGAGGGTACAAGCTCAAAGAGG	478	478	Xylogen	EX151913
58	CEST78	TCCTCAGCCTAAAGAGGAACAC/ GAAAACCACTCCATTCCATCAT	369	369	Nutrient reservoir	EX152107
59	CEST79	AAAATGTCTTGCTGTGGTGGTA/ CTCCAAAAGGGTGTTTAGATT	400	HS	Metalothionin	EX152094
60	CEST80	TCAATCTTCTTCACTTTCGTCTC / CCACGAGATATGCTTCTGCT	350	350	Disesae resistant	EX152090
61	CEST81	CATGTCTTCTGAGACTGTTGTCC/ ACAACAACATTGTCGGATTCAG	407	407	Copper chaperone	EX152083
62	CEST82	GAAAAAGTTGGGTTTGGTTCAG/ TGATCCACTGAAAGTTGTTTGG	462	462	Protein phosphates	EX152052
63	CEST83	CGTCAGGTAGCTCAAGTATATCAG/ GATCTGTTCCTTCCAATTTGCT	428	NA	Iron-superoxide dismutase	EX151999
64	CEST84	TGAAACCAAGCTAGGAGTGTTG/ CAGCTGCTTAAGTTCCTTCACA	407	NA	MAD box family protein	EX151971

65	CEST85	GGAGAGATGCTCCACAAGGTTA/ TCGCTAGAAGGCTTAGAACCAG	410	410	Lhca5 protein	EX151931
66	CEST86	GAATTCAACTCGCCCTGAAA/ GCACCAATAGGCACAACCTT	430	430	Hypothetical protein	EX151902
67	CEST87	CGGAGTGACTGGAGCAATTA/ TCACCAAACCTTCAGGCATA	406	NA	Cationic amino acid transporter	EX151846
68	CEST88	ATGGGTGGCAATTTATCCAA/ GCCTTAACCACGTGCAAAGA	420	520	Thioredoxin	EX151814
69	CEST89	TCGATCGATGAGGATTTTCC/ CCTTAGAAATCCCGTGCCTA	520	520	Unknown protein	EX151807
70	CEST90	ACAGCGATAACTGCGACTGA/ AAACCGTATTCAGGCTTCGT	410	NA	Nicotina induced protein	EX151779
71	CEST91	GCATCTGCACCTGCAAATAAT / TTGGGAATCCCCTAGGTTCT	413	NA	Calcium binding protein	
72	CEST92	AACAGTCTGAAAAGGCCATCAT/ GCCCTTAAGCAGTGGTATCAAC	476	NA	UV-repressor protein	EX567724
73	CEST93	AAGGGAAAGAATGGTAGCAACA/ TAGGGGGAGAATCTTGAATGAA	450	524	Hypotheical protein	EX567750
74	CEST94	TCGAGAAAATCTCATCATGTCG/ TTACACTGCAATCAAGCATTCC	513	NA	Profilin type	EX567708
75	CEST95	GCTTCCATCAATTTCTGGTCA/ AAAAACACCCTCCAAATGACAC	471	471	No homolgy	EY457896
76	CEST96	CTTCTCAGTGTGTGCCTTTCTG/ ACAGCAGCCCTCTTAATCTTTG	398	NA	Hydroxyproline rich protein	EX567638
77	CEST97	GTTGATGTGCTTGGTGGTCTAA/ ACACACGCAATGAAACAAAGAG	520	550	MtN4 protein	EX567699
78	CEST98	TTGGCGAAGGTAATCGAGTATT/ TCAGCACGAATAATCCCATAGA	445	NA	SKP1-component	EX567597
79	CEST99	TCGCATAAGAGGGTTTGGAT/ TTCCATCCAAAGTTGGAACC	433	433	Unknown protein	EX151770
80	CEST100	GTACGAGAGGAACCGTTGATTC/ ACTAGGGCTGAATCTCAGTGGA	354	354	Senescence-associated protein	EX567738

NA: not amplified
HS: >1.0 kb

Table 6.2: List of potential intron polymorphic (PIP) markers developed from chickpea ESTs. The designed primer pair sequences, observed allele size (bp) in *C. arietinum* cv. ICCV2, predicted intron size in *Arabidopsis*, putative functions based on BLASTX results and GenBank accession numbers are mentioned. Serial nos. 1-67 represent markers designed from the chickpea seed unigenes whereas serial nos. 68-110 represent markers designed from drought related ESTs (Chattopdhyay et al. pers. comm.) for which GenBank accession nos. are not available.

S.No.	Primer Name	Primer sequence	Observed size (bp)	Predicated intron size in *Arabidopsis* (bp)	Putative function	GenBank No.
1	PIP1	AAACTCAAGCTTGGAGAAATCG/ CGACATCCCACACAGTGAAG	520	172	ADP-ribosylation	EX567533
2	PIP2	TGAAACCGTGGAGTACAAGAA/ ATGAGACCCTGGGTGTTCTG	223, 635	230	-do-	EX567533
3	PIP3	GCAATGATAGAGACCGAGTGG/ TTTGTTGGCAAAAACAAGCA	422, 635	82	-do-	EX567533
4	PIP4	TGATGTTGGCTGGTTGAAAA/ TATCAGCCAATGATGCTGGA	NA	78	Alpha-mannosidase	EX567545
5	PIP5*	GGAATGCAGAGCACAGTCAA/ CAATATAATGTGTTGCTGCCTCA	100	103	-do-	EX567545
6	PIP6	TGTTTAGTGTAAGTGAGATTGAAGCA/ CGCCAACTGAAATTCTTCCT	306	117	Calcineurin B-like protein	EX567554
7	PIP7*	TGCAGATAGGGGTGTTCGAT/ GTGCGTTTGGATGAAAGACA	100	120	-do-	EX567554
8	PIP8	CCCCTACAAATCTCAAAATCAT/ AAATGGAAATGTTCCAACCAG	754	89	Cell death protein	EX567564
9	PIP9	TGAAGGCTGTAATAGAGAAAACCA/ CAGCCATGCGACATTCACTA	154	92	3-ketoacyl-CoA thiolase	EX567598
10	PIP10*	TGCTGGTTTTCCTGAAACTG/ AAAATCCCAGCCCTTATAGCA	107	167	-do-	EX567598
11	PIP11	ATTCCTCCACAACGGTGAAA/ GAATTGTTTGGAATGAACATTGG	NA	597	Unknown protein	EX567620
12	PIP12	TGGAGCCACTTAAGGAGCAA/ AAGGAACTTCAGGTCCACCA	267	105	Haem peroxidase	EX567627
13	PIP13	CGGAAAGGACTATGGAAAGC/ TCTCTTAAAATTCAATGGTGACCTC	222	222	Hydroxyproline-rich glycoprotein	EX567638
14	PIP14	GAGAGTTCCTAAGAAGAAGGATCTCA/ TTTCTGACAATCAGTTTTCTACCC	241	100	-do-	EX567638
15	PIP15	ATTGCCATGAGTCTGCCATA/ TCTCACAGCCTTTCCAACTG	134	124	Tubulin binding	EX567674
16	PIP16*	ATTGCCATGAGTCTGCCATA/ TCTCACAGCCTTTCCAACTG	107	275	-do-	EX567674
17	PIP17*	CCAAGATCAGATTCAGTTGGTG/ TTTCATGGAGCTTATGAACAGAAA	108	83	MAPK	EX567687
18	PIP18	CAGACATCAATGGCAGCTGTA/ GGTCTTGCTCGTCCAGACAC	NA	88	-do-	EX567687
19	PIP19	GGTCACAATTTGCATTTAACAA/ GAGGAGCATTGTCAGTATCTTTTG	NA	101	-do-	EX567687
20	PIP20	GCCGAAGCAGAAAGTGTCTA/ CTCAGATCTGTCCAAGAGACCA	742	83	chloroplast binding protein	EX567720
21	PIP21	GCCGAAGCAGAAAGTGTCTA/ CTCAGATCTGTCCAAGAGACCA	735	98	-do-	EX567720
22	PIP22	AGTGGCATCAATGGAGAGGT/ TGTAACGAGCAGCCCATATC	182	94	Glutamine synthetase	EX151668
23	PIP23	AGATATGGGCTGCTCGTTACA/ CCAGCTCCATTCCAATCACT	413	89	-do-	EX151668
24	PIP24	TCAAACCCGACACTGTTTCA/ CGGTAATATTTGGCAAACTTCTC	172	74	cytochrome-c oxidase	EX151674
25	PIP25*	AGAGACGGATCAGCATGGAC/	113	518	CGI-126 protein	EX151720

		TTAGAAGCGGAGATCCGAAA				
26	PIP26	TCACTATTCATTTCCCTCCAGA/ GGCAAATGCTTCCATTGCTA	400	162	UBC9	EX151726
27	PIP27*	TGCTCTAAGAAATCTGGGAAGG/ CGGTTGCTTCTCTGGGAGT	100	97	Nucleic acid-binding	EX151729
28	PIP28	AAGACTCCCAGAGAAGCAACC/ GACAAGTGCCAGCTAAGATACG	754	82	-do-	EX151729
29	PIP29	TCTTCATTTCATCAAGAAGTATCAGA/ TGGCCAATAGTAACATGATCTCC	214, 238	305	-do-	EX151729
30	PIP30	CATCCTGCATGCTGTTTACC/ TTCCGGAACATCTCTTTTCCT	203	95	Hypothetical protein	EX151759
31	PIP31	GCCAATGCTTTTCTTGGATT/ GCAAAATGTCCAAAGCCAGT	264	80	-do-	EX151759
32	PIP32	CCTGTTAATCCCAAACCCTTT/ CGAATCAACTGATACAAGATAACCT	245	248	Ribonucleoprotein-related	EX151763
33	PIP33*	AGAACACCAAGGGAGGGAAC/ CCCTTTCCGGTAGCAATTT	100	258	Stress enhanced protein	EX151765
34	PIP34*	TCCAAGATAGCAAAACGATGC/ ACGGTGGAGGTAGCTTCAGA	100	478	Unknown protein	EX151770
35	PIP35	GCCGTTCTAAAGGGTAACTCTG/ GGAGCAAGGCCAGTGATG	495	87	Superoxide dismutase II	EX151780
36	PIP36	TGGCAAGACAGTTCCAAAAA/ GTGGAATGAGCTGCCTTTGT	504	247	-do-	EX151780
37	PIP37	GAGTTTTGGGACACAGCTCCT/ ATTGCCTGTGCTAAGGGATAGA	HS	422	Unknown protein	EX151837
38	PIP38	TCCATGGGGAATAACTGGAG/ ACAGTCTTGAACGGGTTTCC	NA	255	-do-	
39	PIP39	TCATCATGGGGGATAGTCAA/ GTTTGTCCAGAGCCAACAGC	198	78	Proteosome related	EX151886
40	PIP40	CCACATGGATTGTTACGTTG/ TGCTCTGTGCCTCATTGGTA	264	95	Unknown protein	EX151910
41	PIP41	TGTTCATGGTCAATGGAAGC/ CACCCCATGGAATCTCCTC	306	93	Glyceraldehyde-3-phosphate	EX151958
42	PIP42	TCGTTCTCTATTGCTTTCTCAATTT/ AGCTTGGCATCTTTGTTTTCA	200	380	Glutathione transferase	EX151968
43	PIP43	TTGTTGCTCCCACAAAACAC/ ACCAGACTCACCACTCATCATC	718	102	UBC19	EX152062
44	PIP44	TCTCAAAAGGTTGCAGTCTGA/ TCTTTGCTTCCTGTTATTGTTCC	483	108	-do-	EX152062
45	PIP45*	TGCTCAAAGAAATCTGGAAAGG/ CCGTTGCTTCTCTGGGAGT	100	97	Nucleic acid-binding	EX152004
46	PIP46	AGACTCCCAGAGAAGCAACG/ CACTATGACAAGTGCCAGCTAA	820	82	-do-	EX152004
47	PIP47	TCTTCATTTCATCAAGAAGTATCAGA/ TGGCCAATAGTAACATGATCTCC	184	305	-do-	EX152004
48	PIP48*	CACGGTAGTATTGGCGGAGT/ TCTTGAGAAAAGCAAAGTCTGG	114	335	ATVAMP714	EX152023
49	PIP49	CAGACAACAACCCCAACCTT/ GCAGATTGCTTATAGTTGGATGG	764	630	NADH dehydrogenase	EX152089
50	PIP50*	GCTTTTTGCGAAGAAGGAAA/ TGGTGACGATCTCTCCAAGC	101	77	ADP-ribosylation factor	EX152113
51	PIP51	GAAGTCGAGAGCAGCGACAC/ TTCCATCTTCAAGCTGTTTCC	204	78	-do-	EX152113
52	PIP52	CCGGGAAACAGCTTGAAGA/ TTCGATGATACCTCCACGAAG	183	103	-do-	EX152113
53	PIP53	ACCTTCTTTGATGGCTTTGG/ ACATTTCTTTTTGCGGCAGT	388	235	-do-	EX152113
54	PIP54	ACGCGAATTGGAGAACAAAG/ CATTCTCGCCAAGTTGGATT	435	94	Unknown protein	EX151967
55	PIP55	CCAAGTTAGAAAGAAGTACACAATCC/ AACGAGTACAGGGCCAATCA	417	103	-do-	EX151967
56	PIP56	GCTGAGGCCAATGCTAATGT/ GCTTTTCTGCTTCTCTTCCAAA	374	235	-do-	EX151967
57	PIP57	AACCATGTCTCAGACTGTTGTTCT/ TGATTCCACACCATCCAATTT	NA	187	Copper chaperone	EX567583
58	PIP58	TTGAAGGAAGTGACTCACAAGG/ TCGGAAGTTTTCTGCAGTTTT	NA	123	Peptidyl-prolyl cis-trans isomerase	EX151813
59	PIP59	AAAATTCCTGTTGATGGCCTTA/ TCCAGAAATATTGTTGCAGTGAA	615	126	Putative protein	EX151908

60	PIP60	TTCAAAGAGACGATGTTGAACTTG/ CATTGACCATGAACAGTGTCG	203	93	Glyceraldehyde-3-phosphate deh	EX151958
61	PIP61	TAAGCTCAAGCTTGGAGAAATTG/ AACATCCCAGACAGTAAAGCTGA	650	310	-do-	EX151958
62	PIP62	TGATGTTGGCTGGTTGAAAA/ TATCAGCCAATGATGCTGGA	NA	89	ADP-ribosylation factor	EX151850
63	PIP63	GGAATGCAGAGCACAGTCAA/ CAATATAATGTGTTGCTGCCTCA	NA	82	-do-	EX151850
64	PIP64	ATGGCTAAGGGTCCAGGTCT/ GGTGAGTAAGTGGTGATGCTGA	228	499	Outer mitochondrial m	EX151957
65	PIP65	GCCAGTGCCAATCCTCTTAG/ CTGAGCTGAAAAAGGTCCTAAA	264	98	Chlorophyll binding protein	EX151974
66	PIP66	GCCGAAGCAGAAAGTGTCTA/ CTCAGATCTGTCCAAGAGACCA	HS	83	Chloroplast binding protein	EX567720
67	PIP67	GCCGAAGCAGAAAGTGTCTA CTCAGATCTGTCCAAGAGACCA	HS	89	-do-	EX567720
68	PIP68	GCCTCCAACTTTGAGATTTGT/ ACCAGCCACGGATATTTCAC	172	94	Aconitase family	-
69	PIP69	TGGTTTGGGATCTGGAATTT/ CAGCTCCCCGACCACTATTA	215	109	Integral membrane	-
70	PIP70	AGACTGTTCGCCCCACCT/ GGATCAGCTGACATTTCCTCA	450	590	Ceramidase family protein	-
71	PIP71	GAGAAGTATGGCCCTGAGAGG/ CCACAACAGGCTTTAGACCA	750	104	-do-	-
72	PIP72	ACAACAAGCTGGCACCATTC/ TGACCATGCTTCCCAGTTTT	258	271	Seed maturation	-
73	PIP73*	CAAGATGCCATGTCTGTTGC/ TCTGTTTCAAAGTAGCCGACAC	100	291	Chaperonin	-
74	PIP74	CCTCGGTCAAGACGGTAGC/ GTTCCAGGCTCAGCAAAATC	456	92	Profilin	-
75	PIP75	CAAGGAGCTGAAGGATTTGC/ ACTATCAGCCGGACCCATAA	321	350	Ubiquitin	-
76	PIP76	TGGAGGTGTATTTCTGGTCACT/ TGCTATTGATGTTGGGGTGA	344	162	-do-	-
77	PIP77	GAATTGCAGACTCTACTTTTATTCCA/ AAAGTAGTCATAGTGAGGCTCATACTT	367	113	Prolyl-4-hydroxylase	-
78	PIP78	TGATGACTTTAACACACAGAATGG/ CTGGCAGCAGGGAACACT	156	137	-do-	-
79	PIP79	AACTTTGGCTACAATTCCAGTCT/ CCAGGGAAGGGATCCATAAG	162	101	PPF-1 protein	-
80	PIP80	GGTTTCTTATGGATCCCTTCC/ TTCCAACCCAAAGGTGGAT	NA	84	-do-	-
81	PIP81	TTTGGTAGCAGCTGGTTACTG/ CTCCCCAAGAGAAGGATCAA	224	95	Fructose-1, 6-bisphosphatase	-
82	PIP82	TCCTTCTCTTGGGGAGTTCA/ CATCCCAGTTTTTGGCATTT	238	77	-do-	-
83	PIP83	AACTTTATTCTGTGAATGAGGGAAA/CTTTGGTGGTGAACCGTCTT	194	88	-do-	-
84	PIP84	TGAATGGTGCTCTTTGCTCT/ CTCGTTCACCCTAATCACATCA	504	205	Zn protein	-
85	PIP85	TTGAAGAAGTGGCTTCAACG/ GAATCCAGCCTTTTCAGCTC	176	118	Glycolate oxidase	-
86	PIP86	TTGCTCTCACTGTTGATACTCCA/ GACCTTCAAAGTTCTTCAAGTTCA	302	77	-do-	
87	PIP87	TGCCACCATTTTTGAACTTG/ TCCAGCTTAGAGTGCGATCA	216	130	-do-	-
88	PIP88	GATTGGTGCTGGCTTTGTTC/ AGGCCTTCTTTAAGCGCAAG	NA	99	Cysteine synthase	-
89	PIP89	GCATTATTCACCAATTAGGACCA/ CTCTTCTTGTGCAGCGGTTG	NA	78	-do-	-
90	PIP90	CATGTGACCAGCAATGTCTCA/ TGAATTGTGAAGCAGCAGCA	268	85	Omega-6-desaturase	-
91	PIP91	TTGGACAATCAATTCCTATCACTC/ ACTTATGAAGCCGAGCCAAT	776	74	-do-	-
92	PIP92	TTGTGATGATCCACGCAACT/ TCCCTTGGCTCATAATGTCTG	214	111	40 S ribosomal protein	-
93	PIP93*	AAACCTGCACTTCCTGCTCT/ TTGAATTTTATCATTTGTACCATCAGA	114	115	Eukaryotic TEF	-
94	PIP94	GCAGCTCAGGCTTCAAGAAA/	218	98	Rubisco activase	-

		CCAACCTCGCACTGACATTA				
95	PIP95	TAACGGTGGATCACATGCAG AGACTTCCACACCCATCTTCA	207	95	Glycerol-3-phosphate acyltransferase	-
96	PIP96	GAAGCCATGAAGATGGGTGT/ AAGCCACCTTCATCACCAAC	400	87	-do-	-
97	PIP97	AGAGTTCACCGGCTCCTACA/ TGATTCTGGCAAATCCTTCA	NA	243	Putative WD-40 repeat	-
98	PIP98	GAATTAGGGAATTGCCACAGA/ TAGCTCCGAGAACAGCATGA	212	144	Chloroplast FtsH proteinase	-
99	PIP99	GAAAGCCAACCTAACCGTCA/ AGTTGGGCACATTGCAAGA	538	137	-do-	-
100	PIP100	TTGACTGGAATCCCCATGAT/ CCAACCCCATTAGAAGTGAGA	247	74	-do-	--
101	PIP101	CATATTATGGGATGCACGTGTT/ ATCATCATGGGGATTCCAGTC	NA	80	-do-	-
102	PIP102	AATGCTTTCCCGTGGTTTTA/ TGATCTCAAGTGCCTCAGGA	HS	81	Allointoinase	-
103	PIP103	GGTCTCCATGGCTTCCATATT/ AGGGGAACCATGTTCTTTCC	268	132	Superoxide dismutase	-
104	PIP104	ATGCTTACTGGCCCTGTCAC/ CATCCTTGATAGCCAAAGCA	203, 272	84	Methionine synthase	-
105	PIP105	TCCATGGGGAATAACTGGAG/ ACAGTCTTGAACGGGTTTCC	HS	255	Hypothetical protein	-
106	PIP106	CACTAGTAGCTAACACTTCAAACATGC/ TCTGCCATCATACTGACATTG	476	89	Vascular ATP synthase	-
107	PIP107	AACATCTCGATGGGCAGAAG/ AGAGGCTAAACGTGCAGCAA	334	103	-do-	-
108	PIP108	TTGATGGCTCAGAAATTGTCA/ TTAGAGCACCAAATCCAGCA	172	100	Cytochrome synthase	-
109	PIP109	CTGAGCGTAAAGACGCTGCT/ GAGCAAGACCAAGCCTGATT	HS	77	14-3-3 brain protein	-
110	PIP110	GAGATTCTCAACTCTCCCGATA/ CCTTATATGATTCCTCTCCCAATG	485	91	-do-	-

* amplifying products of 100-120bp in chickpea and therefore predictably do not contain introns.

Chapter 7: Exploiting chickpea EST-SSR markers for genetic diversity and cross transferability studies across genus Cicer and related genera

7.1 Introduction

Genetic diversity determines the boundaries of crop productivity and survival. Therefore to increase crop yield, enhanced resistance to pests and diseases and improved grain quality, evaluation of genetic diversity and its effective management is indispensable for any crop improvement programs. In chickpea (*C. arietinum*), despite a large germplasm collection, the crop has a narrow genetic base which can be attributed to a series of genetic bottlenecks occurring in chickpea, including the restricted distribution of progenitors, the founder effect associated with domestication and the shift from winter to summer cropping (Abbo et al. 2003). These factors have substantially hampered the long term goal of chickpea breeders to develop elite cultivars (Croser et al. 2003; Millan et al. 2006). Therefore it is imperative for researchers to look into the wild species that are reported to be a treasure house of agronomically favorable traits which influence chickpea yield and stability. But attempts to introgress genes from wild species to cultivars or to improve chickpea breeding are not possible without the precise knowledge of the nature and extent of genetic relatedness between wild, exotic, cultivated and adapted chickpea germplasm. Thus for effective conservation and use of genetic resources, the evaluation of genetic resources is critical for genetic improvement of chickpea.

In the last two decades, plant breeding has benefited from DNA marker technologies that are being extensively utilized for genetic characterization and diversity analysis in several crop species (Gupta and Varshney 2000). In chickpea also, a variety of genetic markers have been employed to mine superior alleles or exploit the genetic diversity within germplasm collections: These include biochemical and isoenzyme analysis, prior to PCR markers and subsequently DNA-based markers like RFLP (Udupa et al. 1993), RAPD (Sant et al. 1999; Iruela et al. 2002; Sudupak et al. 2002; Chowdhary et al. 2002), ISSRs (Sudupak 2004) and SSRs (Hüttel et al. 1999; Winter et al. 1999; Sethy et al. 2006a). Analysis using these marker techniques have provided considerable insights into the genetic diversity and relatedness among the wild annual *Cicer* species (reviewed by Croser et al. 2003; Sethy et al. 2006b) and have lead to the conclusion that microsatellite markers are the best tool to address the genetic variability and thus may be considered as the most efficient marker system for germplasm evaluation in chickpea. Moreover, chickpea microsatellite- based markers like STMS have also proved to be equally efficient and reliable for use in cross-transferability studies (Choumane et al. 2000, 2004). The cross-transferability of microsatellite loci has been widely applied and has now been well established in many other plant species and

genera (Peakall et al. 1998; Kuleung et al. 2004; Varshney et al. 2005a) thus facilitating their use in elucidating phylogentic relationships, comparative mapping studies, identification of novel genes and understanding the pattern of gene evolution. However, to increase transfer rates across species that are genetically widely distant, microsatellites derived from ESTs are currently being preferred for germplasm evaluation because of their association with the highly conserved transcribed regions (Cordeiro et al. 2001; Varshney et al. 2005a). Though EST-SSR markers have demonstrated low polymorphism within species compared to genomic SSRs, but their high transferability and ability to detect the "true genetic diversity" available within or adjacent to the genes provides an effective approach to identify markers tightly linked to traits or candidate genes in order to expedite the isolation and introgression of important genes. Therefore, EST-SSR markers hold immense potential to add a powerful new dimension to the understanding and improvement of crop gene pools. Consequently, EST-SSR markers have been widely applied for genetic diversity analysis like in wheat (Gupta et al. 2003; Wang et al. 2007), barley (Kota et al. 2001; Thiel et al. 2003), *Crotelaria* (Wang et al. 2006), coffee (Aggarwal et al. 2007) etc. and for cross-transferability studies in several plant families such as cereals (Gupta et al. 2003; Varshney et al. 2005b; Saha et al. 2004; Thiel et al. 2003), legumes (Eujayl et al. 2004; Gutierrez et al. 2005), *Vitaceae* (Decroocq et al. 2003) etc.

So far in chickpea, the available SSR markers were mostly derived from anonymous DNA fragments and no reports of EST-SSRs was available. However, now with the availability of the EST-SSRs developed as a part of this study (Chapter 5) and more so because of they better serve molecular breeding applications, efforts need to be made to utilize them for various applications such as germplasm characterization, varietal identification, phylogenetic analysis and cross-transferability studies for comparative mapping in legumes. Assessing genetic diversity with EST-SSRs might enhance the role of genetic markers by assaying the variation in transcribed and known-function genes and may prove more useful for marker-assisted selection if found to be associated with gene/QTL of agronomic interest (Varshney et al. 2005a; Buhariwalla et al. 2005). Thus, the investigations/results reported here were designed to evaluate the potential of the developed chickpea EST-SSR markers for genetic diversity analysis within *C. arietinum* and cross-transferability studies across genus *Cicer* and seven legume genera namely *Medicago*, Soybean, Pea, *Trifolium*, Lentil, Pigeonpea and Blackgram. Further, molecular analysis of size variant alleles was also carried out.

7.2 Results

7.2.1 Analysis of intra-specific genetic diversity in chickpea using genic markers

To elucidate intra-specific variability, a set of sixty chickpea EST-SSR primers were used to amplify the genomic DNA from 30 chickpea accessions (Table 3.1 of Material and Methods) using the reactions and amplification profile mentioned in section 3.2.13. Representative patterns of amplification obtained using two of the chickpea EST-SSR markers are shown in Fig. 7.1. Of these 60 markers, 10 produced polymorphic amplification profiles in the 30 accessions of *C. arietinum*. These 10 polymorphic markers amplified a total of 129 alleles with a maximum of 5 alleles with the primer pair CESSR73 in the chickpea cultivars (Table 7.1). Variable size alleles ranging from 173-650 bp including multiple alleles at some loci were also observed (Table 7.1). The observed heterzygosity values ranged from 0.0 to 0.6 (CESSR47 and CESSR73) and expected heterzygosity ranged from 0.43 (CESSR43) to 0.76 (CESSR73) with an average of 0.56 (Table 7.1).

To assess the molecular basis of length variation across chickpea, homologous alleles amplified by primer pair CESSSR73 were cloned and sequenced (Fig. 7.2). Sequence analysis revealed the presence of a $(CTT)_n$ compound repeat motif. Allelic size variations were mainly due to the presence of an additional repeat motif (CTT) in some of the cultivars (except in ICC15406 and ICCV2) accompanied by single base insertions/deletions in the MFR regions. Cultivar specific insertions were observed at position 301 in ICCV2 and deletions at 96, 102 in ICC15406 and 286 in ICC7676. In addition, isolated point mutations were also observed in the MFRs such as at position 195 in ICC7676 (Fig. 7.2). However, when similar sized alleles from monomorphic loci such as CESSRDB13, CESSRDB27 and CESSRDB44 were sequenced, point mutations were observed at some of these loci, but no variation in the copy number of repeats was found (data not shown).

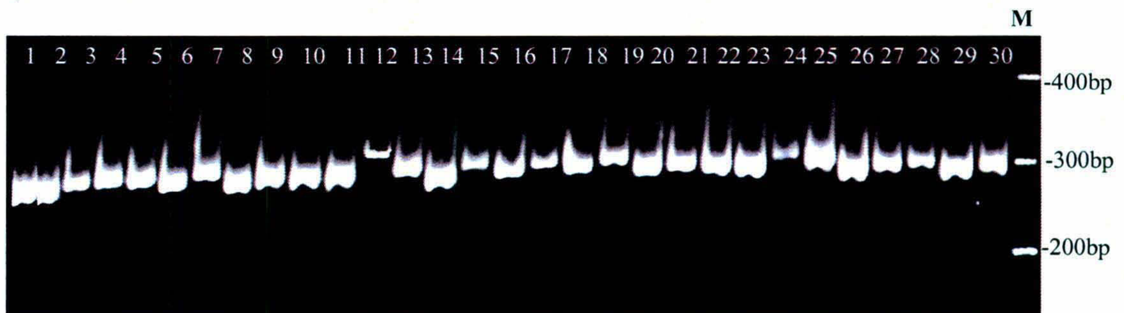
7.2.2 Cross-species transferability across *Cicer*

7.2.2.1 Amplification pattern

In order to assess the transferability rates of the chickpea EST-SSR markers across related *Cicer* species, genomic DNA from nine accessions belonging to five wild, annual *Cicer* species (listed in Materials and Methods; Table 3.2) were amplified with the same 60 EST-SSR primers used for intra-specific analysis. The transferability rates of chickpea EST-SSRs varied from a high of 96.6% in *C. reticulatum* to a low of 68.3% in *C. judaicum* with an average of 82.6% (Table 7.2). Forty-one markers (68.3%) amplified in all the annual species,

Figure 7.1: Amplification profiles of 30 chickpea cultivars with EST-SSR primers **A)** CESS42 and **B)** CESSR72 resolved on 6% PAGE gels. Lanes 1-30: chickpea accessions (listed in Table 3.1) and M: refers to 100bp ladder

A)



B)

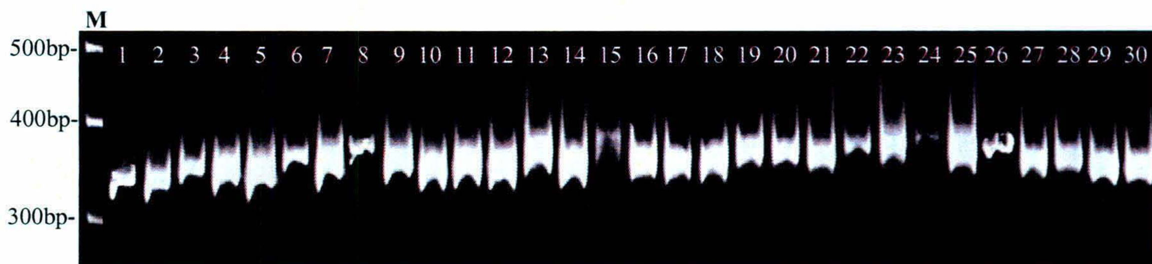


Figure 7.2: Partial sequence alignment of size variant alleles amplified using primer pair CESSR73 across six chickpea accessions (EX151922, EU332161-EU332163, EU332165, EU332166). The asterisks (*) represent similar sequences and dash indicates alignment gaps. Repeat regions are in boldface highlighted with pink color, primer-binding sites are represented by arrows and characters in bold blue shaded boxes indicate point mutations.

```
                          111133333334444444444555555555566666666667777777777
                123456789012344567890123456789012345678901234567890123456789

ICCV2           TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTT---GTTGTTATA
PUSA362         TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC15406        TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTT---GTTGTTATA
ICC283          TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC791          TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC7676         TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTCTTGTTGTTATA
                **********************************************   ***********

                           1111111222222222222222222233333333333333333333333
                888889999990000000966666666668888999999000111444555556666666
                012344567890123456501234567678901234512367834556789012345678

ICCV2           TATATGCGTTGAACTCAGCCCACGG-AGTCATTGGGCCGG-TGG-TAGGAACAGAAGATC  (363bp)
PUSA362         TATATGCGTTGAACTCAGCCCACGG-AGTCATTGG-C-GG-TGG-TAGGAACAGAAGATC  (363bp)
ICC15406        TATATGC-TTGAA-TCAGCCCACGG-AGTCATTGG-C-GGGTGG-TAGGAACAGAAGATC  (359bp)
ICC283          TATATGCGTTGAACTCAGCCCACGG-AGTCATTGG-C-GG-TGG-TAGGAACAGAAGATC  (363bp)
ICC791          TATATGCGTTGAAGTGTGCCCTCCTTAGTCATTGG-C-GG-TGGGTAGGAACAGAAGATC  (365bp)
ICC7676         TATATGCGTTGAACTCAGACCACGGTA-TCATTGGGC-GGGTGGGTAGGAACAGAAGATC  (367bp)
                ******* ***** *  * ** *   * ******* * ** *** ***************
```

of which 27 were polymorphic across the wild *Cicer* species. The representative pattern of amplifications obtained with chickpea EST-SSR primers are shown in Fig. 7.3. Allelic data generated using the 60 EST-SSRs revealed a minimum of 1 and maximum of 9 alleles (CESSRDB47) with a total of 156 alleles at 60 loci leading to an average of 2.6 alleles per locus. Observed heterozygosity (H_o) ranged from 0.15-0.83 with an average of 0.22 (Table 7.2). Of the 60 primer pairs, two (CESSRDB3 and CESSRDB5) amplified only in chickpea (and in no other wild species) and were therefore specific to *C. arietinum*. The EST-SSR markers also exhibited crossability group-specific transferability. Among the first crossability group members (*C. arietinum, C. reticulatum and C. echinospermum*), 55 markers (91.6%) were amplified and 24 markers produced polymorphic amplification profiles. Similarly, 41 markers (58.8%) successfully amplified in all the second crossability group members (*C. bijugum, C. judaicum and C. pinnatifidum*) and 23 primers detected variation between the three species. Five markers (CESSRDB7, CESSRDB23, CESSRDB24, CESSRDB41 and CESSR26) exhibited crossability-group-specific transferability to only first crossability group members.

7.2.2.2 Sequence analysis of size variant alleles from *Cicer* species

To investigate the basis of variation among size variant alleles of six *Cicer* species, fragments amplified in various annual species at the five loci, CESSRDB4, CESSRDB10, CESSRDB26, CESSRDB27 and CESSRDB34, were cloned and sequenced. Multiple alignments of nucleotide sequences from each locus were done and are shown in Fig. 7.4. Sequence comparisons at loci revealed that even though there was overall sequence conservation in the internal microsatellite structure and the primer-binding sites, variations such as differences in the copy number of repeat motifs and repeat interruptions accompanied by indels and point mutations in the microsatellite flanking regions (MFR) frequently occurred, all of which contributed to the allelic length variation (Fig. 7.4 and Table 7.2). For example, at the CESSRDB4 locus, an addition of three repeat motifs (48 to 56 bp) accompanied by mutations at positions 26, 44 and 47 resulted in motifs specific for the first and second crossability group members (Fig. 7.4A). Similarly at the CESSRDB10 locus, a mutation (T→G) at nucleotide position 50 and a TAG repeat expansion among members of first crossability group were observed (Fig. 7.4B). At the CESSRDB26 locus, the second crossability group alleles were much longer due to insertions in the MFR regions (Fig. 7.4C). Changes in the copy number of repeat motifs were also observed at the CESSRDB27 locus (Fig. 7.4D) while for CESSRDB34 (Fig. 7.4E) no change in the number of repeat motifs was

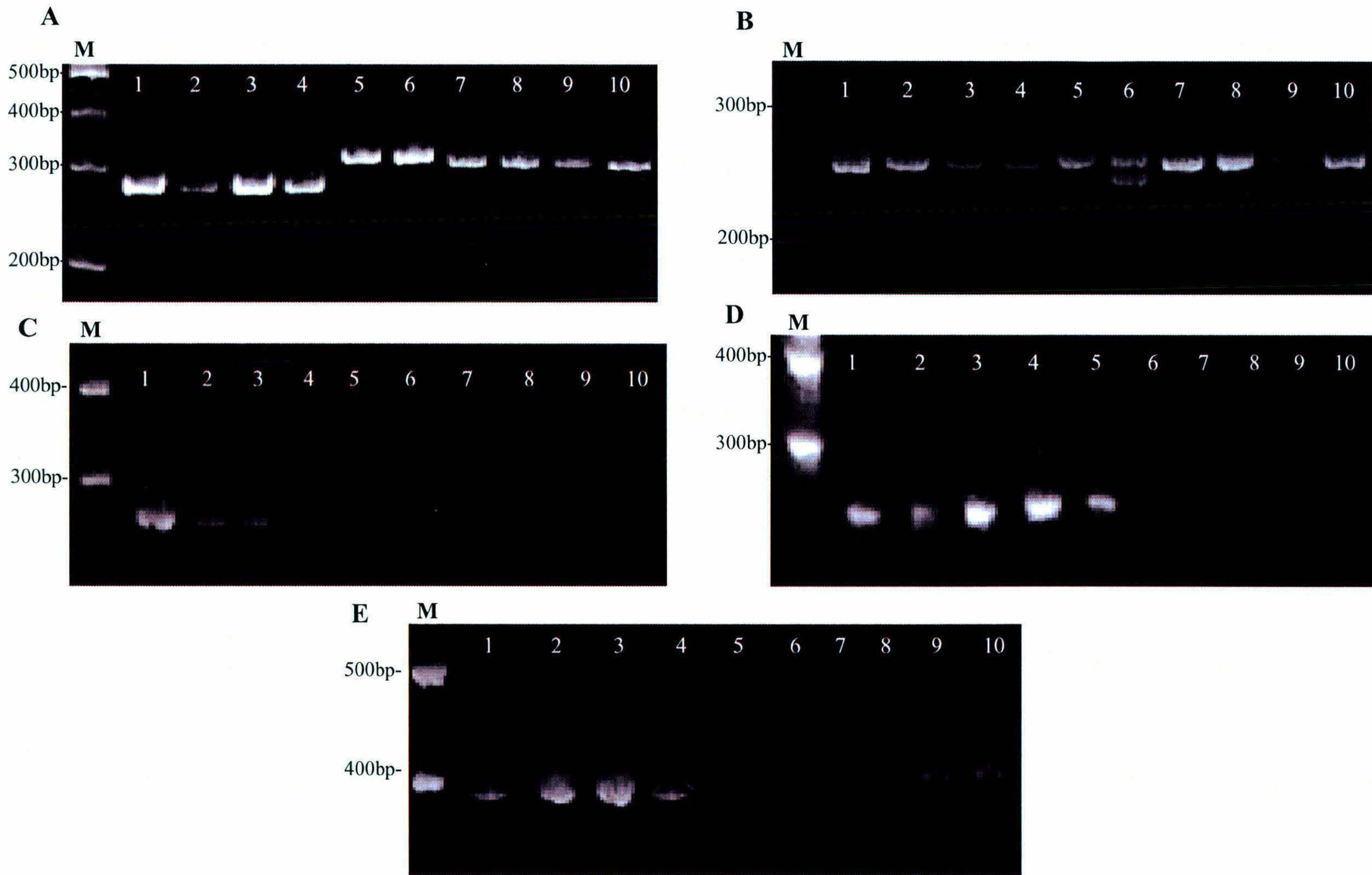


Figure 7.3: Amplification profiles of six annual *Cicer* species with EST-SSR markers **A)** CESSRDB26 **B)** CESSRDB34 **C)** CESSRDB44 **D)** CESSRDB61 and **E)** CESSR20 on a 6% PAGE gel. M: refers to 100bp ladder, 1 :*C. arietinum*, 2-3: *C. reticulatum*, 4: *C. echinospermum,* 5-6: *C. bijugum,* 7-8: *C. judaicum*, 9-10: *C. pinnatifidum*

Figure 7.4: Partial sequence alignment of size variant alleles amplified across six annual *Cicer* species using **(A)** CESSRDB4 primer (EF595573-EF595577) **(B)** CESSRDB10 primer (EF595578- EF595582), **(C)** CESSRDB26 primer (EF595583-EF595587), **(D)** CESSRDB27 primer (EF595588-EF595592) and **(E)** CESSRDB34 primer (EF595593- EF595598). The repeat motifs are represented as bold letters, arrows represent primer-binding sites, ' - ' indicates alignment gaps and ' * ' represents similar sequences. Characters in bold blue shaded boxes indicate base substitutions and group specific mutations are demarcated by green background. C.ari (*C. arietinum)*, C.ret (*C. reticulatum)*, C.ech (*C. echinospermum)*, C.bij (*C. bijugum)*, C.jud (*C. judaicum*), C.pin (*C. pinnatifidum*).

(A) CESSRDB4

```
                1111111111222222222233333333334444444444555555555556
       123456789012345678901234567890123456789012345678901234567890

C.ari  GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.ret  GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.ech  GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.bij  GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
C.jud  GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
C.pin  GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
       ************************* ***************** **            *

        11111111111111111111111111111111111111111222222222222222222
       990000000000111111111128888888889999999999000066666666667777
       890123456789012345678901234567890123456789012312345678901234

C.ari  GGAGAGGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.ret  GGAGAAGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.ech  GGAGAAGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.bij  GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (262bp)
C.jud  GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGCTGGCCAAAACTGTTGCTTG   (262bp)
C.pin  GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (262bp)
       ***** ******************** ************* *******************
```

(B) CESSRDB10

```
                123333344444555555555566666666667777777777888888888
       123456789034567856789012345678901234567890123456789012345678

C.ari  CCCTTAATCATGTT-ATCTAAGAATAATAATAATAATAATAATAGTAGTAGTACCATATA
C.ret  CCCTTAATCATGTT-GTCTAAGAATAATAATAATAATAATAGTAGTAGTAGTACCATATA
C.ech  CCCTTAATCATGTT-GTCTAAGAATAATAATAATAATAATAGTAGTAGTAGTACCATATA
C.bij  CCCTTAATCAACTT-GCCTAATAATAATAATAA---------------TAGTACCATATA
C.jud  CCCTTAATCAACTT-GCCTAATAATAATAATAATAA------------TAGTACCATATA
C.pin  CCCTTAATCATCTTTGCCTAATAATAATAATAATAATAA---------------------
       *********** **   **** ***********
             111111111111111111111111111111111111111111111111111111
       899999000011111111112222222222333333333344444446888888888999
       902345678901234567890123456789012345678901234564678901234567

C.ari  ATCAAC----TACTGTTCCCTGTTGGAGGTTTTGTCCTCCTTTTTTTGTGGTTTGGATAA   (192bp)
C.ret  ATCAAC----TACTGTTCCCTGTTGGAG-TTTTGTCCTC-TTTTTTTGTGGTTTGGATAA   (190bp)
C.ech  ATCAAC----TACTGTTCCCTGTTGGAG-TTTTGTCCTC-TTTTTTTGTGGTTTGGATAA   (190bp)
C.bij  ATAAGCAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (179bp)
C.jud  ATCAGCAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (182bp)
C.pin  -TAAACAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (173bp)
          * *    ***** ***** ****** ********** *** *** ************
```

(C) CESSRDB26

```
                                          111111111111111111111111111111111111111
                           111699999999000000000012222233333333334444444445666
       123456789012923456789012345678905678901234567890123456789012

C.ari  GGTGCATTCTCTAAAA--------------GCGTACTATAC-------TGTTAATACACG
C.ret  GGTGCATTCTCTTAAA--------------GCGTACTATAT-------TGTTAATACACG
C.ech  GGTGCATTCTCTTAAA--------------GCGTACTATAT-------TGTTAATACACG
C.bij  GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTTTCAAGCAAATGCATGTTCATACA-G
C.jud  GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTACTATAT-------TGTTCATACA-G
C.pin  GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTACTATAT-------TGTTCATA---G
       ************  **              **** * *          **** ***   *

       122222222222222222222222222222222222222222222222222227222222
       900000000112222222222333334444444444555555555566666779999999
       303456789090123456789012340123456789012345678906783453456789

C.ari  TGGTGGTGTAATACGGAAAGAAAGTTAGGTGTGTGTGTTGA-----ACGACATATTTGCA    (273bp)
C.ret  TGGTGGTGTAATACGGAAAGAAAGTTAGGTGTTTGTGTTGA-----ACGATATATTTGCA    (273bp)
C.ech  TGGTGGTGTAATACGGAAA--AAGTTGGGTGTTTGTGTTGA-----ACGACATATTTGCA    (271bp)
C.bij  GGGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTAGAATTATAGATTTGCA    (298bp)
C.jud  GAGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTCGAATTATAGATTTGCA    (291bp)
C.pin  GGGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTCGAATTATAGATTTGCA    (289bp)
         ** ****  ***  ***  ****  ***** ***** **     *  * * *******
```

(D) CESSRDB27

```
                                              111111111111111111111
                    44444444666666669999999999000000000011111111112
       123456789012301234567123456780123456789012345678901234567890

C.ari  GGTGAGATTAGGTACTTTGATATGGAGTTTATGTAAGATATTATTATTATTATTATTAT
C.ret  GGTGAGATTAGGTACTTTGATATGGAGTTTATGTATGATATTATTATTATTATTATTAT
C.ech  GGTGAGATTAGGTACTTTGATATGGAGTTTATGTATGATATTATTATTATTATTATTAT
C.bij  GGTGAGATTAGGTACTTTGATATGGAGTTTATGTATGATATTATTATTATTAT------
C.jud  GGTGAGATTAGGTACTTTGATATGAA-TTTATGTAACATATTATTATTAT---------
C.pin  GGTGAGATTAGGTACTTTGATATGAA-TTTATGTAACATATTATTATTAT---------
       ************************ * ********* *************

       111111111111111111111111111111111111111111111111111111222222
       222222222333333333344444444445555555555688889999999999011111
       123456789012345678901234567890123456789067890123456789012345

C.ari  GCAAAAATAAAAGCTTTGGTCTTCATGGGTTGGTTTGTAACATCTTATGGGGATTGGATA    (215bp)
C.ret  GCAAAAATAAAAGCTTTTGTCTTCATGGGTTGGTTTGTAACATCTTATGGGGATTGGATA    (215bp)
C.ech  GCAAAAATAAAAGCTTTTGTCTTCATGGGTTGGTTTGTAACATCTTATGGGGATTGGATA    (215bp)
C.bij  GCAAAAATAAAAGCTTTTGTCTTCATGGGTTGGTTTGTAACATCTTATGGGGATTGGATA    (209bp)
C.jud  GCACAAATAAAAGCTCTTGTTTACATGGGTTGGTTTGCAGCATCTTATGGGGATTGGATA    (205bp)
C.pin  GCAAAAATAAAAGCTTTTGTTTACATGGGTTGGTTTGTAGCACCTTATGGGGATTGGATA    (205bp)
       *** *********** * ******************* **** *****************
```

(E) CESSRDB34

```
               111111111122222222223333333333444444444455555555556
      123456789012345678901234567890123456789012345678901234567890

C.arie      AACCTAAAGCCGAAAAGAAGATCTCGAAGGAAGGAACAAGCGATAAGAAGAAGAAGCGCG
C.reti      AACCTAAAGCCGAAAAGAAGATCTCGAAGGAAGGTACAAGCGATAAGAAGAAGAAGCGCG
C.echi      AACCTAAAGCCGAAAAGAAGATCTCGAAGGAAGGAACAAGCGATAAGAAGAAGAAGCGCG
C.biju_A    AACCTAAAGCCGAAAAGAAGATCTCGAAGGATGGAGCAAGCGATAAGAAGAAGAAGCGCG
C.biju_B    AACCTAAAGCCGAAAAGAAGATATCCAAGGA----GTAAGCGATAAGAAGAAGAAGCGCG
C.juda      AACCTAAAGCCGAAAAGAAGATCTCGAAGGAAGGAGCAAGCGATAAGAAGAAGAAGCGCG
C.pinn      AACCTAAAGCCGAAAAGAAGATCTCGAAGGAAGGAGCAAGCGATAAGAAGAAGAAGCGCG
            ********************* ** *****     ***********************

                  1111111111111111111111111111112222222222222
      666666666677777766666777777777788888888889999993344444444455
      123456789012345667890123456789012345678901234589012345678901

C.arie      TTAAGAAGAGCGTTGATTATTAACGATATTTTTGAGAAACTCGCTCTTACTTCACGGGAG    (251bp)
C.reti      TTAAGAAGAGCGTTGATTATTAACGATATTTTTGAGAAACTCGCTCTTACTTCACGGGAG    (251bp)
C.echi      TTAAGAAGAGCGTTGATTATTAACGATATTTTTGAGAAACTCGCTCTTACTTCACGGGAG    (251bp)
C.biju_A    TTAAGAAGAGCGTTGATCATTAACGATATTTTTGAGAAACTCGCTCTTACTTCACGGGAG    (251bp)
C.biju_B    TT------------GATCATCAACGATATTTTTGAAAAACTCGCCCTTACTTCACGGGAG    (235bp)
C.juda      TTAAGAAGAGCGTTGATCATTAACGATATTTTTGAGAAACTCGCTCTTACTTCACGGGAG    (251bp)
C.pinn      TTAAGAAGAGCGTTGATCATTAACGATATTTTTGAGAAACTAGCTCTTACTTCACGGGAG    (251bp)
```

observed but two deletions were responsible for the occurrence of the smaller allele (B) in *C. bijugum*.

Another interesting feature revealed by sequence comparisons was the crossability-group-specific point mutations and indels. Point mutations (nucleotide positions 26, 44, 47 and 184 in CESSRDB4; positions 34, 45, 50, 115, 121, 143 and 164 in CESSRDB10; positions 92, 145, 193, 205, 210, 219, 223, 224, 234, 252, 266, 267 and 275 in CESSRDB26) and indels [12bp (48 – 59) in CESSRDB4], [9bp (68-76) and 4bp(106-109) in CESSRDB10] and [14bp (95-108), 1bp (161) and (5bp (255-259) in CESSRDB26] were highly crossability group specific. Additionally in the flanking regions, there were species-specific point mutations for example at position 103 in *C. arietinum* and 198 in *C. judaicum* at CESSRDB4 locus (Fig. 7.4A).

7.2.2.3 Phylogenetic analysis

The allelic data obtained from the 60 chickpea EST-SSR markers was used to visualize the genetic relationships among the 30 chickpea accessions and the six annual *Cicer* species. After scoring and computing the allelic data, a dendrogram was constructed that clearly separated the members of the first and second crossability groups into clusters I and II (Fig. 7.5). Cluster I corresponded with the first crossability group members grouping all chickpea accessions into Cluster IA and the *C. reticulatum* and *C. echinospermum* into Cluster IB. The ClusterIA clearly distinguished all the chickpea accessions except ICC15518 and ICC8195, however no correlation between the clustering pattern and geographical location was obtained. Cluster II represented the second crossability group species with *C. judaicum* and *C. pinnatifidum* being closely placed together.

7.2.3 Cross-genera transferability and sequence variation of chickpea EST-SSRs across legumes

Thirty-four of the chickpea functional markers were also utilized to assess their cross-genera transferability across 32 accessions spanning eight legume genera (Fig. 7.6 and Table 3.3 of Material and Methods). This analysis revealed varied levels of marker transferability across legumes ranging from 29.4% in *P. mungo,* 35.2% in *P. sativum,* 41.1% in *G. max* and *T. alexandrinum,* 47.0% in *L. esculenta,* 50.0% in *C. cajan and* 61.7% in *M. truncatula* with an average of 43.6% (Table 7.3). The representative pattern of amplification obtained across legume genera with chickpea EST-SSR primers is shown in Fig. 7.7. Eight markers (23.5%) amplified in all the legume genera though all the accessions

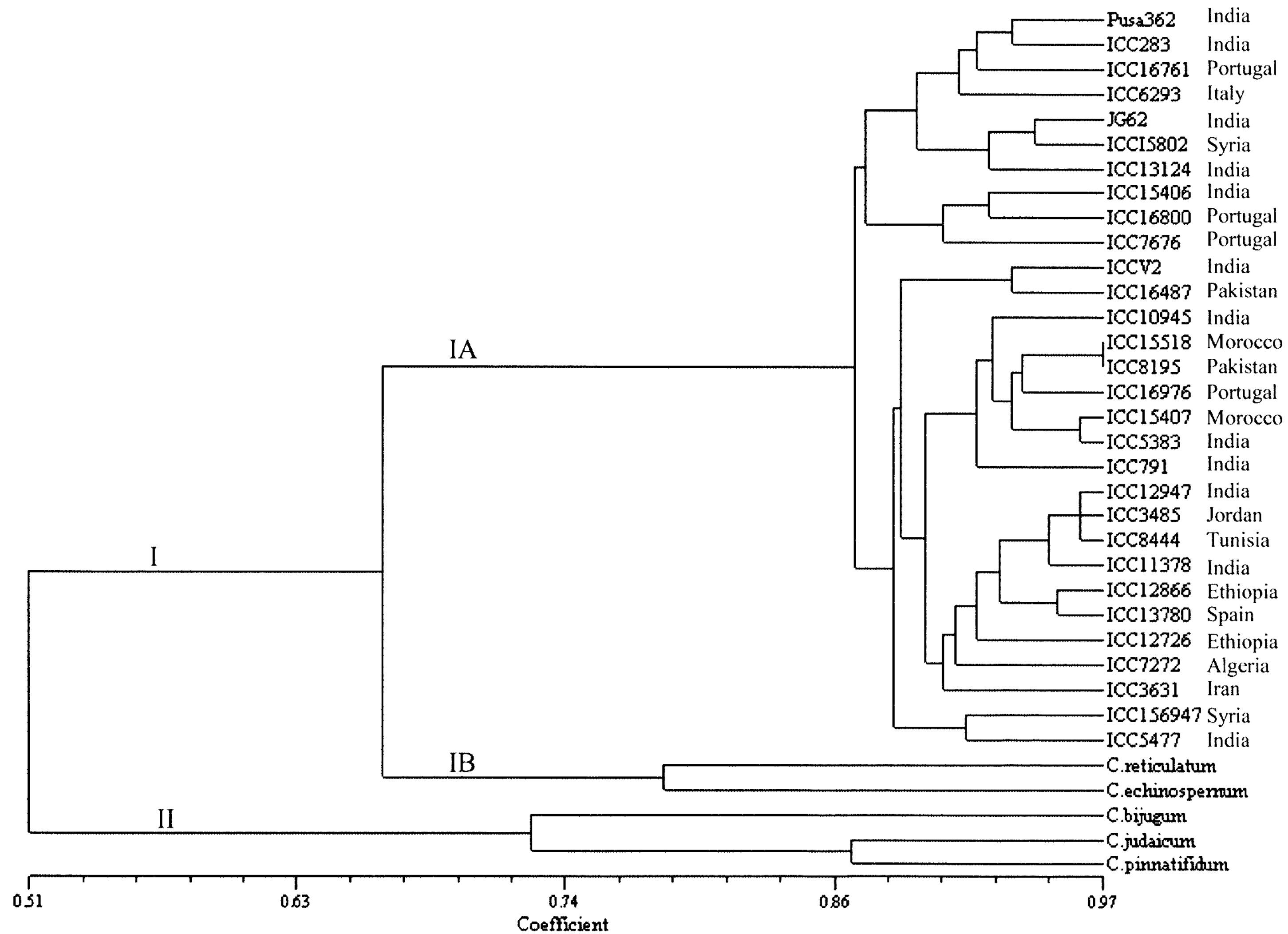


Figure 7.5: UPGMA based dendrogram of thirty chickpea accessions and five annual *Cicer* species was obtained using 60 functional EST-SSR markers and Jaccard's coefficient. Name of cultivars, species and source country are mentioned.

Figure 7.6: The legume species used in this study along with their seeds. **(A)** Chickpea (*C. arietinum*), **(B)** Barrel medic (*Medicago truncatula*), **(C)** Soybean (*Glycine max*), **(D)** Pea (*Pisum sativum*), **(E)** Lentil (*Lens esculenta*), **(F)** Berseem (*Trifolium alexandrinum*), **(G)** Pigeonpea (*Cajanus cajan*), **(H)** Blackgram (*Phaseolus mungo*).

Figure 7.7: Amplification profiles of chickpea EST-SSR primers across seven legume genera A) CESSRDB13 B) CESSRDB34 C) CESSRDB56 and D) CESSRDB40. M: 100 and 50bp ladder on the left and right hand side respectively, 1-4: *C. arietinum* 5-8: *M. truncatula*, 9-12: *G. max*, 13-16: *P. sativum*, 17-20: *L esculenta*, 21-24: *T. alexandrium*, 25-28: *C. cajan* and 29-32: *P. mungo*

A) CESSRDB13

B) CESSRDB34

C) CESSRDB56

D) CESSRDB40

of each species were not amplified. Twenty-five markers (74%) amplified in atleast one legume species other than chickpea whereas nine primers (26%) amplified only in the chickpea accessions and no other legume indicating the uniqueness of these loci to the chickpea genome. Twelve markers produced polymorphic amplification profiles across legumes even though intra-specific polymorphism was not observed (Table 7.3). Even though most markers produced alleles of sizes comparable with the expected sizes, higher sized fragments ranging in size from 0.75 to 1.0 Kb were amplified with some of the markers.

Alleles amplified at various loci of the studied legumes were cloned and sequenced to identify orthologous markers. Among the four markers that produced single monomorphic alleles across all the legumes, loci such as CESSRDB10 (Table 7.3) were sequenced and illustrated that the same sized alleles did not show any sequence variation (data not shown). However, different sized alleles amplified at various loci such as at CESSRDB56 (Fig. 7.8A) demonstrated that although the same sized alleles had high sequence conservation, variable alleles such as the 228bp allele in *Trifolium* revealed polymorphism that was due to both differences of repeat motifs as well as variability of the flanking sequences marked by indels/point mutations. Similar observation was also noted at locus CESSRDB39 amplifying multiple alleles such as the 234bp allele in *L. esculenta* (Fig. 7.8B). However, 309 bp allele in *P. mungo* at the CESSRDB39 locus (Fig. 3B) demonstrated that although conservation was retained in the primer binding sites, the internal sequences were extremely variable, often marked by the absence of the microsatellite motifs.

7.3 Discussion

Chickpea has been shown to exhibit overall low levels of polymorphism with the various molecular markers analysed so far and this has been attributed to its self-pollinating nature as well its recent worldwide dispersal (Udupa et al. 1999). In this study also, the EST-SSR markers displayed a low level of polymorphism (16%) within chickpea accessions in comparison to earlier reports of 40-60% polymorphism detected by genomic SSRs (gSSRs) (Sethy et al. 2006a; Lichtenzveig et al. 2005). This observation is noteworthy as SSRs located in the coding regions are under strong selection pressure and therefore accumulate few mutations (Varshney et al 2005a; Li et al 2004). However, despite the lower polymorphism, the genic-SSRs (EST-SSRs) are preferable over gSSRs as these are associated with the coding regions of the genome and therefore represent "true genetic diversity" that would directly assist in "perfect" marker-trait associations (Eujayl et al. 2002; Thiel et al.

Figure 7.8: Multiple sequence alignment of **(A)** alleles amplified from chickpea and two legumes namely *M. truncatula* (M.tru) and *T. alexandrinum* (T.ale) at CESSRDB56 locus (EF621420 and EF595632) and **(B)** different sized alleles amplified from chickpea and two legumes, *L. esculentum* (L.esc) and *P. mungo* (P.mun) using primer pair CESSRDB39 (EF621418 and EF621419). These primers also amplified alleles which were of the same size as chickpea and also shared high sequence homology, hence are not shown. Only the size variants are shown. The asterisks (*) represent similar sequences, dash indicates alignment gaps. Repeat regions are in boldface, primer-binding sites are represented by underlined letters and characters in bold blue shaded boxes indicate point mutations. 'B' represents the size variant alleles within the same accessions.

(A) CESSRDB56

```
C.ari           TGTCTGGAACAACAAGTGAGTATAATTTGAATGGACAA---TCTGAGTGTTCTGAACAGA
M.tru           TGTCTGGAACAACAAGTGAGTATAATTTGAATGGACAA---TCTGAGTGTTCTGAACAGA
T.ale           TGTCTGGAACAACAAGTGAGTATAATTTGAATGGACAACAATCTGAGTGTTCTGAACAGA
                **************************************   *******************

C.ari           CAACATCAACAGAAGGACCTGTTTTGGAAGAGTTTATTCCAATTAAGAAAAGGGCTTCAC
M.tru           CAACATCAACAGAAGGACCTGTTTTGGAAGAGTTTATTCCAATTAAGAAAAGGGCTTCAC
T.ale           CAACATCAACAGATGGACCTGTTTTGGAAGAGTTTATTCCAATTAAGAAAAGGGCTTCAC
                ************* **********************************************

C.ari           CTTATTGTGAACAAGTGTATGATGATGATGA----AAAGAAGATGATGAACAGC-TTCTC
M.tru           CTTATTGTGAACAAGTGTATGATGATGATGA----AAAGAAGATGATGAACAGC-TTCTC
T.ale           CTTATTGTGAACAAGTTTATGATGATGAGGATGGTGATGAAGATGATGAACAACATTCTC
                **************** *********** **     * ************** * *****

C.ari           ATCATAAGCAGCAAAAGATTTCATCAAATGATAATAATAAGAACAGTGATAAGAGGAAAT
M.tru           ATCATAAGCAGCAAAAGATTTCATCAAATGATAATAATAAGAACAGTGATAAGAGGAAAT
T.ale           ATCATAAACAACAAAAGATTTCA----ATG-----------------GATAAGAGGAAAT
                ******* ** ************    ***                 *************

C.ari           CTGATTGGC (247bp)
M.tru           CTGATTGGC (247bp)
T.ale           CTGATTGGC (228bp)
                *********
```

(B) CESSRDB 39

```
C.ari           CTGAGGTTAATGTGAAAGGCGGCGGAGAGATTGCGCTGGCGAAGCGGTGGCGTTGGGG--
L.esc_B         CTGAGGTTAATGTGAAAGGCGGAGGAGAGACTGCGGTGGCGATGCGGTTGCGTCGGGGGT
P.mun_B         CTGAGGTTAATGTGAAAGGCAGAGTA-ATTTCACATGGGCTTTGGAAAAGCTTAAAAGTT
                ******************** * * * *     *   ***   *     ** *    *

C.ari           ----------ATCATGTGGGATTTGGTACGGTTGGAGGATTGGAGGGAAAACGACGTCGT
L.esc_B         TGTTATCGTTATCGCGTGGGGGATTTATAGGGTTGGAGGATGGAAGAGAAAACGGCGTCGT
P.mun_B         TATTTATAACATC-TGAGGGTGGTCCTAAAGTCATTGTCACTGATCGAGA-----CTTGG
                          ***  * ***   *  **  **    *      *  ** *     * *

C.ari           TTTTGGGGTGAATTAGT-------TTTTGATGCGGTGGTGGTGGTGGAAATGTGGGTCCC
L.esc_B         TTTTGGGAAACGGTGGT-------TT---------TGGTGGTGGTGGATTTGTG---CCA
P.mun_B         CTTTGATGAATGCCATTGCAATTGTATTCCCTGAGTCATATCAGATGTTATGTCGGTTCC
                 ****           *       *          *  *    *  *   ***     *

C.ari           CCACATGGCTGAGTTGTTTTCTTGTGCATAAGAGTCTCTTTGTGTGTCTAGTGTTTTCGA
L.esc_B         CCGCGGGGGTGAGTTACTTTGTTGTGCATAAGAGTGTT----------------------
P.mun_B         ACATCCACAAAAATGTTTTTGCTAAATGCAAAATGTTAGTTCGTTGTAAAGAGGCTTGGG
                 *         * *   ***  *      ** *   *

C.ari           GTGTGTGG-----------------TGTGAGG-----------------------TGTTG
L.esc_B         -TGCGTTG-----------------TGTGAGG-----------------------TGTTG
P.mun_B         ATGGGTTGATGTATGCATGGGAAAATGTGATGGATTGCACTGATGAGAGCTTGTTTGTTG
                 **  ** *                 ***** *                       *****

C.ari           AGCATGTGATGTTGAC            (257bp)
L.esc_B         AGCATGTGATGTTGAC            (234bp)
P.mun_B         AGCATGTGATGTTGAC            (309bp)
                ****************
```

2003). Other species such as rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001) and wheat (Gupta et al. 2003), have also revealed similarly low levels of polymorphism using EST-SSRs compared to genomic SSR markers.

In contrast to the low levels of intra-specific polymorphism with EST-SSRs, the inter-specific polymorphism was significantly higher in the wild relatives of chickpea. Wild genepool is extremely valuable in inter-specific hybridization programmes since they serve as sources of resistance/tolerance to many stresses. Our study with EST-SSR markers will potentially facilitate the transfer of traits of agronomic value into cultivated chickpea thereby leading to the broadening of the narrow genetic base and development of superior genotypes of chickpea. The dendrogram obtained with the EST-SSR markers clearly showed the closeness of *C. judaicum* with *C. pinnatifidum* (Fig. 7.5) which was in agreement with the earlier protein based (Tayyar and Waines 1996) and EST-based studies (Buhariwalla et al 2005), whereas using DNA-based marker systems like AFLP (Shan et al. 2005) and STMS markers (Sethy et al. 2006b) the closeness of *C. pinnatifidum* with *C. bijugum* has been reported. The resemblances of the dendrograms based on protein markers with genic markers suggest that coding sequences of *C. judaicum* and *C. pinnatifidum* may have followed a common evolutionary pathway.

The chickpea EST-SSRs developed in this study revealed much higher rates of transferability (mean 82.6%) across wild annuals than the chickpea-derived gSSRs (68%; Choumane et al. 2000). Higher inter-specific transferability was in accordance with other studies (Scott et al. 2000; Eujayl et al. 2004; Zhang et al. 2005), establishing that functional markers were more transferable and therefore more useful than gSSR markers with the added potential of being used in allele-mining for identification of useful agronomic traits. It was also observed that the mean transferability rates across the primary and secondary crossability groups were an average of 96.0% and 74.9% respectively. This difference could be explained on the basis of an earlier study of Decrooq et al. 2003, which said that the level of sequence conservation of microsatellite loci is inversely proportional to the genetic distance. Similar observations have been made in other species such as wheat (Mc Lauchlan et al. 2001) and sugarcane (Cordeiro et al. 2001) where the genic markers displayed low level of polymorphism in cultivated accessions compared to other members of the genus thereby directing the breeders to look into the related species for introgression of novel genetic material into the germplasm.

Results from the sequencing data also provided evidence for limited sequence variability within the chickpea alleles in comparison to much higher levels of variation across the orthologous alleles from annual species. Sequence comparisons of size variant microsatellite alleles within chickpea accessions illustrated approx. 95% overall sequence conservation with few indels in the repeat as well as the MFR region (Fig. 7.2) suggesting the presence of evolutionary constraints within transcribed regions that limit the mutational events and increase sequence similarity. However, sequence comparison of microsatellite alleles from various *Cicer* species revealed a wide range of length and sequence variability both in terms of fragment size and allele number. Similar results have also been obtained by Buhariwalla et al. 2005 thereby establishing that EST based microsatellite markers of chickpea were not only efficient for marker-assisted introgression programs using wild germplasm but also reliable for synteny studies within the genus *Cicer*. Sequence variations occurred both at the repeat motifs and in the flanking regions and were interestingly found to be crossability-group-specific and therefore phylogenetically highly informative. These would be helpful in understanding the evolution of microsatellites in a phylogenetic context since it has been shown that such events at the genic loci might play an important role in speciation or gene functionality diversification during the process of evolution.

Earlier studies have provided evidences, which show that microsatellites undergo expansion during the course of evolution (Zhu et al. 2000; Peakall et al. 1998). In our study, there was an expansion of the GGT motif at the locus CESSRDB4 (Fig. 7.4A), resulting in the presence of 3 additional repeats in members of the first crossability group. At the locus CESSRDB10, expansion of the TAA motif was accompanied by the birth of a new TAG motif (via a A→G mutation) which later expanded in the members of the first crossability group. Such A→G transition was also observed by Messier et al. 1996 in owl monkey. It has been speculated that base substitution allows the birth of new motifs that subsequently expand by replication slippage (Gordon 1997). Recently, the role of microsatellite expansion/deletion in terms of gene regulation is being investigated well in mammals as well as in plants (Li et al. 2004). The presence of SNPs in the sequence of similar sized alleles from different chickpea cultivars apparently indicated the limitation of scoring the accessions simply based on the amplicon size on gel. Also, this clearly highlights the prospects of SNP mapping in chickpea as these represent the most fundamental source of variation for molecular marker development.

In our study it was observed that the microsatellite motifs were long but punctuated by imperfections which are most often regarded as an effective mechanism for prevention of infinite growth of microsatellites (Kruglyak et al. 1998; Palsboll et al.1999). At the CESSRDB4 and CESSRDB10 loci, base substitutions at positions 44 and 50 respectively in the first crossability group members implied that such interruptions may have a dramatic impact in the long-term evolution of the microsatellite sequence. On the other hand, the phenomenon of microsatellite purification (loss of interruptions), a mechanism counteracting the accumulation of imperfections is also known to occur (Harr et al. 2000). This was observed at locus CESSRDB26 (Fig. 7.4C) where all species, except *C. arietinum*, harbor "T" at position 246 indicating that T represents the ancestral character state. The occurrence of longer motifs in the focal species in comparison to the related species may also be explained by the hypothesis of ascertainment bias (Ellergen et al. 1997; Peakall et al. 1998; Vigouroux et al. 2002) which in our study was demonstrated by sequence comparisons at the CESSRDB4, CESSRDB10 and CESSRDB27 loci (Fig. 7.4A, 7.4B and 7.4D).

The cross-transferability of chickpea EST-SSR markers across legume species was high (mean 43.6%) clearly depicting the conservation of primer-binding sites in genomic DNA over a long evolutionary period. The usefulness of EST-SSR markers over genomic SSRs for transferability across distant relatives has been established in species such as *Medicago* (Gutierrez et al. 2005), wheat (Gupta et al. 2003), barley (Thiel et al. 2003) and grapes (Scott et al 2000). However in chickpea, no extensive study of cross-genera transferability of genomic SSRs was available, except a small study by Pandian et al. 2000, the transferability rates across distant species of genomic vs EST-SSRs could not be compared. Our study showed that the highest rate of transferability of the chickpea EST-SSR markers was to *Medicago* (61.7%), whereas an earlier study by Gutierrez et al. 2005 showed significant, yet lower levels of transferability of the *Medicago* markers to chickpea (36.3%). The difference in the rates of transferability could be attributed to the choice of loci and the overall number of markers analyzed. Our study also demonstrated that the rate of transferability decreases from within the genus *Cicer* (82.6%) to outside the genus (43.6%) which was in agreement with earlier reports in cereals (Thiel et al. 2003, Gupta et al. 2003), grapes and apricot (Decroocq et al. 2003) and *Medicago* (Gutierrez et al. 2005) suggesting that amplification decreases with increasing evolutionary distance from focal species. Overall, the chickpea markers transferred very efficiently to some members of the galegoid legumes (such as *Medicago* and *Trifolium*) as compared to the phaseoloid legumes (such as

P. mungo). However, *Pisum* and *Cajanus* were exceptions to this. The variable marker transferability rate obtained across different legume genera indicated the occurrence of genus-specific evolutionary events. Sequence analysis of size variant alleles revealed variations within the SSRs and in the flanking regions of variant alleles such as in *Trifolium* at CESSRDB56 (Fig. 7.8A) and in lentil at CESSRDB39 (Fig. 7.8B) which concurred with observations of Peakall et al. 1998; Choumane et al. 2004 and Gutierrez et al. 2005 in legumes and Decrooq et al. 2003 and Zhang et al. 2005 in other plants.

In conclusion, the developed chickpea EST-SSR markers were highly informative for detecting polymorphism within chickpea and were able to unequivocally distinguish all accessions even though no correlation between the clustering pattern and geographical location was obtained. Although depicting less polymorphism than gSSRs, EST-SSRs since derived from the transcribed regions could better reflect influences imposed by breeding efforts on the elite germplasm. Further the high rates of transferability of chickpea EST-SSR markers to wild *Cicer* species and among the studied legume genera illustrated that they are a valuable genetic resource for comparative mapping studies, mining of superior alleles and development of candidate gene markers for use in gene introgression programmes. Further, sequence analysis of size variant amplicons at different taxa level revealed that size polymorphism was due to multiple events such as copy number variation, point mutations and insertions/deletions in the microsatellite repeat as well as in the flanking regions. Interestingly, a wide prevalence of crossability-group-specific sequence variations were also observed among *Cicer* species that were phylogenetically informative. Thus, EST-SSR markers would be useful for cross-referring genes among related species for enhancing the resolution of. comparative genomic studies and identifying conserved genomic regions among species.

Table 7.1: Number of alleles (N_a), size range of amplified fragments, observed (Ho) and expected heterozygosity (H_E), Shannon's informative index (I) and Fixation index (F_{IS}) values calculated for 30 chickpea accessions at ten polymorphic EST-SSR loci.

Locus	N_a	Size range (bp)	H_O	H_E	I	F_{IS}
CESSR23	2	399-402	0.00	0.44	0.63	0.08
CESSR42	3	295-303	0.00	0.59	0.94	1.00
CESSR43	2	386-390	0.00	0.43	0.62	1.00
CESSR47	4	539-650	0.60	0.66	1.20	1.00
CESSR61	3	254-262	0.00	0.57	0.92	1.00
CESSR62	3	243-295	0.46	0.66	1.06	0.28
CESSR71	2	295-301	0.00	0.44	0.62	1.00
CESSR72	3	342-348	0.00	0.59	0.95	1.00
CESSR73	5	359-445	0.60	0.76	1.47	0.19
CESSR77	2	173-176	0.00	0.50	0.68	1.00
Average	2.9	-	0.16	0.56	0.91	-
St. Dev.	0.99	-	0.2711	0.1110	0.28	-

Table 7.2: Number of alleles and their sizes (bp) obtained in six annual species of *Cicer* at 60 genic microsatellite loci are mentioned. Observed (H_o) and expected (H_e) heterozygosity values were estimated. Hyphen represents no amplification.

Locus	*C. arietinum*	*C. reticulatum*	*C. echinospermum*	*C. bijugum*	*C. judaicum*	*C. pinnatifidum*			
	No. of Alleles (Size in bp)	No. of Alleles (Size in bp)	No. of Alleles (Size in bp)	No. of Alleles (Size in bp)	No. of Alleles (Size in bp)	No. of Alleles (Size in bp)	% transferability of each marker	H_o	H_e
CESSRDB2	2 (137, 169)	1(169)	1 (169)	3 (129, 169, 222)	1 (129)	1 (129)	100	0.3333	0.7424
CESSRDB3	1 (197)	-	-	-	-	-	0		
CESSRDB4	1 (274)	1 (274)	1 (274)	1 (262)	1 (262)	1 (262)	100	0.0000	0.5455
CESSRDB5	1 (177)	-	-	-	-	-	0		
CESSRDB7	1 (196)	1 (196)	1 (196)	-	-	-	40	0.0000	0.0000
CESSRDB10	1 (192)	1 (192)	1 (192)	1 (179)	1 (182)	1 (173)	100	0.0000	0.7273
CESSRDB11	1 (298)	1 (298)	1 (298)	2 (298, 507)	-	-	60	0.2500	0.2500
CESSRDB13	1 (260)	1 (260)	2 (260, 334)	2 (227, 334)	3 (277, 334, 343)	3 (277, 334, 343)	100	0.6667	0.7576
CESSRDB15	2 (261, 276)	2 (261, 276)	2 (261, 276)	1 (276)	1 (276)	1 (276)	100	0.5000	0.4091
CESSRDB16	2 (295, 346)	2 (323, 374)	2 (323, 374)	2 (323, 374)	-	1 (295)	80	0.8000	0.8000
CESSRDB18	1 (242)	3 (242, 279 315)	3 (252, 308, 315)	3 (252, 300, 324)	2 (252, 291)	2 (252, 291)	100	0.8333	0.8485
CESSRDB21	3 (259, 284, 296)	2 (259, 296)	3 (259, 284, 296)	1 (259)	2 (270, 296)	2 (270, 296)	100	0.8333	0.6818
CESSRDB23	1 (221)	1 (221)	1 (221)	-	-	-	40	0.0000	0.0000
CESSRDB24	1 (283)	1 (283)	1 (283)	-	-	-	40	0.0000	0.0000
CESSRDB26	1 (273)	1 (273)	1 (273)	1 (298)	1 (291)	1 (291)	100	0.0000	0.6667
CESSRDB27	1 (215)	1 (215)	1 (215)	1 (209)	1 (205)	1 (205)	100	0.0000	0.6667
CESSRDB29	1 (176)	1 (176)	1 (176)	1 (176)	1 (176)	1 (176)	100	0.0000	0.0000
CESSRDB33	1 (234)	1 (234)	1 (234)	1 (234)	1 (234)	1 (234)	100	0.0000	0.0000
CESSRDB34	1 (251)	1 (251)	1 (251)	2 (235, 251)	1 (251)	1 (251)	100	0.1667	0.1667
CESSRDB35	1 (272)	1 (272)	1 (272)	1 (272)	1 (272)	1 (272)	100	0.0000	0.0000
CESSRDB38	2 (197, 234)	2 (197, 234)	2 (197, 234)	2 (197, 234)	-	-	60	0.0000	0.0000
CESSRDB39	1 (257)	1 (257)	1 (257)	1 (257)	1 (257)	1 (257)	100	0.0000	0.0000
CESSRDB40	1 (188)	3 (188, 255 266)	2 (212, 272)	3 (199, 255 266)	-	3 (208, 266 272)	80	0.8000	0.9111
CESSRDB41	3 (247, 256, 278)	3 (251, 260, 281)	2 (251, 260)	-	-	-	40	1.0000	0.8667

CESSRDB42	2 (235, 280)	1 (235)	1 (235)	1 (235)	1 (235)	1 (235)	100	0.1667	0.1667
CESSRDB44	1 (267)	1 (267)	-	-	-	-	20	0.0000	0.0000
CESSRDB45	3 (295, 426, 449)	3 (295, 426, 449)	3 (295, 426, 449)	3 (295, 426, 449)	-	-	60	1.0000	0.5714
CESSRDB47	4 (240, 246 253. 276)	3 (240, 253 276)	4 (240, 246 253. 276)	4 (235, 248 258, 281)	3 (248, 268 281)	3 (248, 268 281)	100	1.0000	0.8182
CESSRDB51	1 (297)	1 (297)	1 (297)	1 (297)	1 (297)	1 (297)	100	0.0000	0.0000
CESSRDB53	1 (197)	1 (197)	1 (197)	1 (190)	1 (186)	1 (186)	100	0.0000	0.6667
CESSRDB54	2 (221, 305)	2 (221, 305)	2 (221, 305)	2 (221, 305)	-	-	60	1.0000	0.5714
CESSRDB55	2 (136, 154)	2 (136, 154)	2 (143, 154)	2 (140, 168)	2 (140, 168)	2 (140, 155)	100	0.1515	0.7778
CESSRDB56	1 (247)	1 (247)	1 (247)	1 (247)	1 (247)	1 (247)	100	1.0000	0.8485
CESSRDB61	1 (233)	1 (233)	1 (233)	1 (233)	-	-	60	0.0000	0.000
CESSR14	1 (300)	1 (300)	1 (300)	1 (300)	1 (300)	1 (300)	100	0.0000	0.0000
CESSR15	1 (300)	1 (300)	1 (300)	1 (300)	1 (292)	1 (292)	100	0.0000	0.4848
CESSR20	1 (386)	1 (386)	1 (386)	-	-	1 (390)	60	0.0000	0.5333
CESSR21	1 (233)	1 (233)	3 (233, 300, 324)	3 (233, 300, 324)	1 (245)	1 (245)	100	0.3333	0.6667
CESSR23	1 (399)	1 (391)	1 (391)	1 (395)	1 (391)	1 (395)	100	0.0000	0.6667
CESSR25	1 (363)	1 (363)	1 (363)	1 (363)	1 (363)	1 (363)	100	0.0000	0.0000
CESSR26	1 (275)	1 (275)	1 (275)	-	-	-	40	0.0000	0.0000
CESSR30	1(388)	1(388)	1(388)	1(388)	1(388)	1(388)	100	0.0000	0.0000
CESSR31	1 (113)	1 (113)	2 (113, 108)	1 (113)	1 (113)	1 (113)	100	0.1667	0.16607
CESSR34	1 (294)	1 (287)	2 (294, 280)	1 (294)	2 (294, 280)	1 (294)	100	0.3333	0.4545
CESSR42	1 (298)	1 (305)	1 (298)	1 (305)	1 (305)	1 (298)	100	0.0000	0.5455
CESSR43	1 (386)	1 (395)	1 (395)	1 (395)	1 (395)	1 (395)	100	0.0000	0.3030
CESSR47	1 (541)	1 (541)	1 (552)	1 (552)	-	-	60	0.0000	0.5714
CESSR51	1 (205)	1 (209)	1 (205)	1 (209)	1 (205)	1 (209)	100	0.0000	0.5455
CESSR61	1 (257)	1 (257)	1 (257)	1 (262)	1 (257)	1 (252)	100	0.0000	0.5455
CESSR62	1 (245)	1(249)	1 (249)	1 (254)	1 (254)	1 (260)	100	0.0000	0.7879
CESSR65	1 (352)	1 (352)	1 (370)	1 (366)	1 (366)	1(366)	100	0.0000	0.6667
CESSR68	3 (330, 342, 350)	3 (330, 342, 350)	3 (322, 342, 350))	2 (322, 342)	2 (322, 342)	1 (322)	100	0.8333	0.6898
CESSR71	1 (295)	1 (286)	-	-	-	-	20	0.0000	0.6667
CESSR72	1 (345)	1 (348)	1 (345)	1 (350)	1 (350)	1 (350)	100	0.0000	0.6667

CESSR 73	1 (363)	3 (363, 370, 375)	-	1 (370)	-	-	40	0.3333	0.7333
CESSR 77	1 (173)	1 (170)	1 (170)	1 (170)	1 (170)	1 (170)	100	0.0000	0.3030
CESSR 78	1 (373)	2 (373, 500)	2 (366, 500)	2 (366, 500)	2 (370, 500)	2 (370, 500)	100	0.8333	0.1970
CESSR 80	1 (260)	1 (260)	1 (260)	1 (260)	1 (260)	1 (260)	100	0.0000	0.0000
CESSR85	1 (279)	1 (279)	1 (274)	1 (274)	1 (279)	1 (274)	100	0.0000	0.5455
CESSR93	1 (353)	1 (353)	1 (353)	1 (353)	1 (353)	1 (353)	100	0.0000	0.0000
% of transferable markers	100	96.6%	91.6%	83.3%	68.3%	73.3%		**Avg-0.22**	**Avg-0.35**

Table 7.3: Cross-species amplification of chickpea EST-SSR primers across seven legume genera and *C. arietinum*. The sizes of amplified alleles (N_{bp}) and homologue match (score) and Expected value (E) are mentioned. NA means no amplification.

Locus	*C. arietinum* N_{bp}	*M. truncatula* N_{bp}(score, E)	*G. max* N_{bp}(score, E)	*P. sativum* N_{bp}(score, E)	*L. esculenta* N_{bp} (score, E)	*T. alexandrinum* N_{bp} (score, E)	*C. cajan* N_{bp}(score, E)	*P. mungo* N_{bp}(score, E)	% transferability of each marker
CESSRDB2	137,169	NA	NA	NA(46.1, 0.13)	NA	NA	NA	NA	0%
CESSRDB3	197	NA (42.1, 1.1)	NA	NA	NA	NA	NA	NA	0%
CESSRDB4	274	274	274	274	274	274	274	274	100%
CESSRDB5	177	NA (44.1, 0.14)	NA	177 (105, $4e^{-20}$)	NA	NA	NA	NA	14.3%
CESSRDB7	196	NA (44.1, 0.50)	NA	NA	NA	NA	NA	NA	0%
CESSRDB10	192	190 (38, 6.5)	190	190	190	190	190	190	100%
CESSRDB11	303	303 (46.1, 0.57)	NA	NA	404	303	303	NA	57.1%
CESSRDB13	260,	260 (293, $5e^{-70}$)	260	260	260, 452	260	260	260	100%
CESSRDB15	261, 276	276(289, $6e^{-75}$)	NA	NA	NA	NA	NA	NA	14.3%
CESSRDB16	295, 346	295 (40.1, 6.0)	295, 389	295	845	NA	295	NA	71.4%
CESSRDB18	242	242, (44.1, 0.45)	242	242	242, 389	242, 462	242	242	100%
CESSRDB21	259, 284, 296	NA	259	NA	180	NA	259	NA	42.8%
CESSRDB23	221	221, (222. e^{-54})	NA	NA	NA (85.7, $2e^{-13}$)	NA	NA	NA	14.3%
CESSRDB24	283	283	950 (398, 2e-107)	NA	NA	NA	NA	NA	28.5%
CESSRDB26	273	273 (42.1, 2.5)	273	NA	273	273	273	273	100%
CESSRDB27	215	NA	NA (842, 0.0)	NA	NA	NA	NA	NA	0%
CESSRDB29	176	NA, (88, $7e^{-15}$)	NA	NA	895	NA	NA	NA	14.3%
CESSRDB33	234	234, (1314, 0.0)	234 (311, $5e^{-81}$)	234 (315, $3e^{-82}$)	234	234	234, 915	234	100%
CESSRDB34	251	251, (289, $5e^{-75}$)	251	251	251	251	251	251	100%
CESSRDB35	272	272 (462, $1e^{-126}$)	NA	NA	272	272	272	NA	57.1%
CESSRDB38	197, 234	NA	NA (42.1,	NA (555, 4e-	NA	NA	NA	NA	0%

			1.7)	^{155})					
CESSRDB39	257	257 (40.1, 8.3)	257,	257	257, 234	257	257	257, 309	100%
CESSRDB40	188	147 (44.1, 0.24)	NA (87.7, $2e^{-14}$)	NA (83.8, $3e^{-13}$)	NA	NA	NA	NA	14.3%
CESSRDB41	247, 256, 278	247 (44.1, 0.74)	NA	NA	NA	NA	NA	NA	14.3%
CESSRDB42	235, 280	NA (242, $1e^{-60}$)	NA	NA (266, $9e^{-98}$)	NA	NA	NA	NA	0%
CESSRDB44	267	NA, (1298, 0.0)	NA	NA (48.1, 0.085)	NA	NA	NA	NA	0%
CESSRDB45	295, 426, 449	NA (48.1, 0.04)	NA (48.1, 0.042)	NA	NA	NA	NA	NA	0%
CESSRDB47	240, 246, 253, 276	240	NA	NA (119, $1e^{-23}$)	NA	NA	NA	NA	14.3%
CESSRDB51	297	297	297 (398, $1e^{-107}$)	297 (44.1, 0.82)	297	297	297	NA	85.7%
CESSRDB53	197	NA (42.1, 0.80)	NA	NA	NA	NA	NA	NA	0%
CESSRDB54	221, 305	NA (509, $3e^{-141}$)	NA	221 (424, $1e^{-115}$)	NA	NA	221	NA	28.5%
CESSRDB55	136, 154	136 (42.1, 1.5)	136	NA	NA	136	136	NA	57.1%
CESSRDB56	247	247 (224, $3e^{-55}$)	247	247	247	228	247	247	100%
CESSRDB61	233	233 (198, e^{-47})	NA	NA	233	233	233	233	71.4%
% trans ferability		61.7%	41.1%	35.2%	47.0%	41.1%	50.0%	29.4%	**Avg:** 44.1%

Chapter 8: Construction of an inter-specific linkage map of chickpea

8.1 Introduction

A genetic linkage map, defined as a linear arrangement of genes or genetic markers, is a necessary tool for genome analysis, marker-assisted breeding and map-based cloning. It is therefore most desired in chickpea which is a valuable and important agricultural crop, where yield potential is dramatically affected by several biotic and abiotic stresses. Therefore chickpea breeders throughout the world have been focusing on increasing the yield by pyramiding the desirable genes for resistance/tolerance into agronomically superior varieties through integration of advanced technologies like marker assisted breeding along with conventional approaches (Millan et al. 2006). The usage of molecular markers and linkage maps, prerequisites for MAS (marker-assisted selection), since their introduction, have been well chronicled in several major crop plants such as rice (McCouch et al. 2002), barley (Varshney et al. 2007), soybean (Song et al. 2004) etc. to establish saturated genetic linkage maps. In chickpea, the limited genetic information coupled with the limited genetic polymorphism and scarcity of co-dominant markers have hampered the genome mapping programs. However recently, with the availability of a suite of DNA markers, microsatellite markers have best succeeded in addressing the allelic diversity in chickpea (Sharma et al. 1995b; Hüttel et al. 1999; Sethy et al. 2003, 2006a) and therefore are the preferred marker system for chickpea molecular genetics and breeding. Furthermore, researchers have suggested that the STMS based codominant markers indeed serve as the most elite anchor markers for merging different genetic maps and to set up a high genome coverage consensus map in chickpea (Millan et al. 2006).

For generation of a linkage map, generally intra-specific crosses are preferred owing to low skewed segregation (Becker et al. 1995; Menendez et al. 1997) for direct application in breeding programs (Flandez-Galvez et al. 2003a). However, extremely low level of genetic variation in chickpea has prompted breeders to use inter-specific rather than intra-specific crosses in order to maximize the polymorphism for linkage analysis. Consequently, several skeletal linkage maps have been constructed in inter-specific crosses between chickpea and its wild progenitor *C. reticulatum*, both with morphological and isozyme markers (Gaur and Slinkard 1990; Kazan et al. 1993; Simon and Muehlbauer 1997) and DNA-based markers (Winter et al. 1999, 2000, 2003). Uptill now, the most extensive integrated linkage map of chickpea is available in this population, incorporating approx. 300 markers and covering 2483.3 cM of the chickpea genome (Pfaff and Kahl 2003). Internationally this population is accepted as a 'reference mapping population' by the chickpea research community for further

saturation mapping. Additionally, several intra-specific linkage maps (Cho et al. 2002, 2004; Flandez-Galvez et al. 2003a; Udupa and Baum 2003; Sharma et al. 2004; Lichtenzveig et al. 2006; Radhika et al. 2007; Taran et al. 2007) have also recently become available in chickpea. However, all these maps have been constructed using the only common set of STMS markers available in chickpea (Winter et al. 2000) and therefore resulted in locating the map positions of only these limited number of markers on all maps. Hence the chickpea genomic regions harboring traits of agronomic interest are not sufficiently saturated with codominant markers to implement MAS in plant breeding programs (Millan et al. 2006). Moreover, the STMS marker polymorphism rates observed in chickpea usually range from 30 to 50% (Winter et al. 1999; 2000), therefore several additional SSR markers are needed for mapping with markers originating from different sources in order to provide a better coverage of the whole genome and to construct a high-density genetic map of chickpea.

Recently, development of different types of PCR based functional markers has been possible such as EST-SSRs, ESTPs (Expressed Sequenced Tags Polymorphisms) and ITPs (Intron Targeted Primers). These markers which target genes having known function provide a valuable way of locating them on the genetic map. With the plethora of sequence information available in the databases,. This is a valuable resource for candidate gene approach wherein the opportunity for mapping the gene-rich regions of the genome is provided. Several studies have been conducted for linkage analysis in different crop species using EST-SSRs (Fraser et al. 2004; Yi et al. 2006; Varshney et al. 2007), ESTPs (Temesgen et al. 2001), and ITP markers (Choi et al. 2004a and b). Such gene based maps provide an important opportunity to directly tag genes related to agronomical traits, thus aiding in integrating the information between genes and QTLs and allowing gene cloning and MAS. For example, recently Park et al. 2005 and Guo et al. 2007 constructed the gene-rich linkage map of cotton and successfully tagged genes related to fiber quality. However in chickpea, only few gene-specific markers related to defense response are mapped till now (Pfaff and Kahl 2003; Hüttel et al. 2002). Moreover, marker density on linkage groups where resistance genes and other genes of agronomic importance were mapped, is still too sparse for use in map-based cloning. Therefore in chickpea, several codominant molecular markers especially derived from the expressed regions are sought to enhance the nature and utility of genetic maps.

Therefore in the present study, we set out primarily to exploit all the available chickpea molecular markers originating from the genomic and genic regions of the chickpea

genome for the construction of the most advanced genetic linkage map. For this purpose, all the novel chickpea EST-based molecular markers (EST-SSRs, ESTPs and PIPs) developed in this study (chapters 5 and 6) as well as the genomic-derived STMS markers (simultaneously developed in our laboratory but not yet mapped) were employed in order to localize and identify new positions on the chickpea linkage map.

8.2 Results

Genetic linkage map construction requires a) an appropriate mapping population of sufficient sample size and b) robust molecular markers. In order to develop a mapping population, two genetically divergent parents exhibiting sufficient polymorphism are required. In the present study, an RIL mapping population (comprising of 129 RILs) arising out of an inter-specific cross between *C. arietinum* ICC4958 (cultivated species) x *C. reticulatum* PI489777 (wild species) was utilized for map construction (Section 3.1.2 of Materials and Methods). Genomic DNA from the 129 mapping progeny as well as the mapping parents was isolated as described in section 3.2.1 (Fig. 8.1).

In order to construct detailed genetic maps with high levels of genome coverage, sufficient numbers of molecular markers are required. In the present study, the following different types of third generation codominant molecular markers including genomic-derived SSR markers and EST-based markers were employed for map construction:

1) A total of 272 genomic SSR markers (gSSRs) of chickpea were used which were previously developed in our laboratory (pers. comm.) (section 3.1.4 of Material and Methods)
2) A set of 97 chickpea EST-SSR markers (eSSRs) developed from chickpea seed (chapter 5; Table 5.1) were employed.
3) Other EST-based markers of chickpea (also identified from the same seed library) i.e. 58 ESTPs and 76 PIP functional primers described in (chapter 6; Tables 6.1 and 6.2) were used.
4) From *Medicago*, 15 EST-SSR primers reported by Gutierrez et al. 2005, were also used (see section 3.1.4).

Thus in this study, a total of 503 chickpea molecular markers (272 gSSRs + 97 eSSRs + 58 ESTPs + 76 PIPs) and an additional 15 *Medicago* eSSRs were used for the identification of polymorphic markers between the parental lines of the mapping population. Besides this, the genotyping data of 32 genomic STMS markers and of loci for resistance to fusarium wilt

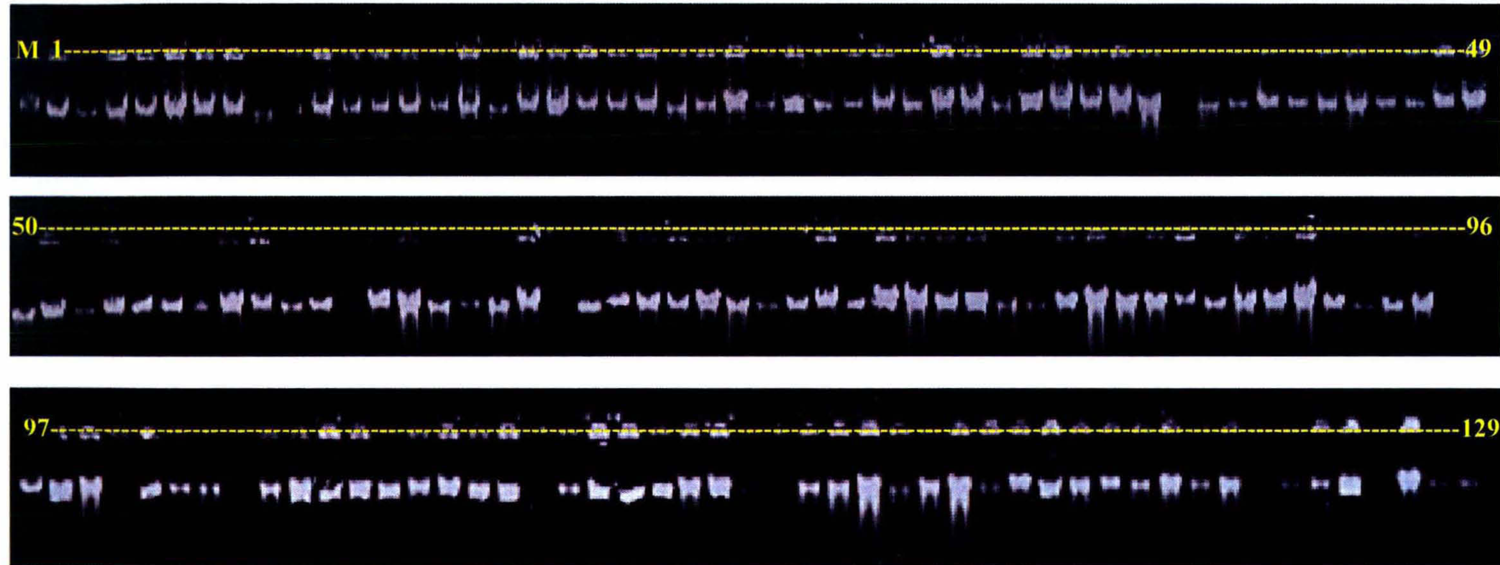


Figure 8.1: Genomic DNA of recombinant inbred lines (RILs) derived from an inter-specific cross of *C. arietinum* ICC4958 x *C. reticulatum* PI489777. DNAs were resolved on 0.8% agarose gel. M: uncut lambda DNA (25ng/μl), Lanes 1-129: genomic DNA of the RILs

races i.e. *Foc*0, *Foc*4 and *Foc*5 obtained from Dr. Fred Muehlbauer, Washington State University, USA (on pers. comm.) were also included in the data analysis for purposes of anchoring and mapping.

8.2.1 Screening for parental polymorphism and genotyping of polymorphic markers

In order to identify polymorphic markers, all the 503 aforementioned molecular markers were screened for polymorphism between the parental lines of mapping population i.e. *C. arietinum* ICC4958 x *C. reticulatum* PI489777. Conditions for PCR amplification are mentioned in section 3.2.13. The amplified products were resolved on 3% Metaphor agarose gels (Fig. 8.2) or 8% PAGE gels (Chapter 6; Fig. 6.2) depending on the resolution pattern. Among the 272 chickpea gSSRs and 97 eSSR markers, 113 and 20 respectively produced clear polymorphic banding patterns between the parental lines. Similarly, among the 134 chickpea EST-based primers (58 ESTPs + 76 PIPs), 10 ESTPs and 24 PIPs were polymorphic. Further, of the 15 *Medicago* EST-SSR primers, 10 produced single band and only two of them were polymorphic.

Additionally, some of the chickpea EST-based primers amplifying large size products possibly due to the presence of intronic sequences produced polymorphic pattern between the mapping parents. However, segregation of the polymorphism was difficult to score in the mapping population after electrophoresis. So to enable polymorphism detection via electrophoresis, alleles amplified by primers namely CESSR19, CESSR52, CESSR69, CEST35 and CEST46 were sequenced directly to reveal the sites of introns using the program Splign (described in section 3.2.18.1) and new primers were designed from the obtained exonic/intronic regions to amplify smaller products (Table 8.1). Sequence analysis revealed that the products were homologous to the EST sequences from which the primers were initially designed. Among these five primer pairs, four (IECESSR19, IECESSR52, IECESSR69, IECEST46) revealed two introns while CEST35 revealed a single intron. Thus, four intron-exon based and two intron-intron based primers were designed from the obtained genomic sequences of the respective primers and are shown in Table 8.1. All the six primers amplified discrete polymorphic PCR products between the parental lines and were genotyped across all the 129 individuals of the RIL population.

Thus, of the 518 (503 chickpea + 15 *Medicago*) new markers screened (as yet unreported and unmapped), 175 (173 + 2) were found to be polymorphic between the mapping parents and are summarized in the table below. These 175 polymorphic primers

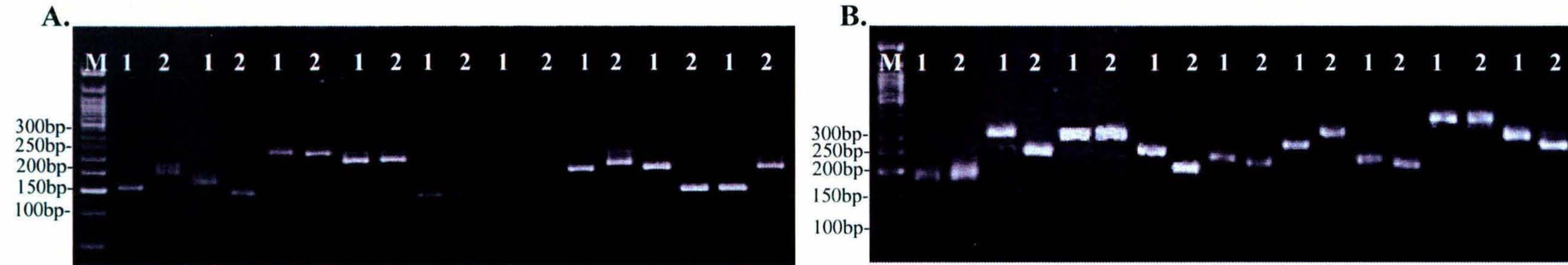


Figure 8.2: Screening of chickpea gSSR primers (NCPGR series) for polymorphism between the parental lines *C. arietinum* ICC4958 (**represented as 1**) and *C. reticulatum* PI489777 (**represented as 2**). The PCR amplified products were resolved on 3 % metaphor agarose gels (**A** and **B**). M represents 100bp ladder.

were genotyped across all the 129 individuals of the RIL population. The representative genotyping gels showing the amplification in the segregating population are depicted in Figs. 8.3 - 8.5. In addition to this, genotyping data of the same 129 RILs using 35 markers from Dr. Fred Muehlbauer was also used for map generation.

S.No.	No. of markers used	No. of polymorphic markers	Percent Polymorphism
1	272 gSSRs	113	41.5%
2	97 eSSRs	20	21.0%
3	58 ESTPs	10	17.2%
4	76 PIPs	24	31.5%
5	15 *Medicago* eSSRs	2	13.3%
6	6 intron-intron/exon based	6	100%
	Total = 524	**175**	**33.3%**

8.2.2 Segregation distortion

The data matrix prepared with the help of genotypic codes for an RIL population, as mentioned earlier in section 3.2.18.2, was used for the construction of a genetic linkage map. Each segregating marker was tested with a chi square test for goodness of fit to the expected 1:1 Mendelian segregation ratio. The values obtained for each marker are tabulated in Table 8.2. Of the 210 polymorphic markers (175 generated in house + 35 obtained from Dr. Fred Muehlbauer), 143 followed the expected segregation ratio, however 67 loci (31.9%) showed segregation distortion (Table 8.2). Among these 67 loci, 37 markers (55.2%) showed slight deviation (* and **) from the ratio while 30 loci (39.59%) exhibited significantly high segregation distortion (*** or above). Of the 67 loci that showed segregation distortion, majority of markers (53) skewed towards the wild annual parent *C. reticulatum* (female) whereas only 12 markers skewed towards cultigen (*C. arietinum*). Thus, on a whole genome basis, the frequency of distorted female markers appeared to be four times more (79.1%) as compared to distorted male markers (19.8%). Further, of the 67 loci, 51 were mapped of which 24 were contributed by loci showing aberrant segregation distortion. Interestingly, most of these distorted markers resided on LG1, LG2, LG3, LG4 and LG12 though LG7 and LG11 harbored single distorted marker (Fig. 8.6). Moreover, clustering of distorted markers often at the centre of the linkage groups was observed at all these linkage groups. Of the mapped markers in LG3, LG4 and LG12, 60% (12 out of 20), 77.7% (14 out of 18) and 75.0% (3 out of 4) of the markers were distorted respectively. These markers together represented 58% of the total distorted markers mapped (Table 8.2).

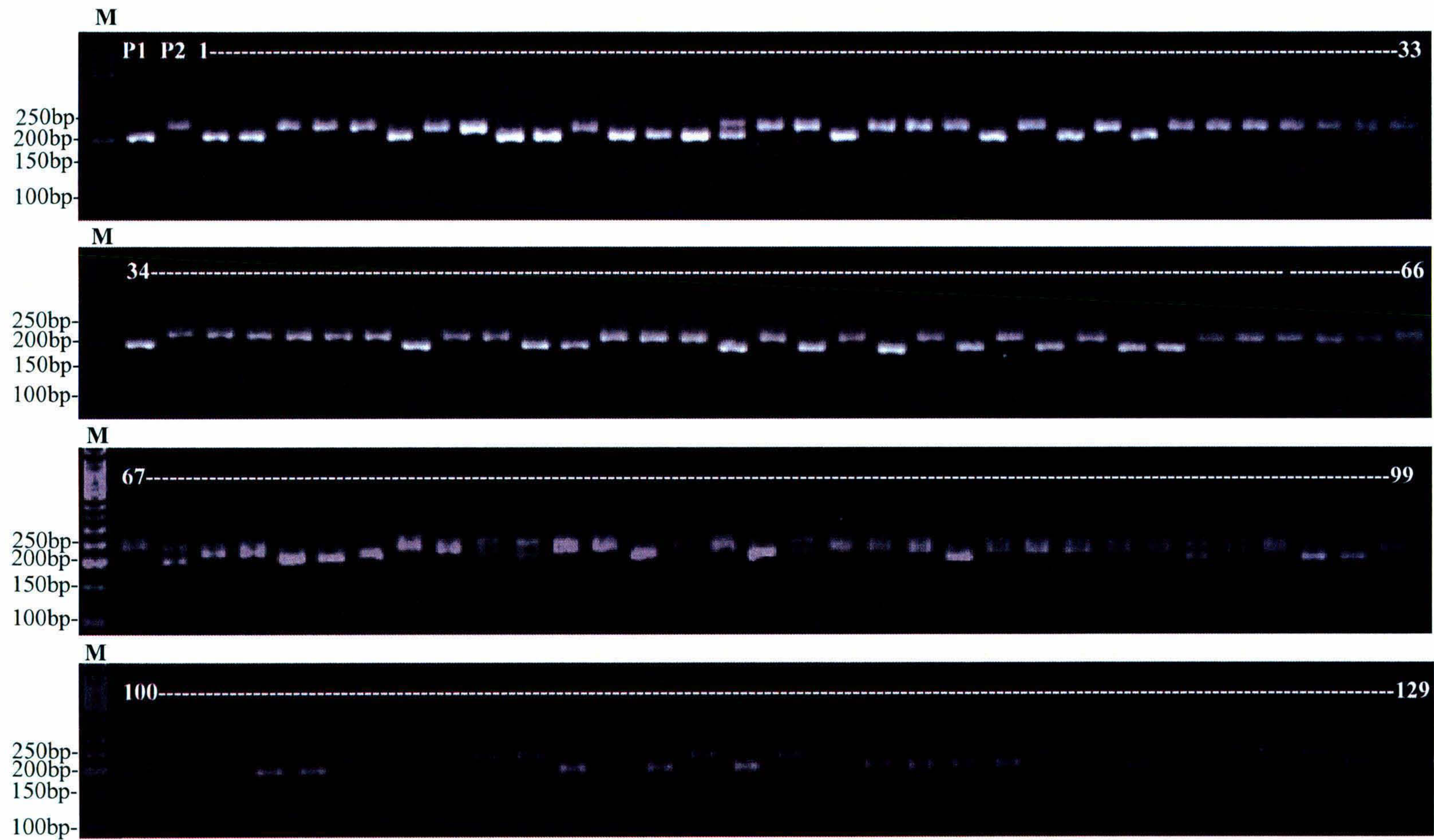


Figure 8.3: Amplification pattern obtained on 3% metaphor agarose gel using STMS primer pair **NCPGR55**. P1 and P2 represent parental lines *C. arietinum* ICC4958 and *C. reticulatum* PI489777, Lanes 1-129: individuals of the RIL mapping population and M indicates 50 bp ladder

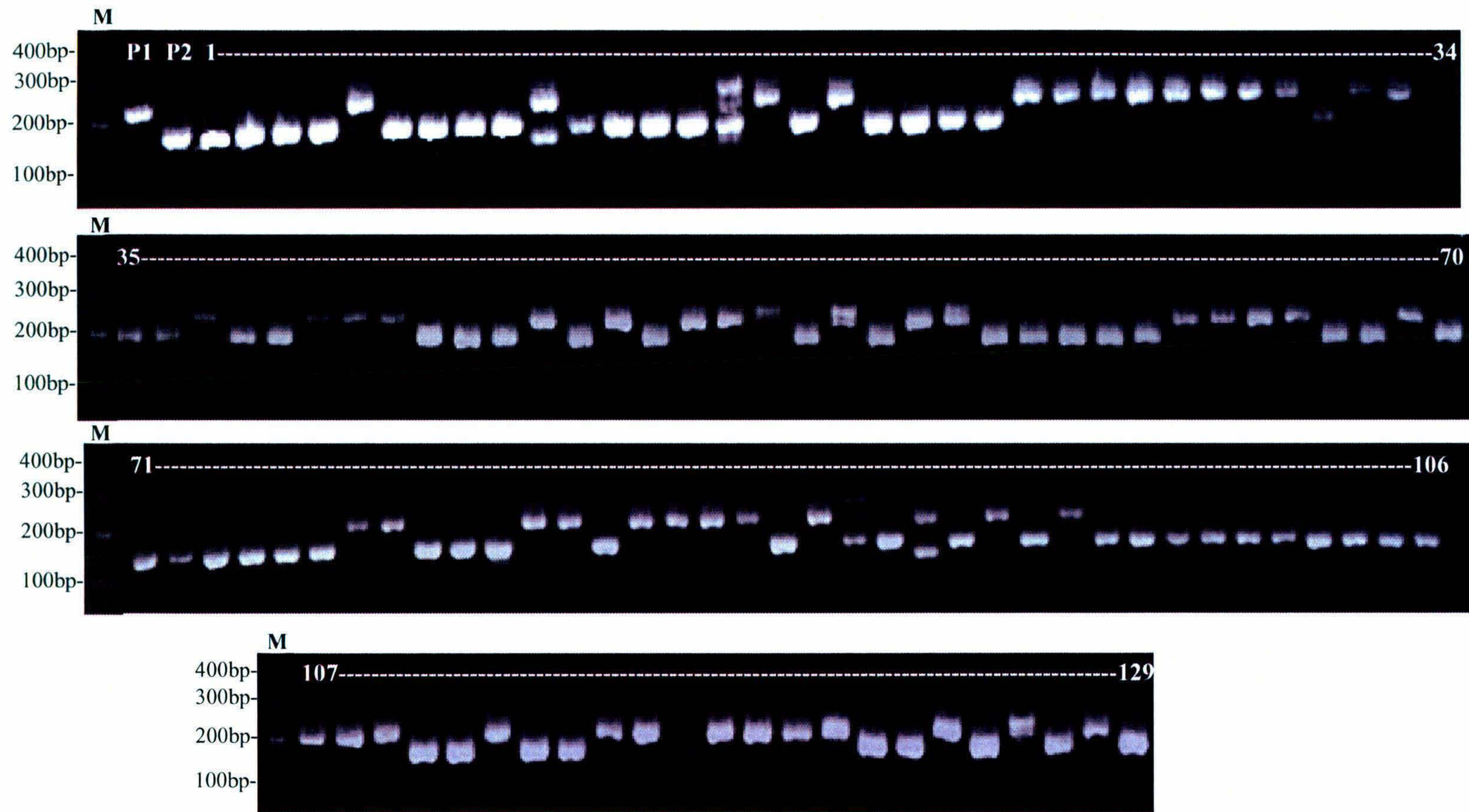


Figure 8.4: Amplification pattern obtained on 3% metaphor agarose gel using **NCPGR86** primer. P1 and P2 indicates parental lines *C. arietinum* ICC4958 and *C. reticulatum* PI489777, Lanes 1-129: individuals of the RIL mapping population and M indicates 100 bp ladder

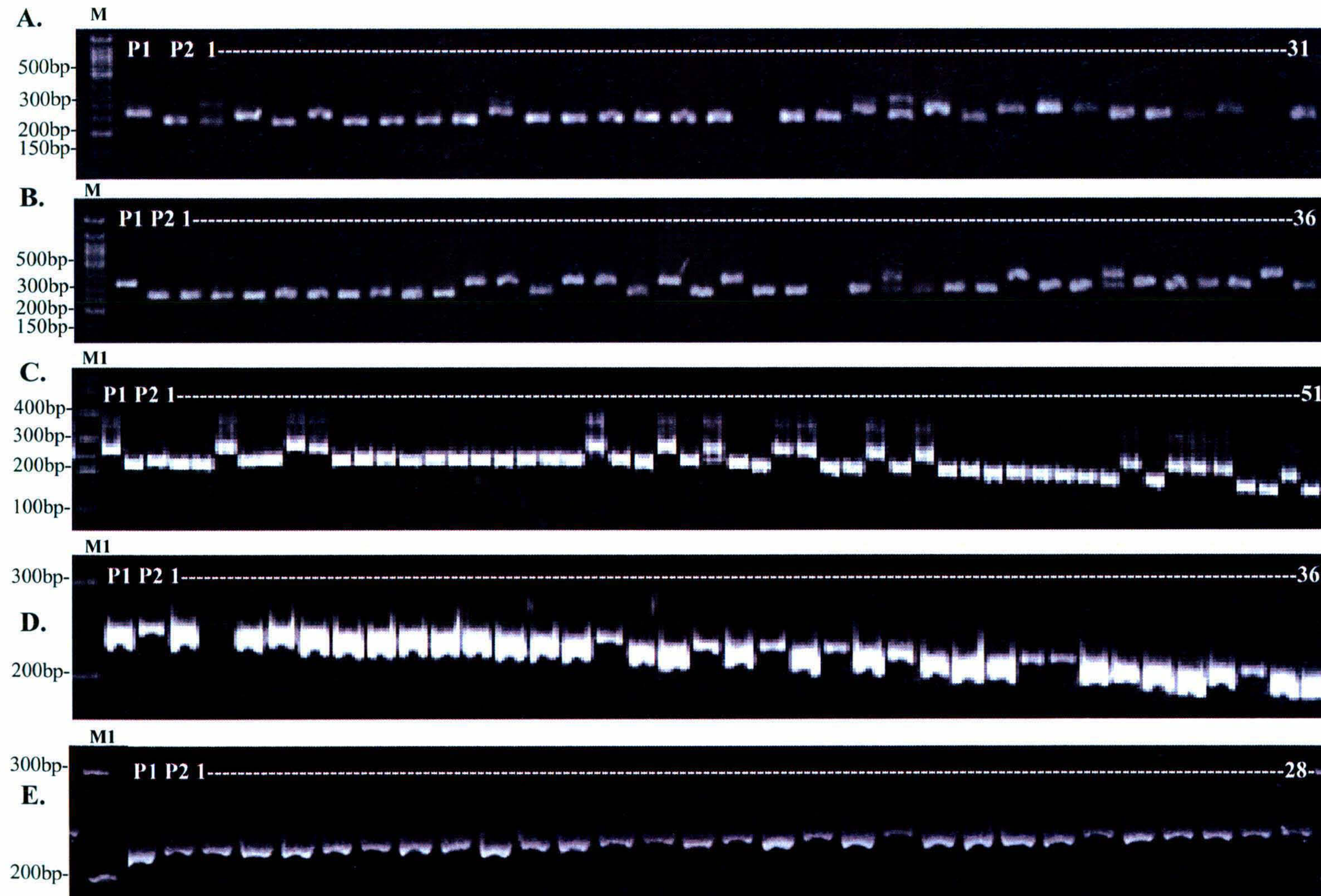


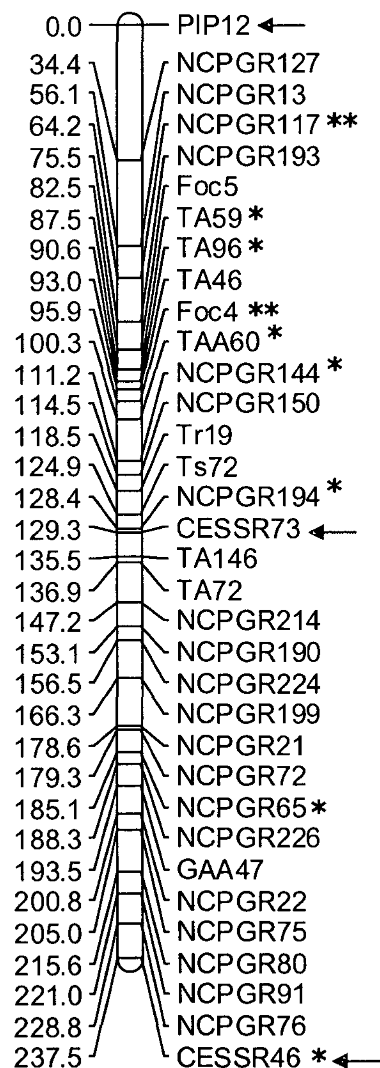
Figure 8.5: PCR amplification with primer pairs **(A) NCPGR127 (B) NCPGR5 (C) CESSR61 (D) IECESSR19 and (E) PIP69** in the genotypes of the RIL mapping population along with the parents as resolved on 3% Metaphor gel (A, B) and 6% PAGE gel (C, D and E). **M:**50bp ladder, **M1:** 100bp ladder, **P1:** allele in Parent1, **P2:** allele in Parent 2 and **Lanes 1-51**: alleles amplified in RILs

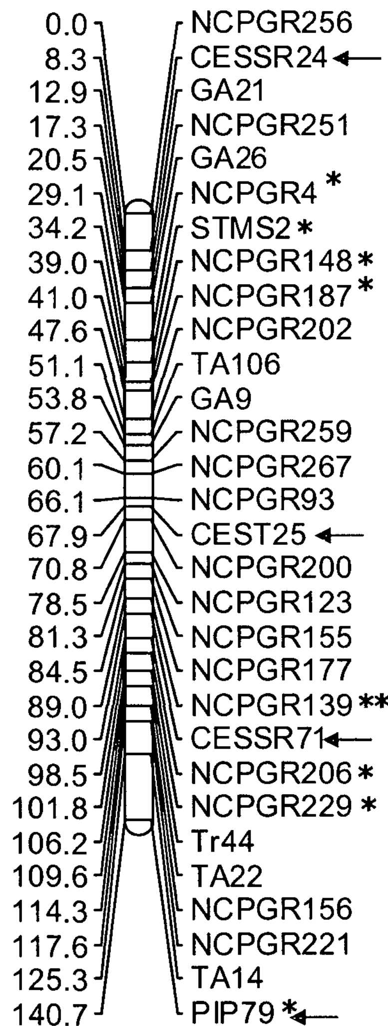
8.2.3 Construction of the inter-specific linkage map of chickpea

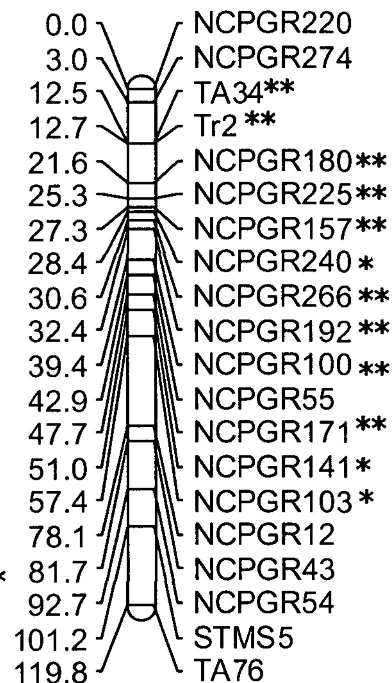
A total of 210 markers were used for the construction of the inter-specific linkage map of chickpea with the help of JoinMap ver 4.0 (van Oojen 2006; section 3.2.18.2). Various permutations and combinations of LOD values were tried for construction of groups; however reliable groups comparable with previous inter-specific linkage maps available in this population were obtained at LOD 3. A total of 146 markers were mapped into 12 linkage groups that spanned 1210.5cM of the chickpea genome at an average marker density of 8.64cM (Fig. 8.6). These 146 markers included 112 new markers (86 gSSRs, 10 EST-SSRs, 7 ESTPs, 8 PIPs and one *Medicago* eSSR) and 34 reported markers of Dr. F. Muehlbauer. The positions of the various mapped markers are depicted in Fig. 8.6 and based on the common markers shared between the recent genetic groups and the previous published map of chickpea (Winter et al. 2000), the LGs of the present map were designated with Arabic numerals, whereas the LGs of the previous map had Roman numerals.

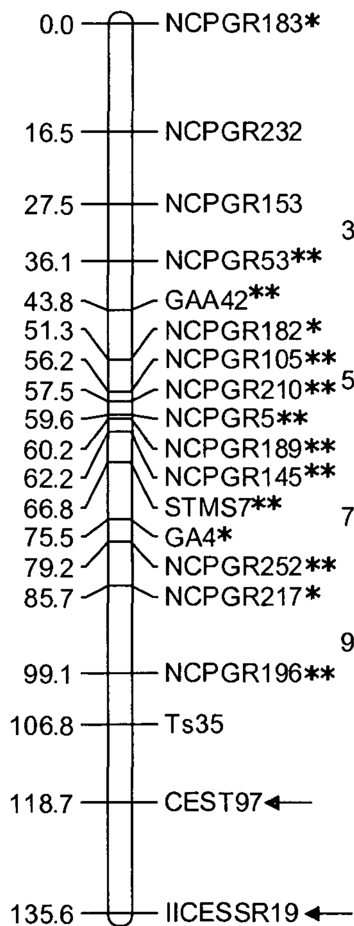
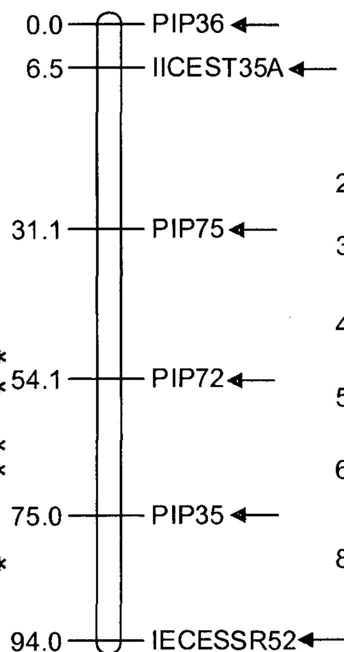
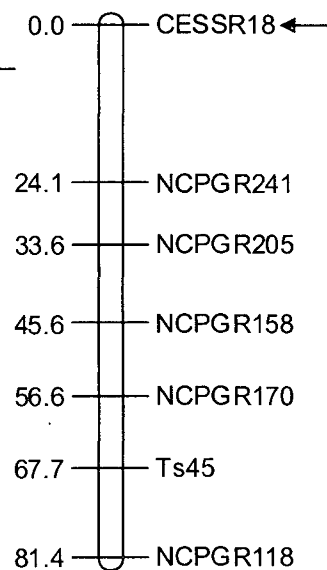
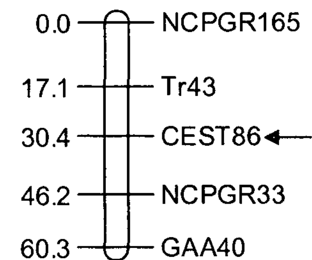
The map spanned 1210.5cM with 50% (LG9, LG11) to 94.2% (LG1) genome coverage (Chakravarti et al. 1991) and a large variation in the lengths of individual linkage groups that varied from 25.0 to 237.5cM, with an average of 100.8cM. The details of number of markers mapped in a linkage group, region they spanned and the average marker density exhibited by them have been summarized in Table 8.3. The markers were not distributed randomly throughout the genome. As shown in Fig. 8.6, some of the linkage groups were densely packed while other markers were sparsely located. LG1 was the largest linkage group both in terms of size (237.5cM) and mapped markers (34). On the other hand, LG12 was the shortest among the linkage groups covering 25.0cM. The sizes of the LGs were not correlated to the number of linked markers in the group. For example, LG8, LG9, LG11 and LG12 harbored same number of markers, but LG8 and LG9 covered 59.7cM and 62.1cM whereas LG11 and LG12 covered 98.5cM and 25.0cM respectively. The average marker density varied from 4.69cM to 32.8cM with an average of 12.27cM. LG2 was the densest with an average marker density of 4.69 followed by LG3 (5.99cM), the second-most dense linkage group. This wide range of marker density indicated differing degrees of saturation of linkage groups with markers.

In the linkage map, the 26 EST based markers mapped to various linkage groups with relatively large gaps. Interestingly 12 of them formed isolated blocks in three linkage groups LG5, LG9 and LG11 indicating a gene rich region that is yet to be unraveled. The remaining 14 markers, including one *Medicago* EST-SSR primer (MTEST), mapped in a backdrop of


LG1 (II +IV)
0.0 PIP12
34.4 NCPGR127
56.1 NCPGR13
64.2 NCPGR117 **
75.5 NCPGR193
82.5 Foc5
87.5 TA59 *
90.6 TA96 *
93.0 TA46
95.9 Foc4 **
100.3 TAA60 *
111.2 NCPGR144 *
114.5 NCPGR150
118.5 Tr19
124.9 Ts72
128.4 NCPGR194 *
129.3 CESSR73
135.5 TA146
136.9 TA72
147.2 NCPGR214
153.1 NCPGR190
156.5 NCPGR224
166.3 NCPGR199
178.6 NCPGR21
179.3 NCPGR72
185.1 NCPGR65 *
188.3 NCPGR226
193.5 GAA47
200.8 NCPGR22
205.0 NCPGR75
215.6 NCPGR80
221.0 NCPGR91
228.8 NCPGR76
237.5 CESSR46 *


LG2 (VI +XIII)
0.0 NCPGR256
8.3 CESSR24
12.9 GA21
17.3 NCPGR251
20.5 GA26
29.1 NCPGR4 *
34.2 STMS2 *
39.0 NCPGR148 *
41.0 NCPGR187 *
47.6 NCPGR202
51.1 TA106
53.8 GA9
57.2 NCPGR259
60.1 NCPGR267
66.1 NCPGR93
67.9 CEST25
70.8 NCPGR200
78.5 NCPGR123
81.3 NCPGR155
84.5 NCPGR177
89.0 NCPGR139 **
93.0 CESSR71
98.5 NCPGR206 *
101.8 NCPGR229 *
106.2 Tr44
109.6 TA22
114.3 NCPGR156
117.6 NCPGR221
125.3 TA14
140.7 PIP79 *


LG3 (III)
0.0 NCPGR220
3.0 NCPGR274
12.5 TA34**
12.7 Tr2 **
21.6 NCPGR180**
25.3 NCPGR225**
27.3 NCPGR157**
28.4 NCPGR240 *
30.6 NCPGR266 **
32.4 NCPGR192 **
39.4 NCPGR100 **
42.9 NCPGR55
47.7 NCPGR171**
51.0 NCPGR141*
57.4 NCPGR103 *
78.1 NCPGR12
81.7 NCPGR43
92.7 NCPGR54
101.2 STMS5
119.8 TA76


LG4(V)
0.0 NCPGR183*
16.5 NCPGR232
27.5 NCPGR153
36.1 NCPGR53**
43.8 GAA42**
51.3 NCPGR182*
56.2 NCPGR105**
57.5 NCPGR210**
59.6 NCPGR5**
60.2 NCPGR189**
62.2 NCPGR145**
66.8 STMS7**
75.5 GA4*
79.2 NCPGR252**
85.7 NCPGR217*
99.1 NCPGR196**
106.8 Ts35
118.7 CEST97
135.6 IICESSR19

LG5
0.0 PIP36
6.5 IICEST35A
31.1 PIP75
54.1 PIP72
75.0 PIP35
94.0 IECESSR52

LG6
0.0 CESSR18
24.1 NCPGR241
33.6 NCPGR205
45.6 NCPGR158
56.6 NCPGR170
67.7 Ts45
81.4 NCPGR118

LG7
0.0 NCPGR165
17.1 Tr43
30.4 CEST86
46.2 NCPGR33
60.3 GAA40

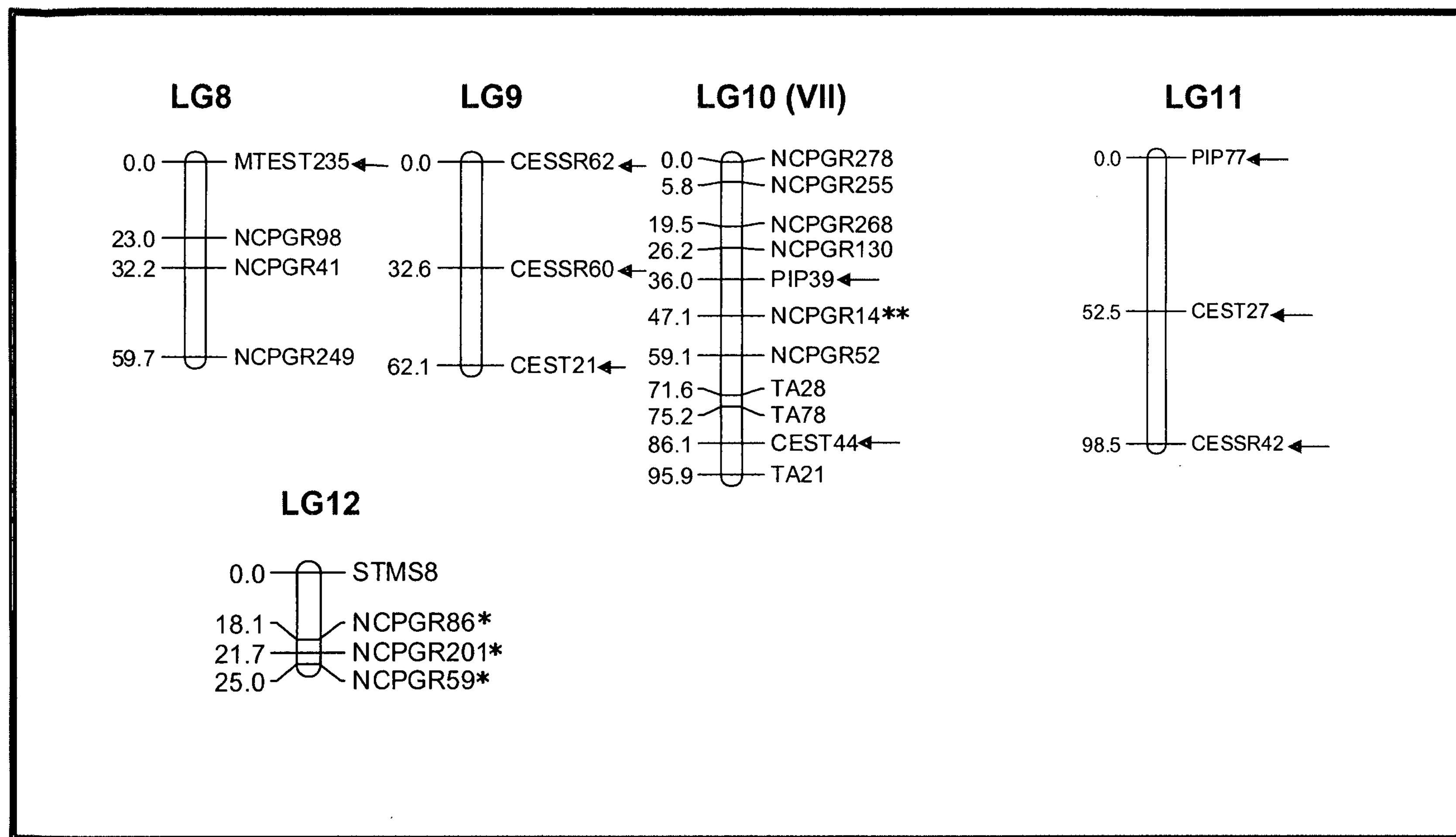


Figure 8.6: The inter-specific linkage map of chickpea constructed using 210 loci of which 146 mapped. Marker distance was set in cM by the Kosambi function (Lander et al. 1987), with 1cM–0.619Mbp. Loci that showed slight segregation distortion are marked with * and aberrant distortion with **. The linkage groups mentioned on the top are numbered in Arabic (LG1 to 12) to differentiate them from the Roman numerals of the previously published map of *C. arietinum* x *C. reticulatum* (LGI to LGXVII; Winter et al. 2000). EST-based markers are marked with purple arrows.

anonymous markers to 7 linkage groups. Sequence homology revealed that of the 26 mapped EST markers, putative function could be assigned to 19 whereas 3 showed homology to hypothetical/unknown protein and remaining three showed 'no hit' (Table 8.2B).

8.3 Discussion

The availability of sufficient number of polymorphic markers is a prerequisite for successful linkage studies. In this regard, the third generation codominant microsatellite based markers emerged as a valuable source for marker generation and map construction. Concomitantly, SSR-based dense linkage maps have been constructed in several economically important crops. However in chickpea, so far only <500 STMS markers have been mapped on the reference mapping population *C. arietinum* ICC4958 x *C. reticulatum* ICC489777 (Winter et al. 1999, 2000) with only a few gene-specific markers (Pfaff and Kahl 2003; Hüttel et al. 2002). To implement marker-assisted pyramiding of genes in chickpea, the pressing need of the hour is to immediately generate a saturated genetic linkage map. Although several intra-specific linkage maps are available for chickpea with various mapping populations (Cho et al. 2002; Flandez-Galvez et al. 2003a; Cobos et al. 2005; Radhika et al. 2007; Taran et al. 2007), but all these maps have been constructed employing the common markers reported by Winter et al. 2000 and lately the markers developed by Lichtenzveig et al. 2005. Furthermore, all these maps also exhibit similar marker order, thus revealing similar genomic locations and therefore are of limited use. Therefore, the present study was undertaken to genetically locate positions of new molecular markers originating from both genomic and genic regions of chickpea including STMS and EST based markers in the chickpea reference mapping population.

As reported in other plant studies, higher level of DNA polymorphism between the parental lines of the chickpea inter-specific cross was obtained with gSSR markers (41.5%) as compared to eSSRs (21.0%). Several studies have compared the level of polymorphism obtained with microsatellites isolated from genomic and EST libraries in different systems and generally observed lower polymorphism in the latter case owing to the conserved nature of the genic regions (Cho et al. 2000; Varshney et al. 2005a). The genomic-SSRs used in the present study yielded 41.5% polymorphism between the parental lines which was comparable with earlier studies carried out at inter/intra level in chickpea demonstrating rates of polymorphism varying from 30% - 50% (Hüttel et al. 1999; Winter et al. 1999, 2000; Udupa and Baum 2003; Cho et al. 2004; Radhika et al. 2007; Taran et al. 2007). Further, the polymorphism rate of 21.0% with chickpea EST-SSRs in the inter-specific cross was

comparable to those reported with cotton EST-SSRs (19.8 %, Park et al. 2005) but lower than those obtained with pepper EST-SSRs (29.2%, Yu et al. 2006). Further among the three kinds of chickpea EST based markers employed for mapping purpose, highest polymorphism was obtained with intron based PIP (31.5.0%), followed by EST-SSR markers (21.0%) and lastly exon based ESTP (17.2%) markers suggesting that introns may serve as an effiecient source of hypervariable markers in chickpea for analysis of genetic diversity, map saturation and comparative mapping studies.

The segregation distortion obtained with the current set of chickpea co-dominant markers (31.9%) was comparable to the marker distortion (38.0%) reported by Winter et al. 2000 on the same mapping population. Generally, a higher percentage of allelic distortion in the RILs of chickpea inter-specific cross was observed than in the intra-specific, an advantageous feature for using intra-specific cross for mapping (Cho et al. 2002; Flandez-Galvez et al. 2003a). This distortion may have been due to recombination suppression at meiosis, a common phenomenon in inter-specific crosses caused by a considerable degree of non-/or partial-homology between the chromosomes of *C. arietinum* and *C. reticulatum* (Winter et al. 1999). Translocations and inversions are common causes of meiotic abnormalities noted in interspecific crosses. Alongwith these, several other factors are also responsible for the segregation distortion such as the types of mapping population, RILs/F2/BC, nature of cross etc. Flandez-Galvez et al. 2003a found only 20.4% distortion in a chickpea F2 intra-specific population. In our study, of the 67 distorted markers, majority (79.1%) skewed in favor of the genome of the female parent (*C. reticulatum*), thus strengthening the observation of Winter et al. 1999; 2000. A possible explanation for this was deduced from the cytological studies which compared the karyotypes of chickpea and *C. reticulatum*, and revealed rearrangements in chickpea relative to the *C. reticulatum* chromosomes. The apparent clustering of distorted loci often at central positions was observed in our study (Fig. 8.6) on LG 1, 3, 4, which is consistent with results obtained earlier in chickpea (Tekeoglu et al. 2002; Winter et al. 1999) and in other crops (Riaz et al. 2004; Kidwell et al. 1993). Inversions and other chromosomal variations are common causes of regional recombination suppression in inter-specific crosses that resulted in clustering of markers on a linkage map, thus leading to less overall coverage of the genome (Tanksley et al. 1992).

The current genomic map of chickpea positioned 146 markers onto 12 linkage groups that spanned 1210.5cM at an average distance of 8.6cM. The number of linkage groups was

more than the chickpea haploid number (n=8) indicating that small groups will coalesce when more new markers from the chickpea genome would be added and integrated to construct the detailed genetic linkage map. If LOD values below 3.0 were selected, all the major linkage groups (LG1-8) fell into one group with more number of markers linked, however the linkage pattern reported earlier was not reconstructed (Winter et al. 2000). Thus LOD 3.0 was selected, in which 69.5% segregating markers were mapped and the 30.5% that remained unlinked were maximally represented by EST-based markers (17.1 %) suggesting that still some gaps and uncovered regions of the genome need to be filled in chickpea.

Relative to the estimated physical size of the genome (750 Mbp; Arumuganathan and Earle 1991), 1-cM distance in this map corresponds to approximately 619Kbp which is roughly double the value of 360kbp obtained by Winter et al. 2000. The extended map length in the current map is primarily due to genic markers suggesting that recombination may be more frequent in gene-rich regions than in non-coding regions (Yi et al. 2006). In the present map, non-random distribution of markers especially genomic-SSRs particularly at the central regions resulted in clustering in some linkage groups (Fig. 8.6). This might be attributed to the fact that microsatellite sequences in the chickpea genome cluster around centromeres (Gortner et al. 1998). Such clustering of microsatellites around centromere has been observed in various plant species like sugarbeet (Schmidt and Heslop-Harrison, 1996), barley (Roder et al. 1993; Pederson and Linde-Laursen 1994), tomato (Arens et al. 1995; Broun and Tanksley 1996) and several other *Triticeae* species (Pederson and Linde-Laursen 1994). Several factors are responsible for this clustering of genomic SSRs on genetic linkage maps, major being their non-random physical distribution in plant genomes (Ramsay et al. 1999; Elsik and Williams 2001), reduced recombination in centromeric regions (Areshechenkova and Ganal 2002; Ramsay et al. 2000) and the genomic origin of DNA sequences used for SSR development (Tanksley et al. 1992).

Further, 26 new genic loci were positioned onto this linkage map, bringing the number of mapped genes of known function to 75 on the chickpea reference mapping population (including 44 mapped by Pfaff and Kahl 2003 and 5 RGAs by Hüttel et al. 2002). Moreover, almost 46% of the mapped EST-based markers showed clustering leading to formation of isolated groups (LG5, LG9 and LG11 Fig. 8.6). This phenomenon has also been observed by King (2002) wherein the clustering of gene based markers resulted in formation of isolated blocks. The other 14 gene-based markers mapped in a backdrop of anonymous markers to 7 linkage groups, and could be utilized in candidate gene approaches

(Pflieger et al. 2001). Generally, in contrast to genomic SSRs, the gene-derived loci in many crops such as wheat (Xue et al. 2008), rice (Wu et al. 2002), white clover (Barrett et al. 2004) are mapped into the distal regions of the map indicating that recombination is more frequent in the telomeric regions than centromeric regions. The present study suggests that EST-SSRs tend to map to regions of high recombination where markers are less likely to be identified using genomic DNA-based markers. The preferential localization in recombination hot spots enhances the value of these markers (Yu et al. 2004), however in future more number of EST markers are required to identify the gene-rich or euchromatic regions of the chickpea genome.

The use of common markers in the present map has enabled us to compare it with the previous maps of Winter et al. 1999 and 2000. The current map revealed linkage conservation in atleast seven linkage groups of Winter et al. 2000 (Fig. 8.6). Those common markers lying adjacent to each other on previous map for example loci TA59, *Foc*5, TA96 and *Foc*4 on LG1; TA106, GA9 on LG2; TA34, TA2 on LG3; GAA42, STMS7 on LG4 and TA28, TA78 on LG10 share almost same marker order and distance in the current map. The development and use of these new chickpea EST-derived markers along with unmapped gSSR markers has enabled us to map new genomic locations in the map. The integration of this genomic information to the available inter-specific chickpea map will substantially increase the marker density, thus leading to a step forward for the development of a saturated linkage map.

Using the 'chickpea reference mapping population', the inter-specific linkage map developed in this study defined the new positions of 112 markers by virtue of utilization of new EST-based markers i.e. EST-SSRs, ESTPs and PIPs and previously developed unmapped genomic STMS markers. Even though, EST-derived markers usually detect less polymorphism than the genomic STMS markers, defining their map positions could provide invaluable information for exploring the expressed region of the chickpea genome and functional analysis of traits of interest. This study demonstrated that the availability of a large number of EST-based markers was a pre-requisite for construction of high-resolution and marker-dense transcriptional map of chickpea which would be invaluable for mapping genes/QTLs for chickpea yield, seed quality and disease resistance and also for integrating physical and genetic maps in the future.

Table 8.1: List of primer pairs designed from the higher sized alleles obtained with the chickpea EST-based markers. The designed primer pair sequences, expected size (bp), putative functions based on BLASTX results and the GenBank accession numbers are mentioned.

S.No.	Primer Name	Primer sequence	Expected size (bp)	Putative function	GenBank no.
1	IECESSR19	TCTCTCTCTCTCTCTGACACAA/ TGTTACCTCAATCTCAGGCTCT	242	DNA-binding	ES544477
2	IECESSSR52	CTTTTGTCTTGTGATTCCAGCA/ CACCACCAACACAATTTTATCC	290	Unknown protein	EX151786
3	IECESSR69	CACTATGATTGTTCTCTTGCTTGG/ AGTGCATCATGGAAAGTCATGT	354	Hypothetical protein	EX151775
4	IICEST35A	ACGGTTTCATTTGTTGTTGG/ TCGTGTAACATAGTGAGTGGGTTG	250	Sugar transporter	EX151642
5	IICEST35B	ACGTCGGAGTTTCAGGTACTTT/ GTCGTCATCTCATAACCACGAA	260	-do-	-do-
6	IECEST46	GGCTACTTAATTCATGTTTGGTG/ CAAAGAGGAACAACTCCTTTCC	375	Disulfide isomerase	EX567537

Table 8.2: Genomic and genic-derived PCR based codominant molecular markers used in the present study for map construction.

A) List of polymorphic chickpea genomic STMS markers with their repeat motifs, primer sequence, product size (bp), chi square value (χ^2), segregation distortion and the map position of the various loci are mentioned.

Sl. No.	Locus	Motifs	Primer Sequence	Product size (bp)	χ 2	Segregation distortion	Map position
1	NCPGR4	$(CT)_{16}$	F: TTACAGCTTGTGCTCAG R: AGTCAGATTCTTATCCGA	195	2.93	*	2
2	NCPGR5	$(GA)_{12}$...$(GA)_3$ ..$(GA)_4$	F: GACAATAATGGTGAACGA R: GGCACAATGTATGTATTG	308	23.61	*******	4
3	NCPGR6	$(CA)_{12}$	F: GACCAAGATTAGTAGAACCT R: TATGTCTACACCTATGCATC	253	11.31	*****	UL
4	NCPGR12	$(CT)_{35}$	F: CCTTGTTAGTGTGTATAGGT R: GTAATGACCAAGTGAACA	251	1.69	-	3
5	NCPGR13	$(CA)_{11}$	F: GTTGTTGCCGTGACTT R: TGAATCGGACTGACACT	313	1.56	-	1
6	NCPGR14	$(GA)_{19}$	F: TCCATTGTAGCTTAGCTTAG R: TCTTACTCTTAGCTTACCTCTT	306	7.94	****	10
7	NCPGR21	$(CT)_{15}$	F: TCTACCTCGTTTTTCGTGCC R: TTGCTCCTTCAACAAAACCC	137	0.01	-	1
8	NCPGR22	$(CT)_{10}N(CT)_3$	F: GTTCTTCTCGCCCAACTACG R: GCAACCACAATAATGTTTTCCA	257	1.22	-	1
9	NCPGR28	$(AT)_6(GT)_5$gc $(GT)_7$	F: TGATGGAAGGTGATGTGGAA R: GAGGGGGAAACGTTTTCTTT	224	0.31	-	UL
10	NCPGR32	(CA)12	F: CGCAGGTAAAGCTCCTCTCA R: CCCCTTTTTCACCCTGTAAG	250	1.40	-	UL
11	NCPGR33	(GA)20	F: ACATCTTGAAGTGCCCCAAC R: TGCAAGCAGACGGTTACAAG	248	0.98	-	8
12	NCPGR34	(CT)17	F: TGGAAGGTGTTTTAGTGGGTG R: GACTAACTGGCCCCCAAAA	240	5.83	**	UL
13	NCPGR40	(GA)12, (GA)4gt gg(GA)10gtN2 gt(GA)6	F: TGAACGAATCATGGCAAGAG R: GCCCTCCTTCTTGCTTACAA	193	4.23	**	UL
14	NCPGR41	(CT)8(CA)17	F: GGGAGGAGGATCAAAATTAC R: CAACTATAAAGAGGCATGTTCC	262	4.30	**	7
15	NCPGR43	(CT)12at(CT)3	F: GAAGTCGAGATGCTGAAAAG R: AATTCTAGAAGGGAAGGGTG	255	1.16	-	3
16	NCPGR46	(CT)6at(CT)6at (CT)(CA)14	F: CCCAAAACTGAAATGGAAAC R: GGCAGTTACTACCAAGGCAT	217	0.52	-	UL
17	NCPGR50	(GA)26	F: ATGATGGATTTTCGGAATGT R: AAAAATGCTGGAAGGAACTG	203	4.37	**	UL
18	NCPGR52	(GA)2aa(GA)25	F: CAAGCTCTTTCAGAATTTGC R: TACTGGTGGAAAAAATGGATG	245	0.67	-	10
19	NCPGR53	(CT)5ca(CT)ca (CT)10ca(CT)4 ca(CT)ta(CT)4 gtca(CT)12	F: CCCTCCTTCTTGCTTACAAA R: TAATGGTGAACGAATCATGG	194	9.17	****	4
20	NCPGR54	(CT)16	F: GAAGTCGAGATGCTGAAAAG R: AATTCTAGAAGGGAAGGGTG	255	0.01	-	3
21	NCPGR55	(GA)16	F: TCCATTGGATACATCACAGG R: GGGCAAATTCAGTATTTTGG	204	8.26	****	3
22	NCPGR56	(GA)12gt(GA)ca (GA)3N6(GA)4gt (GA)9gt(GA)gt (GA)6	F: CATGACAATAATGGTGAACG R: GATCTTGACTTCTGTTTGTGC	162	21.15	********	UL
23	NCPGR59	(GT)12(GA)13	F: CTTGACCAGAGGCATTTATC R: AACATAATGGTGTCCAAAGC	267	4.10	**	12
24	NCPGR62	(GT)14(GA)4gg (GA)2	F: TCTTAGACTCGGACCTGGTA R: TTCGTTTTTCTCTTACGCTC	295	0.93	-	UL
25	NCPGR65	(GA)7gg(GA)29	F: CGTGGACTAACGTTCACTGT R: GCCAAAGCATCTGGAATCTC	243	2.75	*	1
26	NCPGR69	(GA)36	F: GACCGAATGTCCATAAATCA R: GGAGCTGGAAAAACTACAGC	252	3.00	*	UL
27	NCPGR72	(GA)21	F: TTAACCCATTAGCGTGACTT R: GATCAGCTTCTTGCTTTCAT	250	2.10	-	1

28	NCPGR75	(CT)15(CA)14	F: AACTGAAATGGAAACACAGG R: GAAAGCTGACTCCTCTACCA	192	0.13	-	1
29	NCPGR76	(GT)13(GA)2ta (GA)7ta(GA)6	F: GAAAGCTGACTCCTCTACCA R: GAAAATGCTCTCAGTCAAGG	245	2.57	-	1
30	NCPGR80	(AT)4(AC)2at (AC)24	F: TGGACTAACCCTTCTTTCTTC R: TTATATTATGCAGGACCGCT	256	0.45	-	1
31	NCPGR86	(CT)13(CA)11	F: CTACTGCAGAAAAATCAGGG R: ATAGTTCTTGACCAGAGGCA	208	5.65	**	12
32	NCPGR91	(AG)12at(AG)17	F: ATTGAATCCTTTCTGAACCG R: CTGTTCTCTTTTCTCCTCCG	266	0.07	-	1
33	NCPGR93	(CA)2(CT)24 (CA)13	F: CAAAGTTTGTTGCTAGGATTC R: GAAGATCTCCGACGATGATA	299	1.07	-	2
34	NCPGR98	(GA)20gg(GA)14	F: CATCTTATTTTTCATTTTAGAGGAGG R: AGGAAGTGTTATGGAGATGCC	141	1.40	-	7
35	NCPGR100	(CT)15, (CT)5tt(CT)6 at(CT)7	F: CCATTT TCTACAATCTCATGTCT R: GTAGAAAGAGCCAAGAGGCA	263	12.37	******	3
36	NCPGR103	(CT)2tc(CT)21	F: ACAACCATATACTTTTGGCG R: TTAGATGAAAAACGGGAGAA	213	5.53	**	3
37	NCPGR105	(CT)16at(CT)7, (CT)18	F: TTTTTGTTAAGCCATCAAAGT R: TTTCCCTTTTAGAATGATGC	261	27.85	*******	4
38	NCPGR109	(CT)12cccc(CT)10	F: TAGCTCAAAGAGATAACCCG R: AAAACAAATCACCTACCCCT	285	0.77	-	UL
39	NCPGR110	(AT)6(GT)4gc (GT)32at(GT)5 ct(GT)10	F: CAAGGTCAATTCGTAGAAGG R: GAACGAGAGTTGGTATTGTTG	217	0.04	-	UL
40	NCPGR117	(CT)23	F: GAACTTCTTCAATCTCACGG R: CTAGCACGATGAAAGGATTC	199	10.29	****	1
41	NCPGR118	(GT)12(GA)18	F: GAGTCGATTTCGTGTTGATT R: ACGTGAAATTCCACCACTAC	224	1.12	-	6
42	NCPGR123	(CT)25	F: CTCTGCAGACTGAGGGTAAG R: TCTGGAGGAGAAGAGACAAA	273	0.63	-	2
43	NCPGR127	(GA)18	F: CATAATGCAAGGGCAATTAG R: CTCTTATCTTCATGTTGCCG	279	0.09	-	1
44	NCPGR128	(CA)9cg(CA)2 (CGCA)4(CA)2 N42(CG)4(CA)9	F: GCAATGAGCAACTTTTCCTT R: ATTGGTGTAACTTTTCCGCT	290	8.49	****	UL
45	NCPGR129	(GT)21	F: ACGAAGAATTTAATACCGGA R: GAGATTTGAGTTTGACGGTT	293	6.04	**	UL
46	NCPGR130	(CT)24tt(CT)2	F: GATACTGGTGGAAAAAATGGA R: CAAGCTCTTTCAGAATTTGC	245	0.01	-	10
47	NCPGR136	(GT)7gc(GT)ac (GT)gc(GT)gg (GT)10	F: GGACTGAGTGAGTTCGTCTT R: GTATCCTCGGTTTCCCTATC	132	0.62	-	UL
48	NCPGR137	(GT)6ct(GT)3ct (GT)3gg(GT)5	F: GTGATGCGACCATGTGAAAA R: CGTGGACTAACACATGAGGA	287	0.53	-	UL
49	NCPGR139	(GA)40	F: TGGGTCTTATTGGGTTTGAT R: CATGCATTTAGGATGAACCA	245	10.14	****	2
50	NCPGR140	(GT)14gc(GT)gc (GT)gc(GT)10	F: ATTGGTTTGAGAAGTGATGG R: TTTTATTTCTCACCCACCAG	264	0.00	-	UL
51	NCPGR141	(GA)8aa(GA)13 aa(GA)9	F: ACTCAAAAGACAGCAAAGCA R: AGCTTAGAGCACTCACATGC	211	6.02	**	3
52	NCPGR144	(GT)5g(GT)5 (GA)7	F: TCTGAACAAGGTTTTCCTCA R: TTCATTTGTCCATCAACCTC	252	5.63	**	1
53	NCPGR145	(CT)5(CACT)2 (CT)10ca(CT)4	F: CCATATGAAGATATTGTGGCA R: ATCATGGCAAGAGGTAGGTC	316	31.5	*******	4
54	NCPGR148	(GA)12N5(GA)9	F: ACACAAGCCTATGCAATGA R: GCTTGAGTTTATGCTTCTGG	285	4.94	**	2
55	NCPGR150	(AT)5(GT)16	F: GGACCCGACAACACTACTAA R: GGGTTAAAGATGTGCCATAG	287	0.15	-	1
56	NCPGR153	(CT)16	F: TGCCTCAAACTCCTACTCAT R: AGTGGAGCTAGGGAAATACC	281	0.03	-	4
57	NCPGR155	(GA)18	F: GGGAAAAATAATGAGGAGGA R: TGGCTCACAATTTTCTCTCT	281	0.69	-	2
58	NCPGR156	(CA)12(TA)5	F: CGATTATGTGTCATCCCTTT R: ATTTCAACGTCTCAACCATC	261	0.14	-	2
59	NCPGR157	(CA)16(TA)3	F: TCCGTAACAGTGATGAACAA R: TGGGATTACACTGGATAAGG	203	7.14	***	3
60	NCPGR158	(CT)3tc(CT)14n ca(CT)3t(CT)8	F: TAAAGCTGGAAACTCGAAAG R: TAACCTTCCAATACCGAAGA	179	0.68	-	6
61	NCPGR165	(GA)15	F: TCAGAAGAAAACGAAAGAGC R: CAGCAACCTTAATTGGACAC	233	0.55	-	8
62	NCPGR170	(CT)18(CA)12	F: ACGTGAAATTCCACCACTAC	224	0.21	-	6

			R: GAGTCGATTTCGTGTTGATT				
63	NCPGR171	(GA)30	F: AAAGACAGCAAAGCAAAGAG R: AAAACACCATAAATTCCACG	205	7.72	***	3
64	NCPGR177	(GA)19	F: GGGGAAAAATAATGAGGAGG R: GGCACCCAATTTTCTCTTAC	253	0.09	-	2
65	NCPGR180	(CA)4a(CA)10 (TA)4	F: TCCGTAACAGTGATGAACAA R: TGGGATTACACTGGATAAGG	283	6.88	***	3
66	NCPGR182	(CA)12(TA)2	F: CCCAAAGAAGACAAAACAAC R: TCATTTAAGGCAGGTCAGTC	190	623	**	4
67	NCPGR183	(GA)12ggata (GA)9	F: AAAACATTGGTGGCAACTCC R: AGAGTCACACACACACACACA	236	4.65	**	4
68	NCPGR187	(CT)9atc(CT)13	F: CCTTCACTGTCGGTTATGAT R: TAACACAAGCCTATGCAATG	152	3.53	*	2
69	NCPGR188	(TA)2tg(TA)3 (TG)12	F: GTTAATTGAGTTGCGACGAG R: TCTGTTTCCTTCCTTTTTCC	181	1.89	-	UL
70	NCPGR189	(CT)9;(CT)5 (CACT)2 (CT)10ca(CT)4	F: TGGCACAATGTATGTATTGAA R: ATGGCAAGAGGTAGGTCATA	297	20.1	*******	4
71	NCPGR190	(AT)7(GT)13	F: CCTTAGTGTATAAACCCGAAAC R: GACCTGCTTGAGTTAGACCA	289	1.92	-	1
72	NCPGR192	(TA)3(TG)12tt (TG)2	F: TGGGATTACACTGGATAAGG R: TCCGTAACAGTGATGAACAA	203	8.85	****	3
73	NCPGR193	(AT)9gtat(GT)9	F: CCGATAAAATCACAACCGAG R: AAACGGGGTTTACAGAAGG	232	0.09	-	1
74	NCPGR194	(TG)6g(TG)5 (AG)7	F: AGCCAAAAATCGACATAGAA R: ATTTCATTTGTCCATCAACC	190	6.53	**	1
75	NCPGR196	(CT)17	F: TTGGGTCATTACCTTCATCT R: CTCATCCTTGAGAGAAATCG	226	11.13	*****	4
76	NCPGR199	(GA)27	F: GGACATAGTAATCTCCGCTG R: CCAACACCAACACCAACATA	196	0.01	-	1
77	NCPGR200	(GA)24	F: TTCACACAACAACCTTTTCA R: GGTGAGTTTCTTTTTCCCTT	250	1.18	-	2
78	NCPGR201	(CT)13(CA)12	F: TATGCAAGCAATCCTTTAGC R: TCTTTTGGAAACTAAGCCCT	269	6.53	**	12
79	NCPGR202	(CT)25	F: AGGCCTTTTCCTTTTTACCT R: GGAAAAATTCCCGATCATAC	259	2.84	*	2
80	NCPGR205	(CA)17(TA)5	F: AAGCAAAAGGAAGCAAAGAA R: AGTGGGTTGAGAAATTACGG	267	0.14	-	6
81	NCPGR206	(GA)3ta(GA)7aa(G A)8	F: AACAACACTGGGTGAGAGAT R: GATCCACATGCTACCATACC	252	5.17	**	2
82	NCPGR210	(GA)17	F: AAGGTAGACGTGTGCGTG R: CCTGTTATGGAAGATAGGGC	224	22.5	*******	4
83	NCPGR213	(CT)3(CA)12	F: TTCATGGATGTAATTCTCCC R: CCCCACTATTTTCCACATAA	220	7.69	***	UL
84	NCPGR214	(CA)14(TA)5	F: ATTTCCCGTGTCTTTGAGAT R: GGAATTAGTTGATGTGACAATG	225	0.25	-	1
85	NCPGR217	(TG)15	F: GACTACTTGGAATACGTCGC R: CGCGCAGTGATTTAAGCTAT	171	4.68	**	4
86	NCPGR220	(GT)13(GA)4	F: ACTTCTCTACTCAGCCCCTT R: GCCCCTATCTTTCAGACTTT	255	0.00	-	3
87	NCPGR221	(CA)3cga(CA)cg(C A)7(TA)4	F: CATATGCATCATCTCAACCA R: TGTCCTTCGTCTTGTTCTTC	260	0.25	-	2
88	NCPGR224	(AT)6(GT)14	F: TGGAATTAGTTGATGTGACAA R: ATTTCCCGTGTCTTTGAGAT	225	1.26	-	1
89	NCPGR225	(CA)3a(CA)12 (TA)3	F: TCCGTAACAGTGATGAACAA R: TGGGATTACACTGGATAAGG	203	13.8	******	3
90	NCPGR226	(CT)17	F: GACTGCATGTTTTCTTCTCG R: ACCACTTCAAAGCCTATTCA	205	0.40	-	1
91	NCPGR229	(GA)3ta(GA)15	F: CAAATTTTGCGCTGTTGTAG R: ACACCTCATCTCCCTTTGAA	158	4.12	**	2
92	NCPGR232	(GA)34	F: GGACCGAATGTCCATAAATC R: TCTTTTAGGACCCAATGGAG	265	0.01	-	4
93	NCPGR238	(GA)3a(GA)18	F: GTCCGTGACATTGACACTTT R: CATAGTTGGATTGCCTCTCA	273	0.78	-	UL
94	NCPGR240	(GA)17	F: AAGGGGTGAGTTTTTGAGTT R: CCCCTTAATTTCTTTCTCCA	238	6.57	**	3
95	NCPGR241	(TA)5(TG)15	F: GCGTTTTCCAGAGAAATTCA R: GGGAGGAAACATTTTCGTTT	250	0.98	-	6
96	NCPGR242	(CT)11(CA)12	F: TCGTCATATCCACCCGATAA R: TGGATAATGGTGCGAAAGAA	145	1.63	-	UL
97	NCPGR249	(CA)5a(CG)3 (CA)10	F: CTCTTCGATTCGGATAGGTT R: TGTTTTCAGCTAAATTTCACG	231	1.63	-	7
98	NCPGR251	(CA)13	F: AATGGGTTAATTTGACTTGC	282	1.41	-	2

			R: TTAATGGCCACCATAATCTT				
99	NCPGR252	(CA)12	F: TTGCCCTGAGGAATACATTA R: GGTTGTTGAAGGCATAACTG	187	13.7	******	4
100	NCPGR254	(AT)2(GT)11	F: GCCTTTTTCAATTTCTCTCA R: CCCAAAGAAGACAAAACAAC	298	0.00	-	UL
101	NCPGR255	(GT)12	F: TCAGTGGTATTGAGACATCG R: CCATCTTCAAAAGTGAACCT	258	0.79	-	10
102	NCPGR256	(CA)12	F: AATGGGTTAATTTGACTTGC R: TTAATGGCCACCATAATCTT	280	2.25	-	2
103	NCPGR259	(GT)12	F: TATAGCCATAAGGGCAACAT R: TGTGGTAGAATGGGGAATAG	185	1.26	-	2
104	NCPGR266	(CA)12	F: TGTGAAAACTGATGAGGACA R: GTGTGTTGTCGTTTGTCTTG	195	12.5	******	3
105	NCPGR267	(TA)2(CA)13	F: ATTAACTGTGCTGGAGGAAA R: TATAGCCATAAGGGCAACAT	279	0.66	-	2
106	NCPGR268	(GT)11	F: TCAACTAAGGATTTGCTCG R: AGAGCTGAGAGAGTGGACAA	296	0.01	-	10
107	NCPGR272	(AT)4(GT)13	F: TGGACTAACAGCTTTCCATT R: GTCTTCTGTAGATTGAAGTTGTAAA	233	7.13	***	UL
108	NCPGR273	(CA)11	F: CCATCTTCAAAAGTGAACCT R: TCAGTGGTATTGAGACATCG	273	1.71	-	UL
109	NCPGR274	(GT)12	F: GTGTGTTGTCGTTTGTCTTG R: TTTTGAAGAGCAATCAATCC	268	0.33	-	3
110	NCPGR278	(GT)5g(GT)3gc (GT)2	F: TGAGACATCGACTATTGGACA R: GACCATCTTCAAAAGTGAACC	250	2.06	-	10
111	CaSSR2	$(GAAT)_4(GTA_2$	F: GCCTACATTGCTTTCCCTTT R: TCATGTGTGTATGAAGTGGAATGA		14.4	******	UL
112	CaSSR4	$(CAA)A(CAA)_4$ $G(CAA)_4$	F: ATAGTGGCATATTGGGGAGA R: TGAAACCCTAGAGTGGTTGTT		0.42	-	UL
113	CaSSR5	$(TTA)_8$	F: AAACCAAAACTGAAGTTAATAGGG R: GAAAGAAGTGAAAAAGTAGTGGAA		0.23	-	UL
114	TA14	$(TAA)_{22}$	F: TGACTTGCTATTTAGGGAACA R: TGGCTAAAGACAATTAAAGTT	250	1.76	-	2
115	TA21	$(TAA)_{51}$	F: GTACCTCGAAGATGTAGCCGATA R: TTTTCCATTTAGAGTAGGATCTTCTTG	347	2.65	-	10
116	TA 22	$(ATT)_{40}$	F: TCTCCAACCCTTTAGATTGA R: TCGTGTTTACTGAATGTGGA	228	0.73	-	2
117	TA28	$(TAA)_{37}CAA(TAA)_{30}$	F: TAATTGATCATACTCTCACTATCTGCC R: TGGGAATGAATATATTTTTGAAGTAAA	300	0.80	-	10
118	TA34	$(AAT)_{34}$	F: TGGGAATGAATATATTTTTGAAGTAAA R: CCATTATCATTCTTGTTTTCAA	230	14.7	******	3
119	TA46	$(TAA)_{22}$	F: TTTATTGCAATAAAACTCATTTCTTATC R: TTCTTTTTGTGTGAAAAAAAAATA TAGTGA	152	2.18	-	1
120	TA59	$(TAA)_{29}$	F: ATCTAAAGAGAAATCAAAATTGTCGAA R: GCAAATGTGAAGCATGTATAGATAAAG	258	3.77	*	1
121	TA72	$(ATT)_{36}$	F: GAAAGATTTAAAAGATTTTCCACGTTA R: TTAGAAGCATATTGTTGGGATAAGAGT	256	0.42	-	1
122	TA76	$(AAT)_7$	F: TCCTCTTCTTCGATATCATCA R: CCATTCTATCTTTGGTGCTT	206	0.04	-	3
123	TA78	$(TTA)_{30}$	F: CGGTAAATAAGTTTCCCTCC R: CATCGTGAATATTGAAGGGT	205	0.98	-	10
124	TA96	$(AT)_3(TTA)_{30}(AT)3$	F: TGTTTTGGAGAAGAGTGATTC R: TGTGCATGCAAATTCTTACT	275	5.76	**	1
125	TA106	$(TAA)_{26}$	F: CGGATGGACTCAACTTTATC R: TGTCTGCATGTTGATCTGTT	248	1.11	-	2
126	TA146	$(TTA)_{29}$	F:CTAAGTTTAATATGTTAGTCCTTAAATTA T R: ACGAACGCAACATTAATTTTATATT	161	0.40	-	1
127	TR2	$(TTA)_{36}$	F: GGCTTAGAGTTCAAAGAGAGAA R: AACCAAGATTGGAAGTTGTG	210	17.4	*******	3
128	TR19	$(TAA)_{27}$	F: TCAGTATCACGTGTAATTCGT R: CATGAACATCAAGTTCTCCA	227	2.53	-	1
129	TR43	$(TAA)_{24}$	F: AGGACGAAACTATTCAAGGTAAGTAGA R: AATTGAGATGGTATTAAATGGATAACG	297	0.04	-	7
130	TR44	$(TAT)_{16}$	F: TTAATATTCAAAAACTCTCTTGTGCAAT R: TTTACAACAGCGCTTGTATTTAGTAAG	289	0.01	-	2

131	TS35	$(TAA)_9T(A)_3(TAA)_{13}$	F: GGTCAACATGCATAAGTAATAGCAATA R: ACTTTCGCGATTCAGCTAAAATA	247	6.55	**	4
132	TS45	$(TAA)_8(A)_3(TAA)_{18}$	F: TGACACAAAATTGTCTCTTGT R: TGTTCTTAACGTAACTAACCTAA	244	0.04	-	6
133	TS72	$(ATT)_{39}$	F: CAAACAATCACTAAAAGTATTTGCTCT R: AAAAATTGATGGACAAGTGTTATTATG	264	1.14	-	1
134	GA4	$(GA)_9$	F: TTGCGTGTCAATCTCATTGG R: TCAACACCCCTAACTCGGAC	208	5.31	**	4
135	GA9	$(CT)_{17}$	F: GAACGGATTGGATGAAGCAT R: GTGCAAACAACCCTTTTTGG	200	0.11	-	2
136	GA21	$(CT)_{14}$	F: CCCCAGGTGAATTCCTCATA R: CTCAACCTTTGTTCAGCAACAC	238	0.01	-	2
137	GA26	$(CT)_{28}$	F: GATGCTCAAGACATCTGCCA R: TCATACTCAACAAATTCATTTCCC	234	0.60	-	2
138	GAA40	$(CTT)_9$	F: TTGACGCAGAGAACTCTCAA R: ATTGGTGTGATGGGTGGATT	245	1.22	-	8
139	GAA42	$(GAA)_8$	F: CGCTTCAGTGTAGATATTATTCAAACA R: TCTCTCTTTCTCTTCAACACGC	295	10.8	****	4
140	GAA47	$(GAA)_{11}$	F: CACTCCTCATGCCAACTCCT R: AAAATGGAATAGTCGTATGGGG	169	0.19	-	1
141	STMS2	$(TAT)_{25}$	F: ATTTTACTTTACTACTTTTTTCCTTTC R: AATAAATGGAGTGTAAATTTCATGTA	234	3.25	*	2
142	STMS5	$(GA)_{19}$	F:TACAAACTTTTAAGTTCATAAGTTTGA R: AACTTCTCGAATTAGTAAATTAAGTTG	235	0.19	-	3
143	STMS7	$(GA)_{12}$	F: GAGGATTCGGATTCAGAT R: AAAATCTTGGAAGTGATTGAG	161	8.38	****	4
144	STMS8	$(GT)_{10}$	F:GGACTAGAGGCAGAAGCT R: AGCATACAAATAAATAATAATGCATG	100	0.01	-	12
145	TAA60	$(CTT)_6$, $(CTT)_9$	F: TCATGCTTGTTGGTTAGCTAGAACAAA R: GACATAATCGAGTTAAAGAAAA	295	4.46	**	1
146	*Foc*0	-	-		0.68	-	UL
147	*Foc*4	-	-		6.86	***	1
148	*Foc*5	-	-		1.99	-	1

B) List of EST-based markers used in this study. The chi square value (χ^2), segregation distortion, map position and their putative functions based on BLASTX results are mentioned.

S.No.	Locus	$\chi 2$	Segregation distortion	Map position	Putative function
1	CESSR18	1.50	-	UL	Hypothetical protein
2	CESSR20	3.24	*	UL	Unknown protein
3	CESSR22	0.25	-	UL	RNA binding
4	CESSR23	0.40	-	UL	Glutathione peroxidase
5	CESSR24	0.32	-	2	No homology
6	CESSR28	0.32	-	UL	Hypothetical protein
7	CESSR42	040	-	11	Hsr2035 homolog
8	CESSR43	0.04	-	UL	Bimodular protein
9	CESSR46	3.37	*	1	No homology
10	CESSR47	3.83	*	UL	Germin like protein
11	CESSR60	0.01	-	UL	ANAC075 TF
12	CESSR61	2.43	-	UL	Armadillo like helical
13	CESSR62	0.07	-	UL	Receptor-like kinase
14	CESSR66	1.74	-	UL	Homodomain-related
15	CESSR71	0.01	-	2	Unknown protein
16	CESSR72	0.58	-	UL	Unknown protein
17	CESSR73	0.01	-	1	No homology
18	CESSR77	0.03	-	UL	DUF647
19	CESSR88	4.60	**	UL	Histone H2B
20	CESSR103	4.45	**	UL	Seed-alpha amylase
21	IICESSR19	0.14	-	4	DNA-binding
22	IECESSR52	0.01	-	5	Unknown protein
23	IICESSR69	0.14	-	UL	Hypothetical protein
24	CEST21	0.08	-	UL	B-D galactosidase
25	CEST25	0.66	-	2	Legumin
26	CEST27	0.09	-	11	Myoinositol phosphatase
27	CEST44	2.00	-	10	Annexin
28	CEST51	0.34	-	UL	Starch synthase
29	CEST62	0.14	-	UL	Developmental protein
30	CEST86	1.00	-	UL	Hypothetical protein
31	CEST93	1.00	-	UL	Hypotheical protein
32	CEST95	0.08	-	UL	No homolgy
33	CEST97	1.44	-	4	MtN4 protein
34	IICEST35A	0.07	-	5	Sugar transporter
35	IICEST35B	6.23	**	UL	Sugar transporter
36	IECEST46	1.33	-	UL	Disulfide isomerase
37	PIP6	0.03	-	UL	Calcineurin B-like protein

38	PIP12	0.83	-	1	Haem peroxidase
39	PIP23	1.11	-	UL	Glutamine synthetase
40	PIP26	1.40	-	UL	UBC9
41	PIP30	1.83	-	UL	Hypothetical protein
42	PIP35	0.00	-	5	Superoxide dismutase II
43	PIP36		-	5	-do-
44	PIP39	1.35	-	10	Proteosome related
45	PIP41	0.14	-	UL	Glyceraldehyde-3-phosphate
46	PIP44	0.03	-	UL	UBC19
47	PIP54	0.68	-	UL	Unknown protein
48	PIP55	0.32	-	UL	-do-
49	PIP65	0.98	-	UL	Chlorophyll binding protein
50	PIP69	1.37	-	UL	Integral membrane
51	PIP70	0.86	-	UL	Ceramidase family protein
52	PIP72	0.30	-	UL	Seed maturation
53	PIP75	0.85	-	5	Ubiquitin
54	PIP77	0.56	-	11	Prolyl-4-hydroxylase
55	PIP79	5.00	**	2	PPF-1 protein
56	PIP81	0.82	-	UL	Fructose-1, 6-bisphosphatase
57	PIP85	0.29	-	UL	Glycolate oxidase
58	PIP86	2.31	-	UL	-do-
59	PIP96	0.72	-	UL	Glycerol-3-phosphate acyltransferase
60	PIP107	0.42	-	UL	Vascular ATP synthase
61	MTEST235	053	-	8	-
62	MTEST279	0.12	-	UL	-

- (hyphen): indicates Mendelian inheritance (1:1),
*** (*asterisks*):** indicates the degree of distortion (* means low degree of distortion, ** increasing degree of distortion
UL: indicates unlinked markers.

Table 8.3: Distribution of 146 (118 genomic-derived STMS + 26 EST based* + 2 *Foc* loci) markers among the 12 linkage groups of the chickpea genome. The no. of loci, length, density and genome coverage of each linkage group is mentioned.

*: Represents EST based markers

Genome coverage= Maplength/{Maplength X [No. of loci +1/ No. of loci-1]}

Linkage group	No.of loci (EST-based*)	Length (cM)	Density (cM/locus)	Genome coverage (%)
1	34(3*)	237.5	6.98	94.28
2	30(4*)	140.7	4.69	93.55
3	20	119.8	5.99	90.48
4	19(2*)	135.6	7.13	90.00
5	0(6*)	94.0	15.6	71.42
6	7(1*)	81.4	11.6	75.02
7	5(1*)	60.3	12.0	66.66
8	4(1*)	59.7	14.9	59.40
9	0(3*)	62.1	20.7	50.00
10	11(2*)	95.9	8.71	82.81
11	0(3*)	98.5	32.8	50.00
12	4	25.0	6.25	60.00
Total	**146**	**1210.5cM**	**-**	**-**
Mean	**12.16**	**100.87**	**12.27**	**73.63**

Chapter 9: Summary and conclusions

Chickpea is a cool season grain legume crop mainly grown under rain-fed conditions in arid and semi-arid areas around the world, however, the major growing areas include the Indian sub-continent. The short life-cycle of 3-4 months, small genome size (750 Mb) and high economic importance as a food crop makes chickpea an important system for genomics research. Although considerable research efforts have been made worldwide for crop improvement, the impact on chickpea production is marginal. Currently, the productivity of chickpea is very low (world average 0.8t/ha, FAOSTAT, 2005) and has been stagnant for the last few years. The reasons for only marginal improvements are a series of abiotic stresses like drought, salinity, cold and biotic stresses like fusarium wilt, ascochyta blight and pod borer. Therefore, chickpea breeders focus on increasing the yield by pyramiding the genes for resistance/tolerance into agronomically superior varieties through integration of advanced technologies like marker assisted breeding along with conventional approaches. Recently, several genomic tools like DNA based molecular markers, linkage maps, BAC libraries and ESTs have been developed for chickpea. Nevertheless, the progress achieved in chickpea molecular genetics is still lagging far behind other crops for use in enhancement of crop characteristics and for developing elite chickpea germplasm.

Among the popular DNA-based molecular markers available for use, microsatellite-based markers have emerged as the best bet for detecting genetic variation in chickpea and have been implicated for diversity analysis, germplasm characterization, elucidating *Cicer* phylogeny, construction of linkage maps and transferability studies. However, the high developmental costs, species-specificity and their association mostly with non-coding regions have limited the applicability of aforementioned markers referred to as 'random' or 'anonymous' markers for direct tagging of genes, offsetting the gene introgession programs and comparative genomic studies. Therefore, during the past few years, research has shifted towards the generation of functional molecular markers (FMs) instead of anonymous markers by virtue of their association with the transcribed portion of the genome. In this regard, the growing EST datasets of several organisms in conjunction with bioinformatics tools have emerged as a potential source for generation of different kinds of functional molecular markers such as EST-SSRs, EST-SNPs, ESTPs (Expressed Sequence Tag Polymorphisms), COS (Conserved Orthologous Sites) and ITPs (Intron-Targeted Primers). Although known to be less polymorphic compared to anonymous markers, these markers hold immense potential (by virtue of their being associated with the coding region of the genome) to add a powerful new dimension to the understanding and improvement of crop gene pools.

Despite chickpea being an important pulse crop, limited EST resources are publicly available. Uptill now, no major efforts have been undertaken to develop functional markers and utilize them for molecular breeding applications. Therefore in this study, attempts have been made to expand the EST database of chickpea through the construction of cDNA library from developing seeds of chickpea. Moreover, despite the considerable economic importance of chickpea, no earlier attempt had been made to capture the transcriptome associated with stages of seed development. Hence the chickpea EST resources generated here would serve multiple functions – firstly they provide an opportunity for the functional dissection of gene expression during seed development and secondly aid in generation of EST based molecular markers for mapping seed related traits. Therefore, using the bioinformatics tools, the generated chickpea ESTs were assembled and functionally annotated and were also systematically explored for the development and characterization of different types of chickpea functional molecular markers such as EST-SSRs, ESTPs (Expressed Sequence Tag Polymorphisms) and ITPs (intron-targeted primers) which were then utilized for genetic diversity analysis, cross-transferability across related species and genera and in construction of a genetic linkage map. The functional markers developed in the present study would therefore aid in accelerating the chickpea molecular breeding programs.

The results obtained in the present thesis are summarized below:

In the present study, a cDNA library was constructed from 20 DAA developing seeds of chickpea. Large-scale sequencing yielded 1897 ESTs from which 1037 unigenes were identified with overall redundancy of 61.5% obtained using CAP3 program. BLASTX analysis revealed that 58.6% of them had significant homology to previously identified genes whereas approx. 20.0% didn't reveal any homology substantiating the fact that these sequences perform functions that may be of special relevance to developing seeds or may represent the chickpea specific transcriptome. It was observed that the highly abundant ESTs assembled in the contigs comprising of >10 ESTs were those of putative lipid transfer proteins, proteinase inhibitors, seed-specific proteins, Chlorophyll-a/b binding proteins, MAPK, serine carboxypeptidase, photosystem II reaction centre, and broadly represented the degree of expression of the respective genes in developing seeds. Further, the chickpea unigenes were functionally annotated against both KOG (Clusters of Eukaryotic Orthologous groups of proteins) database and Gene Ontology (GO) consortium. Northern analysis of five EST sequences coding for putative functions namely oleosin, conglutin-delta, pectinesterase,

heat-shock binding protein and seed-specific clone revealed that the first two unigenes are expressed at later stages of seed development i.e. at 35-40 DAA whereas the last three are expressed throughout the seed developmental stages in chickpea. Thus the chickpea ESTs generated in this study provides an opportunity in future to analyze a large number of seed related unigenes for in-depth understanding of molecular processes or mechanisms involved during seed development.

For the development of chickpea EST-SSR markers, a total of 2346 chickpea EST sequences (1309 from database + 1037 from inhouse developed ESTs as mentioned above) were employed for the identification of microsatellite motifs. 284 (13.8%) EST sequences were found to contain 324 repeat motifs that mainly comprised of (51.5%) trinucleotide repeats followed by (38.8%) dinucleotides, tetra- (4.9%) and pentanucleotide (4.6%) motifs. Among trinucleotide motifs, AAG (36.0%) was predominant followed by AAT (14.0%) whereas among dinucleotide motifs, GA (75.5%) followed by AT (15.3%) was abundant. Based on the structural organization of repeat motifs, 254 (78.3%) repeats were found to be perfect, 48 (14.8%) were imperfect and 22 (6.7%) were compound. This study for the first time provides an insight into the distribution and composition of different types of SSR motifs in the chickpea transcribed regions. Furthe,r from the 284 microsatellite containing EST sequences (SSR-ESTs) identified, a total of 135 EST-SSR (eSSRs) primers were designed in the present study and of these only 97 markers could be validated for further use as they amplified expected size bands.

To determine the potentiality of the developed chickpea EST-SSR markers for analysis of genetic diversity, a set of sixty chickpea EST-SSR primers were used to amplify genomic DNA of 30 chickpea cultivars for polymorphism analysis. Of these, only 10 markers produced polymorphism across the 30 chickpea cultivars amplifying a total of 129 alleles with an average of 2.7 alleles per locus. The observed heterozygosity and expected heterozygosity values averaged to 0.16 and 0.56. Although these markers displayed a low level of polymorphism (16.0%) compared to earlier reports of 40-50% polymorphism detected by gSSRs (genomic SSRs), the former are preferred owing to their association with coding regions and therefore represent "true genetic diversity". Additionally, the same set of 60 chickpea EST-SSR markers were also assayed for inter-specific transferability studies across six wild, annual *Cicer* species representing the members of first and second crossability group of genus *Cicer*. The transferability rates of chickpea EST-SSR markers

varied from a high of 96.6% in *C. reticulatum* to a low of 68.3% in *C. judaicum* with an average of 82.6% thereby establishing that EST based microsatellite markers of chickpea were not only efficient for marker-assisted introgression programs using wild germplasm but also reliable for synteny studies within the genus *Cicer*. Moreover, these genic markers displayed significantly higher level of polymorphism in the wild relatives of chickpea compared to chickpea accessions and thus could potentially facilitate the transfer of traits of agronomic value into cultivated chickpea thereby leading to the broadening of the narrow genetic base and development of superior genotypes of chickpea. The dendrogram obtained using NTSYSpc software clearly distinguished the chickpea accessions, separating the members of first and second- crossability group and showed the closeness of *C. judaicum* with *C. pinnatifidum* which was in agreement to the earlier protein and EST-based studies carried out in chickpea.

The molecular basis of length variation obtained across chickpea cultivars and wild *Cicer* species was also investigated in the present study. Sequence data demonstrated that in general, limited sequence variability was present within the chickpea alleles in comparison to much higher levels of variation across the orthologous alleles from the wild annual *Cicer* species. Within chickpea accessions, repeat number variation and few isolated point mutations in the MFR were the reasons for allele size differences suggesting the presence of evolutionary constraints within transcribed regions that limit the mutational events and increase sequence similarity. However in the wild species, allelic length variations occurred mainly due to differences in the copy number of repeat motifs and repeat interruptions accompanied by indels and point mutations in the microsatellite flanking regions (MFR). Further, the present study revealed the interesting feature of crossability-group-specific point mutations and indels across annual *Cicer* species that proved to be phylogenetically highly informative in understanding the evolution of microsatellites in a phylogenetic context since it has been shown that such events at the genic loci might play an important role in speciation or gene functionality diversification during the evolutionary process.

Cross-genera transferability was also investigated in the present study using thirty-four chickpea genic-SSR markers (EST-SSRs) across 32 accessions spanning eight legume genera. The markers successfully cross- amplified across the legumes with an average of 43.6% (ranging from 29.4% in *P. mungo* to 61.7% in *M. truncatula*). The study demonstrated that the rate of transferability decreases from within the genus *Cicer* (82.6%) to outside the

genus (43.6%) suggesting that amplification decreases with increasing evolutionary distance from the focal species. Sequencing of the amplified alleles at two loci across studied legumes confirmed the conservation of primer binding sites and moreover showed that the above mentioned factors were responsible for allele size differences.

In order to maximally exploit the available ESTs and to generate the maximum number of chickpea functional molecular markers that could efficiently detect DNA polymorphism in chickpea, these EST sequences were maximally utilized for development of other kinds of PCR-based markers like ESTPs and ITPs. A total of 80 ESTP primers were designed from chickpea seed related unigenes of which 58 produced expected size fragments. Further, using the program PIP (Potential Intron Polymorphism), a total of 110 intron-targeted primers designated as 'PIP' were designed from 1307 chickpea ESTs (1037 inhouse ESTs + 270 other ESTs obtained from Chattopdhyay et al. (pers. comm.). Amplifications of these primers were carried out in chickpea cultivar ICCV2 that yielded 76 functional primers producing alleles larger than expected (>100-120bp) that predictably contained introns. Hence an appreciable number (a total of 231) of new chickpea functional molecular markers including EST-SSRs (97), ESTPs (58) and PIPs (76) were made available in the present study for utilization in assessment of genetic diversity, cross-transferability and linkage map construction.

The developed chickpea ESTP and PIP functional molecular markers were screened to identify polymorphic markers between '*C. arietinum* ICC4958 x *C. reticulatum* PI489777 (considered as chickpea reference mapping population), the parental lines of the inter-specific RIL mapping population used in this study. Of the 58 ESTP and 76 PIP primers analyzed, 34 (10 + 24) were polymorphic in this population. On unraveling the molecular basis of polymorphism, it was found that the indels (insertion and deletions) in the intronic regions were the major factors responsible for allelic polymorphism.

One of the major objectives for which the molecular markers were generated was to construct the chickpea linkage map. Towards this, the parents of the mapping population were screened for polymorphism using 369 chickpea STMS markers that included 272 gSSRs (developed earlier in our laboratory) and 97eSSRs developed in this study. A total of 133 (113+20) polymorphic markers were identified and a high level of polymorphism was achieved with chickpea genomic derived STMS markers (41.5%) compared to EST-SSR markers (21.0%). In addition, the 15 *Medicago* EST-SSR primers (reported by Gutierrez et

al. 2005) were employed for polymorphism analysis of which only two primers produced size variant alleles. All the polymorphic 135 STMS (133 chickpea + 2 *Medicago*), 10 ESTPs and 24 PIPs (a total of 169) were genotyped in the 129 RILs of the mapping population and utilized for map construction. For anchoring purpose, segregating data of 32 previously mapped genomic STMS markers and of loci for resistance to fusarium races i.e. *Foc*0, *Foc*4 and *Foc*5 were utilized. Moreover, six chickpea EST-based primers i.e. CESSR19, CESSR52, CESSR69, CEST35 and CEST46 although amplified large sized products but produced polymorphic pattern between the mapping parents. So to maximally exploit the chickpea available markers, four additional new intron-exon based and two intron-intron based primers were designed from the obtained genomic sequences of these primers and were utilized for segregation analysis. Thus a total of 210 (169 + 32 + 3 *Foc* loci + 6) polymorphic markers were used for genotyping the 129 RILs for construction of a genetic linkage map.

Of these 210 polymorphic markers, 67 loci (31.9%) showed segregation distortion of which 30 loci (39.59%) exhibited significantly high segregation distortion. Interestingly, the majority of markers 53 (79.1%) skewed towards the wild annual parent i.e. *C. reticulatum* (female) and moreover clustering of distorted markers was observed. The map constructed at the LOD 3 value, positioned a total of 146 markers that included 112 new molecular markers (86 gSSRs, 10 EST-SSRs, 7 ESTPs, 8 PIPs and one *Medicago* EST-SSR) and 34 reported markers. The map spanned 1210.5cM of the chickpea genome at an average marker density of 8.64cM. These markers exhibited a nonrandom distribution varying in density from 4.69cM/locus to 32.8cM/locus with an average of 12.27cM/locus. The genome coverage averaged to 73.63% varying from as low as 50.0% (LG9 and LG11) to 94.2% (LG1). The genome size of *C. arietinum* is deciphered to be 750 Mbp, and hence in the present map an average physical equivalent of 1cM would correspond to 619Kbp of *C. arietinum* genome. Moreover, of the 26 mapped EST-based markers, 12 of them formed isolated blocks in three linkage groups whereas 14 markers mapped in a backdrop of anonymous markers to 7 linkage groups.

In conclusion, the study has underpinned the chickpea genomic resources especially ESTs by providing novel 1037 unigenes that would aid in expediting the functional genomic studies and in understanding the complex agronomic traits that affect the productivity and quality of chickpea. A total of 231 new different types of EST-based markers such as EST-SSRs, ESTPs and PIPs were developed that proved to be highly informative for varied

applications spanning analysis of genetic diversity, across genera transferability studies to construction of a linkage map based on functional markers. The present linkage map constructed using these newly developed markers defines the positions of 112 new molecular loci, which will serve as a valuable resource for targeted marker saturation and identification of candidiate genes at agronomically important loci. The gene based map will facilitate the development of a high resolution genetic map of chickpea thereby accelerating map based cloning and genomic-assisted breeding programs for ultimately providing economic benefits to the global producers and consumers of chickpea.

Chapter 10: References

References

Abbo S, Berger J, Turner NC (2003) Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. Funct Plant Biol 30:1081-1087

Abbo S, Molina C, Jungmann R, Grusak M, Berkovitch Z, Reifen R, Kahl G, Winter P, Reifen R (2005) Quantitative trait loci governing carotenoid concentration and weight in seeds of chickpea (*Cicer arietinum* L.). Theor Appl Genet 111:185–195

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao CR. Merril A, Wu B, Olde RF, Moreno AR, Kerlavage WR, McConbie and Venter JC (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor Appl Genet 114:359-372

Agharkar SP (1991) Medicinal Plants of Bombay Presidency, pp 62–63. Scientific Publications, Jodhpur, India

Ahmad F (1999) Random Amplified Polymorphic DNA (RAPD) analysis reveals genetic relationships among annual *Cicer* species. Theor Appl Genet 98:657-663

Ahmad F (2000) A comparative study of chromosome morphology among the nine annual species of *Cicer* L. Cytobios 101:37-53

Ahmad F, Gaur PM, Slinkard AE (1992) Isozyme polymorphism and phylogenetic interpretations in the genus *Cicer* L. Theor Appl Genet 83:620-627

Ahmad F and Slinkard AE (1992) Genetic relationships in the genus *Cicer* L. as revealed by polyacrylamide gel electrophoresis of seed storage proteins. Theor Appl Genet 84:688-692

Akagi H, Yokozeki Y, Inagaki A, Fujimura T (1996) Microsatellite DNA markers for rice chromosomes. Theor Appl Genet 93:1071-1077

Alba MM, Santiba'n˜ez-Koref MF, Hancock. JM (1999) Amino acid reiterations in yeast are overrepresented in particular classes of proteins and show evidence of a slippage-like mutational process. J Mol Evol 49:789–797

Alghanim HJ and Almirall JR (2003) Development of microsatellite markers in *Cannabis sativa* for DNA typing and genetic relatedness analyses. Anal Bioanal Chem 376:1225-33

Amos W, Sawcer SJ, Feakes W, Rubinszstein DC (1996) Microsatellite show mutational bias and heterozygote instability. Nat Genet 13:390–391

Amos W and Rubinstein DC (1996) Microsatellites are subject to directional evolution. Nature Genetics 12:13-14

Andersen JR and Lübberstedt (2003) Functional markers in plants. Trends Plant Sci 8:554–559

Arens P, Odinot P, van Heusden AW, Lindhout P, Vosman B (1995) GATA and GACA repeats are not evenly distributed throughout the tomato genome. Genome 38:84–90

Areshchenkova T and Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229-235

Arnold C, Rossetto M, Mcnally J, Henry RJ (2002) The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in *Vitaceae*. Am J Bot 89:22-28

Arumuganathan K and Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Rep 9:208–218

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski[3] K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene Ontology: tool for the unification of biology. Nature Genetics 25: 25–29

Aubert G, Morin EJ, Jacquin EF, Loridon EK, Quillet MC, Petit EA, Rameau EC, Lejeune-He´naut I, Huguet ET, Burstin EJ (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112:1024-1041

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, Jones CA, Park WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781

Banerjee H, Pai RA, Sharma RP (1999) Restriction fragment length polymorphism and random polymorphic DNA analysis of chickpea accessions. Biologia Plantarum 42:197-208

Bao S, JS, Corke H, Sun M (2002) Microsatellites in starch-synthesizing genes in relation to starch physicochemical properties in waxy rice (*Oryza sativa* L.). Theor Appl Genet 105: 898–905

Barrett B, Griffiths A, Schreiber M, Ellison N, Mercer C, Bouton J, Ong B, Forster J, Sawbridge T, Spangenberg G, Bryan G, Woodfield D (2004) A microsatellite map of white clover. Theor Appl Genet 109:596–608

Batzoglou S, Pachter L, Mesirov JP, Berger B, Lander ES (2000) Human and mouse gene structure: comparative analysis and application to exon prediction. Genome Res 10:950–958

Beavis WD and Grant D (1991) A linkage map based on information from four FP populations of maize (*Zea mays* L.). Theor Appl Genet 82:636-644

Becker J and Heun M (1995a) Barley microstellites: allele variation and mapping. Plant Molecular Biology 27:835–845

Beckmann JS and Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627–631

Benko-Iseppon AM, Winter P, Huettel B, Staginnus C, Muehlbauer FJ, Kahl G (2003) Molecular markers closely linked to fusarium resistance genes in chickpea show significant alignments to pathogenesis-related genes located on *Arabidopsis* chromosomes 1 and 5. Theor Appl Genet 42:379-386

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251–269

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G. Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY (2004) Functional annotation of the *Arabidopsis* genome using controlled vocabularies. Plant Physiol 135:745–755

Bhattramakki D, Ching A, Morgante M, Dolan M, Register J, Smith H, Tingey S, Rafalski AJ (2002) Insertion-deletion polymorphism in 3′ regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48:539–547

Bierne N, Lehnert SA, Be'dier E, Bonhomme F, Moore SS (2000) Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. Mol Ecol 9:233-235

Billotte N, Lagoda PJL, Risterucci AM, Baurens FC (1999) Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. Fruits 54:277–288

Boominathan P, Shukla R, Kumar A, Manna D, Negi D, Verma PK, Chattopadhyay D (2004) Long term transcript accumulation during the development of dehydration adaptation in *Cicer arietinum*. Plant Physiol 135:1608-1620

Brondani RPV, Williams ER, Brondanil C, Grattapaglia D (2006) A microsatellite-based consensus linkage map for species of *Eucalyptus* and a novel set of 230 microsatellite markers for the genus. BMC Plant Biology 6:20

Broun P and Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250:39–49

Brown GR, Kadel III EE, Bassoni DL, Kiehne KL, Temesgen B, Buijtenen van JP, Sewell MM, Marshall KA Neale DB (2001) Anchored reference loci in loblolly Pine (*Pinus taeda* L.) for Integrating Pine Genomics. Genetics 159:799–809

Bruford MW, Cheesman DJ, Coote T, Green HAA, Haines SA, O'Ryan C, William TR (1996) Microsatellites and their application to conservation genetics. In: Smith TB, Wayne RK (eds) Molecular genetic approaches in conservation. Oxford University Press, New York, pp. 278–297

Brunel D, Froger N, Pelletier (1999) Development of amplified consensus genetic markers (ACGM) in *Brassica napus* from *Arabidopsis thaliana* sequence of known biological function. Genome 42:387–402

Bryan GJ, Collins AJ, Stephenson P, Orry A, Smith JB, Gale MD (1997) Isolation and characterisation of microsatellites from hexaploid bread wheat. Theor Appl Genet 94:557–563

Buhariwalla HK, Jayashree B, Eshwar K, Crouch JH (2005) Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biology 5:16

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Wagh R (2000) Computational & experimental characterization of physically clustered SSR in plants. Genetics 156:847-854

Castelo AT, Martins W, Gao GR (2002) TROLL-tandem repeat occurrence locator. Bioinformatics 18:634–636

Cato SA, Gardner RC, Kent J, Richardson TE (2001) A rapid PCR-based method for genetically mapping ESTs. Theor Appl Genet 102:296–306

Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. Genet Res Crop Evol 52:903–909

Chagne D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT, Vendramin GG, Garcia V, Frigerio JMM, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109:1204–14

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-,tri-, and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

Chakraborty R and Nel M (1982) Genetic differentiation of quantitative characters between populations of species: I. Mutation and random genetic drift. Genet Res Camb 39:303–314

Chambers GK and MacAvoy ES (2000) Microsatellites: consensus and controversy. Comparative Biochem Physiol (Part B) 126:455–476

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr. (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248-57

Chetelat RT, DeVerna JW, Bennett AB (1995) Introgression into tomato (*Lycopersicon esculentum*) of the *L. chmielewskii* sucrose accumulator gene (*sucr*) controlling fruit sugar composition. Theor Appl Genet 91:327–333

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866-873

Cho S, Chen W, Muehlbauer F (2004) Pathotype-specific genetic factors in chickpea (*Cicer arietinum* L.) for quantitative resistance to ascochyta blight. Theor Appl Genet 109:733–739

Cho S, Kumar J, Shultz J, Anupama K, Tefera F, Muehlbauer F (2002) Mapping genes for double podding and other morphological traits in chickpea. Euphytica 128:285–292

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713-722

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004a) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289-15294

Choi HK, Kim D, Uhm T,†Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004b) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166:1463–1502

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2006) Development of sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Molecular Ecology Notes 6(1):93-95

Choumane W, Winter P, Baum M, Kahl G (2004) Conservation of microsatellite flanking sequences in different taxa of Leguminosae. Euphytica 138:239–245

Choumane W, Winter P, Weigand F, Kahl G (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer arietinum* L.) within the genus *Cicer*. Theor Appl Genet 101:269-278

Chowdhury MA, Vandenberg V, Warkentin T (2002) Cultivar identification and genetic relationship among selected breeding lines and cultivars in chickpea (*Cicer arietinum* L.). Euphytica 127:317–325

Clarke, H and Siddique KHM (2003) Chilling tolerance in chickpea– Novel methods for crop improvement. In: R.N. Sharma, M. Yasin, S.L. Swami, M.A. Khan & A.J. William (Eds.), International Chickpea Conference, Indira Ghandhi Agricultural University, Raipur, India, pp. 5–12

Condit R and Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34: 66-71

Co^rte-Real HBSM, Dixon DR, Holland PWH (1994) Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. Mar Biol 120:407-413

Cobos M, Fernandez M, Rubio J, Kharrat M, Moreno M, Gil T, Millan T (2005) A linkage map of chickpea (*Cicer arietinum* L.) based on populations from Kabuli and Desi crosses: location of genes for resistance to fusarium wilt race 0. Theor Appl Genet 110:1347–1353

Cobos MJ, Rubio J, Strange RN, Moreno MT, Gil J, Millan T (2006) A new QTL for Ascochyta blight resistance in an RIL population derived from an interspecific cross in chickpea. Euphytica 149:105-111

Collard BCY, Pang ECK, Ades PK, Taylor PWJ (2003) Preliminary investigation of QTLs associated with seedling resistance to ascochyta blight from *Cicer echinospermum*, a wild relative of chickpea. Theor Appl Genet 107:719–729

Coram T and Pang E (2005) Isolation and analysis of candidate ascochyta blight defense genes in chickpea, Part I. Generation and analysis of an expressed sequence tag (EST) library. Physiological and Molecular Plant Pathology 66:192-200

Coram T and Pang E (2007) Transcriptional profiling of chickpea genes differentially regulated by salicylic acid, methyl jasmonate, and aminocyclopropane carboxylic acid to reveal pathways of defence related gene regulation. Functional Plant Biology 34:52–64

Cordeiro GM and Casu R, McIntyr CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Science 160:1115-1123

Crawford A, Knappes SM, Paterson KA, deGotari MJ, Dodds KG, Freking RT, Stone RT, Beattie CW (1998) Microsatellite evolution: testing the ascertainment bias hypothesis. J Mol Evol 46:256-260

Croser JS, Ahmad F, Clarke HJ, Siddique KHM (2003) Utilisation of wild *Cicer* in chickpea improvement - progress, constraints and prospects. Aust J Agric Res 54:429–444

Cullis CA (2002) The use of DNA polymorphisms in genetic mapping. Genet Eng. (N Y) 24:179-89

Daguin C and Borsa P (1999) Genetic characterization of *Mytilus galloprovincialis* Lmk. In North West Africa using nuclear DNA markers. J Exp Mar Biol Ecol 235:55-65.

Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912-922

Deka R, Chakraborty R, Ferrell RE (1991) A population genetic study of six VNTR loci in three ethnically defined populations. Genomics 11:83–92

Deutsch M and Long M (1999) Intron–exon structures of eukaryotic model organisms. Nucleic Acids Research 27:3219–3228

Devos KM and Gale MD (2000) Genome Relationships: The Grass Model in Current Research Plant Cell 12:637-646

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166-3170

Dodgson JB, Cheng HH, Okimoto R (1997) DNA marker technology: A revolution in animal genetics. Poultry Sci 76:1108-1114

Dong J, Keller WA, Yan W, Georges F (2004) Gene expression at early stages of *Brassica napus* seed development as revealed by transcript profiling of seed-abundant cDNAs. Planta 218:483– 491

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

Duke JA 1(981) Handbook of legumes of world economic importance. Plenum Press, New York. pg 52-57

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758

Elsik CG, Williams CG (2001) Families of clustered microsatellites in a conifer genome. MGG 265:535-542

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution- a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol Bio Evol 14:854-860

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross species genetic markers for *Medicago* spp. Theor Appl Genet 108:414-422

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399-407

Feltus FA, Singh HP , Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol140:1183–1191

Firnhaber C, Pühler A, Küster H (2005) EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with developmental expression regulation in flowers and pods. *Planta* 222:269-283

Flandez-Galvez H, Ford R, Pang E, Taylor P (2003a) An intraspecific linkage map of chickpea (*Cicer arietinum* L.) genome based on sequence tagged microsatellite site and resistance gene analog markers. Theor Appl Genet 106:1447–1456

Flandez-Galvez H, Ades PK, Ford R, Pang ECK, Taylor PWJ (2003b) QTL analysis for ascochyta blight resistance in an intraspecific population of chickpea (*Cicer arietinum* L.). Theor Appl Genet 107:1257–1265

Fraser LG, McNeilage EMA, Tsang EGK, Harvey CF, De Silva EHN (2005) Cross-species amplification of microsatellite loci within the dioecious, polyploid genus *Actinidia* (*Actinidiaceae*) Theor Appl Genet 112:149-157

Fraser LG, Harvey CF, Crowhurst RN, De Silva HN (2004) EST derived microsatellites from *Actinidia* species and their potential for mapping. Theor Appl Genet 108:1010–1016

Fredslund J, Madsen LH, Hougaard BK, Nielsen AM, Bertioli D, Sandal N, Stougaard J, Schauser L (2006) GeMprospector–online design of crossspecies genetic marker candidates in legumes and grasses. Nucleic Acids Research 34.

Fujimori S, Washio T, Higo K, Ohtomo Y, Murakami K, Matsubara K, Kawai J, Carninci P, Hayashizaki Y, Kikuchi S, Tomita M (2003) A novel feature of microsatellites in plants: a distribution gradient along the direction of transcription. FEBS Letters 554:17-22

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457–1467

Gale MD and Devos KM (1998) Plant comparative genetics after 10 years. Science 282:656–659

Gallardo K, Firnhaber C, Zuber H, Héricher D, Belghazi M, Henry C, Küster H, Thompson R (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: Evidence for metabolic specialization of maternal and filial tissues. Molecular & Cellular Proteomics 6:2165-21

Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. Plant Physiol 133:664–682

Gao LF, Jing RL, Huo NX, Li Y, Li XP, Zhou RH, Chang XP, Tang JF, Ma ZY, Jia JZ (2004) One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theor Appl Genet 108:1392–1400

Gastier JM, Pulido JC, Sunden S , Brody T, Buetow KH, Murray JC, Weber JL, Hudson TJ, Sheffield VC, Duyk M (1995) Survey of trinucleotide repeats in the human genome: Assessment of their utility as genetic markers. Hum Mol Gen 4:1829-1836

Gaur PM and Slinkard AE (1990a) Inheritance and linkage of isozyme coding genes in chickpea. J Hered 81:455-461

Gaur PM and Slinkard AE (1990b) Genetic control and linkage relations of additional isozyme markers in chickpea. Theor Appl Genet 80:648–656

Gil J and Cubero JI (1993) Inheritance of seed coat thickness in chickpea (*Cicer arietinum* L.) and its evolutionary implications. Plant Breed 111:257-260

Gonzalo MJ, Oliver EM, Garcia-Mas EJ, Monfort EA, Dolcet-Sanjuan R, Katzir EN, Aru′ s EP Monforte EAJ (2005) Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.) Theor Appl Genet 110:802–811

Gordon AJE (1997) Microsatellite birth register. J Mol Evol 45:337-338

Gortner G, Nerino M, Weising K, Zink D, Nagl W, Kahl G (1998) Chromosomal localization and distribution of simple sequence repeats and the *Arabidopsis*-type telomere sequence in the genome of *Cicer arietinum* L. Chrom Res 6:97–104

Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in *Gossypium*. Genetics 176:527-541

Gupta PK and Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica 113:163–185

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4:139–162

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genomics 270:315–323

Gupta PK, Varshney RK, Prasad M (2002) Molecular markers: principles and methodology. In: Jain SM, Ahloowalia BS, Brar DS, (eds) Molecular techniques in crop improvement. Kluwer Academic Publishers, Netherlands pp 9-54

Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 70:45–54

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT, Torres AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theor Appl Genet 110:1210-1217

Habu Y, Fukada-Tanaka S, Hisatomi Y, Iida S (1997) Amplified restriction fragment length polymorphism-based mRNA fingerprinting using a single restriction enzyme that recognizes a 4-bp sequence. Biochem Biophys Res Commun 234:516–521

Hammond RL, Saccheri IJ, Ciofi C, Coote T, Funk SM, McMillan WO, Bayes MK, Taylor E, Bruford MW (1998) Isolation of microsatellite markers in animals. In: Karp A, Issac PG, Ingram DS (eds) Molecular tools for screening biodiversity. Chapman and Hall, London, pp. 279–287.

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Gen Genomics 272:308–327

Hancock JM (1996) Simple sequences and the expanding genome. Bioassays 18:421-425

Hanhan D (1983) Studies on transformation of *Escherichia coli* with plasmid. J Mol Biol 166: 557–579

Harr B and Schlotterer C (2000) Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. Genetics 155:1213–1220

Harry DE and Temesgen B, Neale DB (1998) Codominant PCR-based markers for *Pinus taeda* developed from mapped cDNA clones Theor Appl Genet 97:327-336

Haware MP and Nene YL (1982) Races of *Fusarium oxysporum* f. sp. *ciceris.* Plant Dis 66:809–810

Hawtin GC and Singh KB (1980) Kabuli-Desi introgression: problems and prospects. In: Green JM, Nene YL, Smithson JB (eds) Proc Int Workshop on Chickpea Improvement, ICRISAT, Hyderabad, India, pp 51–60

Hearne CM, Ghosh S, Todd JA (1992) Microsatellites for linkage analysis of genetic traits. Trends Genet 8:288–294

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761–769

Höfte H, Desprez T, Amselem J, Chiapello H, Caboche M, Moisan A, Jourjon MF, Charpenteau JL, Berthomieu P, Guerrier D, Giraudat J, Quigley F, Thomas F, Yu DY, Mache R, Raynal M, Cooke R, Grellet F, Delseny M, Parmentier Y, Marcillac G, Gigot C, Fleck J, Philipps G, Axelos M, Bardet C, Tremousaygue D, Lescure B (1993) An inventory of 1152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana.* Plant J. 4:1051-1061

Holland JB, Helland SJ, Sharopova N, Rhyne DC (2001) Polymorphism of PCR-based markers targeting exons, introns, promoter regions and SSRs in maize and introns and repeat sequences in oat. Genome 44:1065–1076

Hongtrakul V, Slabaugh MB, Knapp SJ (1998) DFLP, SSCP, and SSR markers for delta 9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines. Mol Breed 4:195–203

Huang X and Madan A (1999) CAP3: A DNA Sequence Assembly Program. Genome Res. 9:868-877

Hughes AL, Yeager M (1997) Comparative evolutionary rates of introns and exons in murine rodents. J Mol Evol 45:125–130

Hulse JH (1991) Nature, composition and utilization of grain legumes. p. 11-27. In: Uses of tropical Legumes: Proceedings of a Consultants Meeting, 27-30 March 1989, ICRISAT Center. ICRISAT, Patancheru, A.P. 502 324, India

Hüttel B, Santra D, Muhelbauer FJ, Kahl G (2002) Resistance gene analogues of chickpea (*Cicer arietinum* L.): isolation, genetic mapping and association with *Fusarium* resistance gene cluster. Theor Appl Genet 105:479-490

Hüttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite markers for chickpea (*Cicer arietinum* L.). Genome 42:210-217

Iruela M, Rubio J, Barro F, Cubero JI, Millan T, Gil J (2006) Detection of two quantitative trait loci for resistance to ascochyta blight in an intra-specific cross of chickpea (*Cicer arietinum* L.): development of SCAR markers associated with resistance. Theor Appl Genet 112:278–287

Iruela M, Rubio J, Cubero JI, Gil J, Millan T (2002) Phylogenetic analysis in the genus *Cicer* and cultivated chickpea using RAPD and ISSR markers. Theor Appl Genet 104:643-651

Ishikawa G, Yonemaru J, Saito M, Nakamura T (2007) PCR-based landmark unique gene (PLUG) markers effectively assign homoeologous wheat genes to A, B and D genomes. BMC Genomics 8:135

Jaiswal P, Ni J, Yap I, Ware D, Spooner W, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K, Faga B, Canaran P, Fogleman M, Hebbard C, Avraham S, Schmidt S, Casstevens TM, Buckler ES, Stein L, McCouch S (2006) Gramene: a bird's eye view of cereal genomes. Nucleic Acids Research 34:717–723

Jana S and Singh KB (1993) Evidence on geographical divergence in Kabuli chickpea from germplasm evaluation data. Crop Sci 33:626–632

Jarne P and Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424-429

Jayashree B, Punna R, Prasad P, Bantte K, Hash CT, Chandra S, Hoisington DA, Varshney RK (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined *in silico*: survey and evaluation. *In Silico* Biology 6:0054

Javadi F and Yamaguchi H (2004) Interspecific relationships of the genus *Cicer* L. (*Fabaceae*) based on trnT-F sequences. Theor Appl Genet 109:317–322

Jermstad KD, Bassoni KL, Wheeler NC, Neale DB (1998) A sex-averaged linkage map in coastal Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) based on RFLP and RAPD markers. Theor Appl Genet 97:762–770

Jimenez-Dıaz RM, Alcala-Jimenez AR, Hervas A, Trapero-Casas JL (1993) Pathogenic variability and host resistance in the *Fusarium oxysporum* f. sp. *ciceris/C. arietinum* pathosystem. In: Proceedings of the 3rd European Seminar on Fusarium Mycotoxins, Taxonomy, Pathogenicity and Host Resistance, Plant Breeding and Acclimatization, Inst, Radzikov, Poland, pp 87–94

Jones N, Ougham H, Thomas H (1997) Markers and mapping: we are all geneticists now. New Phytol 137:165-177

Jose`-Estanyol M, Ruiz-Avila L, Puigdome`nech P (1992) A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. Plant Cell 4:413–423

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, sorghum and wheat. Plant Mol Biol 48:501-510

Kashi Y, King D, Soller M (1997) Simple sequence repeats as a source of quantitative genetic variation. Trends Genet 13:74–78

Kazan K and Muehlbauer FJ (1991) Allozyme variation and phylogeny in annual species of *Cicer* (Leguminosae). Plant Syst Evol 175:11-21

Kazan K, Muhelbauer FJ, Weeden NF, Ladizinsky G (1993) Inheritance and linkage relationships of morphological and isozyme loci in chickpea (*Cicer arietinum* L.) Theor Appl Genet 86:417-426

Khlestkina E, Than MHM, Pestsova EG, Röder MS, Malyshev SV, Korzun V, Börner A (2004) Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor Appl Genet 109:725–732

Kidwell KK, Echt CS, Osborn TC, McCoy TJ (1993) RFLP and RAPD linkage map of alfalfa (*Medicago sativa* L.) 2n = 16. In: O'Brien SJ (ed) Genetic maps. Cold Spring Harbor Laboratory Cold Spring Harbor, New York, pp 6.84–85

Kimura M and Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Kimura M and Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868-2872

King GJ (2002) Through a genome, darkly: comparative analysis of plant chromosomal DNA. Plant Mol Biol 48:5-20

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A (2003) Snipping polymorphism from large EST collections in barley (*Hordeum vulgare* L.). Mol Genet Genomics 270:224–233

Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001) Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas 135:145–151

Kruglyak S, Durrett R, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Kuleung C, Baenziger PS, Dweikat I (2004) Transferability of SSR markers among wheat, rye, and triticale. Theor Appl Genet 108:1147–1150

Kumar J and Abbo S (2001) Genetics of flowering time in chickpea and its bearing on productivity in the semi-arid environments. Adv Agron 72:107–138

Kumpatia SP and Mukhopadhyay S (2005) Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985-998

Labdi M, Robertson LD, Singh KB, Charrier A (1996) Genetic diversity and phylogenetc relationships among the annual *Cicer* species as revealed by isozyme polymorphism. Euphytica 88:181-188

Ladizinsky G (1975) A new *Cicer* from Turkey. Notes from the Royal Botanical Garden Edinburgh. 34:201-202

Ladizinsky G and Adler A (1976a) Genetic relationships among the annual species of *Cicer* L. Theor Appl Genet 48:197-203

Ladizinsky G and Adler A (1976b) The origin of chickpea *Cicer arietinum* L. Euphytica 88:181-188

Lagercrantz U and Putterill J, Coupland G, Lydiate D (1996) Comparative mapping of *Arabidopsis* and *Brassica*, fine scale genome colineraity and congruence of genes controlling flowering time. Plant J 9:13-20

Lee JS, Hanford MG, Genova JL, Farber RA (1999) Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. Hum Mol Genet 8:2567-2572

Lessa EP (1992) Rapid survey of DNA sequence variation in natural populations. Mol Biol Evol 9:323-330

Levinson G and Gutman GA (1987) High frequency of short frameshifts in poly CA/GT tandem repeats borne by bacteriophage M13 in *E.coli* K-12. Nucleic Acids Research15:5323-5338

Li Y-C, Korol AB, Fahima T, Beiles AV, Nevo Eviatar (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Eco 11:2453-2465

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: Structure, Function, and Evolution. Mol Biol Evol 21:991-1007

Li WH (1997) Molecular evolution. Sinauer Associates, Sunderland, Mass

Lichtenzveig J, Scheuring C, Dodge J, Abbo S, Zhang HB (2005) Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor Appl Genet 110:492-510

Lichtenzveig J, Bonfil DJ, Zhang HB, Shtienberg D, Abbo S (2006) Mapping quantitative trait loci in chickpea associated with time to flowering and resistance to *Didymella rabiei* the causal agent of Ascochyta blight. Theor Appl Genet 113:1357-1369

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Litt M and Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397-401

Livingstone KD, Lackney VK, Blauth JR, Wijk RJ, Jahn MK (1999) Genome Mapping in *Capsicum* and the Evolution of Genome Structure in the *Solanaceae* Genetics 152:1183–1202

Maestri E, Malcevschi A, Massari A, Marmiroli N (2002) Genomic analysis of cultivated barley (*Hordeum vulgare*) using sequence-tagged molecular markers. Estimates of divergence based on RFLP and PCR markers derived from stress responsive genes, and simple-sequence repeats (SSRs). Mol Genet Genomics 267:186–201

Makova KD, Nekrutenko A, Baker RJ (2000) Evolution of Microsatellite alleles in four species of Mice (Genus *Apodemus*). J Mol Evol 51:166-172

Maynez M, Moreno MT, Gil J (1993) Desi/Kabuli introgression for yield improvement in chickpea (*Cicer arietinum* L.). J Genet Breed 47:61–64

Mantri NL, Ford R, Coram TE, Pang Edwin CK (2007) Transcriptional profiling of chickpea genes differentially regulated in response to high-salinity, cold and drought. BMC Genomics 8:303

Marcotte EM, Pellegrini M, Yeates TO, Eisenberg D (1999) A census of protein repeats. J Mol Biol 293:151–160.

McCouch SR, Teytelman L, Xu YB, Lobos KB, Clare K,Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q,Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D,Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L). DNA Res 9:199–207

McDonald DB and Potts WK (1997) DNA microsatellites as genetic markers for several scales. In: Mindell DP (ed.) Avian molecular evolution and systematics. Academic Press, San Diego, pp. 29–49

McIntosh GH and Topping DL (2000) Food legumes in human nutrition. In: R Knight (Ed.), Linking Research and Marketing Opportunities for Pulses in the 21st Century, pp. 655–666. Kluwer Academic Publishers, Dordrecht, The Netherlands

McLauchlan A, Henry RJ, Isaac PG, Edwards KJ (2001) Microsatellites analysis in cultivated wheat and wild relatives. In: R.J.Henry (ed.) Plant genotyping-the DNA fingerprinting of plants, CABI publishing, Wallingford, UK

Menendez CM, Hall AE, Gepts P (1997) A genetic linkage map of cowpea (*Vigna unguiculata*) developed from a cross between two inbred, domesticated lines. Theor Appl Genet 95:1210-1217

Millan T, Rubio J, Iruela M, Daly K, Cubero JI, Gil J (2003) Markers associated with Ascochyta blight resistance in chickpea and their potential in marker-assisted selection. Field Crops Res 84:373–384

Millan T, Clarke HJ, Siddique Kh, Buhariwalla HK, Gaur PM, Kumar J, Gil J, Kahl G, Winter P (2006) Chickpea molecular breeding: New tools and concepts. Euphytica 147:81-103

Molina C, Rotter B, Horres R, Udupa SM, Besser B, Bellarmino L, Baum M, Matsumura H, Terauchi R, Kahl G, Winter P (2008) SuperSAGE: the drought stress-responsive transcriptome of chickpea roots. BMC Genomics 9:553

Moreno MT and Cubero JI (1978) Variation in *Cicer arietinum* L. Euphytica 27:465-468

Morgante M and Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175–182

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with non-repetitive DNA in plant genomes. Nat Genet 30:194-200

Muehlbauer FJ, Kaiser WJ, Simon CJ (1994) Potential for wild species in cool season food legume breeding. Euphytica 73:109-114

Muehlbauer FJ and Singh KB (1987) Genetics of Chickpea. In: M.C. Saxena and K.B. Singh (eds) The chickpea CAB International. ICARDA, Aleppo, pp. 99–125

Naoumkina M, Torres-Jerez I, Allen S, He J, Zhao P, Dixon AR, May GD (2007). Analysis of cDNA libraries from developing seeds of guar (*Cyamopsis tetragonoloba* (L.) Taub) BMC Plant Biology 7:62

Nene YL and Reddy MV (1987) Chickpea diseases and their control. In: Saxena MC, Singh KB (eds) The Chickpea. CAB International, Oxon, pp 233–270

Nguyen TT, Taylor PWJ, Redden RJ, Ford R (2004) Genetic diversity estimates in *Cicer* using AFLP analysis. Plant Breed 123:173–179

Nimbalkar S, Harsulkar A, Giri A, Sainani M, Franceschi V, Gupta V (2006) Differentially expressed gene transcripts in roots of resistant and susceptible chickpea plant (*Cicer arietinum* L.) upon *Fusarium oxysporum* infection. Physiological and Molecular Plant Pathology 68:176–188.

Nicot NV, Chiquet B, Gandon L, Amilhat F, Legeai P, Leroy M, Bernard P, Sourdille (2004) Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). Theor Appl Genet 109:800–805

Ocampo B, Venora G, Errico A, Singh KB, Saccardo F (1992) Karyotype analysis in the genus *Cicer*. J Genet Breed 46:229-240

Or E, Hovav R, Abbo S (1999) A major gene for flowering time in chickpea. Crop Sci 39:315–322

Ostrander EA, Sprague GF Jr, Rine J (1993) Identification and characterization of dinucleotide repeat $(CA)_n$ markers for genetic mapping in dog. Genomics 16:207-13

Palomino C, Fernández-Romero MD, Rubio J, Torres A, Moreno MT, Millán T (2008) Integration of new CAPS and dCAPS-RGA markers into a composite chickpea genetic map and their association with disease resistance. Theor Appl Genet DOI 10.1007/s00122-008-0928-7

Palsboll PJ, Berube M, Jorgensen H (1999) Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. Genetics 151:285-296

Pandian A, Ford R, Taylor PWJ (2000) Transferability of Sequenced Tagged Microsatellite Site (STMS) primers across four major pulses. PMB Reporter 18:395-395

Panjabi P, Jagannath A,Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. BMC Genomics 9:113

Park YH, Magdy E, Alabady S, Ulloa M, Sickler EB, Thea E, Wilkins A, Yu J, David E, Stelly M, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Gen Genomics 274:428–441

Pashley CH, Ellis JR, Mccauley DE Burke JM (2006) EST Databases as a Source for Molecular Markers: Lessons from *Helianthus*. Journal of Heredity 97:381–388

Paterson AH, Freeling M, Sasaki T (2005) Grains of knowledge: genomics of model cereals. Genome Res 15:1643–1650

Paterson AH (1996a) DNA Marker-Assisted crop improvement. *In* "Genome mapping in plants". (Paterson, A.H. ed). R.G. Landes Co. pp. 71-79

Paterson AH (1996b) Making Genetic Maps. *In* "Genome mapping in plants". (Paterson, A.H. ed). R.G. Landes Co. pp. 23-37

Pe´pin L, Amigues Y, Pingle A, Berthier JL, Bensaids A, Vaiman D (1995) Sequence conservation of microsatellites between *Bos taurus* (cattle), *Capra hircus* (goat) and related species. Examples of use in parentage testing and phylogeny analysis. Heredity 74:53–61

Peakall R, Glimore S, Keys W, Morgante M, Rafalski A (1998) Cross species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume genera: implication for transferability of SSRs in plants. Mol Biol Evol 15:1275-1287

Pedersen C and Linde-Laursen I (1994) Chromosomal location of four minor rDNA loci and a marker microsatellite sequence in barley. Chromosome Res 2:67–71

Peters JL, Constandt H, Neyt P, Cnops G, Zethof J, Zabeau M, Gerats T (2001) A physical amplified fragment-length polymorphism map of *Arabidopsis*. Plant Physiol 127:1579–1589

Pfaff T and Kahl G (2003) Mapping of gene-specific markers on the genetic map of chickpea (*Cicer arietinum* L.). Mol Genet Genome 269:243–251

Pflieger, S., V. Lefebvre, and M. Causse. 2001. The candidate gene approach in plant genetics: A review. Mol Breed 7:275–29

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST Databases Genome Res 9:167-174

Pillen K, Binder A, Kreuzkam B, Ramsay L, Waugh R, Förster J, Leon J (2000) Mapping new EMBL-derived barley microsatellites and their use in differentiating German barley cultivars. Theor Appl Genet 101:652–660

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko AD, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. Theor Appl Genet 276:436-449

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215-222

Powell W, Morgante M, McDevitt R, Vendramin G, Rafalski J (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759-7763

Primmer CR, Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Primmer CR, Saino N, Moller AP, Ellegren H (1998) Unraveling the process of microsatellite evolution through analysis of germ line mutation in barn swallows, *Hirundo rustica*. Mol Biol Evol 15:1047–1054

Putney SD, Herlihy WC, Schimmel P (1983) A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. Nature 302:718–721

Qin L, Prins P, Jones JT, Popeijus H, Smant G, Bakker J, Helder J (2001) GenEST, a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP. Nucleic Acids Res 29:1616–1622

Radhika P, Gowda SJM, Kadoo NY, Mhase LB, Jamadagni BM, Sainani MN, Chandra S, Gupta VS (2007) Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L.) using two recombinant inbred line populations. Theor Appl Genet 115:209-216

Rafalski JA (2002) Applications of single nucleotide polymorphism in crop genetics. Curr Opin Plant Biol 5:94–100

Rafalski JA and Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet 9:275-280

Rajesh P, Tullu A, Gil J, Gupta V, Ranjekar P, Muehlbauer F (2002) Identification of an STMS marker for the double-podding gene in chickpea. Theor Appl Genet 105:604–607

Rajesh PN, O'Bleness M, Roe BA , Muehlbauer FJ (2008) Analysis of genome organization, composition and microsynteny using 500 kb BAC sequences in chickpea. Theor Appl Genet 117:449-458

Ramsay L, Macaulay M, Ivanissevich DS, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari A, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh R (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Ramsay L, Macaulay M, Cardle L, Morangte M, Ivanissevich DS, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analyis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211-1227

Rao SS, Sinha R, Das GK (1994) Genetic variability, heritability, expected genetic advance and correlation studies in chickpea. Indian J Pulses Res 7:25–27

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite marker based framework linkage map of *Vitis vinifera* L. Theor Appl Genet 108:864–872

Richard GF and Dujon B (1997) Trinucleotide repeats in yeast. Res Microbiol 148:731–744

Richards RI and Sutherland GR (1994) Simple repeat DNA is not replicated simply. Nat Genet 6:114-116

Roder MS, Lapitan NLV, Sorrells ME, Tanksley SD (1993) Genetic and physical mapping of barley telomeres. Mol Gen Genet 238:294–303

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007-2023

Rogers SO and Bendich AJ (1987) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol Biol 9:509–520

Rohlf FJ (1998) NTSYS: numerical taxonomy and multivariate analysis system, Version 2.1. Applied Biostatistics, Setauket

Romo S, Labrador E, Dopico B (2004) Water stress-regulated gene expression in *Cicer arietinum* seedlings and plants. Plant Physiol Biochem 39:1017-1026

Rossetto M, McNally J, Henry RJ (2002) Evaluating the potential of SSR: flanking regions examining taxonomic relationships in the *Vitaceae*. Theo Appl Genet 104:61–66

Rowland LZ and Dhanaraj AK (2003) Utility of blueberry-derived EST-PCR primers in related *Ericaceae* species. Hortscience 38:1428-1432

Rozen S and Skaletsky HJ (1997) Primer 3. Code available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith M (1995) Microsatellites are generally longer in humans compared to their homologues in non-human primates: evidence for directional evolution at microsatellite loci. Nature Genet 10:337-343

Rudd S (2003) Expressed sequence tags: alternative or complement to whole genome sequences? Trends Plant Sci 8:321–329

Rungis D, Bérube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K (2004) Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet 109:1283–1294

Saha MC, Rouf Mian MA, Eujayl I, Zwonitzer JC, Wang L, May GD (2004) Tall fescue EST-SSR markers with transferability across several grass species. Theor Appl Genet 109:783–791

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Samson D, Legeai F, Karsenty E, Reboux S, Veyrieras J-B, Just J, Barillot E (2003) GenoPlante-Info (GPI): a collection of databases and bioinformatics resources for plant genomics. Nucleic Acids Res 31:179–182

Sandhu D, Sidhu D, Gill KS (2002) Identification of expressed sequence markers for major gene-rich regions of wheat chromosome group 1 using RNA fingerprinting-differential display. Crop Sci 42:1285–1290

Santra D, Tekeoglu M, Ratnaparkhe M, Kaiser WJ, Muehlbauer FJ (2000) Identification and mapping Of QTLs conferring resistance to *Aschochyta blight* in chickpea. Crop Sci 40:1606-1612

Sargent DJ, Rys A, Nier S, Simpson DW, Tobutt KR (2007) The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera. Theor Appl Genet 114:373–384

Saxena MC (1993) The challenge of developing biotic and abiotic stress resistance in cool-season food legumes. In Breeding for stress tolerance in cool-season food legumes. Singh KB, Saxena MC (eds) pp. 3-14, John Wiley and Sons, Chichester, UK.

Schlotterer C (1998) Microsatellites. In: Hoelzel AR (ed) Molecular genetic analysis of populations: A practical approach. IRL Press, Oxford, pp. 237–261

Schmidt T and Heslop-Harrison JS (1996) The physical and genomic organization of microsatellites in sugar beet. Proc Natl Agri Sci 93:8761-8765

Schubert R, Mueller-starck G, Riegel R (2001) Development of EST-PCR markers and monitoring their intrapopulational genetic variation in *Picea abies* (L.) Karst. Theor Appl Genet 103:1223-1231

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723-726

Semagn K, Bjørnstad A, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. African Journal of Biotechnology 5:2540-2568

Serret MD, Udupa SM, Weigand F (1997) Assesment of genetic diversity of cultivated chickpea using RFLP-derived microsatellite markers: implications for origin. Plant Breed 116:573-578

Sethy NK, Shokeen B, Bhatia S (2003) Isolation and characterization of sequence-tagged microsatellite sites markers in chickpea (*Cicer arietinum* L.). Mol Ecol Notes 3:428-430

Sethy NK, Shokeen B, Edwards KJ, Bhatia S (2006a) Development of microsatellite markers and analysis of intra-specific genetic variability in chickpea (*Cicer arietinum* L.). Theor Appl Genet 112:1416-1428

Sethy NK, Choudhary S, Shokeen B, Bhatia S (2006b) Identification of microsatellite markers from *Cicer reticulatum*: molecular variation and phylogenetic analysis. Theor Appl Genet 112:347-357

Shan F, Clarke HC, Plummer JA, Yan G, Siddique KHM (2005) Geographical patterns of genetic variation in the world collections of wild annual *Cicer* characterized by amplified fragment length polymorphisms. Theor Appl Genet 110:381-391

Sharma PC, Hüttel B, Winter P, Kahl G, Gardner RC, Weising K (1995a) The potential of microsatellites for hybridization- and polymerase chain reaction- based DNA fingerprinting of chickpea (*Cicer arietinum* L.) and related species. Electrophoresis 16:1755-1761

Sharma PC, Winter P, Bunger T, Hüttel B, Weigand F, Weising K, Kahl G (1995b) Abundance and polymorphism of di-, tri- and tetra- nucleotide tandem repeats in chickpea (*Cicer arietinum* L.). Theor Appl Genet 90:90-96

Sharma KD, Winter P, Kahl G, Muehlbauer FJ (2004) Molecular mapping of *Fusarium oxysporum* f. sp. *Ciceris* race 3 resistance gene in chickpea. Theor Appl Genet 108:1243-1248

Sharma KD, Muehlbauer FJ (2007) Fusarium wilt of chickpea: physiological specialization, genetics of resistance and resistance gene tagging. Euphytica 157:1-14

Shoemaker J, Zaitlin D, Horn J, De Mars S, Kirshman J, Pitas J (1992) A comparison of three agrigenetics maize RFLP linkage maps. Maize Genet Coop Newsl 66:65-69

Simon CJ and Muhelbauer FJ (1997) Construction of a chickpea linkage map and its comparison with maps of pea and lentil. J Hered 88:115-119

Singh KB 1987 Chickpea breeding. In: M.C. Saxena & K.B.Singh (Eds.), The Chickpea, pp. 127–162. CAB International,Wallingford, UK.

Singh KB (1997) Chickpea (*Cicer arietinum* L). Fields Crop Res 53:161–170

Singh A, Singh NP, Asthana N (1999) Genetic potential of wide crosses in chickpea. Legume Research 22:19-25

Singh KB, Malhotra RS, Halila MH, Knights EJ, Verma MM (1994) Current status and future strategy in breeding chickpea for resistance to biotic and abiotic stresses. Euphytica 73:137–149

Singh KB and Ocampo B (1997) Exploitation of wild *Cicer* species for yield improvement in chickpea. Theor Appl Genet 95:418-423

Singh KB, Ocampo B, Robertson LD (1998) Diversity for abiotic and biotic stress resistance in the wild annual *Cicer* species. Genet Resour Crop Evol 45:9–17

Sledge MK, Ray IM, Jiang G (2005) An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 111:980-992

Small RL and Wendel JF (2000) Copy number liability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155:1913–1926

Small RL, Cronn RC, Wendel JF (2004) L.A.S. Johnson review, 2: use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany 17:145–170

Smithson JB, Thompson JA, R.J. Summerfield (1985) Chickpea (*Cicer arietinum* L.). p. 312-390. In: R.J. Summerfield and E.H. Roberts (eds.), Grain Legume Crops. Collins, London, UK

Somerville C and S Somerville (1999) Plant functional genomics. Science 285:380–383

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin N, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvork J, Zhang D, Nguyen HT, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339-1348

Sreenivasulu N, Kishor PBK, Varshney RK, Altschmied L (2002) Mining functional information from cereal genomes – the utility of expressed sequence tags. Current Science 83: 965-973

Staginnus C, Winter P, Desel C, Schmidt T, Kahl G (1999). Molecular structure and chromosomal localization of major repetitive DNA families in the chickpea (*Cicer arietinum* L.) genome. Plant Mol Biol 39:1037–1050

Sudupak MA (2004) Inter and intra-species inter simple sequences repeat (ISSR) variations in the genus *Cicer*. Euphytica 135:229–238

Sudupak MA, Akkaya MS, Kence A (2002) Analysis of genetic relationships among perennial and annual *Cicer* species growing in turkey using RAPD markers. Theor Appl Genet 105:1220-1228

Sudupak MA, Akkaya MS, Kence A (2004) Genetic relationships among perennial and annual *Cicer* species growing in Turkey assessed by AFLP fingerprinting. Theor Appl Genet 108:937–944

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132:1141–1160.

Taran B, Warkentin TD, Tullu A, Vandenberg A (2007) Genetic mapping of aschochyta blight resistance in chickpea (*Cicer arietinum* L.) using a simple sequence repeat map. Genome 50:26-34

Tartarini S (2003) Marker-assisted selection in pome fruit breeding. Marker assisted selection: a fast track to increase genetic gain in plant and animal breeding? Session I: MAS in plants. (Conference 10 of the FAO Biotechnology Forum) Molecular marker assisted selection as a potential tool for genetic improvement of crops, forest trees, livestock and fish in developing countries, 17 November to 12 December 2003, Italy

Tautz D and Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acids Res 12:4127-4138

Taylor JS, Durkin JMH, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Evol Biol 16:567–572

Tayyar RI and Waines JG (1996) Genetic relationships among annual species of *Cicer* (Fabaceae) using isozymes variation. Theor Appl Genet 92:245-254

Tekeoglu M, Rajesh PN, Muehlbauer FJ (2002) Integration of sequence tagged microsatellite sites to chickpea genetic map. Theor Appl Genet 105:847-854

Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.) Theor Appl Genet 102:664–675

Temesgen B, Neale DB, Harry DE (2000) Use of haploid mixtures and heteroduplex analysis enhance polymorphisms revealed by denaturing gradient gel electrophoresis. BioTechniques 28:114–122

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch S (2000) Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:697–712

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.) Theor Appl Genet 106:411-422

Tian AG, Wang J, Cui P, Han YJ, Xu H, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ, Wang YJ, Zhang JS, Chen SY (2004) Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903-913

Torado A, Koike M, Mochida K, Ogihara Y (2006) SSR-based linkage map with new markers using an intraspecific population of common wheat. Theor Appl Genet 112:1042-1051

Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different Eukaryotic Genomes: Survey and analysis. Genome Res 10:967-981

Toutenhoofd SLF, Garcia DA, Zacharias R, Wilson A, Strehler EE (1998) Minimum CAG repeat in the human calmodulin-1 gene 5' untranslated region is required for full expression Biochim Biophys Acta 1398:315-320

Tsumura Y, Suyama Y, Yoshimura K, Shirato N, Mukai Y (1997) Sequence-tagged-sites (STSs) of cDNA clones in *Cryptomeria japonica* and their evolution as molecular markers in conifers. Theor Appl Genet 94:764–772

Tuwafe S, Kahler AL, Boe A, Ferguson M (1988) Inheritance and geographical distribution of allozyme polymorphisms in chickpea (*Cicer arietinum*). Heredity 79:170-174

Udupa SM, Sharma A, Sharma RP, Pai RA (1993) Narrow genetic variability in *Cicer arietinum* as revealed by RFLP analysis. J Plant Biochem Biotechnolo 2:83-86

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at (TAA)n microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. Mol Gen Genet 261:354–363

Udupa SM and Baum M (2003) Genetic dissection of pathotype-specific resistance to *aschochyta* blight disese in chickpea (*Cicer arietinum* L.) using microsatellite markers. Theor Appl Genet 106:1196-1202

Udupa SM, Malhotra RS, Baum M (2004) Tightly linked di- and tri-nucleotide microsatellites do not evolve in complete independence: evidence from linked (TA)n and (TAA)n microsatellites of chickpea (*Cicer arietinum* L.). Theor Appl Genet 108:550-557

Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda LL, Hoisington D, Singh S (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.) BMC Plant Biology 8:106

van der Maesen LJG (1972) *Cicer* L., a monograph of the genus, with an special reference to the chickpea (*Cicer arietinum* L.), its ecology and cultivation. In: H. Veenmam and Q. Zonen (Eds.), 342 pp. Mededlingen Landbouwhogeschool (Communication Agricultural University), Wageningen

van der Maesen LJG (1987) Origin, history and taxonomy of chickpea. In: Saxena MC, Singh KB (eds) The Chickpea. CAB Int Publ, UK. pp 11-34

van Ooijen J (2006) JoinMap version 4.0: software for the calculation of genetic linkage maps. Plant Res Int, Wageningen

Varshney RK, Marcel TC, Ramsay L, Russell J, Roder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A (2007a) A high density barley microsatellite consensus map with 775 SSR loci. Theor Appl Genet 114:1091-1113

Varshney RK, Hoisington DA, Upadhyaya HD, Gaur PM, Niggam SN, Saxena K, Vadez V, Sethy NK, Bhatia S, Aruna R, Gowda MVC, Singh NK (2007b) Molecular genetics genetics and breeding of grain legume crops for the semi-arid tropics. R. Varshney and R. Tuberosa (eds.), Genomic Assisted Crop Improvement: Genomics Applications in Crops 2:207–241

Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. Trends in Biotechnology 23:48-55

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Science 168:195-202

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) *IN SILICO* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular and Molecular Biology letters 7:537-546

Venter C, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, et al. (2001) The sequence of the human genome. Science 291:1304-1351

Verma MM, Sandhu JS, Brar HS, Brar JS (1990) Crossability studies in different species of *Cicer* L. Crop Improv 17:179-181

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebly J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for the evidence of selection during domestication. Proc Natl Acad Sci USA 99:9650-9655

Wang HY, Wei YM, Yan ZH, Zheng YL (2007) EST-SSR DNA polymorphism in durum wheat

Wang X, Zhao X, Zhu J, Wu W (2005) Genome-wide Investigation of Intron Length Polymorphisms and their Potential as Molecular Markers in Rice (*Oryza sativa* L.). DNA Research 12:417-427

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1-6

Weber H, Borisjuk L, Heim U, Sauer N, Wobus U (1997) A role for sugar transporters during seed development: molecular characterization of a hexose and a sucrose carrier in fababean seeds. Plant Cell 9:895–908

Weber JL (1990) Informativeness of human $(dC - dA)_n \cdot (dG - dT)_n$ polymorphisms. Genomics 7:524-530

Weber JL and Wong C (1993) Mutation of short tandem repeats. Hum Mol Genet. 2:1123–1128

Weeden NL, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129

Wei H, Fu Y, Arora R (2005) Intron-flanking EST-PCR markers: from genetic marker development to gene structure analysis in *Rhododendron*. Theor Appl Genet 111:1347-1356

Weising K, Kaemmer D, Weigand F, Epplen JT, Kahl G (1992) Oligonucleotide Fingerprinting reveals various probe-dependent levels of informativeness in chickpea (*Cicer arietinum*). Genome 35:436-422

Westman AL and Kresovich S (1998) The potential of cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop *brassicas*. Theor Appl Genet 96: 272-281

White JA, Todd J, Newman T, Focks N, Girke T, de Ila′ rduya OM, Jaworski JG, Ohlrogge JB, Benning C (2000) A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. Plant Physiol 124:1582–1594

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics 146:769-779

Williams PC and Singh U (1987) The Chickpea–Nutritional quality and the evaluation of quality. In: M.C. Saxena & K.B. Singh (Eds.), The Chickpea, pp. 329–356. CAB International, Wallingford, UK

Winter P, Pfaff T, Udupa SM, Hüttel B, Sharma PC, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol Gen Genet 262:90-101

Winter P, Benko-Iseppon AM, Hüttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santara D, Sant VJ, Rajesh PN, Kahl G, Muhelbauer FJ (2000) A linkage map of chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5. Theor Appl Genet 101:1155-1163

Winter P, Staginnus C, Sharma PC, Kahl G (2003) Organisation and genetic mapping of the chickpea genome. In: P.K. Jaiwal and R.P. Singh (Eds.), Improvement Strategies of Leguminosae Biotechnology, pp. 303–351. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Winter P, Molina C, Jungmann R, Horres R, Rotter B, L'taief B, Geistlinger J, Dolezel J, KahlG (2006) Towards expression maps in chickpea. In 'Plant and Animal GenomeXV Conference, San Diego'. (EdsGLazo,DGrant, V Blake) p. W140. (San Diego)

Wolfsberg TG and Landsman D (1997) A comparison of expressed sequence tags (ESTs) to human genomic sequences. Nucleic Acids Research 25:1626-1632

Wren JD, Forgacs E, Fondon 3rd JW, Pertsemlidis A, Cheng SY, Gallardo T, Williams RS, Shohet RV, Minna JD, Garner HR (2000) Repeat polymorphisms within gene regions: phenotypic and evolutionary implications. Am J Hum Genet 67:345–356

Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535

Xue S, Zhang Z, Lin F, Kong Z, Cao Y, Li C, Yi H, Mei W, Zhu H, Wu J, Xu H, Zhao D, Tian D, Zhang C, Ma Z (2008) A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags Theor Appl Genet 117:181-189

Yamamoto K and Sasaki TL (1997) Large-scale EST sequencing in rice. Plant Mol Biol 35:135-144.

Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W (2007) PIP: a database of potential intron polymorphism markers. Bioinformatics 23:2174–2177

Yeh FC and Boyle TJB (1997) Population genetic analysis of codominant and dominant markers and quantitative traits. Belgian J Bot 129:157

Yi G, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST-SSRs and an SSR-based linkage map. Theor Appl Genet 114:113-130

Yu J-K, Dake TM, Singh S, Benscher D, Li W, Gill B, Sorrells ME (2004) Development and mapping of EST-derived simple sequence repeat (SSR) markers for hexaploid wheat. Genome 47:805–818

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1-16

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677-687

Zhang Y, Sledge MK, Bouton Jh (2007) Genome mapping of white clover (*Trifolium repens* L.) and comparative analysis within the *Trifolieae* using cross-species SSR markers. Theor Appl Genet 114:1367-1378

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324-338

Appendices

Appendix I

I. List of publications

1. **Choudhary S**, Sethy NK, Shokeen B, Bhatia S (2008) Development of chickpea EST-SSR markers and analysis of allelic variation across related species Theor Appl Genet 118:591–608

2. **Choudhary S**, Sethy NK, Shokeen B, Bhatia S (2006) Development of sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Molecular Ecology Notes 6(1): 93-95

3. Sethy NK, **Choudhary S**, Shokeen B, Bhatia S (2006) Identification of microsatellite markers from *Cicer reticulatum*: Molecular variation and phylogenetic analysis. Theor. Appl. Genet. 112: 347-357

4. Shokeen B, Sethy NK, **Choudhary S**, Bhatia S (2005) Development of STMS markers from the medicinal plant Madagascar periwinkle [*Catharanthus roseus* (L.) G. Don.]. Mol. Ecol. Notes 5:818-820.

II. Abstract Published

1. **Choudhary S**, Sethy NK and Bhatia S (2005). EST- SSR markers from chickpea. Proceedings of 4th International Food Legumes Research Conference at New Delhi, India, October 2005, pp222

2. Sethy NK, **Choudhary S**, Shokeen B, Bhatia S (2005). Development of microsatellite markers from *Cicer reticulatum*, the wild annual progenitor of chickpea. Proceedings of 4th International Food Legumes Research Conference at New Delhi, India, October 2005, pp222

3. Sethy NK, **Choudhary S**, Nandkeolyar N, Gupta V, Shokeen B, Bhatia S (2005). Identification of microsatellite markers from chickpea (*Cicer arietinum* L.) and analysis of intraspecific sequence variation. Proceedings of 4th International Food Legumes Research Conference at New Delhi, India, October 2005, pp221

Appendix II

LURIA BROTH (LB) 1000 ml	**20x SSC (1000 ml, pH 7-8)**
Tryptone 10gm Yeast Extract 5gm Sodium chloride 5gm *Dissolve in 500ml and adjust to 1000ml.*	Tri sodium citrate 88.23 gm. Sodium chloride 175.32 gm. Water 800 ml. *Adjust to 1000ml after dissolving in 500ml.*
LOADING DYE (10 ml) Glycerol (100%) 5 ml. Water 4 ml SDS (10%) 1 ml Bromophenol blue & 0.5% Xylene cyanol	**LOADING DYE (10ml) (denaturing)** Formamide 9.8ml EDTA(0.5M) 0.2 g Xylene Cyanol 0.025g Bromophenol blue 0.0250g
5X TBE: (1000ml, pH 8.0) Tris 54 g Boric Acid 27.5g EDTA 3.8 g *Add 500ml of RO water, dissolve and make up the volume to 1000ml.*	**RNAse SOLUTION (for 10ml)** RNAse A 100mg 3M sodium acetate pH 5.2 30µl Auto MQ water 9.970µl *Boiled for 15 min, cooled to RT store at –20^{0}C*
SOLUTION I (GET buffer) 50 ml 1M Glucose 2.5 ml 0.5M EDTA 1.0 ml 1 M Tris 1.25 ml Water 45.25 ml	**SOLUTION II (10ml)** 1N NaOH 2ml 10%SDS 1ml Water 7ml
SOLUTION III 100ml (pH4.8) Potassium acetate 29.44 gm Glacial acetic acid 11.5ml Water 28.5ml *Make up the volume up to 60ml with water. Set pH. Make up the volume to 100 ml*	**IPTG (20% w/v, 0.8 M)** IPTG 2 gm dd Water 8 ml *Adjust the volume to 10 ml with dd.water. Sterilize by passing through 0.22µm disposable filter. Dispense the solution into 1 ml aliquot in sterile tubes. Store them at -20°C*
X-gal SOLUTION (2% w/v) 10 ml X-gal 200mg *Dissolve X-gal in 10ml of dimethylformamide. Use glass or polypropylene tube. Wrap the tube in aluminium foil. Store at -20°C.*	**AMPICILLIN 100mg/ml (5ml)** Ampicillin 500mg *Add 5 ml of water and then filter sterilize. Aliquot under Laminar air flow.*
3% METAPHOR AGAROSE GEL (150ml) Metaphor agarose 3.375 g	**AGAROSE GEL (0.8%) :(100 ml)**

Agarose 1.125 g TBE (1X) 150 ml	Agarose 0.8 g TBE Buffer (1X) 100 ml
RNA EXTRACTION BUFFER (50 ml) 3M Sodium acetate pH 5.2 3.3ml 10% SDS 5.0 ml 0.5M EDTA pH 8.0 1.0 ml DEPC water 40.67 ml	**DNA EXTRACTION BUFFER** (250 ml) Tris Cl (1M) 25 ml EDTA (0.5M) 10 ml NaCl 20.45 gm CTAB 5 gm *Make final volume to 250 ml by* Autoclaved Milli Q *Autoclave and store at RT.*
Prehybriziation buffer (10 ml) 0.1M Na phospahte buffer pH 7.2 3 ml 10 % SDS 7 ml 0.5 M EDTA 20 µl	**10X Formaldehyde gel loading Buffer (5ml)** 50% glycerol (diluted in DEPC water) 2.5 ml 10mM EDTA (pH8) 0.1 ml 0.25% Xylene Cyanol 0.0125 g *Make final volume to 5 ml by DEPC water*
6% PAGE (50ml) Milli Q 15 ml 5X TBE Buffer 10 ml 20% Acryl-Bisacrylamide(19:1) sol 15ml 10% APS 120µl TEMED 40µl	**10X MOPS pH 7.0** (1 lit.) Dissolve 41.8 gm MOPS in 700 ml DEPC water 1M DEPC treated Sodium acetate pH 5.2 20 ml 0.5M DEPC treated EDTA pH 8.0 20 ml Adjust pH 7.0 with 2N NaOH *Make final volume to 1 lit. by DEPC water and kept in Al foil wrapped bottle*

GLYCEROL STOCK

Solution A (500ml):Tryptone:5gm, YE: 2.5g, NaCl: 5g, Sodium citrate: 0.25g, Ammoniun sulphate: 0.45g, glycerol: 22ml
Solution B (100ml): KH_2PO_4: 1.8g, K_2HPO_4: 6.27g
Solution C: 1M $MgSO_4$.

Autoclave all the solutions and store at RT

Glycerol Mixture (50ml) with 50.0µl of Ampicillin100mg/ml
Solution A: 45 ml
Solution B: 5ml
Solution C: 20µl

Theor Appl Genet (2009) 118:591–608
DOI 10.1007/s00122-008-0923-z

ORIGINAL PAPER

Development of chickpea EST-SSR markers and analysis of allelic variation across related species

Shalu Choudhary · Niroj Kumar Sethy · Bhumika Shokeen · Sabhyata Bhatia

Received: 4 January 2008 / Accepted: 24 October 2008 / Published online: 20 November 2008
© Springer-Verlag 2008

Abstract Despite chickpea being the third important grain legume, there is a limited availability of genomic resources, especially of the expressed sequence tag (EST)-based markers. In this study, we generated 822 chickpea ESTs from immature seeds as well as exploited 1,309 ESTs from the chickpea database, thus utilizing a total of 2,131 EST sequences for development of functional EST-SSR markers. Two hundred and forty-six simple sequence repeat (SSR) motifs were identified from which 183 primer pairs were designed and 60 validated as functional markers. Genetic diversity analysis across 30 chickpea accessions revealed ten markers to be polymorphic producing a total of 29 alleles and an observed heterozygosity average of 0.16 thereby exhibiting low levels of intra-specific polymorphism. However, the markers exhibited high cross-species transferability ranging from 68.3 to 96.6% across the six annual *Cicer* species and from 29.4 to 61.7% across the seven legume genera. Sequence analysis of size variant amplicons from various species revealed that size polymorphism was due to multiple events such as copy number variation, point mutations and insertions/deletions in the microsatellite repeat as well as in the flanking regions. Interestingly, a wide prevalence of crossability-group-specific sequence variations were observed among *Cicer* species that were phylogenetically informative. The neighbor joining dendrogram clearly separated the chickpea cultivars from the wild *Cicer* and validated the proximity of *C. judaicum* with *C. pinnatifidum*. Hence, this study for the first time provides an insight into the distribution of SSRs in the chickpea transcribed regions and also demonstrates the development and utilization of genic-SSRs. In addition to proving their suitability for genetic diversity analysis, their high rates of transferability also proved their potential for comparative genomic studies and for following gene introgressions and evolution in wild species, which constitute the valuable secondary genepool in chickpea.

Communicated by C. Gebhardt.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0923-z) contains supplementary material, which is available to authorized users.

S. Choudhary · B. Shokeen · S. Bhatia (✉)
National Institute of Plant Genome Research, Post Box Number 10531, Aruna Asaf Ali Marg, Jawaharlal Nehru University Campus, New Delhi 110067, India
e-mail: sabhyata_bhatia@nipgr.res.in; sabhyatab@yahoo.com

Present Address:
N. K. Sethy
Defence Institute of Physiology and Allied Sciences (DIPAS), Defence Research and Development Organization (DRDO), Timarpur, Delhi 110 054, India

Introduction

Extensive efforts at sequencing of expressed genomic regions obtained from tissues under different conditions and developmental stages have led to a large number of EST sequences being deposited in the public database for a number of model species as well as economically important plants. Besides providing an effective approach for gene discovery and transcript pattern characterization, these ESTs emerge as a cost-effective, valuable source for molecular marker generation. These easily accessible sequences provide the advantage of in silico analysis and broaden the field of comparative studies in species where limited or no sequence information is available.

Microsatellites or simple sequence repeats (SSRs) are 1–6 bp iterations of DNA sequences that were earlier known

to occur only in the non-coding regions. However, the occurrence of microsatellites in transcribed sequences is now well established and are commonly known as EST-SSRs or genic SSRs (Morgante et al. 2002; Li et al. 2002, 2004). These have been reported from a number of plant species such as *Oryza* (Cho et al. 2000), *Saccharum* (Cordeiro et al. 2001), *Triticum* (Gupta et al. 2003), *Hordeum* (Thiel et al. 2003), *Medicago* (Eujayl et al. 2004), *Coffea* (Poncet et al. 2006; Aggarwal et al. 2007), *Capsicum* (Yi et al. 2006) and *Citrus* (Chen et al. 2006). Similar to genomic SSRs, the EST-SSRs are useful for many applications in plant genetics and breeding such as molecular mapping, genetic diversity analysis and cross-transferability across related species and genera (Varshney et al. 2005a). Moreover, as a result of their association with coding sequences, they provide the possibility of direct gene tagging for QTL mapping of agronomically important traits. The EST-SSRs find higher levels of cross-species transferability than genomic microsatellite markers (Scott et al. 2000; Eujayl et al. 2004; Zhang et al. 2005) aiding in gene introgression programs, identification of conserved gene order across orthologous linkage groups (Varshney et al. 2005a, b), depiction of gene evolution associated with microsatellites and phylogenetic studies.

In chickpea (*Cicer arietinum* L.), the third most important grain legume crop, research efforts worldwide have led to identification and characterization of a number of microsatellite markers (Hüttel et al. 1999; Sethy et al. 2003; Lichtenzveig et al. 2005; Choudhary et al. 2006, Sethy et al. 2006a) and their utilization for genome mapping (Winter et al. 2000) and phylogenetic analysis of *Cicer* (Sethy et al. 2006a, b). More recently, EST sequences from chickpea have been reported (Boominathan et al. 2004; Romo et al. 2004; Buhariwalla et al. 2005; Coram and Pang 2005). Among these, only the study of Buhariwalla et al. 2005 investigated the use of ESTs as a source of genic markers. But even in this study, of the 106 EST markers developed by them, only 14 contained SSR motifs and these are the only chickpea EST-SSRs reported till date. Hence the need for large scale development of chickpea EST-SSRs was imminent. This would not only help in molecular mapping but would also be of significance in comparative genome analysis in legumes since a high degree of conservation among the genomes of cultivated species and model legumes has been revealed (Weeden et al. 1992; Choi et al. 2004). A recent study on transferability of both genomic and EST-SSR markers of *M. truncatula* to pea, chickpea and lentil revealed a high degree of cross-transferability (Gutierrez et al. 2005). However, to be fully effective, genomic information from one species must be transferred in both directions, i.e. from model species to cultivars and vice versa (Gepts et al. 2005). Thus, there was an urgent need to develop EST-SSR markers in chickpea and assess their transferability to the model as well as to other non-model important legumes and for tracking the introgression of genes from the wild or elite species of chickpea.

The present study was aimed at: (1) development and characterization of chickpea EST-SSRs, (2) assessing the utility of EST-SSRs for genetic diversity analysis, (3) evaluating the cross-transferability of chickpea EST-SSRs among the *Cicer* species and other legumes, and (4) establishing the molecular basis of variation in alleles from related species and genera.

Materials and methods

Plant material and DNA isolation

Chickpea and the wild annual species are classified into various crossability groups. This study includes members of Crossability group I (*C. arietinum, C. reticulatum* and *C. echinospermum*) and Crossability group II (*C. bijugum, C. judaicum and C. pinnatifidum*). Thirty accessions of cultivated chickpea (*C. arietinum*) were used for the analysis of genetic diversity within species (Table 1a). For inter-specific transferability studies, nine accessions belonging to the five wild annual *Cicer* species were used which included a single accession of *C. echinospermum* (ICC17159) and two accessions each of *C. reticulatum* (ICC17121, ICC17164), *C. bijugum* (ICC17125, ICC17122), *C. judaicum* (ICC17148, ICC17150) and *C. pinnatifidum* (ICC17126, ICC17200). For cross-genera studies across legumes, 28 accessions belonging to seven legume genera were used (Table 1b). All accessions used in this study were grown at the field site of NIPGR.

DNA was isolated from fresh, young leaf tissue of chickpea and legume accessions using the CTAB method (Doyle and Doyle 1987). Genomic DNA from the wild *Cicer* accessions was isolated using GenElute genomic DNA miniprep kit (SIGMA Aldrich). The quality and final concentration was estimated by agarose gel electrophoresis using known concentration of uncut λ DNA as a standard.

Construction of a cDNA library and identification of EST-SSRs

Total RNA was isolated from 20 DAF seeds of *C. arietinum* ICCV2; 0.8 g of frozen seeds was ground to a fine powder in liquid nitrogen and transferred to an Eppendorf tube containing 500 µl of extraction buffer (200 mM NaOAc pH −5.2, 1% SDS, 10 mM EDTA pH 8.0) and 500 µl of phenol. This was centrifuged at 14,000×*g* for 10 min at RT. The aqueous phase was separated and extracted twice with phenol:chloroform (1:1) followed by O/N precipitation at 4°C with 0.3 vol of 10 M LiCl for RNA precipitation. The RNA pellet was recovered by centrifugation at 10,000 rpm for 10 min at 4°C, and was washed twice with 2.5 M LiCl

Table 1 A: list of chickpea accessions alongwith their sources, B: list of legume accessions used in this study

S. no.	Acc. no./name	Source
A		
1	ICCV2[a]	India
2	JG62[a]	-do-
3	ICC10945[a]	-do-
4	ICC15406[a]	-do-
5	ICC283[a]	-do-
6	ICC12947[a]	-do-
7	ICC13124[a]	-do-
8	ICC791[a]	-do-
9	ICC5383[a]	-do-
10	ICC11378[a]	-do-
11	ICC5477[a]	-do-
12	ICC15802[a]	Syria
13	ICC156947[a]	-do-
14	ICC16976[a]	Portugal
15	ICC7676[a]	-do-
16	ICC16800[a]	-do-
17	ICC16761[a]	-do-
18	ICC12866[a]	Ethiopia
19	ICC12726[a]	-do-
20	ICC3485[a]	Jordan
21	ICC6293[a]	Italy
22	ICC3631[a]	Iran
23	ICC16487[a]	Pakistan
24	ICC8195[a]	-do-
25	ICC7272[a]	Algeria
26	ICC13780[a]	Spain
27	ICC8444[a]	Tunisia
28	ICC15518[a]	Morocco
29	ICC15407[a]	-do-
30	Pusa362[b]	India

S. no.	Acc. no.	Species	Common name
B			
31	NRC37[c]	*Glycine max*	Soybean
32	MAUS47[c]	-do-	Soybean
33	PRATAP[c]	-do-	Soybean
34	BRAGG[c]	-do-	Soybean
35	IC381277[d]	*Lens esculenta*	Lentil
36	IC334282[d]	-do-	Lentil
37	IC384444[d]	-do-	Lentil
38	IC383609[d]	-do-	Lentil
39	IC411188[d]	*Trifolium alexandrinum*	Berseem Clover
40	IC411189[d]	-do-	Berseem Clover
41	IC508311[d]	-do-	Berseem Clover
42	IC411183[d]	-do-	Berseem Clover
43	IC347150[d]	*Cajanus cajan*	Pigeonpea
44	IC339040[d]	-do-	Pigeonpea
45	IC337447[d]	-do-	Pigeonpea
46	IC396014[d]	-do-	Pigeonpea
47	IC342955[d]	*Phaseolus mungo*	Blackgram
48	IC328538[d]	-do-	Blackgram
49	IC397612[d]	-do-	Blackgram
50	IC362567[d]	-do-	Blackgram
51	IC279013[d]	*Pisum sativum*	Field pea
52	IC356344[d]	-do-	Field pea
53	RFP-19[c]	-do-	Field pea
54	RFP-18[c]	-do-	Field pea
55	SA27783[e]	*Medicago truncatula*	Barrel Medic
56	SA11959[e]	-do-	Barrel Medic
57	SA3235[e]	-do-	Barrel Medic
58	SA3780[e]	-do-	Barrel Medic

[a] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India

[b] Indian Agriculture Research Institute (IARI), India

[c] Maharana Pratap Agriculture University, India (MPAU)

[d] National Bureau of Plant Genetic Resources (NBPGR), India

[e] Australian Medicago Genetic Resource Centre, SARDI, Australia

and once with 70% ethanol. The pellet was air-dried and dissolved in DEPC-treated ddH_2O. One microgram of total RNA was used to construct the cDNA library using the CLONTECH Smart PCR-cDNA synthesis kit according to the manufacturer's protocol. Double-stranded cDNA was introduced into the pCR2.1-TOPO vector for sequencing using the TOPO TA cloning kit (Invitrogen).

Random 5′ sequencing of cDNAs was done using the BigDye Terminator technology (Applied Biosystems) in an ABI Prism 3700 automated DNA sequencer. After sequence trimming (removal of low quality sequences, vector regions and sequences <100 bp), the EST sequences were mined for microsatellites consisting of $\geq$5 dinucleotides and $\geq$4 trinucleotides repeats using the TROLL program (Castelo et al. 2002). To reduce redundancy, cluster analysis was performed on microsatellite containing sequences (EST-SSRs) using the CAP3 program (Huang and Madan 1999). The identified EST-SSR sequences were deposited in the GenBank to obtain the accession numbers (see Table 2). The putative function of the developed chickpea functional markers was found by the BLASTX tool of NCBI, assuming a threshold of $<1e-05$.

Generation of EST-SSRs from database

The 1,309 chickpea EST sequences available in the NCBI nucleotide database (up to January 2007) were screened for

Table 2 Characteristics of the chickpea EST-SSR markers

S. no.	Locus name	Primer sequence (5′ → 3′)	Motif	T_m (°C)	Expected size (bp)	GenBank no.	Putative function
1	CESSRDB2	CGGGCAGGTATTGAATTGTAA/ GAAAGGTTTACAGCCGTTGG	$(CT)_{17}$	59.4	169	CD051322	No homology
2	CESSRDB3	TTATCACTTGTTATTGTCCTCTAAG/ AATTTATGGACCCCATGTAA	$(TAA)_6$	60.5	197	AJ609280	No homology
3	CESSRDB4	GAAGAGGTAGCGGAGGAG/ CAAGCAACAGTTTTCACTCA	$(GGT)_3CGT(GGT)_2N_6(GGT)_3$	61.4	274	AJ609279	RNA and export binding protein
4	CESSRDB5	CCGACATCTCTTCTCAATTC/ CTTTAGGTGGTGGTTGTTGT	$(TCA)_{14}$	61.4	177	AY370650	SAT5 gene
5	CESSRDB7	AAGTGGTGTCGGTAATGGT/ TAATACCAAAAGCATGCACA	$(GGT)_5N_6(GGT)_2$	60.2	196	AJ487469	Glycine-rich protein
6	CESSRDB10	CCCTTAATCAATTCA CCTCA/ TTATCCAAACCAATGATTCC	$(TAA)GAA(TAA)_6$	59.8	192	AJ005947	No homology
7	CESSRDB11	AATCTAACAGCAACGACGAT/ ATCAAGCTTCTTCTGCACAT	$(CCA)CAA(CCA)_3$	58.4	298	AJ006048	Unknown protein
8	CESSRDB13	ATCTGGGAGCTTGTGAGTTA/ TTGTATCTCCTTCAGATGGC	$(AT)_5$	60.0	260	AJ012683	Hypothetical protein
9	CESSRDB15	CTTACGATTTCTCCTCCCTT/ TTTCTCATACCGAATCCTTG	$(GCT)N_6(GCT)_4$	61.7	276	AJ012681	Hypothetical protein
10	CESSRDB16	ATGCTATGCATGATGTTTCA/ GTTCCAAACAAACACAACAA	$(TA)_{10}$, $(TTA)_4G(TTA)A(TTA)_2$,	57.0	295	AJ487472	Invertase inhibitor
11	CESSRDB18	TGCAAATAAAGCCTTCAAGT/ GAAAGTGGGAAAATGCAATA	$(TA)_5CA(TG)_2T(TG)_3$	57.4	242	AJ487042	No homology
12	CESSRDB21	GTGTATCGGTCAGGAAAAGA/ GGTACACACCACAATTCACA	$(ATT)_2AT\ (AAT)_3$	60.0	259	AJ012693	Plantacyanin
13	CESSRDB23	GTGTGGACCTGAAATTGAGT/ GAATATGGGAACAAGTGCAT	$(TA)_5$	59.4	221	AJ012689	Ribonuclease T2
14	CESSRDB24	TGTGCTTGACTTGTTCACAT/ TATGCATCCTCATTTTCTCC	$(GGC)_4$	59.6	283	AJ006763	β-amylase
15	CESSRDB26	GGTGCATTCTCTTCCATAAG/ TGCAAATCT TTAACCAAACA	$(GT)_5$	57.4	273	AJ004959	Expansin
16	CESSRDB27	GGTGAGATTAGGAAGCAATG/ TATCCAATCCCCATAAGATG	$(TAT)_7$	58.7	215	AJ271660	Cationic peroxidase
17	CESSRDB29	TTTAGTTGCACAACAACAGC/ AAATCCACATCCAAAAAGGT	$(TGA)_5$	57.4	176	AJ299064	GTP-binding protein

Table 2 continued

S. no.	Locus name	Primer sequence (5′ → 3′)	Motif	T_m (°C)	Expected size (bp)	GenBank no.	Putative function
18	CESSRDB33	GCTGCACAAAAAGTACATGA/ ATCCATCGAAACACCAATAG	$(GA)T(GA)TT(GA)_5$	59.3	234	AJ250836	Pal gene
19	CESSRDB34	AACCTAAAGCCGAAAAGAAG/ CTCCCGTGAAGTAATAGTCG	$(AAG)N_4(AAG)_4$	61.4	251	AJ400863	Histone H2B
20	CESSRDB35	TCTAGAGCTAGCCAAAGGAA/ GCATCGTAATCATCGGTACT	$(GAT)_7$	61.7	272	AJ400861	UDP-glycose
21	CESSRDB38	GAGTAAGATGGCACAGTGGT/ GTATCTATTAGCGAAGCGGA	$(CCG)_4$	60.4	197	X93220	Cysteine proteinase
22	CESSRDB39	CTGAGGTTAATGTGAAAGGC/ GTCAACATCACATGCTCAAC	$(GGT)_4$	61.0	257	AJ275314	Glycine-rich protein
23	CESSRDB40	GAAATTAGGAAGCATTGTGC/ AATTGATTGAACCCACTTGT	$(TTAT)_4$	57.0	188	AJ275313	Peroxidase
24	CESSRDB41	GAACCAATAAAGCCTTGAAA/ TGACCAATTGATACAATCCA	$(GCT)_4$, $(TTTA)_4$	57.4	247	AJ275307	PM intrinsic polypeptide
25	CESSRDB42	GAGACAAAGATAGTGGCTGG/ TATTAATCACTCGCACGACA	$(TAAAT)_4$ $(GTTT)_3$	61.3	235	AJ275304	ABA-responsive protein
26	CESSRDB44	ATCCTTTCCTTGTTGTGCTA/ TTTAGTGAAGCATTGTTGGA	$(CTTT)_3$, $(TTG)_4$	57.8	267	AJ012581	Cytochrome P450
27	CESSRDB45	AGATGGTTTGAATGTTGAGG/ CACTTGACCCTTTGATTGTT	$(AT)_7(AG)_5$	59.5	295	AJ249802	Cytochrome P450
28	CESSRDB47	ACGAAGAAAGTTCCTGTGAA/ ACCGAAAACCTGATTCATTA	$(TTA)_3TAA(TTA)_4$	57.6	240	AJ006767	Histone H1
29	CESSRDB51	ACTATTACAAGAGCCCACCC/ CATAATGGTAAGGAGGTGGA	$(CAA)_4$	62.4	297	AJ006770	Extensin
30	CESSRDB53	CCCTTAATCAATTCACCTCA/ GCTTCTTATCCAAACCAATG	$(TAA)_6$	59.5	197	AJ005947	No homology
31	CESSRDB54	AGTGTTGTGGGTTTCATTTC/ TTGATTTGCCAAAGTACACA	$(TTA)_5$	59.6	221	AJ005869	Trans-membrane channel protein
32	CESSRDB55	CGATTATCTCAACTTTTGGC/ ACATGCACACGACAAATAAA	$(TA)_5$, $(ACT)_5$	59.0	136	AJ005000	Transcriptional regulator
33	CESSRDB56	TGTCTGGAACAACAAGTGAG/ GCCAATCAGATTTCCTCTTA	$(ATG)_4$	58.4	247	CK149113	Myb family transcriptional factor
34	CESSRDB61	GCAGAATGGGAGATAATGAA/ TGCTGATTCTGATGTCTACG	$(CTT)_7$	60.0	233	CK149116	bZIP transcriptional factor

Table 2 continued

S. no.	Locus name	Primer sequence (5′ → 3′)	Motif	T_m (°C)	Expected size (bp)	GenBank no.	Putative function
35	CESSR14	GGCACAAGGTATCTCCACAA/ ATGCTTGCCTCAACCTCAGA	$(TGC)_6$	56.8	300	ES544474	Unknown protein
36	CESSR15	CATGACATCCTCAATCCTTGG/ TAGCGACAAATCTTAGCCGTAG	$(TGC)_3AGC(TGC)_2$	59.4	300	ES544475	Unknown protein
37	CESSR20	CGA AACTCGAACGTGCAAT/ TTTGGCGAATTTGAAAGGAG	$(GATTC)_6$, $(ATTTA)_5$	58.5	386	ES544478	Unknown protein
38	CESSR21	CCTCAACGCTCATTCTTCTTCT/ CCCCAAGGAACCATTCTAAGAT	$(CTT)_6$	60.5	233	ES544479	OSH1 related protein
39	CESSR23	CGCGTAAACGTTATTCTCTTCA/ CATCATTTCCCTTAGCATCCTT	$(TTC)_7$, $(CTT)_3$	58.5	399	EX151810	Glutathione peroxidase
40	CESSR25	CTATGGCAAAAAGCATCACAAG/ ATAGCCATGGCCACATTAAACT	$(CCG)_6$	60.3	363	ES544482	No homology
41	CESSR26	GGCAAAATCGAAATTCAACC/ TGATCAATGACAGTGTAGAAGG	$(CTT)_4$, $(CT)_{10}$	59.3	275	ES544483	No homology
42	CESSR30	TCGGACCACAAGAGCATCTA/ CGTGGAAGAAAGGAATGTTG	$(CT)_6TT(CT)_2TT(CT)_3$, $(CAC)_6$	57.8	388	ES544486	Actin-binding potein
43	CESSR31	ACGTAGGTTAAGGTTGCTGGTC/ TTCAACGTGTTCGAAAGCTC	$(AAG)_8T(AAG)$	58.0	113	ES544487	Unknown protein
44	CESSR34	CATTGCTCAAAGCCAATTCA/ TCGATGAATCGGAACAAACA	$(TAT)_4$, $(TGT)_6$	56.9	294	ES544488	No homology
45	CESSR42	TGGTTGAAGAAAAGAAGGTAGTG/ CGGTTCACTAATGCAAAAACCT	$(ACC)_5$	59.5	298	ES544489	Hsr203J homolog protein
46	CESSR43	CATTAAAGCTAGGAGTTTGTGCTG/ ACGGTACCATACCCGACTACAT	$(CTA)_4$	55.6	386	EX567535	Bimodular protein
47	CESSR47	GAGTTCCACATTGTCACAGGAA/ AATGCAACAGTCCTTGTGGATA	$(TTC)_5$	57.4	541	EX567643	Germin like protein
48	CESSR51	CACATGAACAGAAAAAGGGACA/ GCATGTTGAGCCAAAGCTAAAT	$(TTTG)_5$	58.9	205	EX567864	Pleckstrin domain related
49	CEESR61	CACTCTTCCCTCCCTTTCTTTA/ GAATCAGGGTAGGTTTGTTTGC	$(CT)_7T(CT)_2$	60.0	257	EX151660	Armadillo like helical protein
50	CESSR62	ACCAGCTGCTAGACCTGATGTT/ GCAATAAAACAAAATCCTCACACC	$(TGA)_5$, $(TAT)_3$	62.5	245	EX567512	RHG1protein
51	CESSR65	CTCCTCCACTCATCT TCATCTTC GAGAAGGTGTTTCCGGTAAAAGT	$(CTT)_2CTC(CTT)_2GG(CTT)$	58.9	352	EX567847	Hypothetical protein

Table 2 continued

S. no.	Locus name	Primer sequence (5′ → 3′)	Motif	T_m (°C)	Expected size (bp)	GenBank no.	Putative function
52	CESSR68	AATGGCCACCATTTTCTCATC/ AAACGTTCTTTCCATCCTTCTG	$(ACC)_6$	57.8	330	EX151762	Hypothetical protein
53	CESSR71	TTGTAGTTCTCCTCTCTCTCTCTC/ CATCAAAACCAAACCTATGGAG	$(CT)C(CT)_8$, $(CT)_6$, $(CT)_{11}$	59.2	295	EX567905	Unknown protein
54	CESSR72	ATTTCACTCCTCACTTCTCACC/ CACGAAAATCGGATGATTCAG	$(CT)_7$	58.5	345	EX151914	Unknown protein
55	CESSR73	TCTTCTCCCATTCGTTGTTGAT/ GATCTTCTGTTCCTCAGCCAAC	$(CTT)_3ATT(CTT)_6(GTT)_2(AT)_4$	57.0	363	EX151922	No homology
56	CESSR77	CCAACTTAAACTCATTTCGTCTCA/ CCAAGATGTGTTTTTGATGATG	$(GA)_2AA(GA)_4$, $(CAT)_4$	56.2	173	EX567970	DUF647 protein
57	CESSR78	ATTGCTGAGGCTGTGAATTGTA/ CCCAATACATCAAAGATAGATCG	$(TTAAA)_3$	55.0	373	EX567577	Lipid transfer protein
58	CESSR80	TCACCCTTTCTTCTTCAACTTC/ GAACGCATAAAATAGTCGCTGA	$(GA)_7$	61.2	260	EX151949	No homology
59	CESSR85	ATGTACTTGGTCTGGTCCGTCT/ ACCTTTTCGGCGTTTCTTTTAC	(GCT)4	62.5	279	EX151920	Hypothetical Protein
60	CESSR93	ACGAGAAGAGCATTGCATTTG/ TAACGGCCTCTTTAGTCTGCTC	$(AAC)_3AAG$	56.4	353	EX151839	Pex19 related protein

The designed primer pairs, microsatellite repeat motifs, annealing temperature, expected allele size (bp), Genbank accession numbers and their putative functions based on BLASTX results are mentioned. Serial nos 1–34 represent markers designed using the database whereas serial nos 35–60 represent markers designed based on the in-house generated ESTs

the presence of microsatellite motifs using the TROLL program (Castelo et al. 2002) and the same criteria as mentioned above was used.

Designing EST-SSR primers

Primers were designed using the Primer3.0 software (Rozen and Skaletsky 1997) and designated as CESSR (for in-house generated ESTs) and CESSRDB (for EST-SSRs generated using the database). The parameters used for primer design were: (1) primer length 18–24 bp with an optimum of 20 bp, (2) annealing temperature 50–60°C with an optimum of 55°C, (3) percentage GC in the range of 40–50, and (4) product size in the range of 100–400 bp. All the oligonucleotides were synthesized from Illumina Inc. (USA).

Amplification and detection of microsatellite alleles

All PCR amplifications of genomic DNA (including *Cicer* and legume species) were carried out in a 15 μl reaction volume in a BIORAD thermal cycler (Icycler) containing 40–50 ng of genomic DNA, PCR buffer (20 mM Tris–HCl, 50 mM KCl), 0.75 μM of each primer, 0.125 mM of each dNTP, 1.5 mM $MgCl_2$ and 0.5 U of *Taq* DNA polymerase (Life Technology, India). The following touchdown amplification profile was used: (1) initial denaturation 94°C 3 min, (2) 18 cycles of 94°C 50 s, 65°C 50 s, decreasing annealing temperature 0.5°C/cycle, 72°C 50 s, (3) 20 cycles of 94°C 50 s, 55°C 50 s, 72°C 50 s, and (4) final extension 72°C 7 min. The amplification products were separated on 6% polyacrylamide gels or 3% Metaphor agarose gels (Cambrex, USA) depending upon the size range, stained with ethidium bromide and analyzed using the gel documentation system AlphaImager 2200 (Alpha Innotech Corp., USA). Fragment sizes for each locus were evaluated using standard size markers. All *Cicer* species were scored in a binary matrix and analyzed using POPGENE version 1.32 (Yeh and Boyle 1997) and the UPGMA based dendrogram was constructed using NTSYS-pc Version 2.1 (Rohlf 1994).

Sequence analysis of amplified fragments

Size variant alleles from different *Cicer* and legume accessions were amplified and resolved on 6% PAGE gels. The bands were eluted, cloned into pGEM-T Vector (Promega) and transformed to DH5α cells. After blue-white selection, plasmids from putative recombinants were isolated using the alkaline lysis method (Sambrook et al. 1989). Sequencing reactions were performed as above. Four to six recombinants from each allele were sequenced and the sequences were aligned with the original chickpea sequence using CLUSTAL W (1.83).

Results

Development of functional EST-SSR markers

ESTs from two sources were used for development of the EST-SSR markers in this study. First, using the 1,309 EST sequences reported in the NCBI database till January 2007, representing approximately 0.76 Mb, 133 microsatellite motifs were identified. Second, a cDNA library from immature seed was constructed and used as a source of EST-SSRs. From the 822 seed ESTs generated in-house, 159 EST-SSRs (19%) were identified that clustered into a total of 99 consensus sequences possessing a total of 113 microsatellite motifs. Sequence analysis of the 246 SSR motifs from the two sources (133 + 113) revealed that 207 (84.1%) were perfect repeats, 29 (11.7%) were imperfect and ten (4.0%) were compound. The copy number of the dinucleotide repeat motifs at the perfect loci varied from 5 to 17 and the trinucleotide motifs from 4 to 14. A diverse range of SSR motifs was present which varied widely with trinucleotide repeats (51.2%) being the most abundant followed by di- (37.3%), tetra- (6.9%) and pentanucleotide (4.4%) motifs. The most frequently occurring dinucleotide motifs were GA followed by TA and GT, whereas among trinucleotides AAG was predominant followed by ATT.

Using the flanking regions of 246 SSR motifs, 183 primers were designed of which 94 could be validated in chickpea acc. Pusa362 and ICCV2. Of these, 34 primers either did not amplify or produced anomalous-sized fragments. Therefore, finally 60 functional EST-SSR markers were developed (described in Table 2) of which 49 primer pairs produced single expected sized alleles, whereas 11 primers amplified 2–4 alleles. Based on the BLASTX analysis, putative functions could be assigned to majority (65.0%) of the EST-SSRs that showed significant homology to reported proteins, whereas 18.4% of them represented unknown/hypothetical proteins and 16.6% to novel sequences (Table 2).

Intra-specific diversity within chickpea revealed by EST-SSR markers and sequence analysis

To elucidate intra-specific variability, 60 chickpea EST-SSR primers (Table 2) were used to amplify the genomic DNA from 30 chickpea accessions listed in Table 1a. Of these 60 markers, ten produced polymorphic amplification profiles in the 30 accessions, amplifying a total of 29 alleles with a maximum of five alleles with the primer pair CESSR73 in the chickpea cultivars (Table 3). The observed heterozygosity

Table 3 Number of alleles (N_a), size range of amplified fragments, observed (H_o) and expected heterozygosity (H_E), Shannon's informative index (I) and fixation index (F_{IS}) values calculated for 30 chickpea accessions at ten polymorphic EST-SSR loci

Locus	N_a	Size range (bp)	H_O	H_E	I	F_{IS}
CESSR23	2	399–402	0.00	0.44	0.63	0.08
CESSR42	3	295–303	0.00	0.59	0.94	1.00
CESSR43	2	386–390	0.00	0.43	0.62	1.00
CESSR47	4	539–650	0.60	0.66	1.20	1.00
CESSR61	3	254–262	0.00	0.57	0.92	1.00
CESSR62	3	243–295	0.46	0.66	1.06	0.28
CESSR71	2	295–301	0.00	0.44	0.62	1.00
CESSR72	3	342–348	0.00	0.59	0.95	1.00
CESSR73	5	359–445	0.60	0.76	1.47	0.19
CESSR77	2	173–176	0.00	0.50	0.68	1.00
Average	2.9	–	0.16	0.56	0.91	–
SD	0.99	–	0.2711	0.1110	0.28	–

values ranged from 0 to 0.6 and expected heterozygosity ranged from 0.43 to 0.76 with an average of 0.56 (Table 3).

To assess the basis of length variation across chickpea, homologous alleles amplified by primer pair CESSSR73 were cloned and sequenced (Fig. 1). Sequence comparisons revealed the presence of a $(CTT)_n$ compound repeat motif. Allelic size variations were mainly due to the presence of an additional repeat motif (CTT) in some of the cultivars accompanied by single base insertions/deletions in the MFR regions. Cultivar specific insertions were observed at position 301 in ICCV2 and deletions at 96, 102 in ICC15406 and 286 in ICC7676. In addition, isolated point mutations were also observed in the MFRs such as at position 195 in ICC7676 (Fig. 1). However, when similar sized alleles from monomorphic loci such as CESSRDB13, CESSRDB27, and CESSRDB44 were sequenced, even though point mutations were observed at these loci, no variation in the copy number of repeats was found (data not shown).

Cross-species transferability across Cicer

In order to assess the transferability rates of the chickpea EST-SSR markers across related *Cicer* species, genomic DNA from nine accessions belonging to five wild, annual *Cicer* species (listed in "Materials and methods") were amplified with the same 60 EST-SSR primers used for intra-specific analysis. The transferability rates of chickpea EST-SSRs varied from a high of 96.6% in *C. reticulatum* to a low of 68.3% in *C. judaicum* with an average of 82.6% (Table 4). Forty-one markers (68.3%) amplified in all the annual species, of which, 27 were polymorphic across the wild *Cicer* species. Allelic data generated using the 60 EST-SSRs revealed a minimum of one and maximum of nine alleles (CESSRDB47) with a total of 156 alleles at 60 loci leading to an average of 2.6 alleles per locus. Observed heterozygosity (H_o) ranged 0.15–0.83 with an average of 0.22 (Table 4). Of the 60 primer pairs, two (CESSRDB3 and CESSRDB5) amplified only in chickpea (and in no other wild species) and were therefore specific to *C. arietinum*. The EST-SSR markers also exhibited crossability group-specific transferability. Among the first crossability group members (*C. arietinum*, *C. reticulatum*, and *C. echinospermum*), 55 markers (91.6%) were amplified and 24 markers produced polymorphic amplification profiles. Similarly, 41 markers (58.8%) successfully amplified in all the second crossability group members (*C. bijugum*, *C. judaicum*, and *C. pinnatifidum*) and 23 primers detected variation between the three species. Five markers (CESSRDB7, CESSRDB23, CESSRDB24, CESSRDB41, and CESSR26) exhibited crossability-group-specific transferability to only first crossability group members.

```
CESSR73
                   1111133333344444444445555555555566666666666777777777
          123456789012344567890123456789012345678901234567890123456789
          ------->
ICCV2     TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTT---GTTGTTATA
PUSA362   TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC15406  TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTT---GTTGTTATA
ICC283    TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC791    TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTGTTGTTATA
ICC7676   TCTTCTCCCATTCGAAACCTTCTTCTTATTCTTCTTCTTCTTCTTCTTGTTGTTATA
          *******************************************   ***********

                   1111111222222222222222222233333333333333333333333
          8888899999000000966666666888899999900011144455555666666666
          012344567890123456501234567678901234512367834556789012345678

ICCV2     TATATGCGTTGAACTCAGCCCACGG-AGTCATTGGGCCGG-TGG-TAGGAACAGAAGATC  (363bp)
PUSA362   TATATGCGTTGAACTCAGCCCACGG-AGTCATTGG-C-GG-TGG-TAGGAACAGAAGATC  (363bp)
ICC15406  TATATGC-TTGAA-TCAGCCCACGG-AGTCATTGG-C-GGGTGG-TAGGAACAGAAGATC  (359bp)
ICC283    TATATGCGTTGAACTCAGCCCACGG-AGTCATTGG-C-GG-TGG-TAGGAACAGAAGATC  (363bp)
ICC791    TATATGCGTTGAAGTGTGCCCTCCTTAGTCATTGG-C-GG-TGGGTAGGAACAGAAGATC  (365bp)
ICC7676   TATATGCGTTGAACTCAGACCACGGTA-TCATTGGGC-GGGTGGGTAGGAACAGAAGATC  (367bp)
          ******* ***** *   * ** *    * ******* * ** *** **************
                                                     <--------------
```

Fig. 1 Partial sequence alignment of size variant alleles amplified using primer pair CESSR73 across six chickpea accessions (EX151922, EU332161-EU332163, EU332165, EU332166). The *asterisks* represent similar sequences and *dash* indicates alignment gaps. Repeat regions are in *boldface*, primer-binding sites are represented by *arrows* and characters in *bold shaded boxes* indicate point mutations

Table 4 Number of alleles and their sizes (bp) obtained in six annual species of *Cicer* at 60 genic microsatellite loci are mentioned

Locus	*C. arietinum* No. of alleles (size in bp)	*C. reticulatum* No. of alleles (size in bp)	*C. echinospermum* No. of alleles (size in bp)	*C. bijugum* No. of alleles (size in bp)	*C. judaicum* No. of alleles (size in bp)	*C. pinnatifidum* No. of alleles (size in bp)	Transferability of each marker (%)	H_o	H_e
CESSRDB2	2 (137, 169)	1(169)	1 (169)	3 (129, 169, 222)	1 (129)	1 (129)	100	0.3333	0.7424
CESSRDB3	1 (197)	–	–	–	–	–	0		
CESSRDB4	1 (274)	1 (274)	1 (274)	1 (262)	1 (262)	1 (262)	100	0.0000	0.5455
CESSRDB5	1 (177)	–	–	–	–	–	0		
CESSRDB7	1 (196)	1 (196)	1 (196)	–	–	–	40	0.0000	0.0000
CESSRDB10	1 (192)	1 (192)	1 (192)	1 (179)	1 (182)	1 (173)	100	0.0000	0.7273
CESSRDB11	1 (298)	1 (298)	1 (298)	2 (298, 507)	–	–	60	0.2500	0.2500
CESSRDB13	1 (260)	1 (260)	2 (260, 334)	2 (227, 334)	3 (277, 334 343)	3 (277, 334, 343)	100	0.6667	0.7576
CESSRDB15	2 (261, 276)	2 (261, 276)	2 (261, 276)	1 (276)	1 (276)	1 (276)	100	0.5000	0.4091
CESSRDB16	2 (295, 346)	2 (323, 374)	2 (323, 374)	2 (323, 374)	–	1 (295)	80	0.8000	0.8000
CESSRDB18	1 (242)	3 (242, 279 315)	3 (252, 308 315)	3 (252, 300 324)	2 (252, 291)	2 (252, 291)	100	0.8333	0.8485
CESSRDB21	3 (259, 284, 296)	2 (259, 296)	3 (259, 284, 296)	1 (259)	2 (270, 296)	2 (270, 296)	100	0.8333	0.6818
CESSRDB23	1 (221)	1 (221)	1 (221)	–	–	–	40	0.0000	0.0000
CESSRDB24	1 (283)	1 (283)	1 (283)	–	–	–	40	0.0000	0.0000
CESSRDB26	1 (273)	1 (273)	1 (273)	1 (298)	1 (291)	1 (291)	100	0.0000	0.6667
CESSRDB27	1 (215)	1 (215)	1 (215)	1 (209)	1 (205)	1 (205)	100	0.0000	0.6667
CESSRDB29	1 (176)	1 (176)	1 (176)	1 (176)	1 (176)	1 (176)	100	0.0000	0.0000
CESSRDB33	1 (234)	1 (234)	1 (234)	1 (234)	1 (234)	1 (234)	100	0.0000	0.0000
CESSRDB34	1 (251)	1 (251)	1 (251)	2 (235, 251)	1 (251)	1 (251)	100	0.1667	0.1667
CESSRDB35	1 (272)	1 (272)	1 (272)	1 (272)	1 (272)	1 (272)	100	0.0000	0.0000
CESSRDB38	2 (197, 234)	2 (197, 234)	2 (197, 234)	2 (197, 234)	–	–	60	0.0000	0.0000
CESSRDB39	1 (257)	1 (257)	1 (257)	1 (257)	1 (257)	1 (257)	100	0.0000	0.0000
CESSRDB40	1 (188)	3 (188, 255 266)	2 (212, 272)	3 (199, 255 266)	–	3 (208, 266 272)	80	0.8000	0.9111
CESSRDB41	3 (247, 256, 278)	3 (251, 260, 281)	2 (251, 260)	–	–	–	40	1.0000	0.8667
CESSRDB42	2 (235, 280)	1 (235)	1 (235)	1 (235)	1 (235)	1 (235)	100	0.1667	0.1667
CESSRDB44	1 (267)	1 (267)	–	–	–	–	20	0.0000	0.0000
CESSRDB45	3 (295, 426, 449)	3 (295, 426, 449)	3 (295, 426, 449)	3 (295, 426, 449)	–	–	60	1.0000	0.5714
CESSRDB47	4 (240, 246 253. 276)	3 (240, 253 276)	4 (240, 246 253. 276)	4 (235, 248 258, 281)	3 (248, 268 281)	3 (248, 268 281)	100	1.0000	0.8182
CESSRDB51	1 (297)	1 (297)	1 (297)	1 (297)	1 (297)	1 (297)	100	0.0000	0.0000
CESSRDB53	1 (197)	1 (197)	1 (197)	1 (190)	1 (186)	1 (186)	100	0.0000	0.6667
CESSRDB54	2 (221, 305)	2 (221, 305)	2 (221, 305)	2 (221, 305)	–	–	60	1.0000	0.5714

Table 4 continued

Locus	*C. arietinum* No. of alleles (size in bp)	*C. reticulatum* No. of alleles (size in bp)	*C. echinospermum* No. of alleles (size in bp)	*C. bijugum* No. of alleles (size in bp)	*C. judaicum* No. of alleles (size in bp)	*C. pinnatifidum* No. of alleles (size in bp)	Transferability of each marker (%)	H_o	H_e
CESSRDB55	2 (136, 154)	2 (136, 154)	2 (143, 154)	2 (140, 168)	2 (140, 168)	2 (140, 155)	100	0.1515	0.7778
CESSRDB56	1 (247)	1 (247)	1 (247)	1 (247)	1 (247)	1 (247)	100	1.0000	0.8485
CESSRDB61	1 (233)	1 (233)	1 (233)	1 (233)	–	–	60	0.0000	0.000
CESSR14	1 (300)	1 (300)	1 (300)	1 (300)	1 (300)	1 (300)	100	0.0000	0.0000
CESSR15	1 (300)	1 (300)	1 (300)	1 (300)	1 (292)	1 (292)	100	0.0000	0.4848
CESSR20	1 (386)	1 (386)	1 (386)	–	–	1 (390)	60	0.0000	0.5333
CESSR21	1 (233)	1 (233)	3 (233, 300, 324)	3 (233, 300, 324)	1 (245)	1 (245)	100	0.3333	0.6667
CESSR23	1 (399)	1 (391)	1 (391)	1 (395)	1 (391)	1 (395)	100	0.0000	0.6667
CESSR25	1 (363)	1 (363)	1 (363)	1 (363)	1 (363)	1 (363)	100	0.0000	0.0000
CESSR26	1 (275)	1 (275)	1 (275)	–	–	–	40	0.0000	0.0000
CESSR30	1(388)	1(388)	1(388)	1(388)	1(388)	1(388)	100	0.0000	0.0000
CESSR31	1 (113)	1 (113)	2 (113, 108)	1 (113)	1 (113)	1 (113)	100	0.1667	0.16607
CESSR34	1 (294)	1 (287)	2 (294, 280)	1 (294)	2 (294, 280)	1 (294)	100	0.3333	0.4545
CESSR42	1 (298)	1 (305)	1 (298)	1 (305)	1 (305)	1 (298)	100	0.0000	0.5455
CESSR43	1 (386)	1 (395)	1 (395)	1 (395)	1 (395)	1 (395)	100	0.0000	0.3030
CESSR47	1 (541)	1 (541)	1 (552)	1 (552)	–	–	60	0.0000	0.5714
CESSR51	1 (205)	1 (209)	1 (205)	1 (209)	1 (205)	1 (209)	100	0.0000	0.5455
CESSR61	1 (257)	1 (257)	1 (257)	1 (262)	1 (257)	1 (252)	100	0.0000	0.5455
CESSR62	1 (245)	1(249)	1 (249)	1 (254)	1 (254)	1 (260)	100	0.0000	0.7879
CESSR65	1 (352)	1 (352)	1 (370)	1 (366)	1 (366)	1(366)	100	0.0000	0.6667
CESSR68	3 (330, 342, 350)	3 (330, 342, 350)	3 (322, 342, 350))	2 (322, 342)	2 (322, 342)	1 (322)	100	0.8333	0.6898
CESSR71	1 (295)	1 (286)	–	–	–	–	20	0.0000	0.6667
CESSR72	1 (345)	1 (348)	1 (345)	1 (350)	1 (350)	1 (350)	100	0.0000	0.6667
CESSR 73	1 (363)	3 (363, 370, 375)	–	1 (370)	–	–	40	0.3333	0.7333
CESSR 77	1 (173)	1 (170)	1 (170)	1 (170)	1 (170)	1 (170)	100	0.0000	0.3030
CESSR 78	1 (373)	2 (373, 500)	2 (366, 500)	2 (366, 500)	2 (370, 500)	2 (370, 500)	100	0.8333	0.1970
CESSR 80	1 (260)	1 (260)	1 (260)	1 (260)	1 (260)	1 (260)	100	0.0000	0.0000
CESSR85	1 (279)	1 (279)	1 (274)	1 (274)	1 (279)	1 (274)	100	0.0000	0.5455
CESSR93	1 (353)	1 (353)	1 (353)	1 (353)	1 (353)	1 (353)	100	0.0000	0.0000
Transferable markers (%)	100	96.6	91.6	83.3	68.3	73.3		Avg – 0.22	Avg – 0.35

Observed (H_o) and expected (H_e) heterozygosity values were estimated. Hyphen represents no amplification

Sequence analysis of size variant alleles from *Cicer* species

To investigate the basis of variation among size variant alleles of six *Cicer* species, fragments amplified in various annual species at the five loci, CESSRDB4, CESSRDB10, CESSRDB26, CESSRDB27 and CESSRDB34, were cloned and sequenced. Multiple alignments of nucleotide sequences from each locus were done and the results for three loci (CESSRDB4, CESSRDB10 and CESSRDB26) are shown in Fig. 2. Sequence comparisons at loci revealed that even though there was overall sequence conservation in the internal microsatellite structure and the primer-binding sites, variations such as differences in the copy number of repeat motifs and repeat interruptions accompanied by indels and point mutations in the microsatellite flanking regions (MFR) frequently occurred, all of which contributed to the allelic length variation (Fig. 2; Table 4). For example, at the CESSRDB4 locus, an addition of three repeat motifs (48–56 bp) accompanied by mutations at positions 26, 44 and 47 resulted in motifs specific for the first and second crossability group members (Fig. 2a). Similarly at the CESSRDB10 locus, a mutation (T → G) at nucleotide position 50 and a TAG repeat expansion among members of first crossability group were observed (Fig. 2b). At the CESSRDB26 locus the second crossability group alleles were much longer due to insertions in the MFR regions (Fig. 2c).

Another interesting feature revealed by sequence comparisons was the crossability-group-specific point mutations and indels. Point mutations (nucleotide positions 26, 44, 47 and 184 in CESSRDB4; positions 34, 45, 50, 115, 121, 143 and 164 in CESSRDB10; positions 92, 145, 193, 205, 210, 219, 223, 224, 234, 252, 266, 267 and 275 in CESSRDB26) and indels [12 bp (48–59) in CESSRDB4], [9 bp (68–76) and 4 bp (106–109) in CESSRDB10] and [14 bp (95–108), 1 bp (161) and (5 bp (255–259) in CESSRDB26] were highly crossability group specific. Additionally in the flanking regions, there were species-specific point mutations for example at position 103 in *C. arietinum* and 198 in *C. judaicum* at CESSRDB4 locus (Fig. 2).

Phylogenetic analysis

The allelic data obtained from the 60 chickpea EST-SSR markers were used to visualize the genetic relationships among the 30 chickpea accessions and the six annual *Cicer* species. After scoring and computing the allelic data, a dendrogram was constructed that clearly separated the members of the first and second crossability groups into clusters I and II (Fig. 3). Cluster I corresponded with the first crossability group members grouping all chickpea accessions into Cluster IA and the *C. reticulatum* and *C. echinospermum* into Cluster IB. The ClusterIA clearly distinguished all the chickpea accessions except ICC15518 and ICC8195; however, no correlation between the clustering pattern and geographical location was obtained. Cluster II represented the second crossability group species with *C. judaicum* and *C. pinnatifidum* being closely placed together.

Cross-genera transferability and sequence variation of chickpea EST-SSRs across legumes

Thirty-four of the chickpea functional markers were also utilized to assess their cross-genera transferability across 32 accessions spanning eight legume genera (ESM S1). This analysis revealed varied levels of marker transferability across legumes ranging from 29.4% in *P. mungo*, 35.2% in *P. sativum*, 41.1% in *G. max* and *T. alexandrinum*, 47.0% in *L. esculenta*, 50.0% in *C. cajan and* 61.7% in *M. truncatula* with an average of 43.6%. Eight markers (23.5%) amplified in all the legume genera though all the accessions of each species were not amplified. Twenty-five markers (74%) amplified in at least one legume species other than chickpea, whereas nine primers (26%) amplified only in the chickpea accessions and no other legume indicating the uniqueness of these loci to the chickpea genome. Twelve markers produced polymorphic amplification profiles across legumes even though intra-specific polymorphism was not observed.

Different sized alleles amplified at various loci across legumes were cloned and sequenced. For example, sequence analysis at CESSRDB56 (Fig. 4) demonstrated that although the same-sized alleles had high sequence conservation, variable alleles such as the 228 bp allele in *Trifolium* revealed polymorphism that was due to both differences of repeat motifs as well as variability of the flanking sequences marked by indels/point mutations. A similar observation was also noted at locus CESSRDB39 amplifying multiple alleles (data not shown).

Discussion

This study reports for the first time, development of a large number of EST-SSR markers in chickpea and assesses their transferability across a wide-spectrum of related species and genera, thereby establishing that the chickpea EST-SSR markers are a valuable genetic resource for investigating species relationships and comparative mapping in legumes. The easiest way to develop genic markers is by screening of EST sequences for the presence of hypervariable SSR motifs. In the publicly available EST database of legumes (Fabaceae) nearly 1 million EST

Fig. 2 Partial sequence alignment of size variant alleles amplified using (**a**) CESSRDB4 primer across six annual *Cicer* species (EF595573-EF595577) (**b**) CESSRDB10 primer across six annual *Cicer* species (EF595578-EF595582) and (**c**) CESSRDB26 primer across six annual *Cicer* species (EF595583-EF595587). The repeat motifs are represented as *bold letters*, *arrows* represent primer-binding sites, – Indicates alignment gaps and * represents similar sequences. Characters in *bold shaded boxes* indicate base substitutions and group specific mutations are demarcated by grey background. C.ari (*C. arietinum)*, C.ret (*C. reticulatum)*, C.ech (*C. echinospermum)*, C.bij (*C. bijugum)*, C.jud (*C. judaicum*), C.pin (*C. pinnatifidum*)

(A) CESSRDB4

```
               111111111122222222223333333333444444444455555555556
      123456789012345678901234567890123456789012345678901234567890
C.ari GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.ret GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.ech GAAGAGGTAGCGGAGGAGGTCGTGGTGGTGGTCGTGGTGGTGGCCGTGGTGGTGGTCGAG
C.bij GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
C.jud GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
C.pin GAAGAGGTAGCGGAGGAGGTCGTGGCGGTGGTCGTGGTGGTGGTCGA------------G
      ************************* ***************** **             *

        11111111111111111111111111111111111111111111222222222222222
      990000000000111111111128888888889999999999000066666666677777
      890123456789012345678901234567890123456789012312345678901234
C.ari GGAGAGGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.ret GGAGAAGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.ech GGAGAAGTATCATGCAGAGGCCAAGGAGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (274bp)
C.bij GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (262bp)
C.jud GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGCTGGCCAAAACTGTTGCTTG   (262bp)
C.pin GGAGAAGTATCATGCAGAGGCCAAGGGGGCATTGTGATGGATGGCCAAAACTGTTGCTTG   (262bp)
      ***** ******************** ************* *******************
```

(B) CESSRDB10

```
              1233333344444555555555566666666667777777777888888888
      123456789034567856789012345678901234567890123456789012345678
C.ari CCCTTAATCATGTT-ATCTAAGAATAATAATAATAATAATAGTAGTAGTACCATATA
C.ret CCCTTAATCATGTT-GTCTAAGAATAATAATAATAATAGTAGTAGTAGTACCATATA
C.ech CCCTTAATCATGTT-GTCTAAGAATAATAATAATAATAGTAGTAGTAGTACCATATA
C.bij CCCTTAATCAACTT-GCCTAATAATAATAATAA---------------TAGTACCATATA
C.jud CCCTTAATCAACTT-GCCTAATAATAATAATAATAA------------TAGTACCATATA
C.pin CCCTTAATCATCTTTGCCTAATAATAATAATAATAATAA---------------------
      *********** **   **** ***********
            111111111111111111111111111111111111111111111111111111
      899999000011111111112222222222333333333344444468888888888999
      902345678901234567890123456789012345678901234564678901234567
C.ari ATCAAC----TACTGTTCCCTGTTGGAGGTTTTGTCCTCCTTTTTTTGTGGTTTGGATAA   (192bp)
C.ret ATCAAC----TACTGTTCCCTGTTGGAG-TTTTGTCCTC-TTTTTTTGTGGTTTGGATAA   (190bp)
C.ech ATCAAC----TACTGTTCCCTGTTGGAG-TTTTGTCCTC-TTTTTTTGTGGTTTGGATAA   (190bp)
C.bij ATAAGCAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (179bp)
C.jud ATCAGCAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (182bp)
C.pin -TAAACAATCTACTGCTCCCTATTGGAG-TTTTGTCCTC-TTTGTTTATGGTTTGGATAA   (173bp)
         * *    ***** ***** ****** ********** *** *** ************
```

(C) CESSRDB26

```
                           111111111111111111111111111111111111111
               111699999999000000000001222223333333333444444455666
      123456789012923456789012345678905678901234567890123456789012
C.ari GGTGCATTCTCTAAAA--------------GCGTACTATAC-------TGTTAATACACG
c.ret GGTGCATTCTCTTAAA--------------GCGTACTATAT-------TGTTAATACACG
C.ech GGTGCATTCTCTTAAA--------------GCGTACTATAT-------TGTTAATACACG
C.bij GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTTTCAAGCAAATGCATGTTCATACA-G
C.jud GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTACTATAT-------TGTTCATACA-G
C.pin GGTGCATTCTCTTCAAATCATGAATTTCAAGCGTACTATAT-------TGTTCATA---G
      ************  **              **** * *          **** ***   *

      122222222222222222222222222222222222222222222222222227222222
      900000000112222222222333334444444444555555555566666779999999
      303456789090123456789012340123456789012345678906783453456789
C.ari TGGTGGTGTAATACGGAAAGAAAGTTAGGTGTGTGTGTTGA-----ACGACATATTTGCA   (273bp)
C.ret TGGTGGTGTAATACGGAAAGAAAGTTAGGTGTTTGTGTTGA-----ACGATATATTTGCA   (273bp)
C.ech TGGTGGTGTAATACGGAAA--AAGTTGGGTGTTTGTGTTGA-----ACGACATATTTGCA   (271bp)
C.bij GGGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTAGAATTATAGATTTGCA   (298bp)
C.jud GAGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTCGAATTATAGATTTGCA   (291bp)
C.pin GGGTCGTGTTCTACATAAAGAAAGTCAGGTGTTTGTGTCGAGTCGAATTATAGATTTGCA   (289bp)
        ** ****  ***  ***  ****  ***** ***** **     *  * * *******
```

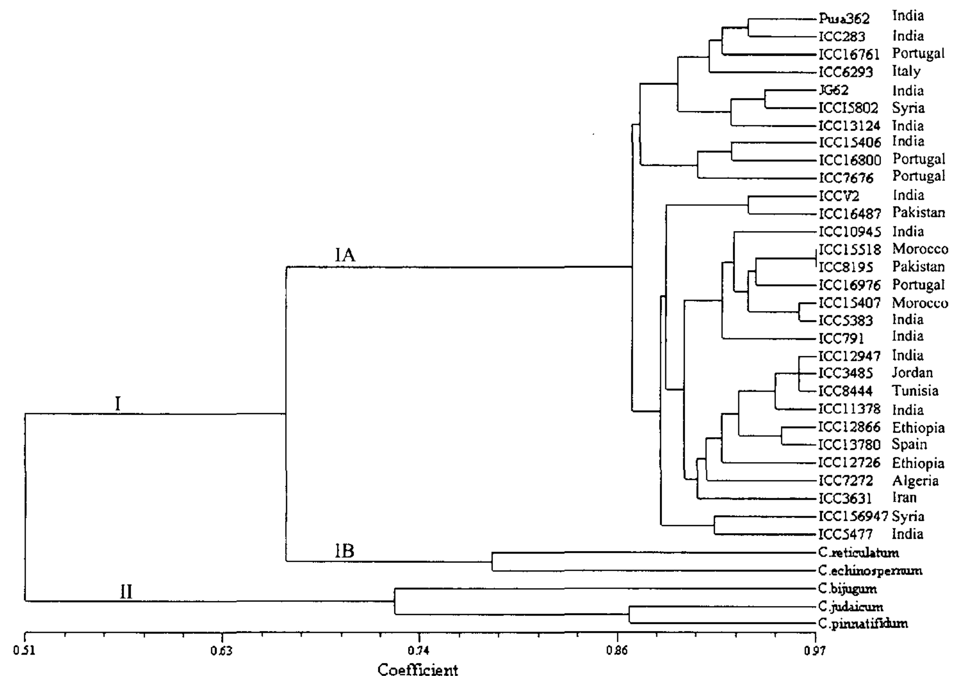


Fig. 3 UPGMA based dendrogram of thirty chickpea accessions and five annual *Cicer* species was obtained using 60 functional EST-SSR markers and Jaccard's coefficient. Name of cultivars, species and source country are mentioned

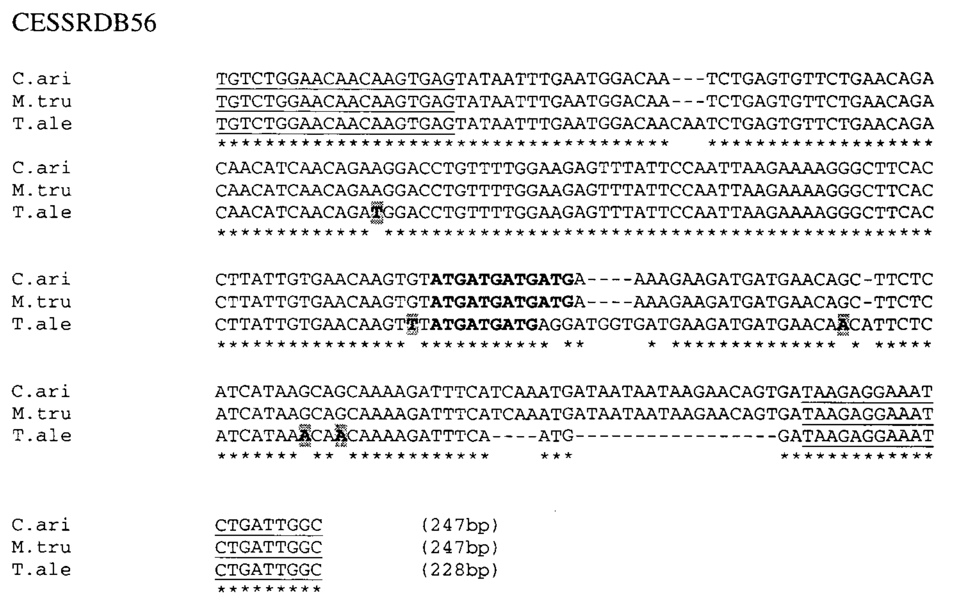


Fig. 4 Multiple sequence alignment of alleles amplified from chickpea and two legumes namely *M. truncatula* (M.tru) and *T. alexandrinum* (T.ale) at CESSRDB56 locus (EF621420 and EF595632). These primers also amplified alleles from legumes which were of the same size as chickpea and also shared high sequence homology, hence are not shown. Only the size variants are shown. The *asterisks* represent similar sequences, *dash* indicates alignment gaps. Repeat regions are in *boldface*, primer-binding sites are represented by *underlined letters* and characters in *bold shaded* boxes indicate point mutations

sequences are available, of which over 92% represent ESTs derived from *M. truncatula*, *L. japonicus* and *G. max* (Ramírez et al. 2005). For chickpea, only about 1,300 ESTs were publicly available (upto January 2007). Hence, our study utilized this resource for developing EST-SSR markers. Moreover, since only a limited number of ESTs were available, our study also undertook to generating new EST sequences and using them for the development of a novel set of functional markers. These markers will not only be a significant addition to the limited set of SSR markers available in chickpea, but will have the added advantage of marker-trait associations.

From the publicly available database and in-house ESTs, a total of 246 SSRs were identified which represented 11.5% of the screened ESTs. This SSR frequency was comparable with those obtained in citrus (10.6%, Chen et al. 2006), pepper (10.7%, Yi et al. 2006) and in other dicot species (Kumpatia and Mukhopadhyay 2005).

However, Kantety et al. 2002 obtained comparatively lower frequency of EST-SSRs ranging from 1.5 to 4.7% in monocots. The abundance of SSRs mined from a sequence database depends on the SSR search criteria, the size of the dataset and the database mining tools (Varshney et al. 2005a). On applying stringent SSR criteria with a minimum of 20 bp, about 5% of ESTs have been shown to contain SSRs in plants (Varshney et al. 2005a) whereas the same when applied to this set of chickpea EST sequences, only 3.1% sequences contained SSRs. The abundance of trinucleotide motifs in the chickpea coding sequences (51.2%) was in close agreement with observations in monocot and dicot plants (Kantety et al. 2002; Tian et al. 2004; Yi et al. 2006) establishing the need of the coding regions to maintain the reading frame (Varshney et al. 2002; Li et al. 2004). The predominance of GA motifs among dinucleotides in the chickpea ESTs was similar to reports in cereals (Varshney et al. 2002) and dicots like *Medicago*, soybean and *Arabidopsis* (Tian et al. 2004). Similarly among trinucleotides, the abundance of AAG motifs in chickpea was quite consistent with the findings of Li et al. (2004) and Kumpatia and Mukhopadhyay (2005). However, earlier studies on chickpea microsatellites have reported the $(TAA)_n$ motif to be most abundant (Udupa et al. 1999). Moreover, it was observed that even though the EST-microsatellites contained less number of repeat motifs than the genomic microsatellites (gSSRs) reported earlier (Sethy et al. 2006a), they proved to be highly informative in the genetic diversity and cross-species transferability studies (Scott et al. 2000; Thiel et al. 2003).

Chickpea has been shown to exhibit overall low levels of polymorphism with the various molecular markers analyzed so far and this has been attributed to its self-pollinating nature as well its recent worldwide dispersal (Udupa et al. 1999). In this study also, the EST-SSR markers displayed a low level of polymorphism (16%) within chickpea accessions in comparison to earlier reports of 40–60% polymorphism detected by gSSRs (Sethy et al. 2006a; Lichtenzveig et al. 2005). This observation is noteworthy as SSRs located in the coding regions are under strong selection pressure and therefore accumulate few mutations (Varshney et al. 2005a; Li et al. 2004). However, despite the lower polymorphism, the genic-SSRs are preferable over gSSRs as these are associated with the coding regions of the genome and therefore represent "true genetic diversity" that would directly assist in "perfect" marker-trait associations (Eujayl et al. 2002; Thiel et al. 2003). Other species such as rice (Cho et al. 2000), sugarcane (Cordeiro et al. 2001) and wheat (Gupta et al. 2003), have also revealed similarly low levels of polymorphism using EST-SSRs compared to genomic SSR markers.

In contrast to the low levels of intra-specific polymorphism with EST-SSRs, the inter-specific polymorphism was significantly higher in the wild relatives of chickpea. Wild genepool is extremely valuable in inter-specific hybridization programs since they serve as sources of resistance/ tolerance to many stresses. Our study with EST-SSR markers will potentially facilitate the transfer of traits of agronomic value into cultivated chickpea thereby leading to the broadening of the narrow genetic base and development of superior genotypes of chickpea. The dendrogram obtained with the EST-SSR markers clearly showed the closeness of *C. judaicum* with *C. pinnatifidum* (Fig. 3) which was in agreement with the earlier protein based (Tayyar and Waines 1996) and EST-based studies (Buhariwalla et al. 2005), whereas using DNA-based marker systems like AFLP (Shan et al. 2005) and STMS markers (Sethy et al. 2006b) the closeness of *C. pinnatifidum* with *C. bijugum* has been reported. The resemblances of the dendrograms based on protein markers with genic markers suggest that coding sequences of *C. judaicum* and *C. pinnatifidum* may have followed a common evolutionary pathway.

The chickpea EST-SSRs developed in this study revealed much higher rates of transferability (mean 82.6%) across wild annuals than the chickpea-derived gSSRs (68%; Choumane et al. 2000). Higher inter-specific transferability was in accordance with other studies (Scott et al. 2000; Eujayl et al. 2004; Zhang et al. 2005), establishing that functional markers were more transferable and therefore more useful than gSSR markers with the added potential of being used in allele-mining for identification of useful agronomic traits. It was also observed that the mean transferability rates across the primary and secondary crossability groups were an average of 96.0 and 74.9%, respectively. This difference could be explained on the basis of an earlier study of Decroocq et al. (2003), which said that the level of sequence conservation of microsatellite loci is inversely proportional to the genetic distance. Similar observations have been made in other species such as wheat (McLauchlan et al. 2001) and sugarcane (Cordeiro et al. 2001) where the genic markers displayed low level of polymorphism in cultivated accessions compared to other members of the genus thereby directing the breeders to look into the related species for introgression of novel genetic material into the germplasm.

Results from the sequencing data also provided evidence for limited sequence variability within the chickpea alleles in comparison to much higher levels of variation across the orthologous alleles from annual species. Sequence comparisons of size variant microsatellite alleles within chickpea accessions illustrated approx. 95% overall sequence conservation with few indels in the repeat as well as the MFR region (Fig. 1) suggesting the presence of evolutionary constraints within transcribed regions that limit the mutational events and increase sequence similarity. However, sequence comparison of microsatellite

alleles from various *Cicer* species revealed a wide range of length and sequence variability both in terms of band size and allele number. Similar results have also been obtained by Buhariwalla et al. 2005 thereby establishing that EST based microsatellite markers of chickpea were not only efficient for marker-assisted introgression programs using wild germplasm but also reliable for synteny studies within the genus *Cicer*. Sequence variations occurred both at the repeat motifs and in the flanking regions and were interestingly found to be crossability-group-specific and therefore highly phylogenetically informative that could help in understanding the evolution of microsatellites in a phylogenetic context since it has been shown that such events at the genic loci might play an important role in speciation or gene functionality diversification during the evolutionary process.

Earlier studies have provided evidence, which shows that microsatellites undergo expansion during the course of evolution (Zhu et al. 2000; Peakall et al. 1998). In our study, there was an expansion of the GGT motif at the locus CESSRDB4 (Fig. 2a), resulting in the presence of three additional repeats in members of the first crossability group. At the locus CESSRDB10, expansion of the TAA motif was accompanied by the birth of a new TAG motif (via a A → G mutation) which later expanded in the members of the first crossability group. Such A → G transition was also observed by Messier et al. (1996) in owl monkey. It has been speculated that base substitution allows the birth of new motifs that subsequently expand by replication slippage (Gordon 1997). Recently, the role of microsatellite expansion/deletion in terms of gene regulation is being investigated well in mammals as well as in plants (Li et al. 2004). The presence of SNPs in the sequence of similar sized alleles from different chickpea cultivars apparently indicated the limitation of scoring the accessions simply based on the amplicon size on gel. Also, this clearly highlights the prospects of SNP mapping in chickpea as these represent the most fundamental source of variation for molecular marker development.

In our study it was observed that the microsatellite motifs were long but punctuated by imperfections which are most often regarded as an effective mechanism for prevention of infinite growth of microsatellites (Kruglyak et al. 1998; Palsboll et al. 1999). At the CESSRDB4 and CESSRDB10 loci, base substitutions at positions 44 and 50, respectively, in the first crossability group members implied that such interruptions may have a dramatic impact in the long-term evolution of the microsatellite sequence. On the other hand, the phenomenon of microsatellite purification (loss of interruptions), a mechanism counteracting the accumulation of imperfections is also known to occur (Harr et al. 2000). This was observed at locus CESSRDB26 (Fig. 2c) where all species, except *C. arietinum*, harbor "T" at position 246 indicating that T represents the ancestral character state. The occurrence of longer motifs in the focal species in comparison to the related species may also be explained by the hypothesis of ascertainment bias (Ellegren et al. 1997; Peakall et al. 1998; Vigouroux et al. 2002) which in our study was demonstrated by sequence comparisons at the CESSRDB4 and CESSRDB10 loci (Fig. 2a, b). The cross-transferability of chickpea EST-SSR markers across legume species was high (mean 43.6%) clearly depicting the conservation of primer-binding sites in genomic DNA over a long evolutionary period. The usefulness of EST-SSR markers over genomic SSRs for transferability across distant relatives has been established in species such as *Medicago* (Gutierrez et al. 2005), wheat (Gupta et al. 2003), barley (Thiel et al. 2003) and grapes (Scott et al. 2000). However in chickpea, no extensive study of cross-genera transferability of genomic SSRs was available, except for a small study by Pandian et al. (2000), the transferability rates across distant species of genomic versus EST-SSRs could not be compared. Our study showed that the highest rate of transferability of the chickpea EST-SSR markers was to *Medicago* (61.7%), whereas an earlier study by Gutierrez et al. (2005) showed significant, yet lower levels of transferability of the *Medicago* markers to chickpea (36.3%). The difference in the rates of transferability could be attributed to the choice of loci and the overall number of markers analyzed. Our study also demonstrated that the rate of transferability decreases from within the genus *Cicer* (82.6%) to outside the genus (43.6%) which was in agreement with earlier reports in cereals (Thiel et al. 2003; Gupta et al. 2003), grapes and apricot (Decroocq et al. 2003) and *Medicago* (Gutierrez et al. 2005) suggesting that amplification decreases with increasing evolutionary distance from focal species. Overall, the chickpea markers transferred very efficiently to some members of the galegoid legumes (such as *Medicago* and *Trifolium*) as compared to the phaseoloid legumes (such as *P. mungo*). However, *Pisum* and *Cajanus* were exceptions to this. The variable marker transferability rate obtained across different legume genera indicated the occurrence of genus-specific evolutionary events.

In conclusion, our study was the first attempt at characterization of a large number of SSRs from the coding regions of the chickpea genome. This study not only contributed to strengthening the chickpea EST database but also provided the first set of functional SSR markers for evaluating the chickpea germplasm and molecular mapping. In this study, it was established that the chickpea EST-SSRs were highly transferable across a number of distantly related species thereby providing ample opportunity for mining of superior alleles and development of candidate gene markers for use in gene introgression programs and comparative genomics in legumes. Further, our study also provided the molecular evidence for understanding the basis of allelic variation

within and across species, which demonstrated the presence of complex mutational processes, highlighting the evolution of microsatellites in a phylogenetic context within the genus *Cicer*.

Acknowledgments We are grateful to the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) India, National Bureau for Plant Genetic Resources (NBPGR) India, Maharana Pratap Agriculture University (MPAU) India, and Australian Medicago Genetic Resource Centre, SARDI, Australia for providing us the seed material of different accessions of cultivated chickpea, wild *Cicer* species and other legumes. Financial assistance for this work was provided by National Institute for Plant Genome Research (NIPGR), New Delhi, India and also the Department of Biotechnology (DBT), Government of India by means of a project grant. The fellowship provided to SC by University Grants Commission (UGC), India and to NK and BS by Council for Scientific and Research (CSIR), India is gratefully acknowledged.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic markers for genome analyses of coffee and related species. Theor Appl Genet 114:359–372

Boominathan P, Shukla R, Kumar A, Manna D, Negi D, Verma PK, Chattopadhyay D (2004) Long term transcript accumulation during the development of dehydration adaptation in *Cicer arietinum*. Plant Physiol 135:1608–1620

Buhariwalla HK, Jayashree B, Eshwar K, Crouch JH (2005) Development of ESTs from chickpea roots and their use in diversity analysis of the *Cicer* genus. BMC Plant Biol 5:16

Castelo AT, Martius W, Gao GR (2002) TROLL—Tandem repeat occurrence locator. Bioinformatics 18:634–636

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006) Mining and characterizing microsatellites from citrus ESTs. Theor Appl Genet 112:1248–1257

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). Theor Appl Genet 100:713–722

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci USA 101:15289–15294

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2006) Development of sequence-tagged microsatellites site markers for chickpea (*Cicer arietinum* L.). Mol Eco Notes 6:93–95

Choumane W, Winter P, Weigand F, Kahl G (2000) Conservation and variability of sequence-tagged microsatellite sites (STMSs) from chickpea (*Cicer arietinum* L.) within the genus *Cicer*. Theor Appl Genet 101:269–278

Coram T, Pang E (2005) Isolation and analysis of candidate ascochyta blight defense genes in chickpea, Part I. Generation and analysis of an expressed sequence tag (EST) library. Physiol Mol Plant Pathol 66:192–200

Cordeiro GM, Casu R, McIntyr CL, Manners JM, Henry RJ (2001) Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci 160:1115–1123

Decroocq V, Favé MG, Hagen L, Bordenave L, Decroocq S (2003) Development and transferability of apricot and grape EST microsatellite markers across taxa. Theor Appl Genet 106:912–922

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

Ellegren H, Moore S, Robinson N, Byrne K, Ward W, Sheldon BC (1997) Microsatellite evolution—a reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol Bio Evol 14:854–860

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross species genetic markers for *Medicago* spp. Theor Appl Genet 108:414–422

Gepts P, Beavis WD, Brummer EC, Shoemaker RC, Stalker HT, Weeden NF, Young ND (2005) Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through Genomics Conference. Plant Physiol 137:1228–1235

Gordon AJE (1997) Microsatellite birth register. J Mol Evol 45:337–338

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSRs markers for the study of polymorphism and genetic diversity in bread wheat. Mol Gen Genet 270:315–323

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT, Torres AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. Theor Appl Genet 110:1210–1217

Harr B, Zangerl B, Schlötterer C (2000) Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. Mol Bio Evol 17:1001–1009

Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. Genome Res 9:868–877

Hüttel B, Winter P, Weising K, Choumane W, Weigand F, Kahl G (1999) Sequence-tagged microsatellite markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217

Kantety RV, Rota ML, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, sorghum and wheat. Plant Mol Biol 48:501–510

Kruglyak S, Durrett RT, Schug M, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Kumpatia SP, Mukhopadhyay S (2005) Miming and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48:985–998

Li Y-C, Korol AB, Fahima T, Beiles AV, Eviatar Nevo (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Eco 11:2453–2465

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Lichtenzveig J, Scheuring C, Dodge J, Abbo S, Zhang HB (2005) Construction of BAC and BIBAC libraries and their applications for generation of SSR markers for genome analysis of chickpea, *Cicer arietinum* L. Theor Appl Genet 110:492–510

McLauchlan A, Henry RJ, Isaac PG, Edwards KJ (2001) Microsatellites analysis in cultivated wheat and wild relatives. In: Henry RJ (ed) Plant genotyping-the DNA fingerprinting of plants. CABI publishing, Wallingford

Messier W, Li SH, Stewart CB (1996) The birth of microsatellites. Nature 381:483

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Palsboll PJ, Berube M, Jorgensen H (1999) Multiple levels of single-strand slippage at cetacean tri-and tetra nucleotide repeat microsatellite loci. Genetics 151:285–296

Pandian A, Ford R, Taylor PWJ (2000) Transferability of sequenced tagged microsatellite site (STMS) primers across four major pulses. PMB Reporter 18:395a–395h

Peakall R, Glimore S, Keys W, Morgante M, Rafalski A (1998) Cross species amplification of soybean (*Glycine max*) simple sequence repeat (SSRs) within the genus and other legume genera: implication for transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko AD, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. Theor Appl Genet 276:436–449

Ramírez M, Graham MA, Blanco-López L, Silvente S, Medrano-Soto A, Blair MW, Hernández G, Vance CP, Lara M (2005) Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol 137:1211–1227

Rohlf FJ (1994) NTSYS-pc: numerical taxonomy and multivariate analysis system. Applied Biostatistics, New York

Romo S, Labrador E, Dopico B (2004) Water stress-regulated gene expression in *Cicer arietinum* seedlings and plants. Plant Physiol Biochem 39:1017–1026

Rozen S, Skaletsky HJ (1997) Primer 3. Code available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, New York

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Sethy NK, Shokeen B, Bhatia S (2003) Isolation and characterization of sequence-tagged microsatellite sites markers in chickpea (*Cicer arietinum* L.). Mol Ecol Notes 3:428–430

Sethy NK, Shokeen B, Edwards KJ, Bhatia S (2006a) Development of microsatellite markers and analysis of intra-specific genetic variability in chickpea (*Cicer arietinum* L.). Theor Appl Genet 112:1416–1428

Sethy NK, Choudhary S, Shokeen B, Bhatia S (2006b) Identification of microsatellite markers from *Cicer reticulatum*: molecular variation and phylogenetic analysis. Theor Appl Genet 112:347–357

Shan F, Clarke HC, Plummer JA, Yan G, Siddique KHM (2005) Geographical patterns of genetic variation in the world collections of wild annual *Cicer* characterized by amplified fragment length polymorphisms. Theor Appl Genet 110:381–391

Tayyar RI, Waines JG (1996) Genetic relationships among annual species of *Cicer* (Fabaceae) using isozymes variation. Theor Appl Genet 92:245–254

Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422

Tian AG, Wang J, Cui P, Han YJ, Xu H, Cong LJ, Huang XG, Wang XL, Jiao YZ, Wang BJ, Wang YJ, Zhang JS, Chen SY (2004) Characterization of soybean genomic features by analysis of its expressed sequence tags. Theor Appl Genet 108:903–913

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*C. arietinum* L.) germplasm. Mol Gen Genet 261:354–363

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol Biol Lett 7:537–546

Varshney RK, Graner A, Sorrells ME (2005a) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Sigmund R, Börner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005b) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195–202

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19:1251–1260

Weeden NF, Muehlbauer FJ, Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83:123–129

Winter P, Benko-Iseppon AM, Hüttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santra D, Sant VJ, Rajesh PN, Kahl G, Muehlbauer FJ (2000) A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5. Theor Appl Genet 101:1155–1163

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157

Yi G, Lee JM, Lee S, Choi D, Kim BD (2006) Exploitation of pepper EST-SSRs and an SSR-based linkage map. Theor Appl Genet 114:113–130

Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. Theor Appl Genet 111:677–687

Zhu Y, Queller DC, Strassmann JE (2000) A phylogenetic perspective on sequence evolution in microsatellite loci. J Mol Evol 50:324–338

Molecular Ecology Notes (2006) 6, 93–95 doi: 10.1111/j.1471-8286.2005.01150.x

PRIMER NOTE

Development of sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.)

SHALU CHOUDHARY, NIROJ K. SETHY, BHUMIKA SHOKEEN and SABHYATA BHATIA
National Centre for Plant Genome Research, Jawaharlal Nehru University Campus, Post Box Number 10531, New Delhi 110067, India

Abstract

Microsatellite loci were identified from chickpea (*Cicer arietinum* L.), the third most important grain legume crop in the world. A total of 13 sequence-tagged microsatellite markers were developed using two different approaches: (i) amplification using degenerate primers and (ii) cloning of intersimple sequence repeat (ISSR)-amplified fragments. Thirty-five chickpea accessions were analysed, which resulted in a total of 30 alleles at the 13 loci. The observed heterozygosity ranged from 0.1143 to 0.4571 with an average of 0.2284. The cross-species transferability of the sequence-tagged microsatellite site (STMS) markers was checked in *Cicer reticulatum*, the wild annual progenitor of chickpea. These microsatellite markers will be useful for assessing the genetic diversity patterns within chickpea as well as aid in construction of intra- and interspecific genetic linkage maps.

Keywords: chickpea, *Cicer reticulatum*, mapping, microsatellites, STMS

Received 16 June 2005; revision received 8 July 2005; accepted 25 July 2005

Chickpea (*Cicer arietinum* L.) is a major source of plant-based dietary protein and ranks third in terms of world grain legume production. It is majorly cultivated in the Mediterranean region, Indian subcontinent, West Asia, North Africa and more recently in Canada and Australia. The plant fixes atmospheric nitrogen and is a suitable rotation crop for agricultural practices. Despite its agronomic importance, the productivity of chickpea is still low and the conventional breeding methods have failed to bring any significant increase in the crop yield. Marker-assisted breeding holds the promise of potential crop improvement in chickpea. In this context, microsatellite markers have proved their usefulness in genotype identification, germplasm classification (Udupa *et al.* 1999) and genetic linkage map construction (Winter *et al.* 1999) in chickpea. However, the number of available microsatellite markers for chickpea is still less as compared to other important legume crops and there is a need to isolate more markers for chickpea genome analysis and map saturation. In this study, we report the isolation and characterization of 13 new microsatellite markers from chickpea using two approaches. The cross-species transferability of these markers to *Cicer reticulatum*, the wild annual progenitor of chickpea, has also been demonstrated.

The first approach used for developing the sequence-tagged microsatellite site (STMS) markers consisted of polymerase chain reaction (PCR) amplification of *C. arietinum* cv. Pusa362 genomic DNA using a single degenerate primer (5′-NNNMMHYHYHGGTTGGTTGGTT-3′) where N = A/C/G/T, M = A/C, H = A/C/T and Y = T/C (Kumar *et al.* 2002). The amplified product was cloned into pGEM-T vector (Promega) and transformed into DH5α competent cells. In the second approach, the genomic DNA of *C. arietinum* cv. Pusa362 was amplified using six 3′-anchored intersimple sequence repeat (ISSR) primers namely UBC810, UBC825, UBC855, UBC856, UBC858 and UBC859 (Ratnaparkhe *et al.* 1998). The 16 major bands ($UBC810_{343}$, $UBC825_{692,708,736}$, $UBC855_{302,661,747}$, $UBC856_{315,507,560,746}$, $UBC858_{420,650}$, $UBC859_{448,596,752}$) were eluted from 1.25% agarose gels and cloned as mentioned earlier. The plasmids from putative recombinants obtained after blue-white selection with both the methods were isolated using alkaline lysis method (Sambrook *et al.* 1989) and sequenced with BigDye Terminator Reaction Kit (Applied Biosystems) on the ABI PRISM 3700 Automated DNA Sequencer. The 45 recombinants sequenced using the first approach and 16 recombinants from the second approach contained

Shalu Choudhary and Niroj K. Sethy equally contributed to this work.
Correspondence: Sabhyata Bhatia, Fax: +91-11-26716658; E-mail: sabhyatab@yahoo.com

© 2006 Blackwell Publishing Ltd

Table 1 Characteristics of the *Cicer arietinum* microsatellite loci. Number of alleles, observed (H_O) and expected (H_E) heterozygosities values calculated for chickpea accessions ($H_E = 1 - \Sigma(p_i)^2$). Primers CaSSR1–CaSSR7 were developed using degenerate primer method and CaSSR8–CaSSR13 using ISSR band cloning method

Locus	Primer sequence (5′–3′)	Motif	T_a	No. of alleles	Expected size (size range)	H_O	H_E	GenBank no.
CaSSR1	F: TCAAAGGCAGACGTGTAGAA R: TAGAGGAAGATTTCGGAGGA	$(GA)_3(CAA)_2CG$ $(TA)_2(TAA)_2$	55.4	3	233 (209–233)	0.2571	0.6679	DQ071843
CaSSR2	F: GCCTACATTGCTTTCCCTTT R: TCATGTGTGTATGAAGTGGAATGA	$(GAAT)_4(GTAT)_2$	54.6	4	244 (244–290)	0.1765	0.1901	DQ071844
CaSSR3	F: CTGTTTATTTGGCCCATTCT R: CGACAATTTCAACTTTTGGT	$(ATT)_8$	55.7	2	296 (296–383)	0.1143	0.1093	DQ071845
CaSSR4	E: ATAGTGGCATATTGGGGAGA R: TGAAACCCTAGAGTGGTTGTT	$(CAA)A(CAA)_4$ $G(CAA)_4G(CAA)$	56.0	3	187 (187–255)	0.1143	0.1093	DQ071846
CaSSR5	F: AAACCAAAACTGAAGTTAATAGGG R: GAAAGAAGTGAAAAAGTAGTGGAA	$(TTA)_8$	55.3	3	292 (292–304)	0.3571	0.2987	DQ071847
CaSSR6	F: TTGGAATTGAGGGACAGATT R: TGTGCCCTACTGAAAATCAA	$(CTAT)_3$	57.2	1	166	0.0000	0.0000	DQ071848
CaSSR7	F: GCTCAAGGCTGAAGGAGATA R: ACCCTGCAAGTCAAGTCTTC	$(TA)_2(GT)_2(GA)_2$ $GTT(ATT)_3$	56.8	2	223 (223–278)	0.1714	0.1590	DQ071849
CaSSR8	F: TTAAGCTTCCATTCGAGAAG R: CGTAACTCAACCACATCAAA	$(AT)_5(GTT)_4$	55.3	1	277	0.0000	0.0000	DQ071850
CaSSR9	F: TGAGAATCACTTTTGGCTCT R: ATCCTGATGAAGGTCGTATG	$(GA)_2C(GA)$ $(CAT)_2$	56.5	2	233 (204–233)	0.4571	0.3578	DQ071851
CaSSR10	F: TCCTATCACTCACCAGAAGG R: TTAGGATTACGGCAAGTAGC	$(CTT)CAT(CTT)_3$	55.9	4	300 (281–318)	0.2941	0.7256	DQ071852
CaSSR11	F: GGATGCTATCAATTTATGATGTGTG R: CTCGTTGGCCTATGTTACTC	$(GT)_2(GAT)TAAT$ $(GAT)TGT(GAT)_2$	57.6	2	151 (138–151)	0.1429	0.1346	DQ071853
CaSSR12	F: TGGGTCACTGATACAAGGTG R: TGTTGACCTAAGCAAGCAAG	$(TA)ATTTG(TA)_3$ $TGT(TA)$	55.3	1	174	0.0000	0.0000	DQ071854
CaSSR13	F: GCTTTTTGGAAGCTGAAGTG R: CCCTTTTCTGTTTCCATTTG	$(GCA)_5(GCT)_7$	55.2	2	177 (177–257)	0.2000	0.1826	DQ071855

microsatellite motifs at both 5′ and 3′ termini. However, seven and six internal microsatellite motifs were identified from the first and second approaches, respectively (Table 1). Homology searches revealed that CaSSR8 shared homology with *Arabidopsis thaliana* putative glucose 6-phosphate isomerase (BT000953, score 105, e-value 2e-19), while CaSSR13 was homologous with *A. thaliana* HECT ubiquitin–protein ligase 3 (AY265959, score 77.8, e-value 4e-11). Primers were designed based on the microsatellite-flanking regions with the help of PRIMER 3 software (Rozen & Skaletsky 1997).

The STMS markers developed using the degenerate primer method (CaSSR1–CaSSR7) and ISSR band cloning method (CaSSR8–CaSSR13) were used to assess the level of genetic polymorphism within a set of 35 chickpea accessions obtained from International Crop Research Institute for the Semi-Arid Tropics (ICRISAT), India. Amplification of the microsatellite loci was carried out in a 20 µL reaction volume in a Bio-Rad thermal cycler containing 30 ng genomic DNA, 1 × PCR buffer, 0.75 µM of each primer, 0.125 mM each dNTP and 0.5 U of *Taq* DNA polymerase (NEB). The following touchdown amplification profile was used: an initial denaturation at 94 °C for 3 min followed by 18 cycles consisting of denaturation at 94 °C for 50 s, annealing at 64 °C for 50 s with a 0.5 °C decrease in every subsequent cycle and extension at 72 °C for 50 s. This was followed by 20 cycles of amplification with denaturation at 94 °C for 50 s, annealing at 55 °C for 50 s and extension at 72 °C for 50 s. A final extension step was carried out at 72 °C for 7 min. Amplification products were electrophoresed on 6% polyacrylamide gels in 1× TBE, stained with ethidium bromide and analysed using the ALPHAIMAGER 2200 (Alpha Innotech) gel documentation system. The presence and absence of the band was scored as 0 or 1 and analysed using POPGENE version 1.32 computer software (Yeh & Boyle 1997).

All the 13 primer pairs produced bands of expected sizes in *C. arietinum* cv. Pusa362 (Table 1). Ten primer pairs produced polymorphic amplification profiles whereas CaSSR6, CaSSR8 and CaSSR12 produced monomorphic profiles in the 35 accessions of *C. arietinum* (Table 1). A minimum of two alleles for CaSSR3, CaSSR7, CaSSR9, CaSSR11 and CaSSR13 loci and maximum of four alleles for CaSSR2 and CaSSR10 loci were obtained, generating a total of 30 alleles with an average of 2.3 alleles per locus. The observed

Table 2 Cross-species amplification of microsatellite loci in three *Cicer reticulatum* accessions ICC17121, ICC17123 and ICC17163. Amplified fragment sizes were mentioned. NA indicates no amplification

Microsatellite locus	ICC17121	ICC17123	ICC17163
CaSSR1	233	233	233
CaSSR2	300	318	318
CaSSR3	NA	296	296
CaSSR4	187	187	187
CaSSR5	292	292	292
CaSSR6	166	166	166
CaSSR7	223,259,326	223,259,326	223,259,326
CaSSR8	277	277	277
CaSSR9	204	204	233
CaSSR10	318	287,318	318,345
CaSSR11	138	138	138
CaSSR12	174	174	174
CaSSR13	177	177,210	210

heterozygosity at these loci ranged from 0.1143 (CaSSR3, CaSSR4) to 0.4571 (CaSSR9) with an average of 0.2284, making these markers suitable for chickpea genotyping and mapping programs. All loci were screened for deviations from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium using 1000 randomizations in the program FSTAT version 2.9.3.2 (Goudet 2001). Three loci, CaSSR1, CaSSR2 and CaSSR10, deviated significantly from HWE ($P < 0.001$), indicating heterozygote deficit and therefore, the possibility of null alleles at these loci could not be ruled out. Several departures from linkage equilibrium were also found at CaSSR1/CaSSR7, CaSSR3/CaSSR4, CaSSR3/CaSSR7, CaSSR3/CaSSR10, CaSSR4/CaSSR7, CaSSR7/CaSSR9 and CaSSR11/CaSSR13.

The amplification pattern of all the 13 primer pairs was also evaluated in the three accessions of *C. reticulatum*, ICC17121, ICC17123 and ICC17163, using conditions optimized for *C. arietinum*. All the primers amplified in the three *C. reticulatum* accessions indicating 100% transferability except CaSSR3, which failed to amplify only in one of the accessions namely ICC17121 (Table 2). Out of the 13 *C. arietinum* STMS markers, five markers (CaSSR2, CaSSR3, CaSSR9, CaSSR10 and CaSSR13) were able to detect polymorphism within the *C. reticulatum* accessions while 10 primers produced polymorphism between *C. arietinum* and *C. reticulatum* (Tables 1 and 2), indicating their applicability in interspecific mapping. The polymorphic microsatellite markers reported here will be useful for diversity analysis, mapping in chickpea as well as for cross-species transferability studies.

Acknowledgements

We gratefully acknowledge ICRISAT and Indian Agricultural Research Institute (IARI) for the plant material used in this study. Financial assistance for this work was provided by Department of Biotechnology, Government of India (BT/PR3518/AGR/02/179/2002) and National Centre for Plant Genome Research. The fellowships provided to SC by University Grants Commission (UGC) and to NKS, BS by Council of Scientific and Industrial Research (CSIR), India are duly acknowledged.

References

Goudet J (2001) *FSTAT version 2.9.3.2003: a program to estimate and test gene diversities and fixation indices*. http://www2.unil.ch/izea/softwares/fstat.html.

Kumar SV, Tan SG, Quah SC, Yusoff K (2002) Isolation and characterization of seven tetranucleotide microsatellite loci in mungbean, *Vigna radiata*. *Molecular Ecology Notes*, **2**, 293–295.

Ratnaparkhe MB, Santra DK, Tullu A, Muehlbauer FJ (1998) Inheritance of inter simple-sequence-repeat polymorphism and linkage with fusarium wilt resistance gene in chickpea. *Theoretical Applied Genetics*, **96**, 348–353.

Rozen S, Skaletsky HJ (1997) PRIMER 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1527–1574. Code available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Udupa SM, Robertson LD, Weigand F, Baum M, Kahl G (1999) Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. *Molecular Genetics and Genomics*, **261**, 354–363.

Winter P, Pfaff T, Udupa SM *et al.* (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Molecular Genetics and Genomics*, **262**, 90–101.

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany*, **129**, 157.