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.