ORIGINAL PAPER

Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.)

**Neha Gujaria · Ashish Kumar · Preeti Dauthal · Anuja Dubey · Pavana Hiremath ·
A. Bhanu Prakash · Andrew Farmer · Mangla Bhide · Trushar Shah · Pooran M. Gaur ·
Hari D. Upadhyaya · Sabhyata Bhatia · Douglas R. Cook · Greg D. May · Rajeev K. Varshney**

Abstract A transcript map has been constructed by the development and integration of genic molecular markers (GMMs) including single nucleotide polymorphism (SNP), genic microsatellite or simple sequence repeat (SSR) and intron spanning region (ISR)-based markers, on an inter-specific mapping population of chickpea, the third food legume crop of the world and the first food legume crop of India. For SNP discovery through allele re-sequencing, primer pairs were designed for 688 genes/expressed sequence tags (ESTs) of chickpea and 657 genes/ESTs of closely related species of chickpea. High-quality sequence data obtained for 220 candidate genic regions on 2–20 genotypes representing 9 *Cicer* species provided 1,893 SNPs with an average frequency of 1/35.83 bp and 0.34 PIC (polymorphism information content) value. On an average 2.9 haplotypes were present in 220 candidate genic regions with an average haplotype diversity of 0.6326. SNP2CAPS analysis of 220 sequence alignments, as mentioned above, provided a total of 192 CAPS candidates. Experimental analysis of these 192 CAPS candidates together with 87 CAPS candidates identified earlier through in silico mining of ESTs provided scorable amplification in 173 (62.01%) cases of which predicted assays were validated in 143 (82.66%) cases (CGMM). Alignments of chickpea unigenes with *Medicago truncatula* genome were used to develop 121 intron spanning region (CISR) markers of which 87 yielded scorable products. In addition, optimization of 77 EST-derived SSR (ICCeM) markers provided 51 scorable markers. Screening of easily assayable 281 markers including 143 CGMMs, 87 CISRs and 51 ICCeMs on 5 parental genotypes of three mapping populations identified 104 polymorphic markers including 90 markers on the inter-specific mapping population. Sixty-two of these GMMs together with 218 earlier published markers (including 64 GMM loci) and 20 other unpublished markers

N. Gujaria · A. Kumar · P. Dauthal · A. Dubey · P. Hiremath ·
A. Bhanu Prakash · T. Shah · P. M. Gaur · H. D. Upadhyaya ·
R. K. Varshney (✉)
International Crops Research Institute for the Semi-Arid
Tropics (ICRISAT), Patancheru 502324,
Andhra Pradesh, India
e-mail: r.k.varshney@cgiar.org

N. Gujaria · M. Bhide
Dr. Hari Singh Gaur University,
Sagar 470003, Madhya Pradesh, India

A. Farmer · G. D. May
National Centre for Genome Resources (NCGR),
Santa Fe, NM 87505, USA

S. Bhatia
National Institute for Plant Genome Research (NIPGR),
New Delhi 110067, India

D. R. Cook
University of California, Davis (UC-Davis),
Davis, CA 95616, USA

R. K. Varshney
CGIAR Generation Challenge Programme (GCP),
c/o CIMMYT, 06600 Mexico, DF, Mexico

could be integrated into this genetic map. A genetic map developed here, therefore, has a total of 300 loci including 126 GMM loci and spans 766.56 cM, with an average inter-marker distance of 2.55 cM. In summary, this is the first report on the development of large-scale genic markers including development of easily assayable markers and a transcript map of chickpea. These resources should be useful not only for genome analysis and genetics and breeding applications of chickpea, but also for comparative legume genomics.

Introduction

Chickpea (*Cicer arietinum* L.), the only cultivated species within the genus *Cicer*, is a self-pollinated diploid ($2n = 2x = 16$) crop with a relatively small genome size of 740 Mbp (Arumuganathan and Earle 1991). It ranks third among food legumes in terms of production after common bean (*Phaseolus vulgaris*) and pea (*Pisum sativum*). Total annual world production of chickpea is 9.8 million tonnes, and major producers India and Pakistan contribute 65 and 10%, respectively, to the world harvest (FAOSTAT database, 2008, http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567#ancor). Chickpea seed contains 20–30% protein, approximately 40% carbohydrates and only 3–6% oil (Gil et al. 1996) and, moreover, it is a good source of different minerals such as calcium, magnesium, potassium, phosphorus, iron, zinc and manganese (Ibrikci et al. 2003).

Majority of chickpea crop grown area falls under semi-arid tropic (SAT) region of the world where the crop is exposed to several biotic and abiotic stresses. As a result, the crop productivity is seriously challenged in SAT regions. Although chickpea breeders have been working tirelessly to develop improved chickpea varieties with enhanced resistance or tolerance to biotic and abiotic stresses, modern biotechnological tools offer opportunities to accelerate such activities. Molecular markers, genetic maps and markers associated with traits are pre-requisites for undertaking molecular breeding leading to crop improvement (see Varshney et al. 2005a). Many kinds of molecular markers including RFLP (restriction fragment length polymorphism), RAPD (random amplification of polymorphic DNA), AFLP (amplified fragment length polymorphism), and microsatellite or simple sequence repeat (SSR) have been developed in different crops and applied for a variety of studies in the field of genetics and plant breeding. However, the choice of the marker system is largely dependent on the intended application, costs involved in development and genotyping and the ease of use. Among the different classes of molecular markers, SSRs have been proven useful for a variety of applications in plant genetics and breeding because of their reproduc-

ibility, multi-allelic nature, co-dominant inheritance, relative abundance and genome-wide coverage (Gupta and Varshney 2000). In case of chickpea, several hundred SSR markers have been developed (Winter et al. 1999; Choudhary et al. 2006, 2009; Sethy et al. 2006; Nayak et al. 2010). However, most of these markers were designed from genomic sequences and do not essentially represent the functionally important coding regions.

Due to emphasis on functional genomics and advent of next generation sequencing (NGS) technologies, it has become possible to develop the markers from genes or coding regions (Varshney 2010). As these markers are derived from genes and a putative function is known or can be deduced for the corresponding genes majority of times, these markers are popularly referred as 'genic molecular markers (GMMs)' (Varshney 2010) or 'functional markers' (Andersen and Lübberstedt 2003). A number of GMMs have several intrinsic advantages over genomic DNA markers as they serve as a useful source for identification of 'perfect marker' for marker-assisted selection (MAS), estimating the functional genetic diversity present in germplasm collection, comparative mapping among related species and identification of chromosome duplication events. Genetic maps developed based on GMMs are popularly called as 'transcript maps' or 'functional maps'.

A number of methods have been used for developing GMMs in past for several crop species (Gupta and Rustgi 2004; Varshney 2010). Some of these methods include: (a) identification of SNPs (single nucleotide polymorphisms) by allele re-sequencing for candidate genes across different genotypes (e.g. Kota et al. 2008), (b) development of SSR markers from genes or expressed sequence tags (ESTs), called EST-SSR markers (see Varshney et al. 2005b); (c) identification of SNPs through in silico mining of ESTs coming from different genotypes and development of markers based on such SNPs, often referred as EST-SNP markers (e.g. Kota et al. 2008), (d) designing the primers from exonic regions to amplify the intronic region and detect either length or sequence polymorphism in introns (Feltus et al. 2006), referred here as intron spanning region (ISR) markers. While EST-SSR markers can be assayed on traditional agarose or polyacrylamide or capillary electrophoresis, a number of SNP genotyping platforms, e.g. pyrosequencing (Alderborn et al. 2000; Ching and Rafalski 2002; Varshney et al. 2008), mass spectrometry (Rodi et al. 2002), Affymetrix chips (Borevitz et al. 2003), GoldenGate assays (Fan et al. 2003, Rostoks et al. 2006), are available. However, one of the most cost-effective SNP genotyping approaches is conversion of SNPs into cleaved amplified polymorphic sequence (CAPS) assay which can be done by using agarose gel electrophoresis (Varshney et al. 2007). ISR markers are generally assayed on denaturing MDE (mutation detection enhancement) gel (Thudi et al. 2010).

Using above-mentioned methodologies and several others, a large number of GMMs as well as transcript maps have been developed in several crop species such as rice (Wu et al. 2002), wheat (Qi et al. 2004), barley (Stein et al. 2007; Kota et al. 2008; Sato et al. 2009), soybean (Choi et al. 2007) etc.

In case of chickpea, though few reports are available on identification of SNPs based on re-sequencing of some genes (Rajesh and Muehlbauer 2008; Nayak et al. 2010), no comprehensive study has been undertaken so far on identification of SNPs at large scale or development of a detailed transcript map. Therefore, this study was undertaken with following objectives: (a) SNP discovery and sequence diversity analysis through allele re-sequencing approach, (b) development of CAPS markers for assaying SNPs, (c) development of ISR markers, (d) optimization of SSR markers derived from ESTs, and (e) construction of transcript map of chickpea.

Materials and methods

Plant material and DNA extraction

A set of 22 diverse chickpea genotypes originating from 7 countries was collected from Genebank and Chickpea Breeding Divisions of ICRISAT for the development of SNP markers through allele-specific sequencing. These chickpea genotypes represent 9 *Cicer* species including 11 cultivated and 11 wild chickpea genotypes (Table 1). All the newly developed markers were optimized for amplification initially on two genotypes (ICC 4958 and ICC 1882).

Total genomic DNA was extracted from leaves of 2-week-old seedlings using high-throughput mini DNA extraction protocol as mentioned in Cuc et al. (2008). The quality and quantity of extracted DNA was checked on 0.8% agarose gel. The DNA was normalized to 5 ng/μl for further use.

Designing of primer pairs

For SNP discovery based on allele re-sequencing approach, primer pairs were designed using PRIMER3 for: (a) 688 chickpea ESTs/genes, (b) 657 heterologous genes or transcription factors (TFs) (identified in *Medicago truncatula*, *Medicago sativa*, *Lotus japonicus*, *Lupinus* spp., *Arachis hypogaea*, *P. sativum*, *Crotolaria tenuifolia*, *P. vulgaris*, *Phaseolus coccineus*, *Glycine max*, *Glycine soja*, *Robinia pseudoacacia* and *Trifolium pratense*), (c) 77 chickpea EST-SSRs (Varshney et al. 2009a), (d) 87 EST contigs containing SNPs that could be optimized for CAPS assay (Varshney et al. 2009a) and (e) 121 intronic regions identified based on comparison of chickpea ESTs with *Medicago* genome.

Table 1 List of chickpea genotypes used for allele re-sequencing for identification of SNPs

S. no.	Accession	Species	Geographical origin	Type
1	ICC 1882	*C. arietinum*	India	Desi
2	ICC 283	*C. arietinum*	India	Desi
3	ICC 3137	*C. arietinum*	Iran	Desi
4	ICC 4958	*C. arietinum*	India	Desi
5	ICC 506EB	*C. arietinum*	India	Desi
6	ICC 8261	*C. arietinum*	Turkey	Kabuli
7	ICCV2	*C. arietinum*	India	Kabuli
8	Annigeri	*C. arietinum*	India	Desi
9	ICCC 37	*C. arietinum*	India	Desi
10	JG 62	*C. arietinum*	India	Desi
11	Vijay	*C. arietinum*	India	Desi
12	ICC 17162	*C. cuneatum*	Ethiopia	Wild
13	ICC 17148	*C. judaicum*	Lebanon	Wild
14	ICC 17248	*C. microphyllum*	Pakistan	Wild
15	ICC 17152	*C. pinnatifidum*	Turkey	Wild
16	ICC 17123	*C. reticulatum*	Turkey	Wild
17	IG 72933	*C. reticulatum*	Turkey	Wild
18	IG 72953	*C. reticulatum*	Turkey	Wild
19	PI 489777	*C. reticulatum*	Turkey	Wild
20	ICC 17116	*C. yamashitae*	Afghanistan	Wild
21	ICC 17122	*C. bijugum*	Turkey	Wild
22	ICC 17138	*C. pungens*	Afghanistan	Wild

Polymerase chain reaction (PCR)

PCR for all above-mentioned markers was carried out in 20 μl reaction in a GeneAmp® PCR System 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA) containing 5 ng template DNA, 0.2 mM dNTPs, 2 mM $MgCl_2$, 2 pmol of forward and reverse primer, 1 U of *Taq* DNA polymerase (Sib enzyme), and 1× PCR buffer. The amplification cycles were: initial denaturation of 5 min at 94°C followed by 10 cycles of denaturation for 15 s at 94°C, touchdown from 61 to 51°C with 1°C decrease in each cycle for 20 s followed by extension at 72°C for 30 s. The next 40 cycles included denaturation at 94°C for 15 s, annealing at 54°C for 30 s and extension at 72°C for 30 s followed by final extension of 30 min at 72°C and left at 4°C until further use. For testing the amplification, PCR products were separated on 1.2% agarose gel electrophoresis.

Allele re-sequencing and SNP detection

For allele re-sequencing-based SNP discovery, PCR products obtained using 688 chickpea and 657 heterologous genes/transcription factors were sequenced in both directions using Sanger sequencing methodology. Initially, PCR

amplicons producing solitary bands were purified by treating with 1 U each of Exonuclease I (Exo) and shrimp alkaline phosphatase (SAP). Subsequently, the EXO/SAP-treated PCR products were incubated at 37°C for 45 min followed by denaturation at 80°C for 15 min to deactivate remaining EXO activity. All the Exo/SAP-treated products were sequenced from both ends using respective forward and reverse primers at Macrogen Inc., Seoul, South Korea (http://www.macrogen.com/).

Sequencing data were inspected manually for possible sequencing error. The forward and reverse sequences for the given gene/EST were merged to prepare the contig for each genotype using DNA Baser v 2.9 software (http://dnabaser.com) under the following parameters: minimum match percentage, 80%: minimum of overlap, 25 bases; quality value of 25. Subsequently, contigs for all genotypes were aligned using Clustal W (Thompson et al. 1994; http://www.ebi.ac.uk/Tools/clustalw2/index.html) and aligned files are saved with .aln extension. The .aln files of each gene/marker were subsequently opened in BioEdit version 7.0.5.3. SNP identification among different genotypes was undertaken manually after considering the base quality.

The multiple sequence alignment files for each gene/marker after manual confirmation of identified SNPs in BioEdit programme were saved as FASTA file. Subsequently, FASTA files for different genes were put together in a single file. This file was used in SNP DIVersity ESTimator module (*divest.pm*) developed at ICRISAT (Jayashree et al. 2009) for calculating the polymorphism information content (PIC) value of individual SNPs as well as gene/marker, nucleotide diversity (π), number and PIC value of haplotypes for each gene/marker.

Genotyping assays

Three genotyping assays were employed for generating marker genotyping data: (a) CAPS assays, (b) capillary electrophoresis and (c) polyacrylamide gel electrophoresis using MDE gel.

CAPS assays

All the identified SNPs through allele re-sequencing undertaken in this study and in silico mining approach of ESTs, described in Varshney et al. (2009a), were subjected to 'SNP2CAPS' program (http://hpc.icrisat.cgiar.org/Pise/5.a/data_manipulation/snp2caps.html; Thiel et al. 2004) which employed restriction enzymes Rebase database (810.gcc version) containing recognition sequence information for 725 commercially available restriction enzymes. The programme predicted possible conversion of several SNP-containing markers/genes into CAPS assays.

All the predicted CAPS candidates were amplified using the same PCR conditions as mentioned above. Amplicons were then subjected to restriction digestion with the corresponding restriction enzyme followed by electrophoretic separation on agarose gel electrophoresis (3% agarose, 1× TBE buffer, 1 h, 120 V) and visualized by means of ethidium bromide staining (Varshney et al. 2007).

Polyacrylamide gel electrophoresis using MDE gel

PCR products, generated using ISR markers, were denatured and separated on single-strand confirmation polymorphism (SSCP) gels containing MDE solution as given in Thudi et al. (2010).

Capillary electrophoresis

PCR products generated using EST-SSR markers were analyzed on capillary electrophoresis using ABI PRISM® 3730 DNA analyzer, and allele calling was carried out as given in Varshney et al. (2009b).

Genetic mapping and map construction

Genotyping data for polymorphic markers on inter-specific mapping population (ICC 4958 × PI 489777) were generated on 131 recombinant inbred lines (RILs). In addition, marker genotyping data for 218 marker loci were compiled from Winter et al. (1999, 2000) and Nayak et al. (2010) and 20 DArT (Diversity Array Technology) marker loci (unpublished). The map distances were calculated by applying the 'Kosambi' mapping function (Kosambi 1944) as per MAPMAKER/EXP 3.0 programme (Lander et al. 1987). For integrating new marker loci into different linkage groups (LGs), a minimum logarithm of odds (LOD) of 3.0 was used. With an objective of developing the high-quality genetic map, JoinMap® 4 programme (Van Ooijen 2006; http://www. kyazma.nl) was used to construct high-quality genetic map using the same set of marker loci determined in MAPMAKER/EXP 3.0 programme. Maps were drawn with MapChart, version 2.1 (Voorips 2002). The relative mapping positions of the unmapped markers from JoinMap® 4 programme have been shown on the right side of the linkage groups.

Results

Development of GMMs

For identification of SNPs in chickpea, a total of 1,345 primer pairs were designed for genes or transcription factors identified in chickpea or other legume species, e.g. *M. truncatula*, *M. sativa*, *L. japonicus*, *Lupinus* spp.,

A. hypogaea, *P. sativum*, *C. tenuifolia*, *P. vulgaris*, *P. coccineus*, *G. max*, *G. soja*, *R. pseudoacacia* and *T. pratense*. Subsequently, these primer pairs were used to amplify 2–20 genotypes given in Table 1. All successful amplicons generated were used for sequencing in both directions. Sequence data obtained on all the genotypes for a given gene were aligned and compared to identify the SNPs.

Chickpea ESTs/genes-derived primer pairs

A total of 688 primer pairs were designed as follows: (a) 409 for SNPs identified based on Illumina sequencing of four genotypes namely ICC 4958, ICC 1882, ICC 506EB and ICCC 37 (Ca2C, Ca2S, CaESTs and Ca series; this study), (b) 217 ESTs (AGLC series; Buhariwalla et al. 2005, this study), and (c) 62 candidate genes (CaHa series; Singh et al. 2008, Rajesh and Muehlbauer 2008). Primer sequences for all above-mentioned genes/ESTs are given in ESM Table 1. After screening these primer pairs on two chickpea genotypes namely ICC 4958 and ICC 1882, only 587 primer pairs showed scorable amplification. Subsequently, all 587 primer pairs were used on 2–19 genotypes. Sequencing of amplicons in both directions provided high-quality sequences for 305 markers/genic regions on 2–19 genotypes.

Heterologous genes-/transcription factors-derived primer pairs

A total of 657 primer pairs were designed based on transcription factor/gene sequences of chickpea-related legume species. For instance, 479 primer pairs were designed for genes/transcription factors of *M. truncatula* and *M. sativa* (297), *L. japonicus* (144) and *Lupinus albus* and *L. luteus* (38). Some of these markers were taken from Nelson et al. (2006). Another set of 178 primer pairs was designed based on gene sequences of *G. max* and *G. soja* (97), *T. pratense* (38) and other legume species (43), e.g. *A. hypogaea*, *P. sativum*, *C. tenuifolia*, *P. vulgaris*, *P. coccineus* and *R. pseudoacacia* that had significant similarity with Illumina sequence reads and SNPs between chickpea genotypes ICC 4958 and ICC 1882 identified in another study (unpublished results). Primer sequences for all these genes/markers are given in ESM Table 1. Screening of these 657 primer pairs on two genotypes showed scorable amplification with 401 primer pairs. Subsequently, sequencing of these amplicons provided high-quality sequences for 134 markers/genes on 3–20 genotypes (Table 2).

In summary, a total of 988 markers showed scorable amplification and high-quality sequences were generated for 439 markers.

Identification of SNPs and estimates on sequence diversity in chickpea germplasm

Sequence data generated for 439 candidate genic regions across 2–20 genotypes were aligned and these alignments were edited manually by trimming low-quality sequences at both ends. The analysis of the sequence alignments across the genotypes showed occurrence of SNPs in case of 220 candidate genic regions (ESM Fig. 1). Among these genes,

Table 2 Development, amplification and sequencing status of gene sequences based primers derived from chickpea and heterologous species

Marker series	Marker type	Primer Id series/ source of genes	Markers designed	Markers amplified	Number of genotypes surveyed		High-quality sequences
					Cultivated	Wild	
Chickpea	Illumina/454 chickpea sequences (409)	Ca2C, Ca2S	272	214	1–5	1–2	104
		CaESTs, Ca	137	115	2	1	45
		Total	409	329			149
	Chickpea ESTs (279)	AGLC	217	214	5–9	1–11	145
		Rajesh and Muehlbauer (2008)	12	7	2	1	5
		Singh et al. (2008)	50	37	4	2	6
		Total	279	258			156
	Total chickpea (688)		688	587			305
Heterologous species	*Medicago truncatula*		297	155	2–9	1–11	57
	Lupinus spp.		38	14	5–9	2–11	14
	Lotus spp.		144	54	8–9	6–11	15
	Glycine max		97	97	2	1	24
	Trifolium pratense		38	38	2	1	9
	Others		43	43	2	1	15
	Total heterologous (657)	Total	657	401			134
Total		1,345	1,345	988			439

122 have been originated from chickpea, while 98 have come from heterologous species (35 from *Medicago* spp., 19 from soybean, 13 each from *L. japonicus*, *Lupinus* spp., 6 from *T. pratense* and 12 genes from other legumes). DIVEST analysis of above-mentioned sequence alignments has shown a total of 1,893 SNPs in 67,828 bp sequence data generated for 220 candidate genic regions with a SNP frequency of 1/35.83 bp. Among 1,893 SNPs identified, 1,089 accounted for the transition and 888 for transversion. In addition, a total of 63 *InDel* were also observed.

As the genes surveyed for sequence diversity were originally derived from gene/transcription factor sequences of chickpea and related legume (heterologous) species and examined on cultivated and wild species, sequence diversity was estimated as per origin of the gene/TF sequence as well as in terms of the cultivated versus wild species.

Sequence diversity based on genes from chickpea versus heterologous species

Although almost equal numbers of genes were derived from chickpea (688) and heterologous species (657), a slightly higher proportion of chickpea genes (17.73%) as compared to heterologous species (14.92%) showed SNPs. As shown in Table 3, 122 genic regions coming from chickpea showed 416 SNPs with a frequency of 1/88.15 bp, while 98 genic regions coming from heterologous species provided 1,477 SNPs with 1/21.09 bp. The nucleotide diversity index (π) for the chickpea genes ranged from 0.6×10^{-3} to 25.9×10^{-3} (mean = 5.3×10^{-3}), while heterologous species genes showed a higher nucleotide diversity index of 1.0×10^{-3} to 74.0×10^{-3} (mean = 14.6×10^{-3}; ESM Table 2).

With an objective to identify the informative set of SNPs for chickpea genetics and breeding, PIC values were calculated and these ranged from 0.21 to 0.50 (average 0.34) in chickpea genes, while 0.11 to 0.46 (average 0.34) in heterologous species genes. As PIC values of the bi-allelic SNPs cannot exceed 0.50, sequence data for these candidate genic regions were analyzed in terms of haplotypes as well. Number of haplotypes observed varied from one to nine with an average of 0.41 (range 0.22–0.97) haplotype PIC value in chickpea genes, while heterologous species genes showed one to ten haplotypes with an average PIC of 0.61 (range 0.20–1.0). Haplotype diversity calculated was higher with heterologous species genes (0.7704) as compared to chickpea genes (0.5220).

It is interesting to note that at least eight genic regions coming from heterologous species (LG80, LG99, LG101, LG103, LG104, LG111, LUP120, TC77515) and one candidate genic region (AGLC212) from chickpea showed >50 SNPs across the genotypes examined. However, higher PIC values (>0.50) were observed with three chickpea-derived

Table 3 Sequence diversity analysis based on source of genes and species types

	Number of genotypes surveyed (average)	Total length of sequence surveyed (average in bp)	Number of genes showing SNPs	Number of SNPs identified (average)	PIC value of individual SNP (average)	Nucleotide diversity range (average)	Number of haplotypes (average)	Haplotype diversity range (average)	PIC value of haplotypes (average)
Source of genes									
Chickpea (688)	2–19 (6.29)	36,672 (300.59)	122	416 (3.41)	0.21–0.50 (0.34)	0.0006–0.0259 (0.0053)	1–9 (2.3)	0.2500–1.1670 (0.5220)	0.22–0.97 (0.41)
Heterologous species (657)	3–20 (9.68)	31,156 (317.92)	98	1,477 (15.07)	0.11–0.46 (0.34)	0.0010–0.0740 (0.0146)	1–10 (3.7)	0.2230–1.1670 (0.7704)	0.20–1.00 (0.61)
Within species type									
Cultivated (1–9)	220	67,828 (308.31)	66	132 (2.000)	0.20–0.50 (0.41)	0.0000–0.0270 (0.0051)	1–3 (1.3)	0.0000–1.0840 (0.6671)	0.00–0.91 (0.46)
Wild species (0–11)	220	67,828 (308.31)	66	1,445 (21.894)	0.19–0.50 (0.36)	0.0000–0.0884 (0.0235)	1–10 (2.0)	0.0000–1.1670 (0.9759)	0.00–1.00 (0.76)
Across/total									
220 genes across 2–20 genotypes	220	67,828 (308.31)	220	1,893 (8.605)	0.11–0.50 (0.34)	0.0006–0.0740 (0.0094)	1–10 (2.9)	0.2230–1.1670 (0.6326)	0.20–1.00 (0.50)

candidate genic regions (Ca2C21276, Ca2S126415_1648_0587, Ca2C3599).

Sequence diversity among cultivated versus wild species genotypes

While analyzing the sequence data as per cultivated species and wild species, a higher level of sequence diversity was observed in wild species as compared to cultivated species. On surveying 220 candidate genic regions in cultivated and wild species, the number of SNPs present in wild species (1,445 SNPs) was more than 10 times higher as compared to cultivated species (132 SNPs). Similarly, nucleotide diversity was higher (23.5×10^{-3}) in case of wild species as compared to cultivated species (5.1×10^{-3}). In terms of PIC value of SNPs, cultivated species depicted average PIC value of 0.41 (ranging from 0.20 to 0.50) and wild species showed average PIC value of 0.36 (ranging from 0.19 to 0.50).

As the number of SNPs are much higher in case of wild species, the number of haplotypes ranged from 1 to 10 (average 2.0) in wild species in contrast to 1 to 3 (average 1.3) in cultivated species. Haplotype diversity was also observed high in case of wild species (0.9759) as compared to cultivated species (0.6671). Similarly, a higher PIC value of haplotype was observed in wild species (average 0.76) as compared to cultivated species (0.46).

In total, 1,893 SNPs with an average SNP frequency of 1 SNP per 35.83 bp were observed on surveying 67,828 bp sequence data across all the 220 candidate genic regions. PIC values of SNP ranged from 0.11 to 0.50 with an average of 0.34 across all the 220 genic regions. Number of haplotypes across 220 genic regions ranged from 1 to 10 with PIC value ranging from 0.20 to 1.00 (avg. 0.50). On an average 2.9 haplotypes (ranged from 1 to 10) were present in 220 genic regions with an average haplotype diversity of 0.6326 (ranged from 0.2230 to 1.1670).

Development of CAPS markers for assaying SNPs

SNPs identified from allele re-sequencing were converted into potential CAPS markers by relating the SNP position to the presence/absence of a restriction site in amplicon using 'SNP2CAPS' programme. After analyzing 220 sequence alignments, as mentioned above, with SNP2CAPS programme, a total of 192 (87%) genic regions showed the presence of a putative restriction site in the sequence alignments. In addition, another potential set of 87 CAPS-based SNP markers was available from Varshney et al. (2009a).

With an objective of validation of in silico predicted CAPS markers, a total of 279 primer pairs including 192 identified in this study and 87 identified in Varshney et al. (2009a) were used for CAPS assays on 5 chickpea genotypes (ICC 4958, ICC 1882, ICC 283, ICC 8261 and PI 489777). Details about these 279 CAPS candidates are given in Table ESM 3. While scorable amplification was observed in 173 (62.01%) cases out of 279 CAPS candidate, CAPS assays were succeeded in 143 cases (82.66%) out of 173 scorable amplified CAPS candidates. These gene-based markers have been designated as *C*hickpea *G*enic *M*olecular *M*arker (CGMM). While details of the validated 143 CGMMs are presented in ESM Table 4, a representative CAPS profile for 6 CGMMs on a panel of 5 chickpea genotypes has been shown in Fig. 1. Validated CAPS includes 115 (59.90%) out of 192 candidates identified based on allele re-sequencing data and 28 (32.18%) out of 87 candidates identified through mining of ESTs (Varshney et al. 2009a). This clearly indicates that allele re-sequencing approach is more effective than EST mining for conversion of SNPs into CAPS.
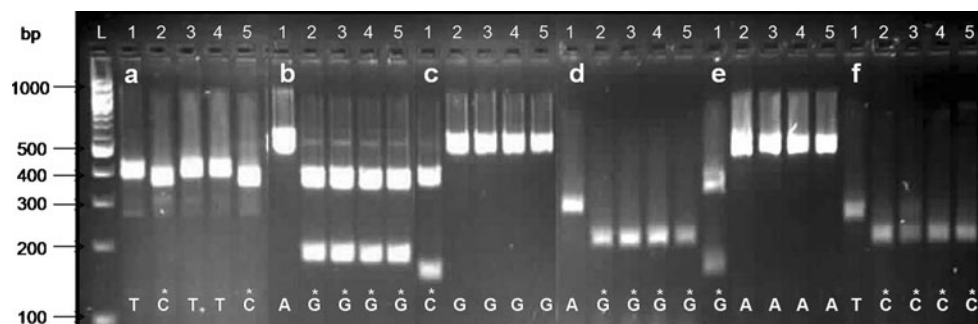

Fig. 1 Some selected examples of assaying SNPs via CAPS markers. Restricted digested products for six CAPS candidate markers have been shown on 1.2% agarose gel as following: *a* CGMM002-*Hpy*CH4IV (recognition site A/CGT), *b* CGMM020-*Hin*cII (recognition site GTY/RAC), *c* CGMM009-*Hin*fI (recognition site G/ANTC), *d* CGMM023-*Hpy*ch4IV (recognition site A/CGT), *e* CGMM051-*Mbo*I (recognition site /GATC), and *f* CGMM041-*Hha*I (recognition site GCG/C). Nucleotide variation has been depicted in the picture and the *asterisk* (*) marked nucleotide identified in the recognition site. Names of genotypes for DNA samples in each panel (*a* to *f*) are as following: *Lane 1* PI 489777, *Lane 2* ICC 4958, *Lane 3* ICC 1882, *Lane 4* ICC 8261, *Lane 5* ICC 283, *Lane L* 100 bp ladder

Development of ISR markers

Another set of gene-based markers was developed after aligning the chickpea unigenes with *M. truncatula* genome sequence. Subsequently, the exonic regions were used for designing the primer pairs to amplify the intronic sequence and assaying the polymorphism. In this context, alignment of 9,569 unigenes of chickpea was performed with available genome sequence of *M. truncatula* (http://www.medicago.org/genome) using BLASTN ($E \leq 1e{-}10$ and >90% identity; Table 4). As a result, a total of 267 chickpea unigenes showed the presence of 784 intronic regions (ESM Fig. 2). However, only 144 unigenes were selected for designing primer pairs based on the criteria of potential of generation of amplicons ranging from 200 to 2,000 bp. In total, 121 primer pairs were generated for 92 chickpea unigenes indicating that 29 primer pairs were generated for more than one intronic region in the same gene. These markers have been referred as *C*hickpea *I*ntron *S*panning *R*egion (CISR) markers and their primer sequences have been given in ESM Table 5. The analysis of these 121 CISR markers on MDE gels showed amplification in case of 87 (71.90%) CISR markers representing 74 unigenes.

Validation of EST-derived SSR (EST-SSR) markers

Based on a comprehensive EST dataset, 77 EST-SSR markers were reported in an earlier study (Varshney et al. 2009a; ESM Table 6). Screening of these EST-SSR markers on 5 parental genotypes (ICC 4958, ICC 1882, ICC 283, ICC 8261 and PI 489777) of three mapping populations in this study showed amplification in case of 51 EST-SSR (66.23%) markers. These markers have the designation *IC*RISAT *C*hickpea *e*ST *M*icrosatellite (ICCeM) markers.

Marker polymorphism

As mentioned above, successful assays were developed for 143 CGMM, 87 CISR and 51 ICCeM markers. All these markers were screened on a set of five chickpea genotypes that represent parents of one inter-specific (ICC 4958 × PI 489777) and two intra-specific mapping populations (ICC 283 × ICC 8261; ICC 4958 × ICC 1882).

In case of CGMMs, 53 (37.06%) CGMMs showed polymorphism for the inter-specific mapping population, while 6 (4.2%) CGMMs and 2 (1.4%) CGMMs were polymorphic in two intra-specific mapping populations namely ICC 4958 × ICC 1882 and ICC 283 × ICC 8261, respectively (ESM Table 4). Out of 87 CISR markers tested, 21 (24.14%) markers showed polymorphism across 5 genotypes (ESM Table 5). While 18 (20.69%) CISR markers showed polymorphism in the parental lines of the inter-specific mapping population, 3 (3.45%) CISR markers showed polymorphisms for each of two intra-specific mapping populations. In case of ICCeMs, 19 (37.25%) markers showed polymorphism for the inter-specific mapping population, while 9 (17.65%) markers showed polymorphism for each of two intra-specific mapping populations (ESM Table 6).

In summary, out of 281 markers screened, 104 (57 CGMM, 21 CISR and 26 ICCeM) markers showed polymorphism in 5 genotypes. Majority of the markers (90) showed polymorphism between parents of inter-specific mapping populations (Table 5), while parents of intra-specific populations showed polymorphism with less than 10% markers. It is also important to mention that four markers (CGMM002, CISR117, ICCeM033 and ICCeM035) showed polymorphism in all three mapping populations and the other set of four markers (CGMM003, ICCeM050, ICCeM051 and ICCeM058) were polymorphic for both

Table 4 Details on development, amplification and polymorphism assessment of CISR markers

Medicago truncatula (Mt) chromosome number	Number of corresponding chickpea unigenes having intronic region	Number of primer pair designed for amplification of intronic regions	Number of primer pairs yielding scorable amplification	Number of primer pairs showing polymorphism across 5 chickpea genotypes
Chromosome (0) + strand	13	3	2	1
Chromosome (1) + strand	9	3	1	0
Chromosome (2) + strand	13	2	2	1
Chromosome (3) + strand	36	20	18	4
Chromosome (4) + strand	40	15	10	5
Chromosome (5) + strand	78	43	30	5
Chromosome (6) + strand	15	8	5	1
Chromosome (7) + strand	38	18	12	3
Chromosome (8) + strand	25	9	7	1
Total	267	121	87	21

Table 5 Polymorphism status of easily assayable GMMs in three mapping populations

Marker series	Primer pairs designed	Marker assays optimized	Number of polymorphic markers		
			ICC 4958 × PI 489777 (inter-specific population)	ICC 4958 × ICC 1882 (intra-specific population)	ICC 283 × ICC 8261 (intra-specific population)
CGMM	279	143	53	6	2
CISR	121	87	18	3	3
ICCeM	77	51	19	9	9
Total	477	281	90	18	14

intra-specific mapping populations. On the other hand, another set of four markers (CGMM007, ICCeM018, ICCeM054 and CISR002) were polymorphic for the inter-specific and one intra-specific (ICC 4958 × ICC 1882) mapping population, while two markers (ICCeM015 and ICCeM040) were polymorphic for the inter-specific and one intra-specific (ICC 283 × ICC 8261) mapping population.

Construction and features of genetic map

As a large number of GMMs showed polymorphism for the inter-specific mapping population (ICC 4958 × PI 489777) and this population has been used as a reference mapping population in several studies (e.g. Winter et al. 1999, 2000; Nayak et al. 2010), this population was selected for developing a transcript map based on GMMs. As a result, genotyping data were collected for all 90 (53 CAPS-SNP, 19 EST-SSR and 18 ISR) polymorphic markers on the mapping population.

For development of a genetic map, marker genotyping data for 238 marker loci were compiled from other sources (Winter et al. 1999, 2000; Nayak et al. 2010; ICRISAT, unpublished). Genotyping data of these markers together with 90 newly developed markers in the present study were used with MAPMAKER/EXP version 3.0. As a result, only 79 (out of 90) new marker loci were integrated into this genetic map. The JoinMap® 4 analysis on these marker genotyping data, however, could integrate only 62 GMM (32 CGMM, 15 CISR and 15 ICCeM) loci into this genetic map of chickpea. The putative mapping positions of the remaining 15 CGMM and 2 CISR loci have been shown to the right of linkage groups. The linkage groups have been designated according to Nayak et al. (2010).

As this genetic map has 64 gene-based loci published in Nayak et al. (2010), in total, this transcript map has a total 126 GMM loci with an average of about 15.75 ± 4.6 GMM loci per linkage group. However, the integrated genetic map has a total 300 marker loci on 8 linkage groups spanning 766.56 cM with an average inter-marker distance of 2.552 cM (Fig. 2). The average number of marker loci per

linkage group was 37.5 ± 9.6, with a maximum of 57 marker loci on linkage group 2 and a minimum of 24 marker loci in linkage group 7.

Discussion

Molecular markers are important genetic tools for understanding genome dynamics and facilitating molecular breeding. In case of chickpea, the progress in the area of development of molecular markers and genetic map, however, has remained slow. Nevertheless, in the past few years, significant progress has been made in the area of development of molecular markers (Hüttel et al. 1999; Sethy et al. 2006; Lichtenzveig et al. 2005; Buhariwalla et al. 2005; Choudhary et al. 2006, 2009; Hyten et al. 2010) and genetic maps (Winter et al. 2000; Pfaff and Kahl 2003; Choi et al. 2007; Radhika et al. 2007; Millan et al. 2010; Nayak et al. 2010). However, majority of these markers have derived from genomic DNA library and, therefore, developed markers or integrated marker loci onto the genetic map do not essentially represent genes. While gene-based molecular markers, popularly called GMMs, have been developed in large number in several crop species including some legumes like soybean (Choi et al. 2007), concerted efforts were not undertaken in case of chickpea. The present study, therefore, reports development of a comprehensive set of GMMs and integration of some of these GMM loci into this genetic map. In this context, three main approaches have been used: (a) allele re-sequencing approach to identify SNPs and conversion of some of these SNPs into CAPS markers (CGMMs), (b) aligning the chickpea unigenes to *Medicago* genome to develop intron-based markers (CISRs), and (c) optimizing of EST-derived SSR markers (ICCeMs).

SNPs and sequence diversity in chickpea germplasm

Although 988 candidate genes-based primer pairs showed scorable amplification, high-quality sequences were obtained for only 439 candidate genic regions. The analysis
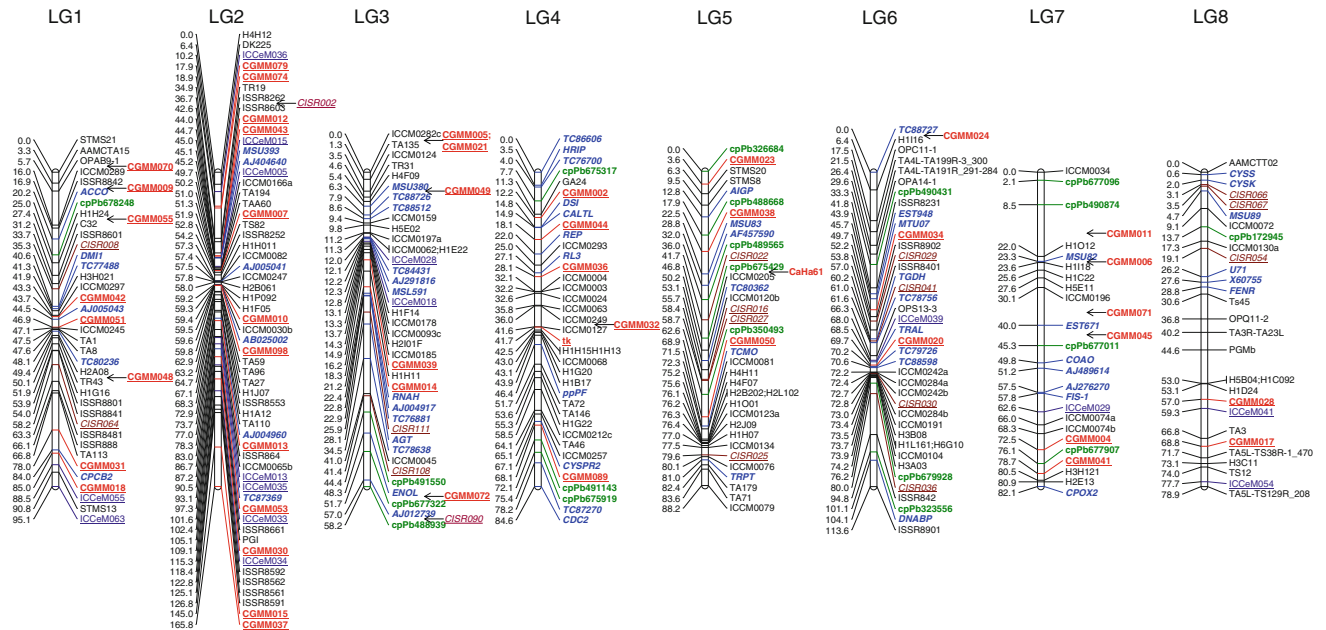
Fig. 2 A transcript map of chickpea based on recombinant inbred lines of *C. arietinum* (ICC 4958) × *C. reticulatum* (PI 489777). Distances between the mapped loci (in cM) are shown to the *left* of the linkage group and all the loci are to the *right side* of the map. Newly developed and integrated marker loci have been shown in *bold, underlined* and *colored fonts*: *red color* CGMM loci, *brown color* CISR loci,

gray color ICCeM loci. In addition to these new loci, gene-based marker loci mapped by Nayak et al. (2010) have been shown *italicized* and in *blue color*. The marker loci, otherwise as mentioned above, are either published earlier (Winter et al. 1999, 2000; Nayak et al. 2010) or unpublished (e.g. cpPb series DArT marker loci) (color figure online)

of sequence alignment for these genic regions showed 1,893 SNPs in 220 genes and 63 *InDel* across a set of 2–20 genotypes. SNP and *InDel* markers have been proven very useful for estimating the linkage disequilibrium and association mapping for crop improvement (Rafalski 2002).

Scanning of 67,828 bp sequence data (220 candidate genic regions) led to the identification of 1,893 SNPs with an average SNP frequency of 1 SNP per 35.83 bp in the chickpea germplasm surveyed. However, earlier SNP report in chickpea presented SNP frequency as 1/61 bp in coding region while 1/71 bp in genomic region (Rajesh and Muehlbauer 2008). The SNP frequency in the present study is even higher than the earlier report and it can be attributed to use of higher number of genotypes and especially more wild species. Interestingly, SNP frequency in chickpea seems to be higher as compared to other crop species like: wheat (1/1,000 bp, Bryan et al. 1999); rice (1/89 bp, Nasu et al. 2002); barley (1/300 bp, Kota et al. 2008); common bean (1/76 bp, Gaitán-Solís et al. 2008) and soybean (1 in 277 bp, Zhu et al. 2003). However, this is an outcome of using 11 wild species representing the secondary as well as tertiary genepool in the present study. In case SNP frequency is considered only in germplasm of only cultivated species, the SNP frequency is 1/513 bp. In this scenario, it is evident that the cultivated genepool of chickpea is narrow as compared to other legume or cereal species mentioned above.

In general, SNPs are bi-allelic, 76 SNPs for 25 candidate genic regions showed three alleles that enhance the value of SNPs identified in the present study. For instance, the marker CGMM101 (primer pair TC77707) is having 2 bp substitution at 128 and 136 bp length from G → T → A and C → A → G, respectively, showing haplotype diversity of 0.70 across all accessions, but 0.95 across wild chickpea accessions (ESM Fig. 1).

The chickpea lines coming from wild species had >10× higher SNPs as compared to lines coming from the cultivated species. Similarly, nucleotide diversity (π) was about 5× higher in wild species (23.5×10^{-3}) as compared to the cultivated species (5.1×10^{-3}). While comparing this π with the earlier report of 14×10^{-3} by Rajesh and Muehlbauer (2008), it is clear that π in cultivated species is lower and is higher in wild species in the present study. It is interesting to note that the LG101 marker derived from *L. japonicus* showed the highest π as 74.0×10^{-3} (cultivated species, 4.8×10^{-3}; wild species, 88.4×10^{-3}), whereas genic region for AGLC57, derived from chickpea EST, showed minimum π as 0.6×10^{-3}.

However, unlike number and frequency of SNPs and π, the PIC value of SNPs in the wild species is lower (average 0.36) than the cultivated species (average 0.41). As PIC value is a direct function of allelic frequency for the given SNP in the germplasm collection and as compared to few genotype per wild species (4 for *C. reticulatum* and 1 each

for 7 wild species), 11 lines were examined for cultivated species in the present study, a higher PIC value in cultivated species is not unexpected.

Comprehensive set of GMMs for chickpea genetics and breeding

In total, primer pairs were designed for 1,630 genes; however, 1,184 primer pairs showed scorable amplification in the two genotypes examined. Thus, this study reports the first comprehensive set of 1,184 GMMs comprising 988 candidate genes including 220 genic regions with SNPs or 143 CAPS (CGMMs), 87 intron-based markers (CISRs) and 51 EST-SSRs (ICCeMs) for chickpea. The primer sequence information provided in ESM Tables 1, 3, 5 and 6 will be of great value for chickpea genetics and breeding community.

In terms of development of CAPS markers, a total of 279 CAPS candidates including 192 identified in this study using allele re-sequencing approach and 87 identified in an earlier study by in silico mining of ESTs (Varshney et al. 2009a) were assayed on five chickpea genotypes. While scorable amplification was observed in 173 (62.01%) cases, CAPS assays were observed in 143 cases (82.66%). A higher success rate was observed for conversion of SNPs identified through allele-specific sequencing (59.90%) as compared to those that were identified through EST mining (32.18%) approach. This can be attributed for the possible sequencing error instead of presence of true SNPs. Also, CAPS identified based on in silico mining of ESTs did not show polymorphism in these five genotypes. This is possible as the genotypes deployed for CAPS validation in the present study are different than those from ESTs that were used for mining (Raju et al. 2010).

A first generation 'transcript map' of chickpea

With an objective of integrating the GMMs into genetic map, easily assayable 281 GMMs were screened on 5 parental genotypes of 3 mapping populations. As a maximum number of markers (90) showed polymorphism in the inter-specific mapping population, genotyping data collected for these markers were attempted to integrate the corresponding marker loci into genetic map. Although analysis of these marker data using MAPMAKER/EXP version 3.0 integrated 79 GMM loci, the map distance observed was high. Therefore, to develop a high-quality map, Join-Map® 4 programme was used but 62 GMM loci could be integrated into this genetic map. Therefore, the relative mapping positions of the remaining 17 markers have been shown on right side of the linkage groups.

Out of 218 framework markers used in the present study, 64 marker loci were gene-derived (Nayak et al. 2010), the

present map has a total of 126 GMM loci. Probably, this is the first genetic map of chickpea that contained the highest number of GMM loci integrated. The present transcript map has 15.75 ± 4.6 GMM loci per linkage group with an average marker density of 1 marker per 6.08 cM. As compared to other legume crop species, transcript maps have been developed recently in cowpea (Muchero et al. 2009) and soybean (Choi et al. 2007). It is evident that integration of more GMM loci is required to enhance the density of GMM loci on a transcript map. This first generation transcript map will be useful for trait mapping, comparative mapping with other legume species as well as linking genetic map with physical map of chickpea as the GMM loci integrated in this map are sequence and based and represent the genes. It is also anticipated that the integrated GMM loci will serve as anchor markers for other chickpea maps so that those maps can be aligned with the reference genetic map.

In summary, the present study provides primer sequences for 1,184 amplifiable markers, 281 easily assayable GMMs and a transcript map of the chickpea comprising 126 GMM loci. In addition, the study also reports detailed features of SNPs in both cultivated as well as wild species. The GMMs and this transcript map should be useful on one hand to enhance our understanding in the area of genome dynamics of chickpea as well as comparative legume genomes, on the other hand these resources may be proven very useful for chickpea genetics and breeding applications.

References

Alderborn A, Kristofferson A, Hammerling U (2000) Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res 10:1249–1258

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Arumuganathan K, Earle ED (1991) Estimation of nuclear DNA content of plants by some important plant species. Plant Mol Biol Rep 9:208–218

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res 13:513–523

Bryan GJ, Stephenson P, Collins A, Kirby J, Smith JB, Gale MD (1999) Low levels of DNA sequence variation among adapted genotypes of hexaploid wheat. Theor Appl Genet 99:192–198

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

Ching A, Rafalski A (2002) Rapid genetic mapping of ESTs using SNP pyrosequencing and indel analysis. Cell Mol Biol Lett 7:803–810

Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassel CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Choudhary S, Sethy NK, Shokeen BA, Bhatia S (2006) Development of sequence-tagged microsatellite site markers for chickpea (*Cicer arietinum* L.). Mol Ecol Notes 6:3–95

Choudhary S, Sethy NK, Shokeen BA, Bhatia S (2009) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. Theor Appl Genet 118:591–608

Cuc LM, Mace E, Crouch J, Quang VD, Long TD, Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol 8:55

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen M, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, Mcbride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B, Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2003) Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 68:69–78

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 Physiol 140:1183–1191

Gaitán-Solís E, Choi IY, Quigley C, Cregan P, Tohme J (2008) Single nucleotide polymorphisms in common bean: their discovery and genotyping using a multiplex detection system. Plant Genome 1:125–134

Gil J, Nadal S, Luna D, Moreno MT, De Haro A (1996) Variability of some physico-chemical characters in Desi and Kabuli chickpea types. J Sci Food Agric 71:179–184

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, 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

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

Hyten DL, Song O, Fickus EW, Quigley CV, Lim JS, Choi IY, Hwang EY, Pastor Corrales M, Cregan PB (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475

Ibrikci H, Knewtson SJB, Grusak MA (2003) Chickpea leaves as a vegetable green for humans: evaluation of mineral composition. J Sci Food Agric 83:945–950

Jayashree B, Bhanuprakash A, Jami A, Reddy SP, Nayak S, Varshney RK (2009) Perl module and PISE wrappers for the integrated analysis of sequence data and SNP features. BMC Res Notes 2:92

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics 8:223–233

Lander ES, Green P, Abrahamson J, Barlow A, Daley M, Lincoln S, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

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

Millan T, Winter P, Jungling R, Gil J, Rubio J, Cho S, Cobos MJ, Iruela M, Rajesh PN, Tekeoglu M, Kahl G, Muehlbauer FJ (2010) A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations. Euphytica 175:175–189

Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ (2009) A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci USA 106:18159–18164

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Res 9:163–171

Nayak SN, Zhu H, Varghese N, Datta S, Choi H-K, Horres R, Jüngling R, Singh J, Kavi Kishor PB, Sivaramakrihnan S, Hoisington DA, Kahl G, Winter P, Cook DR, Varshney RK (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. Theor Appl Genet 120:1415–1441

Nelson M, Phan H, Ellwood S, Moolhuijzen P, Hane J, Williams A, O'Lone C, Fosu-Nyarko J, Scobie M, Cakir M, Jones M, Bellgard M, Ksiazkiewicz M, Wolko B, Barker S, Oliver R, Cowling W (2006) The first gene-based map of *Lupinus angustifolius* L.—location of domestication genes and conserved synteny with *Medicago truncatula*. Theor Appl Genet 113:225–238

Ooijen Van (2006) JoinMap® 4: software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, The Netherlands

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

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorak J, Linkiewicz AM, Ratnasiri A, Bermudez-Kandianis J, Dubcovsky CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma, Miftahudin XF, Gustafson JP, Conley EJ, Nduati V, Gonzalez Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004) A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 168:701–712

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) Novel genetic mapping tools in plants SNPs and LD-based approaches. Plant Sci 162:329–333

Rajesh PN, Muehlbauer FJ (2008) Discovery and detection of single nucleotide polymorphism (SNP) in coding and genomic sequences in chickpea (*Cicer arietinum* L.). Euphytica 162:291–300

Raju NL, Gnanesh BN, Lekha P, Jayashree B, Pande S, Hiremath PJ, Byregowda M, Singh NK, Varshney RK (2010) The first set of

EST resource for gene discovery and marker development in pigeonpea (*Cajanus cajan* L.). BMC Plant Biol 10:45

Rodi CP, Darnhofer PB, Stanssens P, Zabeau M, van den Boom D (2002) A strategy for the rapid discovery of disease markers using the MassARRAY system. BioTechniques 32:62–69

Rostoks N, Ramsay L, Mackenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. Proc Natl Acad Sci USA 103:18656–18661

Sato K, Nankaku N, Takeda K (2009) A high-density transcript linkage map of barley derived from a single population. Heredity 103:110–117

Sethy N, Shokeen B, Edwards K, Bhatia S (2006) Development of microsatellite markers and analysis of intraspecific genetic variability in chickpea (*Cicer arietinum* L.). Theor Appl Genet 112:1416–1428

Singh A, Singh IK, Verma PK (2008) Differential transcript accumulation in *Cicer arietinum* L. in response to a chewing insect *Helicoverpa armigera* and defence regulators correlate with reduced insect performance. J Exp Bot 59:2379–2392

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney RK, Perovic D, Grosse I, Graner A (2007) A 1,000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

Thiel T, Kota R, Grosse I, Stein N, Graner A (2004) SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. Nucleic Acids Res 32:e5

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thudi M, Senthilvel S, Bottley A, Hash CT, Reddy AR, Feltus AF, Paterson AH, Hoisington DA, Varshney RK (2010) A comparative assessment of the utility of PCR-based marker systems in pearl millet. Euphytica 174:253–260

Varshney RK (2010) Gene-based marker systems in plants: high throughput approaches for discovery and genotyping. In: Jain SM, Brar DS (eds) Molecular techniques in crop improvement. Springer, The Netherlands, pp 119–142

Varshney RK, Graner A, Sorrells ME (2005a) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630

Varshney RK, Graner A, Sorrells ME (2005b) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23:48–55

Varshney RK, Nayak S, Jayashree B, Eshwar K, Upadhyaya HD, Hoisington DA (2007) Development of cost-effective SNP assays for chickpea genome analysis and breeding. SAT eJournal 3:1–3

Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S, Graner A (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. Mol Breed 22:1–13

Varshney RK, Hiremath P, Lekha P, Kashiwagi J, Balaji J, Deokar AA, Vadez V, Srinivasan R, Xiao Y, Gaur PM, Siddique KHM, Town CD, Hoisington DA (2009a) A comprehensive resource of drought- and salinity-responsive ESTs for gene discovery and marker development in chickpea (*Cicer arietinum* L.). BMC Genomics 10:523

Varshney RK, Hoisington DA, Nayak SN, Graner A (2009b) Molecular plant breeding: methodology and achievements. In: Somers D, Langridge P, Gustafson PJ (eds) Plant genomics: methods and protocols. The Humana Press, Totowa, pp 283–304

Voorips RE (2002) Mapchart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Winter P, Pfaff T, Udupa SM, Hüttel B, Sharma PC, Sahi S, Arrequin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*C. arietinum* L.). Mol Gen Genet 262:90–101

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 the recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for *Fusarium* races 4 and 5. Theor Appl Genet 101:1155–1163

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

Zhu YL, Song QJ, Hyten DL, Tassell CPV, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134