CNMS: The preferred genic markers for comparative genomic, molecular phylogenetic, functional genetic diversity and differential gene regulatory expression analyses in chickpea

DEEPAK BAJAJ, SHOUVIK DAS and SWARUP K PARIDA*

National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India

**Corresponding author (Fax, +91-11-26741658; Email, swarup@nipgr.ac.in; swarupdbt@gmail.com)*

The intra/inter-genomic comparative mapping-based phylogenetic footprinting identified 5 paralogous and 656 orthologous genome-wide CNMS markers in the upstream sequences of chickpea genes. These CNMS markers revealed a high-degree of gene-based syntenic relationship between chickpea and *Medicago* genomes while minimum between chickpea and *Vitis* genomes. The time of divergence and duplication estimated using CNMS markers highlight the expected phylogenetic relationships between chickpea and six dicot (legume) species as well as occurrence of ancient genome (~53 Mya) with small-scale recent segmental (~10 Mya) duplication events in chickpea. A wider level of functional molecular diversity (14 to 88%) and admixed population genetic structure was detected among *desi*, *kabuli* and wild genotypes by genic CNMS markers at a genome-wide scale suggesting their utility in large-scale genetic analysis in chickpea. The subfunctionalization at the *cis*-regulatory element region and TFBS (transcription factor binding site) motif levels in the upstream sequences of CNMS marker-associated orthologous genes than the paralogues was predominant. Functional constraint might have considerable effect on these CNMS-containing regulatory elements controlling consistent orthologous gene expression in dicots. A rapid subfunctionalization based on diverge differential expression of paralogous CNMS marker-associated genes particularly those that underwent recent small-scale segmental duplication events in chickpea was apparent. The differential regulation of expression and subfunctionalization potential of Ultra CNMS marker-associated genes suggest their utility in deciphering the complex gene regulatory function as well as identification and targeted mapping of potential genes/QTLs governing vital agronomic traits in chickpea. The gene-based CNMS markers with desirable inherent genetic attributes like higher degree of comparative genome mapping, functional genetic diversity and differential gene regulatory expression potential can significantly propel the genomics-assisted chickpea crop improvement.

[Bajaj D, Das S and Parida SK 2015 CNMS: The preferred genic markers for comparative genomic, molecular phylogenetic, functional genetic diversity and differential gene regulatory expression analyses in chickpea. *J. Biosci.* **40** 579–592] **DOI** 10.1007/s12038-015-9545-1

1. Introduction

Microsatellites/simple sequence repeats (SSRs) are widely distributed in diverse coding and non-coding (introns, and upstream and downstream regulatory regions) sequence components of crop plant genomes (Li *et al.* 2004; La Rota and Sorrells 2004; Varshney *et al.* 2005; Parida *et al.* 2006, 2009a). The microsatellite markers designed from the conserved coding sequences flanking the repeat-motifs usually exhibit a higher degree of cross-transferability across related species and even distantly related genera compared with markers derived from the genomic sequences (Yu *et al.*

Keywords. Chickpea; CNMS; comparative genome mapping; microsatellites; orthologous; paralogous; phylogeny; subfunctionalization

Supplementary materials pertaining to this article are available on the Journal of Biosciences Website at *http://www.ias.ac.in/jbiosci/sep2015/supp/Bajaj.pdf*

Published online: 20 August 2015

2004; Saha *et al.* 2005, 2006; Wang *et al.* 2005; Varshney *et al.* 2005; Gutierrez *et al.* 2005; Parida *et al.* 2006; Aggarwal *et al.* 2007; Hackauf *et al*. 2009; Parida *et al.* 2010a, b; Kumari *et al.* 2013). These useful characteristics of genic microsatellite markers evolutionarily conserved across plant taxa in both sequence and copy number led to identify numerous gene-based COS (conserved orthologous set) markers for the plant genomes (Fulton *et al.* 2002; Parida *et al.* 2006). These genic microsatellite-based COS markers, therefore, can be utilized as anchor markers for comparative genome mapping and evolutionary studies in crop plants. The genomic microsatellite markers being derived from the less conserved genomic sequences, under moderate selection pressure are more efficient than the genic microsatellite markers in detecting high degree of polymorphism and thereby distinguishing closely related genotypes (Cho *et al.* 2000; Coburn *et al.* 2002; Ni *et al.* 2002; Chen *et al.* 2002; Li *et al.* 2004; Garris *et al.* 2005; La Rota and Sorrells 2004; Parida *et al.* 2009a, b). Considering the diverse inherent genetic attributes of genic and genomic SSR markers, it is now desirable to develop such informative gene-based markers revealing higher level of cross-transferability and comparative genome mapping as well as molecular diversity potential across genotypes to be employed for genomics-assisted crop improvement. In this context, CNMS (conserved non-coding microsatellite) markers recently developed at a genome-wide scale by targeting the microsatellite repeat-motifs within the regulatory sequence element regions/transcription factor-binding sites (TFBS) of the non-coding upstream sequence components of protein-coding genes would certainly be relevant in plant species, including rice and chickpea (Parida *et al.* 2009a; Bajaj *et al.* 2015). Association of these markers with evolutionary conserved and functionally more constrained non-coding sequences of monocot and dicot plant genomes further implicates the significance of CNMS markers in defining gene regulatory functions of plant specific pathways and comprehending the non-coding regulatory sequence evolution in multiple crop plants (Zhang *et al.* 2006; Parida *et al.* 2009a; Bajaj *et al.* 2015). Henceforth, these genic CNMS markers could prove useful for further studies on comparative genome analysis, intra-/inter-genomic phylogenetic footprinting and development of syntenic networks for understanding the evolution of genes and genomes among crop species (Koch *et al.* 2000; Levy *et al.* 2001; Colinas *et al.* 2002; Kaplinsky *et al.* 2002; Morishige *et al.* 2002; Santi *et al.* 2003; Guo and Moose 2003; Lockton and Gaut 2005; Zhang *et al.* 2006; Creux *et al.* 2008; Parida *et al*. 2009a; Spensley *et al*. 2009; Freeling and Subramaniam 2009; Baxter *et al.* 2012; Spangler *et al.* 2012; Haudry *et al.* 2013; Spangler and Feltus 2013; Bajaj *et al.* 2014; Hettiarachchi *et al.* 2015). A higher degree of syntenic and colinearity potential of CNMS markers could also assist us to exploit small diploid genome (such as *Medicago*) as an inter-genomic cloning vehicle for facilitating map-based isolation/positional cloning of several useful trait-associated genes/QTLs and targeted sequencing of gene-rich regions of large genome crop species like chickpea. The most comprehensive application of marker-based comparative mapping approach is exemplified from the studies that isolated and cloned specific barley disease resistance and chickpea nodulation genes by transferring the genetic information from small diploid genomes like rice and *Medicago*, respectively (Kilian *et al.* 1995, 1997; Han *et al*. 1999; Bruggeman *et al*. 2002; Ali *et al.* 2014). A number of expression profiling and transgenic studies involving promoter-reporter gene constructs have demonstrated the significance of CNMS marker-associated genes in revealing highly regulated conserved patterns of expression among orthologous genes of related plant species (Zhang *et al.* 2006). This implicates the potential of CNMS markers in defining the regulatory networks of gene expression as well as their functional role and patterns of regulatory or promoter sequences evolution in related dicot plant species, including chickpea. Furthermore, the utility of CNMS markers for assaying high intra-specific polymorphic potential (37.6%) are well demonstrated in *desi* and *kabuli* chickpea (Bajaj *et al.* 2015). Significant association of CNMS markers based on their alteration of microsatellite repeat-length variations in the functional regulatory element/TFBS of genes with diverse agronomic traits (Bao *et al.* 2002; Zhang *et al.* 2006; Bajaj *et al.* 2015) suggest the utility of these functional genetic markers in fast establishment of marker-trait linkages and identifying genes/QTLs regulating important agronomic traits in chickpea. Based on the aforementioned possibilities, it would be interesting to evaluate the potential of genome-wide CNMS markers as developed by us currently from *desi* and *kabuli* genes for diverse applications of chickpea genomics and breeding, including comparative genome mapping, and understanding the molecular phylogeny, functional genetic diversity, population structure and complex gene regulatory function in chickpea.

2. Materials and methods

2.1 *Comparative mapping and phylogenetic analyses of CNMS markers*

For CNMS marker-based comparative genome mapping in paralogous and orthologous genes, the amino acid sequences encoded by the CNMS marker-associated chickpea genes (developed recently by Bajaj *et al.* 2015) were BLASTP ($1e \leq -10$) searched against each other and amino acid sequences of protein-coding genes annotated from *Medicago truncatula*, *Glycine max*, *Lotus japonicus*, *Cajanus cajan*,

Arabidopsis thaliana and *Vitis vinifera* (*http://www.plantgdb.org; http://www.phytozome.net; http://www.arabidopsis.org*; Singh *et al.* 2012; Varshney *et al.* 2012) using the reciprocal best hit method of OrthoMCL (Li *et al.* 2003; Parida *et al.* 2010b). To infer the CNMS marker-based syntenic relationships especially between chickpea and *Medicago* genomes, the CNMS marker loci in the paralogous and orthologous genes physically mapped on each of the eight chromosomes of chickpea were compared individually with the genes annotated on the chromosome pseudomolecules of *Medicago truncatula* (*http://www.plantgdb.org*) based on their ascending order of physical positions (bp) using Circos 0.55 (*http://circos.ca*) and MapChart 2.2 (Voorrips 2002).

For estimation of synonymous (Ks) and non-synonymous (Ka) substitution rates, the aligned amino acid sequences and their corresponding cDNA sequences of CNMS marker-associated paralogous and orthologous chickpea genes conserved across each of seven dicot species, including chickpea were analyzed using the CODEML program in PAML interface tool of PAL2NAL (Suyama *et al.* 2006; *http://www.bork.embl.de/pal2nal*). The time (million years ago, Mya) of duplication and divergence of each paralogous and orthologous genes, respectively among dicot species were calculated employing a synonymous mutation rate of λ substitutions per synonymous site per year as $T = Ks/2\lambda$ ($\lambda = 6.5 \times 10^{-9}$) (Lynch and Conery 2000).

2.2 *Assessment of CNMS marker-based functional genetic diversity and population structure*

The validated polymorphic CNMS markers distributed over eight chromosomes of chickpea (Bajaj *et al.* 2015) were utilized to determine the molecular diversity, population structure and genetic relationships among 25 *desi*, *kabuli* and wild (*Cicer reticulatum*) chickpea genotypes (supplementary table 1). The cluster analysis among genotypes was performed based on Nei and Li similarity coefficient (Nei and Li 1979) using the neighbor joining (NJ) method (with 1000 bootstrap replicates) in PowerMarker v3.51 (Liu and Muse 2005) and unrooted phylogenetic tree was constructed. The marker genotyping data of 25 chickpea genotypes was used in a model based program STRUCTURE (Pritchard *et al.* 2000) for determination of population structure using the admixture and correlated allele frequency with varying levels of K (number of populations) = 2 to 4 (burn-in of 50000 iterations, run length of 100000 and 20 independent replications of K). Using the optimum K, the population structure model representing better relationships among the genotypes was constructed and various population genetic parameters, including genetic variability (F_{ST}) and degree of admixture within and between population groups were estimated.

2.3 *Expression analysis of CNMS marker-associated genes*

To evaluate the utility of CNMS markers for differential gene regulation, expression profiling of 256, including two Ultra CNMS marker-associated orthologous and paralogous chickpea genes was performed. The vegetative leaf tissues from 15-20 days old seedlings of genotypes belonging to chickpea (*desi*: ICC4958 and JG62, and *kabuli*: ICCV2), *Glycine*, *Cajanus* and *Arabidopsis* grown in controlled environment growth chamber with preferred photoperiod (200 kilolux), temperature (21–30°C) and relative humidity (75-80%) conditions were collected. For salinity stress, 15–20 days old seedlings of chickpea at vegetative stage were placed into a beaker containing 150–200 mM NaCl solution in the growth chamber with 200 kilolux photoperiod, 22 ± 1°C temperature and 80% relative humidity. Three individual biological replicates of each tissue sample were collected. The total RNA was isolated from these leaf tissues using TRIzol Reagent (Life Technologies, USA) according to manufacturer's instructions. The quantity and quality of RNA was assessed using NanoDrop Spectrophotometer (NanoDrop Products, USA) and Qubit Fluorometer (Life technologies, USA). The equally quantified high-quality RNA isolated from leaf tissues was amplified with CNMS marker-associated gene-specific primers and internal control gene [elongation factor 1-alpha (*EF1α*)] using semi-quantitative and quantitative RT-PCR assays following Kujur *et al.* (2013) and Bajaj *et al.* (2015). The level of expression assayed by individual CNMS marker-associated genes in diverse genotypes of dicot species as well as chickpea genotypes (under salinity stress) was normalized against internal control gene *EF1α*. The expression pattern of these CNMS marker-associated genes was further compared and correlated among the genotypes under study.

3. Results and discussion

3.1 *Comparative genome mapping and evolutionary significance of CNMS markers*

Using intra/inter-genomic comparative mapping-based phylogenetic footprinting, we identified 5 (8.9%) CNMS markers in the upstream sequences of 56 paralogous chickpea genes and 656 (4.2%) markers in 15441 orthologous chickpea genes that are conserved in at least one of the six dicot genomes (*Medicago*, *Lotus*, *Glycine*, *Cajanus*, *Arabidopsis* and *Vitis*) under study. The identified orthologous CNMS markers included maximum (497, 3.2%) in the upstream sequences between chickpea and *Medicago* gene orthologs followed by chickpea and *Lotus* orthologs (463, 3%) and minimum (326, 2.4%) between

chickpea and *Vitis* gene orthologs (table 1). Based on these analyses, we identified 68 CNMS marker-associated orthologous genes conserved among chickpea and four legume species and 17 orthologous genes conserved among legume and two non-legume dicot species. However, the identification of higher proportion of CNMS marker-associated paralogous chickpea genes compared with orthologous genes is most likely due to the presence of ancient genome duplication events in chickpea (Jain *et al.* 2013; Varshney *et al.* 2013). This also suggests the effect of evolutionary pressure for retention of more CNMS across paralogs within a chickpea genome rather than orthologs among chickpea and six dicot genomes. The intra- and inter-genomic phylogenetic footprinting comparisons interestingly, identified three genes with both paralogous and orthologous CNMS. In two of such genes, the CNMS repeat-motifs carrying regulatory elements/TFBS were conserved as well as overlapped at the same positions in the sequences upstream to the initiation codons of both orthologous and paralogous chickpea genes and hence, being referred as 'Ultra-CNMS' (Zhang *et al.* 2006; Freeling *et al.* 2007). The GO (gene ontology) analysis and determination of putative functions of regulatory elements and their corresponding orthologous and paralogous, including Ultra CNMS marker-associated genes revealed their significant enrichment for transcriptional regulation and developmental processes. This is possibly due to greater regulatory constraint on transcriptional level of the transcription factors and genes regulating developmental processes that not only exhibit conserved functions of their common ancestor but also retain them even after chickpea duplication and speciation events vis-à-vis other genes.

The comprehensive sequence level analysis and comparison of orthologous and paralogous CNMS marker-associated genes gave insight into their involvement in potential subfunctionalization (partitioning of functions) at the *cis*-regulatory element region and TFBS motif levels in the upstream sequences of genes. For instance, a $(AGC)_5$ CNMS repeat-motif containing ANAERO2CONSENSUS regulatory elements present in the sequences upstream to the initiation codons of a paralogous chickpea gene-pairs shared common orthologous conserved gene regulatory elements across four legumes (*Medicago*, *Cajanus*, *Lotus* and *Glycine*) (figure 1). Interestingly, the detailed sequence level analysis within the specified regulatory element regions/TFBS of these orthologous and paralogous chickpea genes depicted a specific pattern of nucleotide conservation in each of the five legume species by the expansion/contraction of $(AGC)_n$ CNMS repeats and occurrence of mutations (single nucleotide substitutions and insertions/deletions) (figure 1) and thus might be involved in potential subfunctionalization of TFBS motifs. It could be due to reduced purifying selective pressure at CNMS repeats-containing regulatory element regions/TFBS of such chickpea gene paralogs that possibly underwent recent segmental duplication event relative to speciation of orthologs (Varshney *et al.* 2013). If subfunctionalization between duplicate CNMS marker-associated gene-pairs holds true, then these paralogous genes are expected to diverge through expression and also exhibit functional diversification by partitioning of regulatory gene functions through modification of regulatory elements/TFBS present in the upstream gene sequences. Likewise, we observed converse differential expression pattern of these two paralogous CNMS marker-associated genes at different

Table 1. Molecular dating of duplication and divergence of CNMS marker-associated paralogous and orthologous chickpea genes

Species	Number (%) of paralogous[*] and orthologous[#] CNMS marker-associated genes	Synonymous substitution rate (Ks)	Non-synonymous substitution rate (Ka)	Non-Synonymous to synonymous substitution rate (Ka/Ks)	Approximate duplication event/time of divergence (Mya)
Chickpea vs. Chickpea	3 (5.4)[a]	0.69	0.60	0.87	53
	2 (3.6)[b]	0.13	0.06	0.46	10
Chickpea vs. *Medicago*	497 (3.2)	0.17	0.14	0.82	13
Chickpea vs. *Lotus*	463 (3.0)	0.28	0.20	0.71	21
Chickpea vs. *Cajanus*	447 (2.9)	0.61	0.35	0.57	47
Chickpea vs. *Glycine*	445 (2.8)	0.64	0.33	0.52	49
Chickpea vs. *Arabidopsis*	397(2.6)	1.08	0.34	0.32	83
Chickpea vs. *Vitis*	326 (2.4)	1.22	0.23	0.19	94

[*] CNMS proportion estimated from 56 paralogous chickpea genes

[#] CNMS proportion estimated from 15441 orthologous chickpea genes

[a] Three paralogous chickpea genes showing ancient duplication events

[b] Two paralogous chickpea genes showing recent segmental duplication events

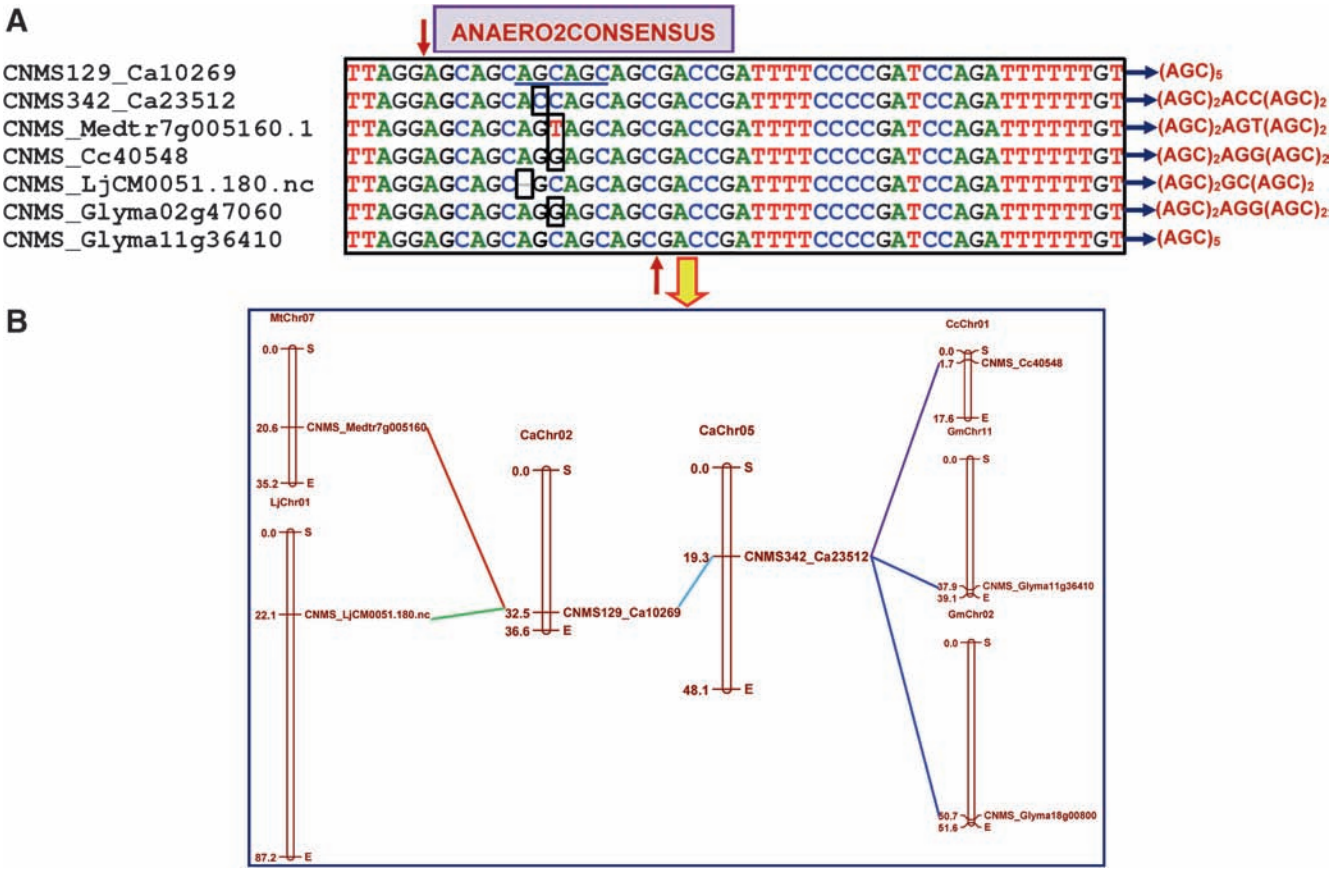

Figure 1. (**A**) Multiple sequence alignment of paralogous and orthologous Ultra CNMS marker-associated genes among *Cicer arietinum* (Ca), *Medicago truncatula* (Medtr), *Cajanus cajan* (Cc), *Lotus japonicus* (Lj) and *Glycine max* (Glyma) showed expansion/contraction of CNMS repeat-units, including occurrences of mutations (single nucleotide substitutions and insertions/deletions) at the signal sequence (AGCAGC) binding site of regulatory element (ANERO2CONSENSUS) of peptidase-C48 chickpea gene. (**B**) Comparative genome mapping of this Ultra CNMS marker-associated gene revealed syntenic relationships among the genes annotated on the chromosomes of chickpea and four legume species. The physical distance (bp) and identity of the marker loci integrated on the chromosomes of four legumes, including chickpea are indicated on the left and right side of the chromosomes, respectively. *Ca*: *Cicer arietinum*, *Mt*: *Medicago truncatula*, *Cc: Cajanus cajan*, Gm: *Glycine max* and *Lj*: *Lotus japonicus*.

vegetative and reproductive tissues and/or seed developmental stages of contrasting low and high seed weight chickpea genotypes based on our earlier expression profiling studies (Bajaj *et al.* 2015). It suggests that the fractionation of regulatory element regions/TFBS might contribute to the diverse expression pattern and rapid subfunctionalization of duplicated genes particularly those underwent recent small-scale segmental duplication events in dicots, including chickpea (Blanc and Wolfe 2004; Lockton and Gaut 2005; Freeling *et al.* 2007; Freeling and Subramaniam 2009; Baxter *et al.* 2012; Varshney *et al.* 2013). The subfunctionalization at TFBS motif levels was also predominant in case of orthologous CNMS marker-associated chickpea genes than the paralogues. For example, two orthologous LOB-domain (figure 2) and KANADI (figure 3) protein-encoding genes among chickpea and four legumes displayed the expansion/contraction of microsatellite repeats, including the presence of mutations (single nucleotide substitutions and insertions/deletions) in the $(GA)_{13}$ and $(CAA)_7$ CNMS repeat-motifs containing GAGA8HVBKN3 and RAV1AAT regulatory elements/TFBS present in the sequences upstream to these genes, respectively. However, comparative genome mapping of each of these two CNMS marker-associated orthologous genes in chickpea chromosomes expectedly showed synteny with single gene regions of individual *Medicago*, *Cajanus* and *Lotus* chromosomes, in contrast to two gene regions of *Glycine* chromosomes (figures 2 and 3). This overall ascertains that the elimination/modification of deleterious alleles by purifying selection and specific pattern of nucleotide conservation at regulatory elements/TFBS in paralogous and orthologous CNMS marker-associated genes of five diverse dicot species might combinedly cause functional constraint to these CNMS-containing regulatory elements/TFBS for controlling gene expression.

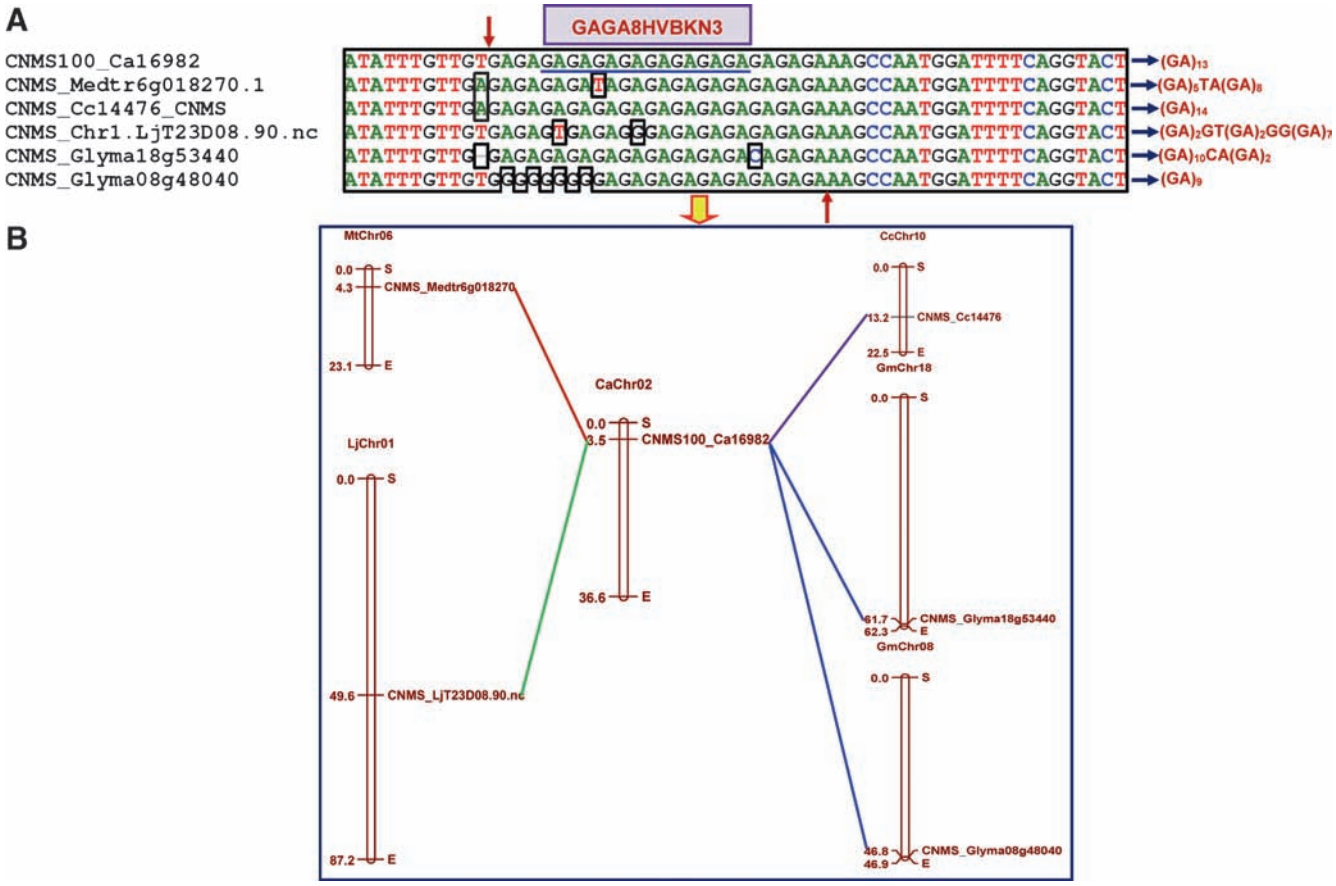

Figure 2. (**A**) Multiple sequence alignment of orthologous CNMS marker-associated genes across *Cicer arietinum* (Ca), *Medicago truncatula* (Medtr), *Cajanus cajan* (Cc), *Lotus japonicus* (Lj) and *Glycine max* (Glyma) showed expansion/contraction of CNMS repeat-units, including occurrences of mutations (single nucleotide substitutions and insertions/deletions) at the signal sequence $(GA)_8$ binding site of regulatory element (GAGA8HVBKN3) of LOB-domain protein encoding chickpea gene. (**B**) Comparative genome mapping of these orthologous CNMS marker-associated gene revealed syntenic relationships among the genes annotated on the chromosomes of chickpea and four legume species. The physical distance (bp) and identity of the marker loci integrated on the chromosomes of four legumes, including chickpea are indicated on the left and right side of the chromosomes, respectively. *Ca*: *Cicer arietinum*, *Mt*: *Medicago truncatula*, *Cc: Cajanus cajan*, Gm: *Glycine max* and *Lj*: *Lotus japonicus*.

The estimation of proportionate distribution of non-synonymous (Ka) to synonymous (Ks) substitution rate across orthologous and paralogous CNMS marker-associated chickpea genes depicted that a relatively large fraction of such genes are under negative/purifying selection pressure (Ka/Ks < 1.0) (table 1) and appears to be conserved substantially. This could be due to the representation of CNMS-containing regulatory elements as signature of functionally constrained elements, which are maintained by purifying selection. Therefore, the intra/inter-genomic phylogenetic footprinting-based comparative genomics approaches would be useful to successfully guide the prediction of potential *cis*-regulatory sequences of genes regulating differential expression for particular traits of agricultural importance in crop plants (Freeling and Subramaniam 2009; Baxter *et al.* 2012). The average Ka/Ks ratio was maximum between chickpea and *Medicago* orthologous CNMS marker-associated chickpea genes and minimum between chickpea and *Vitis* gene orthologs (table 1). This suggests that the efficiency of purifying selection against non-synonymous sequence polymorphisms was highest for the chickpea and *Medicago* orthologous genes compared with other chickpea-legume and/or non-legume orthologues. This could be the results of phylogenetic relationships among chickpea and six legume and non-legume species as observed in earlier genome-wide gene evolution studies (Young and Bharti 2012; Varshney *et al.* 2013; Jain *et al.* 2013). However, the rate of purifying selection based on Ka/Ks value was more in the paralogous CNMS marker-associated chickpea genes than that of orthologous genes among chickpea and six dicot genomes (table 1). Therefore, identification of higher proportion of paralogous CNMS marker-associated genes in contrast to orthologous genes from chickpea genome is expected.

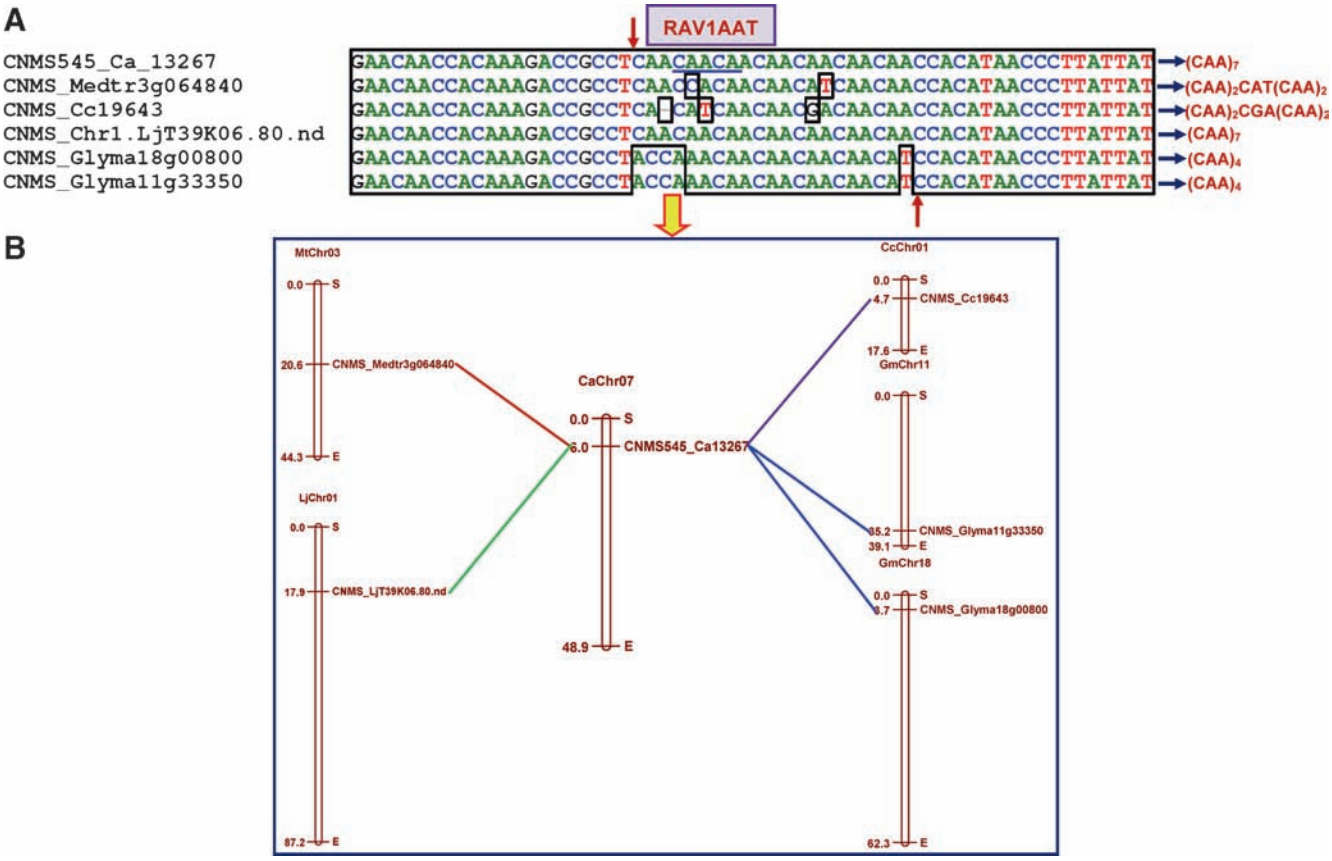

Figure 3. (**A**) Multiple sequence alignment of orthologous CNMS marker-associated genes across *Cicer arietinum* (Ca), *Medicago truncatula* (Medtr), *Cajanus cajan* (Cc), *Lotus japonicus* (Lj) and *Glycine max* (Glyma) showed expansion/contraction of CNMS repeat-units, including occurrences of mutations (single nucleotide substitutions and insertions/deletions) at the signal sequence (CAACA) binding site of regulatory element (RAV1AAT) of KANADI protein encoding chickpea gene. (**B**) Comparative genome mapping of these orthologous CNMS marker-associated gene revealed syntenic relationships among the genes annotated on the chromosomes of chickpea and four legume species. The physical distance (bp) and identity of the marker loci integrated on the chromosomes of four legumes, including chickpea are indicated on the left and right side of the chromosomes, respectively. *Ca*: *Cicer arietinum*, *Mt*: *Medicago truncatula*, *Cc: Cajanus cajan*, Gm: *Glycine max* and *Lj*: *Lotus japonicus*.

The estimation of time of divergence and duplication based on synonymous substitution rates (K_S) in the paralogous and orthologous CNMS marker-associated chickpea genes revealed maximum Ks values between chickpea and *Vitis* (1.22) giving their approximate divergence time at ~94 Mya (table 1). This was followed by chickpea-*Arabidopsis* (1.08 modal Ks, with divergence time ~83 Mya), chickpea-*Glycine* (0.64, ~49 Mya), chickpea-*Cajanus* (0.61, ~47 Mya), chickpea-*Lotus* (0.28, ~21 Mya) and chickpea-*Medicago* (0.17, ~13 Mya). The estimation of Ks and time of duplication (Mya) in the five paralogous CNMS marker-associated chickpea gene-pairs indicated that ancient duplication event (0.69, ~53 Mya) of three gene-pairs within chickpea was in the midst of average divergence time of chickpea from four legumes (*Medicago*, *Lotus*, *Cajanus* and *Glycine*, ~32 Mya) and two non-legume species (*Arabidopsis* and *Vitis*, ~88 Mya) (table 1). The remaining two paralogous CNMS marker-associated chickpea gene-pairs having altered TFBS (subfunctionalization) at upstream sequences showed recent segmental duplication event near about 10 Mya (Ks: 0.13) after the divergence between chickpea and *Medicago* (~13 Mya). The time of divergence and duplication estimated for the orthologous and paralogous CNMS marker-associated chickpea genes are consistent with the limits of ~100 Mya divergence time as documented for reliable and significant CNS identification in plant genomes based on intra-/inter-genomic phylogenetic footprinting (Freeling and Subramaniam 2009; Reineke *et al.* 2011). This is further clearly evident from the estimation of divergence and duplication time of functionally important Ultra-CNMS marker-associated genes, which are highly conserved under strong purifying selective pressure from a common ancestor over ~90 Mya. The time of divergence

and duplication further emphasizes the known phylogenetic relationships among chickpea and six legume and non-legume species and occurrence of ancient genome duplication event (~58 Mya) along with small-scale recent segmental duplication events in chickpea (Young and Bharti 2012; Jain *et al.* 2013; Varshney *et al.* 2013).

The comparative mapping of 666 CNMS marker loci physically mapped on eight chickpea chromosomes with their physical positions on the genes annotated on the pseudomolecules of *Medicago* chromosomes illustrated a high degree of syntenic relationships (497 markers, 74.6%) among two genomes (table 2). High degree of synteny was observed between the chickpea chromosome 2 and *Medicago* chromosome 5 (56.3%, 27 markers) followed by those between chickpea chromosome 5 and *Medicago* chromosome 3 (52.3%, 34 markers) (supplementary figures 1 and 2). The genic CNMS marker-based comparative maps identified many conserved collinear chromosomal regions among chickpea and *Medicago* (supplementary figure 1). The observed syntenic relationships between chickpea and *Medicago* chromosomes are similar to the previous marker-based comparative genome mapping studies (Nayak *et al.* 2010; Gaur *et al.* 2012; Hiremath *et al.* 2012). Striking synteny between chickpea and *Medicago* chromosomes is expected keeping in view their evolutionary closeness as they belong to the same clade Galegoid (Choi *et al.* 2004; Zhu *et al.* 2005; Young *et al.* 2011; Young and Bharti 2012). The CNMS marker-based comparative physical map constructed among chickpea and *Medicago* chromosomes and genome-wide sequence level conservation observed at regulatory element regions/TFBS of orthologous genes particularly those lying within the conserved chromosomal collinear regions between these two legume species would increasingly guide cloning and targeted mapping of genes/QTLs as well as sequencing of gene-rich regions of chickpea genome using the positional information of candidate genes from completely sequenced model legume species like *Medicago*.

3.2 *CNMS marker-based functional molecular diversity and population genetic structure in chickpea*

The estimation of pair-wise similarity among 24 *desi* and *kabuli* and one wild chickpea genotypes (supplementary table 1) using 256 polymorphic CNMS markers distributed over eight chromosomes (Bajaj *et al.* 2015) depicted a wide range of similarity coefficient from 0.12 to 0.86 with an average of 0.44. The similarity among *desi* and *kabuli* genotypes ranged from 0.23 to 0.71 with an average of 0.56. The level of genetic diversity (14 to 88%) estimated among chickpea genotypes using the CNMS markers is higher than that detected previously with the EST-derived

Table 2. Comparative genome mapping of chickpea CNMS marker-associated genes on the *Medicago* chromosomes

Chromosomes	Physically mapped chickpea CNMS markers	*Medicago* chromosomes								
		*Mt*Chr1	*Mt*Chr2	*Mt*Chr3	*Mt*Chr4	*Mt*Chr5	*Mt*Chr6	*Mt*Chr7	*Mt*Chr8	Total
*Ca*Chr1	80	2 (3.2)	**23 (37.1)**	7 (11.3)	10 (16.1)	9 (14.5)	0 (0.0)	9 (14.5)	2 (3.2)	**62** (77.5)
*Ca*Chr2	63	4 (8.3)	3 (6.3)	5 (10.4)	2 (4.2)	**27 (56.3)**	3 (6.3)	4 (8.3)	0 (0.0)	**48** (76.2)
*Ca*Chr3	70	6 (11.5)	5 (9.6)	5 (9.6)	6 (11.5)	4 (7.7)	0 (0.0)	**23 (44.2)**	3 (5.8)	**52** (74.3)
*Ca*Chr4	116	**34 (40.5)**	6 (7.1)	12 (14.3)	9 (10.7)	8 (9.5)	3 (3.6)	6 (7.1)	6 (7.1)	**84** (72.4)
*Ca*Chr5	85	2 (3.1)	3 (4.6)	**34 (52.3)**	4 (6.2)	6 (6.2)	1 (1.5)	9 (13.8)	6 (6.2)	**65** (76.5)
*Ca*Chr6	108	11 (12.8)	7 (8.1)	7 (8.1)	**28 (32.6)**	14 (16.3)	4 (4.7)	5 (5.8)	10 (11.6)	**86** (79.6)
*Ca*Chr7	107	3 (4.1)	9 (12.2)	8 (10.8)	16 (21.6)	10 (13.5)	3 (4.1)	7 (9.5)	**18 (24.3)**	**74** (69.2)
*Ca*Chr8	37	1 (3.8)	1 (3.8)	4 (15.4)	4 (15.4)	**13 (50.0)**	3 (11.5)	0 (0.0)	0 (0.0)	**26** (70.3)
Total	**666**	**63** (12.7)	**57** (11.5)	**82** (16.5)	**79** (15.9)	**91** (18.3)	**17** (3.4)	**63** (12.7)	**45** (9.1)	**497** (74.6)

Boldface digits indicate the maximum number (%) of CNMS markers exhibiting higher degree of syntenic relationship between *Medicago* and chickpea chromosomes

(3 to 49%; Choudhary *et al.* 2009) and genomic (37 to 80%; Sethy *et al.* 2006 and 32 to 80%, Bharadwaj *et al.* 2011) microsatellite markers. The higher efficiency of CNMS markers in assaying functional molecular diversity in the regulatory sequence component of the genome suggested their utility in establishing distinctness among the *desi* and *kabuli* genotypes and thus would have significance in the perspective of varietal improvement in chickpea. The genetic relationships among 25 genotypes was depicted in a dendrogram (figure 4). The CNMS markers clearly discriminated all the genotypes from each other resulting in a definite grouping among *desi* and *kabuli* and wild chickpea. As expected, the correspondence of clustering pattern obtained among the genotypes using genic CNMS markers with the pedigree relationship and parentage further suggests that the functional diversity assayed by these markers developed from the upstream regulatory regions of genes is realistic and thus would be useful in chickpea genomics-assisted breeding.

The analysis of population genetic structure among the 25 chickpea genotypes (supplementary table 1) using the STRUCTURE with varying levels of K (K = 2 to 4) with 10 replications (figure 5A, B, C) revealed that at K value of 3, all the genotypes were classified into three distinct populations, *kabuli*, *desi* and wild as per expected pedigree relationships and parentage with high resolution population structure (figure 5B). This was further comparable to the clustering pattern as obtained among the 25 chickpea genotypes by the neighbor-joining tree analysis (figure 4) using pair-wise genetic distances. Based on population structure analysis, the population group I consisting of 7 genotypes of *kabuli*, group II with 17 genotypes of *desi* and group III having one genotype of the wild species *C. reticulatum* (figure 5B). The molecular genetic variation among and within the three populations based on 256 polymorphic CNMS markers revealed a wider level of quantitative genetic differentiation (F_{ST} varied from 0.08 to 0.73 with an average of 0.46) among three population groups. Among the three population groups, the divergence was maximum between *kabuli* and wild (F_{ST} = 0.53) and minimum between *desi* and *kabuli* (0.21). The proportion of F_{ST} and thus diversity between population groups (43%) was higher as compared to that estimated within populations (29%). Higher genetic differentiation was observed in *desi* (F_{ST} 0.28) than that of *kabuli* (F_{ST} 0.19). All the 25 genotypes clearly belonged to a single population in which about 92.4% of their inferred ancestry was derived from one of the model-based population and remaining ~7.6% contained admixed ancestry (supplementary table 1). The existence of admix ancestry might have resulted from complex breeding history involving intercrossing and introgression among *desi*, *kabuli* and wild chickpea genotypes coupled with different strong adaptive

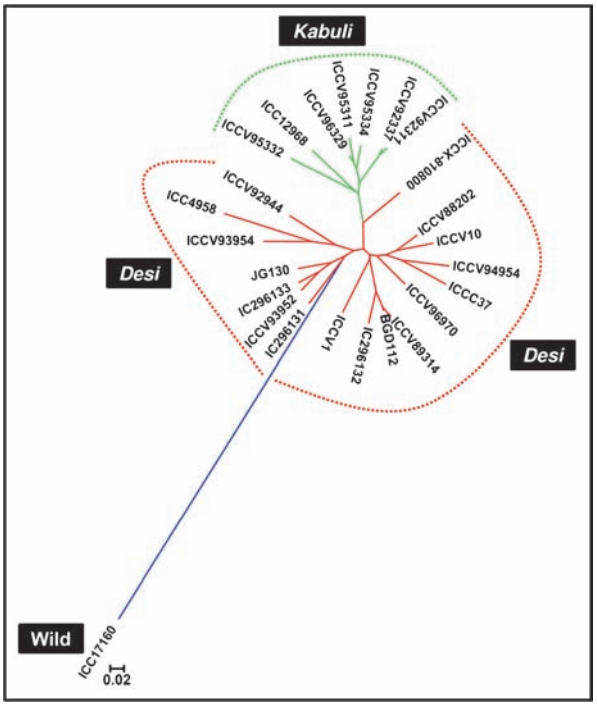

Figure 4. Unrooted phylogenetic tree depicting the genetic relationships among 24 *desi* and *kabuli* and one wild chickpea genotypes based on Nei and Li's similarity coefficient using 256 polymorphic gene-based CNMS markers. Molecular classification corresponded to the known evolutionary and pedigree relationship as well as the parentage.

selection pressure during their domestication. Maximum admixture (~4%) was observed between *desi* and *kabuli* populations followed by *desi* and wild (~2%) and *kabuli* and wild (~0.2%) populations (figure 5B). The observed results stand parallel with the earlier documentation on genetic diversity and phylogenetic relationship studies among *desi*, *kabuli* and wild chickpea using genic and genomic microsatellite markers (Sethy *et al.* 2006; Upadhyaya *et al.* 2008; Bharadwaj *et al.* 2011; Kujur *et al.* 2013).

3.3 *Differential regulation of CNMS marker-associated gene expression*

To evaluate the potential of identified CNMS markers in gene regulation, differential expression profiling of 256 informative CNMS marker-associated genes showing polymorphism among chickpea genotypes was performed. One Ultra CNMS marker-associated orthologous and paralogous peptidase-C48 chickpea gene was also included for expression analysis to determine its potential role in sub-functionalization. The differential expression profiling of all these genes was analyzed in the leaves of genotypes representing four dicot species (*Glycine*, *Cajanus*, *Arabidopsis*

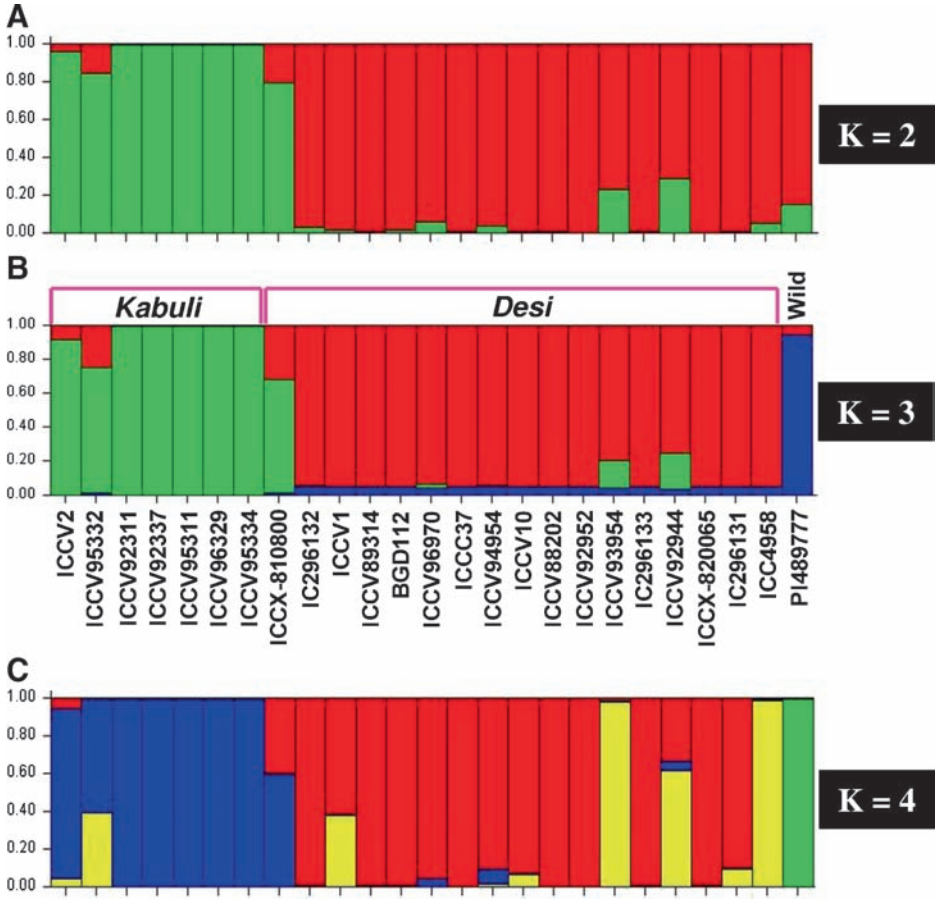

Figure 5. Optimization of number of populations (K value) varying from K = 2 to K = 4 to determine the best possible population genetic structure among 25 chickpea genotypes using 256 polymorphic genic CNMS markers distributed across eight chromosomes. These mapped genic markers assigned chickpea genotypes into three distinct *desi*, *kabuli* and wild population groups. The chickpea genotypes represented by vertical bars along the horizontal axis were classified into K colour segments based on their estimated membership fraction in each K cluster.

and *C. arietinum*) as well as between control and salinity stressed leaves of two contrasting tolerant (JG62) and sensitive (ICCV2) chickpea genotypes using semi-quantitative and quantitative RT-PCR assays. One-hundred and ninety-eight (77.3%) genes were expressed consistently without any significant quantitative variation of expression value in leaf tissues of all the genotypes belonging to four dicot species as well as control and salinity stressed leaves of two contrasting tolerant and sensitive chickpea genotypes. This functional coherence of CNMS marker-associated genes across dicots, including legumes is in line with that documented earlier in the grass species (Kaplinsky *et al.* 2002). Forty-five CNMS marker-associated genes were differentially up- (29 genes) and down-regulated (16 geners) (≥2-fold) in the control (non-stressed) and salinity stressed leaves of tolerant (JG62) and sensitive (ICCV2) chickpea genotypes. Of these, 13 genes exhibited converse differential regulation between tolerant and sensitive chickpea genotypes under salinity stress. Interestingly, fragment length polymorphism between ICCV2 and JG62 based on CNMS repeat-unit variations within the regulatory elements/TFBS of all these 13 genes was evident. Therefore, functionally relevant 13 CNMS marker-associated genes chosen by us could be utilized as potential candidates in understanding the salinity stress-mediated gene regulatory functions in chickpea.

One Ultra CNMS marker-associated orthologous and paralogous peptidase-C48 chickpea gene with ANAERO2CONSENSUS regulatory element-containing $(AGC)_5$ CNMS repeat-motif was up-regulated (~5-fold) in leaves of tolerant (JG62) chickpea genotype under salinity (NaCl) stress as compared to its non-stress control (figure 6). In contrast, no quantitative variation of the transcript expression of its paralogous chickpea gene with $(AGC)_2ACC(AGC)_2$ CNMS repeat-motif was observed between control and salt stressed leaves of tolerant JG62 and sensitive ICCV2 chickpea genotypes (figure 6). However, in the four dicot species irrespective of CNMS repeat-motif length variations, the expression of the transcripts for this peptidase-C48 gene was consistent (figure 6) suggesting that the Ultra CNMS marker-associated gene is

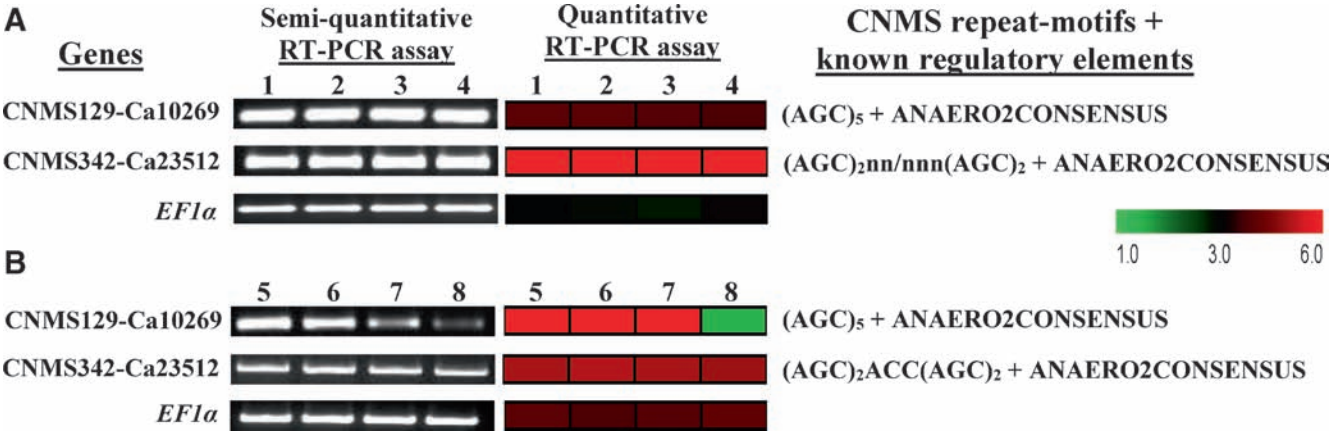

Figure 6. Differential expression profiling of one paralogous and orthologous Ultra CNMS marker-associated peptidase-C48 chickpea gene in the leaf tissues of genotypes belonging to four dicot species as well as control (non-stressed) and salt stressed leaves of tolerant (JG62) and sensitive (ICCV2) chickpea genotypes using the semi-quantitative and quantitative RT-PCR assays. The expression values across leaf tissues of different samples used, were normalized by endogenous control *EF1α* in quantitative RT-PCR assay. The colour scale represents the average log of the gene signal expression values in diverse samples, in which green, black and red colours denote the low, medium and high expression levels, respectively. Details on CNMS marker-associated gene is illustrated in the Figure 1. The orthologous (**A**) Ultra CNMS marker-associated peptidase-C48 chickpea gene expression pattern in the leaf tissues of genotypes belonging to *Glycine* (lane 1), *Cajanus* (lane 2), *Arabidopsis* (lane 3) and *C. arietinum* (lane 4), whereas its paralogous (**B**) gene indicates the differential expression pattern in leaves of salt stressed (lane 5) and control (lane 6) sensitive chickpea genotype (ICCV2) as well as in leaves of salt stressed (lane 7) and control (lane 8) tolerant chickpea genotype (JG62).

functional in all these species, but differentially expressed in response to salinity stress in chickpea. It would be thus interesting to understand the subfunctionalization characteristics of Ultra CNMS marker-associated genes based on differential transcript profiling in various tissues/developmental stages of contrasting genotypes under stress to decipher the regulatory complexities of genes controlling diverse important agronomic traits in chickpea.

The differential regulation of Ultra CNMS marker-associated peptidase-C48 chickpea gene in response to salt stress is consistent with the up-regulated pattern of expression as assayed by its gene homolog (*OTS1*: *OVERLY TOLERANT TO SALT 1*) earlier in *Arabidopsis thaliana* under salinity stress. Mutant complementation analysis of this gene homolog have further proved its involvement in regulation of salinity stress tolerance in *Arabidopsis* (Conti *et al.* 2008). These findings thus suggest a possible correlation between the expansion/contraction of numbers of CNMS repeat-units and regulation of gene expression in response to abiotic stress in chickpea, which needs further validation. Alteration of CNMS repeats in the *cis*-regulatory element-binding regions of genes is known to control light and salicylic acid responses, respectively in *Arabidopsis* and *Brassica* (Zhang *et al.* 2006). The CNMS identified in this study thus would be of immediate use as markers for identifying genes harboring QTLs and (expressed) eQTLs that regulate various qualitative and quantitative traits such as flowering time QTLs in maize (Salvi *et al.* 2007) and seed weight QTLs/eQTLs in chickpea (Bajaj *et al.* 2015).

4. Conclusion

Collectively, our study ascertained the potential of CNMS markers for comparative genome mapping and understanding phylogeny, functional genetic diversity and population structure in chickpea. These desirable inherent genetic attributes establish CNMS as preferred informative gene-based markers for various large-scale genotyping applications in chickpea. The differential regulation of expression and subfunctionalization potential of CNMS (Ultra CNMS) marker-associated orthologous and parlaogous genes imply their substantial involvement in elucidating complex gene regulatory function and identifying potential genes/QTLs governing useful agronomic traits in chickpea.

Acknowledgments

The authors gratefully acknowledge the financial support for this research study provided by a research grant (102/IFD/SAN/2161/2013-14) from the Department of Biotechnology (DBT). SD acknowledges the DBT for Junior Research Fellowship award.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V and 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

Ali L, Madrid E, Varshney RK, Azam S, Millan T, Rubio J and Gil J 2014 Mapping and identification of a *Cicer arietinum NSP2* gene involved in nodulation pathway. *Theor. Appl. Genet.* **127** 481–488

Bajaj D, Saxena MS, Kujur A, Das S, Badoni S, Tripathi S, Upadhyaya HD, Gowda CLL, *et al*. 2015 Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. *J. Exp. Bot.* **66** 1271–1290

Bao S, Corke H and 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

Baxter L, Jironkin A, Hickman R, Moore J, Barrington C, Krusche P, Dyer NP, Buchanan-Wollaston V, *et al*. 2012 Conserved noncoding sequences highlight shared components of regulatory networks in dicotyledonous plants. *Plant Cell* **24** 3949–3965

Bharadwaj C, Srivastava R, Chauhan SK, Satyavathi CT, Kumar J, Faruqui A, Yadav S, Rizvi AH, *et al*. 2011 Molecular diversity and phylogeny in geographical collection of chickpea (*Cicer* sp.) accessions. *J. Genet.* **90** e94–e100

Blanc G and Wolfe KH 2004 Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16** 1679–1691

Bruggeman FJ, Westerhoff HV, Hoek JB and Kholodenko BN 2002 Modular response analysis of cellular regulatory networks. *J. Theor. Biol.* **218** 507–520

Chen X, Cho YG and McCouch SR 2002 Sequence divergence of rice microsatellites in *Oryza* and other plant species. *Mol. Gen. Genomics* **268** 331–343

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, Park WD, Ayres N, Cartinhour S, *et al*. 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, *et al*. 2004 Estimating genome conservation between crop and model legume species. *Proc. Natl. Acad. Sci. USA* **101** 15289–15294

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

Coburn R, Temnykh SV, Paul EM and McCouch SR 2002 Design and application of microsatellite marker panels for semi-automated genotyping of rice (*Oryza sativa* L.). *Crop Sci.* **42** 2092–2099

Colinas J, Birnbaum K and Benfey PN 2002 Using cauliflower to find conserved non-coding regions in *Arabidopsis*. *Plant Physiol.* **129** 451–454

Conti L, Price G, O'Donnell E, Schwessinger B, Dominy P and Sadanandom A 2008 Small ubiquitin-like modifier proteases OVERLY TOLERANT TO SALT1 and -2 regulate salt stress responses in *Arabidopsis*. *Plant Cell* **20** 2894–2908

Creux NM, Ranik M, Berger DK and Mybrug AA 2008 Comparative analysis of orthologous cellulose synthase promoters from *Arabidopsis*, *Populus* and *Eucalyptus*: evidence of conserved regulatory elements in angiosperms. *New Phytol.* **179** 722–737

Freeling M and Subramaniam S 2009 Conserved noncoding sequences (CNSs) in higher plants. *Curr. Opin. Plant Biol.* **12** 126–132

Freeling M, Rapaka L, Lyons E, Pedersen B and Thomas BC 2007 G-boxes, bigfoot genes, and environmental response: characterization of intragenomic conserved noncoding sequences in *Arabidopsis*. *Plant Cell* **19** 1441–1457

Fulton TM, Hoeven RV, Eannetta NT and Tanksley SD 2002 Identification, analysis and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell* **14** 1457–1467

Garris AJ, Tai TH, Coburn J, Kresovich S and McCouch SR 2005 Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169** 1631–1638

Gaur R, Azam S, Jeena G, Khan AW, Choudhary S, Jain M, Yadav G, Tyagi AK, *et al*. 2012 High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). *DNA Res.* **19** 357–373

Guo H and Moose SP 2003 Conserved non-coding sequences among cultivated cereal genomes identify candidate regulator sequence elements and patterns of promoter evolution. *Plant Cell* **15** 1143–1158

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

Hackauf B, Rudd S, van der Voort JR, Miedaner T and Wehling P 2009 Comparative mapping of DNA sequences in rye (*Secale cereale* L.) in relation to the rice genome. *Theor. Appl. Genet.* **118** 371–384

Han F, Kilian A, Chen JP, Kudrna D, Steffenson B, Yamamoto K, Matsumoto T, Sasaki T, *et al*. 1999 Sequence analysis of a rice BAC covering the syntenous barley *Rpg1* region. *Genome* **42** 1071–1076

Haudry A, Platts AE, Vello E, Hoen DR, Leclercq M, Williamson RJ, Forczek E, Joly-Lopez Z, *et al*. 2013 An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat. Genet.* **45** 891–898

Hettiarachchi N, Kryukov K, Sumiyama K and Saitou N 2014 Lineage specific conserved noncoding sequences of plant genomes: their possible role in nucleosome positioning. *Genome Biol. Evol.* **6** 2527–2542

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, *et al*. 2012 Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* **10** 716–732

Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, Shah N, Singh VK, *et al*. 2013 A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* **74** 715–729

Kaplinsky NJ, Braun DM, Penterman J, Goff SA and Freeling M 2002 Utility and distribution of conserved noncoding sequences in the grasses. *Proc. Natl. Acad. Sci. USA* **99** 6147–6151

Kilian A, Kudrna DA, Kleinhofs A, Yano M, Kurata N, Steffenson B and Sasaki T 1995 Rice-barley synteny and its application to

saturation mapping of the barley *Rpg1* region. *Nucl. Acids Res.* **23** 2729–2733

Kilian A, Chen J, Han F, Steffenson B and Kleinhofs A 1997 Towards map-based cloning of the barley stem rust resistance genes *rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. *Plant Mol. Biol.* **35** 187–195

Koch MA, Haubold B and Mitchell-Olds T 2000 Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (*Brassicaceae*). *Mol. Biol. Evol.* **17** 1483–1498

Kujur A, Bajaj D, Saxena MS, Tripathi S, Upadhyaya HD, Gowda CLL, Singh S, Jain M, *et al*. 2013 Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* **20** 355–373

Kumari K, Muthamilarasan M, Misra G, Gupta S, Subramanian A, Parida SK, Chattopadhyay and Parsad M 2013 Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PLoS One* **8** e67742

La Rota M and Sorrells ME 2004 Comparative DNA sequence analysis of mapped wheat ESTs reveals complexity of genome relationships between rice and wheat. *Funct. Integr. Genomics* **4** 34–46

Levy S, Hannennalli S and Workman C 2001 Enrichment of regulatory signals in the conserved non-coding genomic sequences. *Bioinformatics* **17** 871–877

Li L, Stoeckert CJ Jr and Roos DS 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13** 2178–2189

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

Liu K and Muse SV 2005 PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21** 2128–2129

Lockton S and Gaut BS 2005 Plant conserved non-coding sequences and paralogue evolution. *Trends Genet.* **21** 60–65

Lynch M and Conery JS 2000 The evolutionary fate and consequences of duplicate genes. *Science* **290** 1151–1155

Morishige DT, Childs KL, Moore LD and Mullet JE 2002 Targeted analysis of orthologous phytochrome A regions of the sorghum, maize, and rice genomes using comparative gene-island sequencing. *Plant Physiol.* **130** 1614–1625

Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R, Jüngling R, Singh J, *et al*. 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

Nei M and Li WH 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76** 5269–5273

Ni J, Colowit PM and Mackill DJ 2002 Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Sci.* **42** 601–607

Parida SK, Raj Kumar KA, Dalal V, Singh NK and Mohapatra T 2006 Unigene derived microsatellite markers for the cereal genomes. *Theor. Appl. Genet.* **112** 808–817

Parida SK, Dalal V, Singh NK and Mohapatra T 2009a Genic non-coding microsatellites in the rice genome: characterization, marker design and use in assessing genetic and evolutionary relationships among domesticated groups. *BMC Genomics* **10** 140

Parida SK, Kalia SK, Kaul S, Dalal V, Hemaprabha G, Selvi A, Pandit A, Singh A, *et al*. 2009b Informative genomic microsatellite markers for efficient genotyping applications in sugarcane. *Theor. Appl. Genet.* **118** 327–338

Parida SK, Pandit A, Gaikwad K, Sharma TR, Srivastava PS, Singh NK and Mohapatra T 2010a Functionally relevant microsatellites in sugarcane unigenes. *BMC Plant Biol.* **10** 251

Parida SK, Yadava DK and Mohapatra T 2010b Microsatellites in *Brassica* unigenes: Relative abundance, marker design and use in comparative physical mapping and genome analysis. *Genome* **53** 55–67

Pritchard JK, Stephens M and Donnelly PJ 2000 Inference of population structure using multilocus genotype data. *Genetics* **155** 945–959

Reineke AR, Bornberg-Bauer E and Gu J 2011 Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. *Nucl. Acids Res.* **39** 6029–6043

Saha MC, Mian MAR, Zwonitzer JC, Chekhovskiy K and Hopkins AA 2005 An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). *Theor. Appl. Genet.* **110** 323–336

Saha MC, Cooper JD, Mian MA, Chekhovskiy K and May GD 2006 Tall fescue genomic SSR markers: development and transferability across multiple grass species. *Theor. Appl. Genet.* **113** 1449–1458

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, *et al*. 2007 Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl. Acad. Sci. USA* **104** 11376–11381

Santi L, Wang Y, Stile MR, Berendzen K, Wanke D, Roig C, Pozzi C, Muller K, *et al*. 2003 The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene *Bkn3*. *Plant J.* **34** 813–826

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

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, Bhutani S, Dogra V, *et al*. 2012 The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.* **21** 98–112

Spangler JB and Feltus FA 2013 Conserved non-coding sequences are associated with rates of mRNA decay in *Arabidopsis*. *Front. Plant Sci.* **4** 129

Spangler JB, Subramaniam S, Freeling M and Feltus FA 2012 Evidence of function for conserved noncoding sequences in *Arabidopsis thaliana*. *New Phytol.* **193** 241–252

Spensley M, Kim JY, Picot E, Reid J, Ott S, Helliwell C and Carré IA 2009 Evolutionarily conserved regulatory motifs in the promoter of the *Arabidopsis* clock gene *LATE ELONGATED HYPOCOTYL*. *Plant Cell* **21** 2606–2623

Suyama M, Torrents D and Bork P 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl. Acids Res.* **34** W609–W612

Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda CLL, Hoisington D and Singh S 2008 Genetic structure,

diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* **8** 106

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

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, *et al*. 2012 Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* **30** 83–89

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, *et al*. 2013 Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* **31** 240–246

Voorrips RE 2002 MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93** 77–78

Wang ML, Barkley NA, Yu JK, Dean RE, Newman ML, Sorrells ME and Pederson GA 2005 Transfer of simple sequence repeat (SSR) markers from major cereal crops to minor grass species for germplasm characterization and evaluation. *Plant Genet. Resour.* **3** 45–57

Young ND and Bharti AK 2012 Genome-enabled insights into legume biology. *Annu. Rev. Plant Biol.* **63** 283–305

Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, *et al*. 2011 The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480** 520–524

Yu JK, La Rota M, Kantety RV and Sorrells ME 2004 EST derived SSR markers for comparative mapping in wheat and rice. *Mol. Genet. Genomics* **271** 742–751

Zhang L, Zuo K, Zhang F, Cao Y, Wang J, Zhang Y, Sun X and Tang K 2006 Conservation of noncoding microsatellites in plants: implication for gene regulation. *BMC Genomics* **7** 323

Zhu H, Choi HK, Cook DR and Shoemaker RC 2005 Bridging model and crop legumes through comparative genomics. *Plant Physiol.* **137** 1189–1196

MS received 26 December 2014; accepted 03 July 2015

Corresponding editor: Rajeev Kumar Varshney