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 hypervariability 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.