



A Genome-wide Combinatorial Strategy Dissects Complex Genetic Architecture of Seed Coat Color in Chickpea

OPEN ACCESS

Edited by:

Maria J. Monteros,
The Samuel Roberts Noble
Foundation, USA

Reviewed by:

Steven B. Cannon,
United States Department of
Agriculture - Agricultural Research
Service, USA
Jianping Wang,
University of Florida, USA
Bharadwaj Chellapilla,
Indian Agricultural Research Institute,
India

*Correspondence:

Swarup K. Parida
swarup@nipgr.ac.in;
swarupdbt@gmail.com

†These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 21 February 2015

Accepted: 26 October 2015

Published: 17 November 2015

Citation:

Bajaj D, Das S, Upadhyaya HD,
Ranjan R, Badoni S, Kumar V,
Tripathi S, Gowda CLL, Sharma S,
Singh S, Tyagi AK and Parida SK
(2015) A Genome-wide Combinatorial
Strategy Dissects Complex Genetic
Architecture of Seed Coat Color in
Chickpea. *Front. Plant Sci.* 6:979.
doi: 10.3389/fpls.2015.00979

Deepak Bajaj^{1†}, Shouvik Das^{1†}, Hari D. Upadhyaya^{2†}, Rajeev Ranjan¹, Saurabh Badoni¹, Vinod Kumar³, Shailesh Tripathi⁴, C. L. Laxmipathi Gowda², Shivali Sharma², Sube Singh², Akhilesh K. Tyagi¹ and Swarup K. Parida^{1*}

¹ National Institute of Plant Genome Research, New Delhi, India, ² International Crops Research Institute for the Semi-Arid Tropics, Telangana, India, ³ National Research Centre on Plant Biotechnology, New Delhi, India, ⁴ Division of Genetics, Indian Agricultural Research Institute, New Delhi, India

The study identified 9045 high-quality SNPs employing both genome-wide GBS- and candidate gene-based SNP genotyping assays in 172, including 93 cultivated (*desi* and *kabuli*) and 79 wild chickpea accessions. The GWAS in a structured population of 93 sequenced accessions detected 15 major genomic loci exhibiting significant association with seed coat color. Five seed color-associated major genomic loci underlying robust QTLs mapped on a high-density intra-specific genetic linkage map were validated by QTL mapping. The integration of association and QTL mapping with gene haplotype-specific LD mapping and transcript profiling identified novel allelic variants (non-synonymous SNPs) and haplotypes in a MATE secondary transporter gene regulating light/yellow brown and beige seed coat color differentiation in chickpea. The down-regulation and decreased transcript expression of beige seed coat color-associated MATE gene haplotype was correlated with reduced proanthocyanidins accumulation in the mature seed coats of beige than light/yellow brown seed colored *desi* and *kabuli* accessions for their coloration/pigmentation. This seed color-regulating MATE gene revealed strong purifying selection pressure primarily in LB/YB seed colored *desi* and wild *Cicer reticulatum* accessions compared with the BE seed colored *kabuli* accessions. The functionally relevant molecular tags identified have potential to decipher the complex transcriptional regulatory gene function of seed coat coloration and for understanding the selective sweep-based seed color trait evolutionary pattern in cultivated and wild accessions during chickpea domestication. The genome-wide integrated approach employed will expedite marker-assisted genetic enhancement for developing cultivars with desirable seed coat color types in chickpea.

Keywords: chickpea, GBS, GWAS, haplotype, QTL, seed coat color, SNP, wild accession

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is the third largest produced food legume crop globally that serves as a vital human dietary source of protein enriched with essential amino acids (Kumar et al., 2011; Gaur et al., 2012; Varshney et al., 2013). The seeds are defining characteristics of chickpea aside from having high economic value. The physical appearance of seeds such as light colored and large sized seeds of chickpea has always been a trait of consumer preference and trade value, besides an important target quality component and adaptation traits (Cassells and Caddick, 2000; Graham et al., 2001; Hossain et al., 2010). The seed coat color is an important agronomic trait in chickpea that varies widely among core and mini-core germplasm collections, cultivated *desi* and *kabuli* types, landraces and wild species accessions (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2002, 2006). Considering the agronomic importance and wider level of existing phenotypic variation of seed coat color in chickpea, it is essential to identify the underlying heritable forces and potential genes/QTLs (quantitative trait loci) influencing this large trait variation in *desi* and *kabuli* chickpea. Based on defined agromorphological descriptors, the large-scale phenotyping of core, mini-core and reference core chickpea accessions have classified and characterized these accessions into 24 diverse classes of seed coat color types (like brown, light brown, green, yellow-brown and creamy) (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2002, 2008). This revealed the predominance of yellow/light brown and beige seed coat color in *desi* and *kabuli* chickpea.

Tremendous efforts have been made for studying the genetic inheritance pattern as well as identification and molecular cloning of multiple genes harboring the major QTLs regulating seed coat color trait in diverse crop plants, including rice, soybean, sesame, *Brassica* and grape (Furukawa et al., 2006; Liu et al., 2006; Sweeney et al., 2006, 2007; Yang et al., 2010; Rahman et al., 2011; Huang et al., 2012; Zhang et al., 2013). The formation of different seed coat color is governed by organ/tissue-specific accumulation of plant secondary metabolites like flavonoid compounds, including anthocyanin, flavonols, isoflavones, flavonol glucosides, phlobaphenes, proanthocyanidin, leucoanthocyanidin, and proanthocyanidin in the seed coat (Furukawa et al., 2006). A complex biochemical pathway like phenylpropanoid (flavonoid and anthocyanin biosynthetic pathway) pathway is known to be involved in synthesis of these secondary metabolites in the specific tissues of crop plants (Dixon et al., 2002). The advances in structural and functional genomics have greatly catalyzed isolation of most of the transcriptionally regulated loci/genes that encode enzymes of this important biosynthetic pathway governing seed coat color in rice, maize, *Arabidopsis*, *Brassica*, *Medicago*, common bean, pea *Antirrhinum* and *Petunia* (Ludwig et al., 1989; Mol et al., 1998; Hu et al., 2000; Forkmann and Martens, 2001; Sakamoto et al., 2001; Winkel-Shirley, 2001; Saitoh et al., 2004; Xie and Dixon, 2005; Furukawa et al., 2006; Lepiniec et al., 2006; Quattrocchio et al., 2006; Zhao and Dixon, 2009; Golam Masum Akond et al., 2011; Li et al., 2012; Ferraro et al., 2014; Liu et al., 2014). However, the genetic analysis of seed color trait has mostly been restricted to estimation of heritability and molecular mapping

of only a few seed coat color-associated QTLs because of its quantitative and polygenic nature of inheritance (epistatic and pleiotropic effects) both in *desi* and *kabuli* chickpea (Hossain et al., 2010). However, the markers associated with these QTLs controlling seed color are yet to be validated across diverse genetic backgrounds and environments for their deployment in marker-assisted breeding program to develop improved chickpea cultivars with desirable seed coat color.

A combinatorial approach of SNP (single nucleotide polymorphism) and SSR (simple sequence repeat) marker-based genome-wide association study (GWAS), candidate gene-based association analysis and traditional bi-parental QTL mapping in conjunction with differential gene expression profiling and gene-based molecular haplotyping/LD (linkage disequilibrium) mapping is a widely recognized strategy for dissecting the complex yield and quality component quantitative traits in different crop plants, including chickpea (Konishi et al., 2006; Sweeney et al., 2007; Tian et al., 2009; Mao et al., 2010; Li et al., 2011; Kharabian-Masouleh et al., 2012; Parida et al., 2012; Kujur et al., 2013, 2014; Negrão et al., 2013; Saxena et al., 2014a; Zuo and Li, 2014; Bajaj et al., 2015). Natural variation and artificial selection in many cloned genes (e.g., red pericarp color-regulating *Rc* gene in rice) underlying the QTLs regulating seed coat color during domestication have established seed color as a common target trait for both domestication and artificial breeding in crop plants (Sweeney et al., 2007; Hossain et al., 2011; Meyer et al., 2012). Henceforth, the above-said integrated strategy can be deployed for identification of functionally relevant novel molecular tags (markers, genes, QTLs, and alleles) controlling seed coat color and unraveling the molecular genetic basis of such natural trait variation in various cultivated and wild accessions of chickpea adapted to diverse agro-climatic regions. The resultant findings would assist us to genetically select *desi* and *kabuli* accessions for diverse desirable seed coat color, driving genomics-assisted breeding and ultimately development of tailored chickpea cultivars with preferred seed color.

In view of the aforementioned prospects, our study integrated GWAS and candidate gene-based association analysis with traditional QTL mapping, differential gene expression profiling, gene haplotype-specific LD mapping and proanthocyanidin (PA) quantitation assay to identify novel potential genomic loci (genes/gene-associated targets) and alleles/haplotypes regulating seed coat color and understand their haplotype-based evolutionary/domestication patterns in cultivated and wild chickpea.

MATERIALS AND METHODS

Discovery, Genotyping and Annotation of Genome-wide SNPs

For large-scale discovery and high-throughput genotyping of SNPs at a genome-wide scale, 93 phenotypically and genotypically diverse *desi* (39 accessions) and *kabuli* (54) chickpea accessions (representing diverse eco-geographical regions of 21 countries of the world) were selected (Table S1) from the available chickpea germplasm collections (16991,

including 211 minicore germplasm lines, Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2001, 2008). In addition, 79 diverse *Cicer* accessions belonging to five annual wild species, namely *C. reticulatum* (14), *C. echinospermum* (8), *C. judaicum* (22), *C. bijugum* (19), and *C. pinnatifidum* (15) and one perennial species *C. microphyllum* were chosen for haplotype-based evolutionary study (Table S1). The genomic DNA from the young leaf samples of 172 cultivated and wild chickpea accessions were isolated using QIAGEN DNeasy 96 Plant Kit (QIAGEN, USA).

The genomic DNA of accessions was digested with *Ape*KI, ligated to adapters-carrying unique barcodes and 2 × 96-plex GBS libraries were constructed. Each of two 96-plex GBS libraries were pooled together and sequenced (100-bp single end) individually using Illumina two-lane flow cell of HiSeq2000 NGS (next-generation sequencing) platform following Elshire et al. (2011) and Spindel et al. (2013). The unique barcodes attached to each of 172 accessions were used as reference for demultiplexing of their high-quality sequence reads. The decoded high-quality FASTQ sequences were mapped to reference *kabuli* draft chickpea genome (Varshney et al., 2013) employing Bowtie v2.1.0 (Langmead and Salzberg, 2012). The sequence alignment map files generated from *kabuli* genome were analyzed using the reference-based GBS pipeline/genotyping approach of STACKS v1.0 (<http://creskolab.uoregon.edu/stacks>) for identifying the high-quality SNPs (SNP base-quality, 20 supported by minimum sequence read-depth, 10) from 172 accessions (as per Kujur et al., 2015). The *kabuli* genome annotation (Varshney et al., 2013) was utilized as an anchor for structural and functional annotation of GBS-based SNPs in different coding (synonymous and non-synonymous SNPs) and non-coding sequence components of genes and genomes (chromosomes/pseudomolecules and scaffolds) of *kabuli* chickpea.

Mining and Genotyping of Candidate Gene-based SNPs

For large-scale mining and high-throughput genotyping of candidate gene-based SNPs in chickpea, a selected set of 28 cloned genes/QTLs known to be involved in regulation of seed pericarp color in related dicot species, *Arabidopsis thaliana*, *Medicago*, and soybean (Lepiniec et al., 2006; Zhao and Dixon, 2009; Yang et al., 2010) were acquired. This includes a tandem array of three seed color candidate co-orthologous *tt12*-MATE family transporter repeated genes located at the selective sweep target region of *kabuli* chromosome 4 (reported previously by Varshney et al., 2013). The coding sequences (CDS) of these known/candidate seed color-regulating genes were BLAST searched against the CDS of *kabuli* genes to identify the best possible gene orthologs in chickpea. The CDS (functional domains) and 1000-bp upstream regulatory regions (URRs) of identified true *kabuli* chickpea gene orthologs (*E*-value: 0 and bit score ≥ 500) were targeted to design (Batch Primer3, <http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>) the forward and reverse primers with expected amplification product size of 500–700 bp. For mining the gene-based SNPs, primarily 24 *desi*, *kabuli*, and wild chickpea accessions were selected randomly from 172 accessions (utilized for genome-wide GBS-based SNP genotyping) (Table S1). The gene-based primers were

PCR amplified using the genomic DNA of all these 24 accessions and resolved in 1.5% agarose gel. The amplified PCR products were sequenced, the high-quality sequences aligned and SNPs were detected in diverse sequence components of orthologous chickpea genes among accessions following Saxena et al. (2014a). The gene-derived coding and URR-SNPs discovered were further genotyped in the genomic DNA of 172 cultivated and wild chickpea accessions employing Sequenom MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) MassARRAY (<http://www.sequenom.com>) as per Saxena et al. (2014a). According to allele-specific mass differences between extension products, the genotyping information of different coding and URR-SNPs among accessions were recorded.

Estimation of Polymorphism and Diversity Statistics

To measure the PIC (polymorphism information content) and genetic diversity coefficient (Nei's genetic distance; Nei et al., 1983), and construct an unrooted neighbor-joining (NJ)-based phylogenetic tree (with 1000 bootstrap replicates) among chickpea accessions, the high-quality SNP genotyping data were analyzed with PowerMarker v3.51 (Liu and Muse, 2005), MEGA v5.0 (Tamura et al., 2011) and TASSEL v5.0 (<http://www.maizegenetics.net>). A 100-kb non-overlapping sliding window approach of TASSEL v5.0 was implemented to estimate the various nucleotide diversity measures ($\theta\pi$, $\theta\omega$, and Tajima's *D*) following Xu et al. (2011) and Varshney et al. (2013).

Determination of Population Genetic Structure and LD Patterns

A model-based program STRUCTURE v2.3.4 was utilized to determine the population genetic structure among chickpea accessions adopting the methods of Kujur et al. (2013, 2014) and Saxena et al. (2014a). The high quality genome-wide and candidate gene-based genotyping data of SNPs physically mapped on eight *kabuli* chromosomes were analyzed using a command line interface of PLINK and the full-matrix approach of TASSELv5.0 (Saxena et al., 2014a; Kumar et al., 2015) to evaluate the genome-wide LD patterns (r^2 , frequency correlation among pair of alleles across a pair of SNP loci) and LD decay (by plotting average r^2 against 50 and 20 kb uniform physical intervals across chromosomes) in chromosomes and population of chickpea.

Phenotyping for Seed Coat Color

One hundred seventy-two chickpea accessions were grown in the field [following randomized complete block design (RCBD)] during crop growing season for two consecutive years (2011 and 2012) at two diverse geographical locations (New Delhi; latitude/longitude: 28.4°N/77.1°E and Hyderabad; 17.1°N/78.9°E) of India. The seeds of all accessions were phenotyped and classified/characterized into different seed color types by visual estimation of seed coat/pericarp color of mature seeds (stored no longer than 5 months) representing each accession with at least three replications (Upadhyaya

and Ortiz, 2001; Upadhyaya et al., 2002, 2006; Hossain et al., 2011). Based on these large-scale replicated multi-location/years field phenotyping, the 172 chickpea accessions were broadly categorized/characterized in accordance with 24 predominant seed coat color types (BL: Black, B: Brown, LB: Light brown, DB: Dark brown, RB: Reddish brown, GB: Grayish brown, SB: Salmon brown, OB: Orange brown, GR: Gray, BB: Brown beige, Y: Yellow, LY: Light yellow, YB: Yellow brown, OY: Orange yellow, O: Orange, YE: Yellow beige, I: Ivory, G: Green, LG: Light green, BR: Brown reddish, M: Variegated, BM: Black brown mosaic, LO: Light orange and BE: Beige) (Figure S1). The frequency distribution and broad-sense heritability (H^2) of seed coat color were measured using SPSSv17.0 (<http://www.spss.com/statistics>) as per Kujur et al. (2013, 2014) and Saxena et al. (2014a).

Trait Association Mapping

The genome-wide and candidate gene-based SNP genotyping information of 93 cultivated *desi* and *kabuli* accessions was integrated with their seed coat color phenotyping, ancestry coefficient (Q matrix obtained from population structure) and relative kinship matrix (K) data using GLM (Q model)- and MLM (Q+K model)-based approaches of TASSELv5.0 (Kujur et al., 2013, 2014; Kumar et al., 2015). Furthermore, mixed-model (P+K, K, and Q+K) association (EMMA) and P3D/compressed mixed linear model (CMLM) interfaces of GAPIT (Lipka et al., 2012) relying upon principal component analysis (PCA) were employed for GWAS. The relative distribution of observed and expected $-\log_{10}(P)$ -value for each SNP marker-trait association were compared individually utilizing quantile-quantile plot of GAPIT. According to false discovery rate (FDR cut-off ≤ 0.05) (Benjamini and Hochberg, 1995), the adjusted P -value threshold of significance was corrected for multiple comparisons. By integrating all four model-based outputs of TASSEL and GAPIT, the potential SNP loci in the target genomic (gene) regions revealing significant association with seed coat color trait variation at highest R^2 (degree of SNP marker-trait association) and lowest FDR adjusted P -values (threshold $P < 1 \times 10^{-6}$) were identified.

QTL Mapping

For validation of seed color-associated genomic loci through QTL mapping, an intra-specific 190 F₇ RIL mapping population (ICC 12299 \times ICC 8261) was developed. The chickpea accessions ICC 12299 (originated from Nepal) and ICC 8261 (originated from Turkey) are LB/YB and BE seed coat colored *desi* and *kabuli* landraces, respectively. A selected set of 384 SNPs showing polymorphism between mapping parental accessions (ICC 12299 and ICC 8261) and 96 previously reported SSR markers (Winter et al., 1999, 2000; Hossain et al., 2011; Thudi et al., 2011) physically/genetically mapped on LGs (linkage groups)/chromosomes (considered as anchors) were genotyped using MALDI-TOF mass array SNP genotyping assay and fluorescent dye-labeled automated fragment analyser (Kujur et al., 2013, 2014; Saxena et al., 2014a,b; Bajaj et al., 2015). The genotyping data of parental polymorphic SNP and SSR markers mapped on LGs (chromosomes) of an intra-specific

genetic map and replicated multi-location seed coat color field phenotyping information (following aforesaid phenotyping methods of 172 accessions) of 190 RIL mapping individuals and parental accessions were analyzed using the composite interval mapping (CIM) function of MapQTL v6.0 (Van Ooijen, 2009, as per Saxena et al., 2014a). The phenotypic variation explained (PVE) by QTLs (R^2 %) was measured at significant LOD (logarithm of odds threshold > 5.0) to identify and map the major genomic loci underlying robust seed color QTLs in chickpea.

Differential Expression Profiling

To infer differential gene regulatory function, the expression profiling of seed color-associated genes validated by both GWAS/candidate gene-based association analysis and QTL mapping was performed. For expression analysis, 11 contrasting chickpea accessions as well as parental accessions and two homozygous individuals of a RIL mapping population (ICC 12299 \times ICC 8261) representing two predominant seed coat color types, LB/YB (*desi*: ICCX810800, *desi*: ICC 12299, *desi*: ICCV 10, *desi*: ICC 4958, *desi*: ICC 15061, *kabuli*: ICC 6204, and *kabuli*: Annigeri) and BE (*kabuli*: ICC 20268, *kabuli*: ICC 8261, *kabuli*: Phule G0515, and *desi*: ICC 4926) were selected (Table S1). The vegetative leaf tissues, including seeds without seed coats and seed coats scrapped from the immature (15–25 days after podding/DAP) and mature (26–36 DAP) seeds (30–40 seeds per accession in replicates) of all these accessions and mapping parents/individuals were used for RNA isolation. The isolated RNA was amplified with the gene-specific primers using the semi-quantitative and quantitative RT-PCR assays as per Bajaj et al. (2015). Three independent biological replicates of each sample and two technical replicates of each biological replicate with no template and primer as control were included in the quantitative RT-PCR assay. A house-keeping gene elongation factor 1-alpha (*EF1 α*) was utilized as internal control in RT-PCR assay for normalization of expression values across various tissues and developmental stages of accessions and mapping parents/individuals. The significant difference of gene expression in immature and mature seed coats as compared to vegetative leaf tissues and seeds without seed coats (considered as control) of each accessions and mapping parents/individuals was determined and compared among each other to construct a heat map using TIGR MultiExperiment Viewer (MeV, <http://www.tm4.org/mev>).

Molecular Haplotyping

For gene-based marker haplotyping/LD mapping, the 2 kb URR, exons, introns and 1 kb DRR (downstream regulatory region) of one strong seed color-associated gene (validated by GWAS, QTL mapping, and differential expression profiling) amplified from 172, including 93 cultivated and 79 wild chickpea accessions (Table S1) were cloned and sequenced. For these, the gene amplified PCR products were purified (QIAquick PCR Purification Kit, QIAGEN, USA), ligated with pGEM-T Easy Vector (Promega, USA) and transformed into competent DH5 α *E. coli* strain by electroporation (BIO-RAD, USA) to screen blue/white colonies. The plasmids isolated from 20 random positive white clones carrying the desired inserts were sequenced

in both forward and reverse directions twice on a capillary-based Automated DNA Sequencer (Applied Biosystems, ABI 3730xl DNA Analyzer, USA) using the BigDye Terminator v3.1 sequencing kit and M13 forward and reverse primers. The trace files were base called separately for all the 10 clones of individual accessions and/primers, checked for quality using *phred* and then assembled into contigs using *phrap*. The high-quality sequences generated for the gene were aligned and compared among 172 chickpea accessions using the CLUSTALW multiple sequence alignment tool in MEGA 6.0 (Tamura et al., 2013) to discover the SNP and SSR allelic variants among these accessions. The SNP-SSR marker-based haplotypes were constituted in a gene to determine the haplotype-based LD patterns among accessions following Kujur et al. (2013, 2014) and Saxena et al. (2014a). To evaluate the trait-association potential and determine the haplotype-based evolutionary pattern of the gene, the marker-based gene haplotype-derived genotyping information was correlated with seed color field phenotyping data of accessions used. To assess the potential of seed color-regulatory haplotypes constituted in a gene, differential expression profiling in the seed coats dissected from mature seeds of aforesaid six contrasting chickpea accessions as well as mapping parents and two homozygous RIL individuals representing LB/YB (ICCX-810800, ICC 12299, and Annigeri)- and BE (ICC 20268, ICC 8261, and ICC 4926)-specific haplotype groups was performed using their respective gene haplotype-based primers (following aforementioned methods).

Estimation of Proanthocyanidin (PA)

The seed coats (~50 mg) (three independent biological replicates per accession) from matured seeds of aforementioned six contrasting chickpea accessions, mapping parents and two homozygous RIL individuals representing LB/YB and BE seed coat color haplotypes were scraped. These seed coat tissues were used for soluble (with 0.2% DMACA) and insoluble (with butanol-HCl) PA extraction and quantification following the methods of Pang et al. (2008) and Liu et al. (2014).

RESULTS AND DISCUSSION

Large-scale Discovery, High-throughput Genotyping and Annotation of Genome-wide GBS- and Candidate Gene-derived SNPs

The sequencing of 2×96 -plex *ApeKI* GBS libraries generated on an average of 200.7 million high-quality sequence reads that are evenly distributed (mean: 2.1 million reads per accession) across 172, including 93 cultivated and 79 wild chickpea accessions (Table S1). Notably, on an average 87% of these sequence reads were mapped to unique physical positions on *kabuli* reference genome. The sequencing data generated from 93 cultivated (*desi* and *kabuli*) and 79 wild chickpea accessions in this study have been submitted to a freely accessible NCBI-short read archive (SRA) database (www.ncbi.nlm.nih.gov/sra) with accession numbers SRX845396 (<http://www.ncbi.nlm.nih.gov/gquery/?term=SRX845396>) and SRX971856, respectively. In

total, 8673 high-quality SNPs (with read-depth 10, SNP base quality ≥ 20 , <1% missing data and ~2% heterozygosity in each accession) differentiating 172 accessions were discovered using *kabuli* reference genome-based GBS assay (Table S2). This highlights the advantages of GBS assay in rapid large-scale mining and high-throughput genotyping of accurate and high-quality SNPs simultaneously at a genome-wide scale in chickpea. The successful applications of GBS assay particularly in construction of high-density intra- and inter-specific genetic linkage maps, high-resolution GWAS and fine mapping/positional cloning of genes/QTLs governing important agronomic traits have been demonstrated recently in diverse legume crops, including chickpea (Sonah et al., 2013, 2014; Deokar et al., 2014; Jaganathan et al., 2014; Liu et al., 2014). In this context, GBS-based genome-wide SNPs identified by us have potential to be utilized for genomics-assisted breeding applications in chickpea.

A number of diverse known cloned genes governing complex transcriptional regulatory anthocyanin and flavonoid biosynthetic pathways to accumulate specific secondary metabolites (majorly anthocyanin and proanthocyanidin) in the seed coat for its coloration/pigmentation have been identified and characterized in *Arabidopsis*, soybean and *Medicago* (Focks et al., 1999; Debeaujon et al., 2001; Sagasser et al., 2002; Saitoh et al., 2004; Baxter et al., 2005; Xie and Dixon, 2005; Furukawa et al., 2006; Lepiniec et al., 2006; Liu et al., 2006, 2014; Quattrocchio et al., 2006; Sweeney et al., 2006, 2007; Zhao and Dixon, 2009; Erdmann et al., 2010; Yang et al., 2010; Golam Masum Akond et al., 2011; Rahman et al., 2011; Bowerman et al., 2012; Chen et al., 2012, 2014; Huang et al., 2012; Li et al., 2012; Routaboul et al., 2012; Nguyen et al., 2013; Pourcel et al., 2013; Wang et al., 2013; Zhang et al., 2013; Ichino et al., 2014; Pesch et al., 2014; Wroblewski et al., 2014; Xu et al., 2014). Recently, a tandem array of three seed color-regulating candidate co-orthologous *tt12*-MATE family transporter repeated genes of *M. truncatula* primarily affecting proanthocyanidin biosynthesis under strong purifying selection pressure has been identified at the selective sweep target region of *kabuli* chromosome 4 (Varshney et al., 2013). Besides, the detailed genetic regulation of seed color in chickpea is poorly understood vis-a-vis other legumes. Therefore, the seed color-related known cloned and candidate gene orthologs of legumes (*Medicago* and soybean) and *Arabidopsis* were targeted in our study for genetic association mapping of seed color traits in chickpea. To access the potential of aforesaid known/candidate genes for seed color trait association in chickpea, genetic association mapping was performed by high-throughput genotyping of novel SNP allelic variants mined from diverse sequence components of these genes in 93 cultivated and 79 wild accessions. The PCR amplicon-based sequencing of 28 seed color-related known/candidate *kabuli* gene (CDS and 1 kb-URRs) orthologs in 24 selected chickpea accessions and further comparison of their high-quality gene amplicon sequences detected numerous coding and non-coding SNP allelic variants. The subsequent large-scale validation and high-throughput genotyping of these mined SNP alleles in 172 cultivated and wild accessions by MALDI-TOF mass array successfully identified 372 genic SNPs (233 coding

non-synonymous/synonymous and 139 URR SNPs) with an average frequency of 2.82 SNPs/kb (varied from 1.58 to 3.65 SNPs/kb) (Table 1).

The genome-wide GBS- and candidate gene-based SNP genotyping overall identified 9045 SNPs showing polymorphism among 172 cultivated and wild chickpea accessions. The 7488 and 1557 SNPs of these, were physically mapped on eight chromosomes and scaffolds of *kabuli* genome, respectively (Figure 1A). A maximum number of 1628 SNPs (21.7%) were mapped on *kabuli* chromosome 4, whereas minimum of 321 SNPs (4.3%) were mapped on chromosome 8. The structural annotation of 9045 SNPs identified 5903 (65.3%) and 3142 (34.7%) SNPs in the 1859 genes and intergenic regions of *kabuli* genome (Figure 1B). The abundance of gene-based SNPs was observed in the CDS (2385 SNPs, 40.4%), followed by introns (1576, 26.7%) and least (928, 15.7%) in the URRs. In total, 1205 (50.5%) and 1180 (49.5%) coding SNPs in the 756 and 743 genes exhibited synonymous and non-synonymous substitutions, respectively (Figure 1B). The relative distribution of GBS-based SNPs, including non-synonymous SNPs physically mapped on eight *kabuli* chromosomes showing polymorphism among 93 cultivated (*desi* and *kabuli*) and 79 wild accessions are depicted individually in a Circos circular ideogram (Figure 1C). The functional annotation of 1859 SNP-carrying genes exhibited their maximum correspondence to growth, development and metabolism-related proteins (43%), followed by transcription factors (22%) and signal transduction proteins (11%). The structurally and functionally annotated genome-wide and gene-derived SNPs physically mapped on *kabuli* chromosomes and novel SNP allelic variants mined from diverse seed color-regulating known/cloned and candidate genes can be utilized for various large-scale genotyping applications particularly in complex seed color quantitative trait dissection of chickpea.

Polymorphism and Molecular Diversity Potential of SNPs

The identified 9045 genome-wide GBS- and candidate gene-based SNPs revealed polymorphism among 172 cultivated and wild chickpea accessions with a higher PIC (0.13–0.41 with a mean 0.32) and nucleotide diversity ($\theta\pi$: 2.14, $\theta\omega$: 2.19, and Tajima's D: -1.59) potential (Table S3). The intra-specific polymorphic and nucleotide diversity potential detected by SNPs within wild and cultivated (*desi* and *kabuli*) chickpea species was lower as compared to that of inter-specific polymorphism and nucleotide diversity among species (Table S3).

The molecular diversity potential detected by 9045 SNPs among 172 *desi*, *kabuli* and wild chickpea accessions exhibited a broader range of genetic distances varying from 0.15 to 0.83 with a mean of 0.52. A lower genetic diversity level (mean genetic distance: 0.32) was observed among 93 *desi* and *kabuli* chickpea accessions than that of 79 wild chickpea accessions (0.47). The unrooted neighbor-joining phylogenetic tree construction using 9045 genome-wide and candidate gene-based SNPs depicted a distinct differentiation among 172 cultivated and wild chickpea accessions and further clustered these accessions into two major groups (cultivated and wild) as expected (Figure S2). A relatively higher intra- and inter-specific polymorphism and nucleotide

diversity potential as well as broader allelic (functional) diversity detected by 9045 SNPs among 172 cultivated (*desi* and *kabuli*) and wild chickpea accessions is much higher compared to that of earlier similar documentation (Nayak et al., 2010; Gujaria et al., 2011; Roorkiwal et al., 2013; Varshney et al., 2013). This suggests the utility of our developed genome-wide and gene-based SNP markers for faster screening of preferable diverse accessions/inter-specific hybrids in cross (introgression)-breeding program for chickpea varietal improvement.

Association Analysis for Seed Coat Color Trait

For GWAS and candidate gene-based association mapping, 8837 SNPs showing polymorphism among 93 *desi* and *kabuli* chickpea accessions based on genome-wide GBS- and candidate gene-based SNP genotyping were utilized. The neighbor-joining phylogenetic tree, high-resolution population genetic structure and PCA differentiated all 93 chickpea accessions from each other and clustered into two distinct populations; POP I and POP II representing mostly the *kabuli* and *desi* chickpea accessions with BE and LB/YB seed coat color, respectively (Figures 2A–C). The implication of chromosomal LD patterns, including LD decay in GWAS to identify potential genomic loci associated with diverse agronomic traits have been well demonstrated in many crop plants, including chickpea (Zhao et al., 2011; Thudi et al., 2014). The determination of LD patterns in a population of 93 accessions using 7488 SNPs (physically mapped across eight *kabuli* chromosomes) demonstrated a higher LD estimate (r^2 : 0.68) and extended LD decay (r^2 decreased half of its maximum value) approximately at 350–400 kb physical distance in *kabuli* chromosomes (Figure 2D). This extensive LD estimates and extended LD decay in the domesticated self-pollinating crop plant like chickpea in contrast to other domesticated self-pollinated crops like rice and soybean (Hyten et al., 2007; Mather et al., 2007; McNally et al., 2009; Huang et al., 2010; Lam et al., 2010; Zhao et al., 2011) is quite expected. This could be due to extensive contribution of four sequential bottlenecks during the domestication of chickpea (Lev-Yadun et al., 2000; Abbo et al., 2003; Berger et al., 2005; Burger et al., 2008; Toker, 2009; Jain et al., 2013; Kujur et al., 2013; Varshney et al., 2013; Saxena et al., 2014b) leading to reduction of genetic diversity in cultivated chickpea than that of other domesticated selfing plant species. The longer LD in chickpea chromosomes requires fewer markers to saturate the genome resulting in low resolution and thus indicating low significant association between the genetic markers and genes controlling the phenotypes in chickpea. In the present context due to insufficient and non-uniform marker coverage on a larger genome of chickpea with low intra-specific polymorphism, the GWAS integrated with candidate gene-based association analysis will be of much relevance in efficient quantitative dissection of complex traits in chickpea (Neale and Savolainen, 2004). This strategy of trait association mapping has been successfully implemented in a domesticated as well as larger genome self-pollinated crop species like barley, where extended LD decay was observed at genome level (Haseneyer et al., 2010; Varshney et al., 2012). Henceforth, the combinatorial approach of GWAS

TABLE 1 | Twenty-eight seed color-related known/candidate gene orthologs of chickpea selected for genetic association mapping.

Crop plants	Gene identity	Characterized known cloned genes			Kabuli chickpea gene homologs				
		Gene accession IDs	References	Gene identity	Chromosomes	Sequenced gene amplicon size (bp)	Coding and upstream regulatory SNPs discovered	Average SNP frequency (SNP/kb)	
Soybean	AR (Anthocyanidin reductase)	Glyma08g06630	Yang et al., 2010	Ca05336	Ca_Kabuli_Chr06	3793	12	3.16	
Soybean	ANS (Leucoanthocyanidin dioxygenase)	Glyma01g42350		Ca02395	Ca_Kabuli_Chr08	2179	6	2.75	
Soybean	DFR1 (Dihydroflavonol 4-reductase 1)	Glyma14g07940		Ca10786	Ca_Kabuli_Chr04	5497	15	2.73	
Soybean	DFR2 (Dihydroflavonol 4-reductase 2)	Glyma17g37060		Ca10787	Ca_Kabuli_Chr04	4719	13	2.75	
Soybean	Myb TF (Myeloblastoma Transcription Factor)	Glyma13g09980		Ca11427	Ca_Kabuli_Chr05	1816	5	2.75	
Soybean	F3H1 (Flavanone 3-hydroxylase 1)	Glyma16g23880		Ca09972	Ca_Kabuli_Chr07	3822	13	3.40	
Soybean	F3H2 (Flavanone 3-hydroxylase 2)	Glyma02g05450		Ca09973	Ca_Kabuli_Chr07	3872	10	2.58	
Arabidopsis	tt3-DFR (Dihydroflavonol 4-reductase)	AT5G42800	Wang et al., 2013	Ca10786	Ca_Kabuli_Chr04	5497	20	3.64	
Arabidopsis	tt4-CHS (Chalcone Synthase)	AT5G13930	Wroblewski et al., 2014	Ca08294	Ca_Kabuli_Chr03	3010	11	3.65	
Arabidopsis	tt5-CHI (Chalcone isomerase)	AT3G55120	Pourcel et al., 2013	Ca03250	Ca_Kabuli_Chr07	11604	37	3.19	
Arabidopsis	tt7-F3'H (Flavonoid 3'-hydroxylase)	AT5G07990	Routaboul et al., 2012	Ca13196	Ca_Kabuli_Chr07	1443	4	2.77	
Arabidopsis	tt10-PPO (Polyphenol oxidase)	AT5G48100	Zhang et al., 2012	Ca23340	Ca_Kabuli_Chr03	3106	10	3.22	
Arabidopsis	tt12-MATE secondary transporter (Multi antimicrobial extrusion)	AT3G59030	Debeaujon et al., 2001	Ca10158	Ca_Kabuli_Chr02	3676	9	2.45	
Arabidopsis	tt15-GT (Glucosyl transferase)	AT1G43620	Focks et al., 1999	Ca04376	Ca_Kabuli_Chr04	6919	21	3.04	
Arabidopsis	tt18-LDOX (Leucoanthocyanidin dioxygenase)	AT4G22880	Liu et al., 2014	Ca02395	Ca_Kabuli_Chr08	2179	4	1.84	
Arabidopsis	tt19-GST (Glutathione S-transferase)	AT5G17220	Xu et al., 2014	Ca03442	Ca_Kabuli_Chr04	1903	3	1.58	
Arabidopsis	ban-AMR (Anthocyanin reductase)	AT1G61720	Bowerman et al., 2012	Ca05336	Ca_Kabuli_Chr06	3793	9	2.37	
Arabidopsis	aha10 (Autoinhibited H ⁺ ATPase isoform 10)	AT1G17260	Baxter et al., 2005	Ca01370	Ca_Kabuli_Chr03	10262	35	3.41	
Arabidopsis	tt1-TF WIP (Wiskott-aldrich syndrome protein) type zinc finger	AT1G34790	Sagasser et al., 2002	Ca27078	Ca_Kabuli_Scaffold-2269	2324	5	2.15	
Arabidopsis	tt2-TF AMYB123 (Myeloblastoma transcription factor)	AT5G35550	Chen et al., 2012	Ca12762	Ca_Kabuli_Chr07	2226	5	2.25	
Arabidopsis	tt8-TF AtbHLH042 (Basic helix-loop-helix)	AT4G09820	Chen et al., 2014	Ca14089	Ca_Kabuli_Chr01	2283	6	2.63	
Arabidopsis	tt16-TF MADS AtAGL32 (Agamous like-32)	AT5G23260	Erdmann et al., 2010	Ca15760	Ca_Kabuli_Chr07	3027	8	2.64	
Arabidopsis	ttg1-WD40	AT5G24520	Nguyen et al., 2013	Ca04058	Ca_Kabuli_Chr05	2028	4	1.97	
Arabidopsis	ttg2-TF WRKY	AT2G37260	Pesch et al., 2014	Ca17273	Ca_Kabuli_Chr02	4380	16	3.65	
Arabidopsis	tt9-Unknown	AT3G28430	Ichino et al., 2014	Ca22116	Ca_Kabuli_Chr01	16647	47	2.82	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108990	Zhao and Dixon, 2009	Ca05556 ^a	Ca_Kabuli_Chr04	3936	10	2.54	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108840		Ca05557 ^a	Ca_Kabuli_Chr04	5315	19	3.57	
Medicago	MATE secondary transporter (TRANSPARENT TESTA)	MTR1g108810		Ca05558 ^a	Ca_Kabuli_Chr04	4540	15	3.30	

^a Tandem array of three seed color candidate co-orthologous tt12-MATE family transporter repeated genes located at the selective sweep target region of kabuli chromosome 4 reported by Varshney et al. (2013).

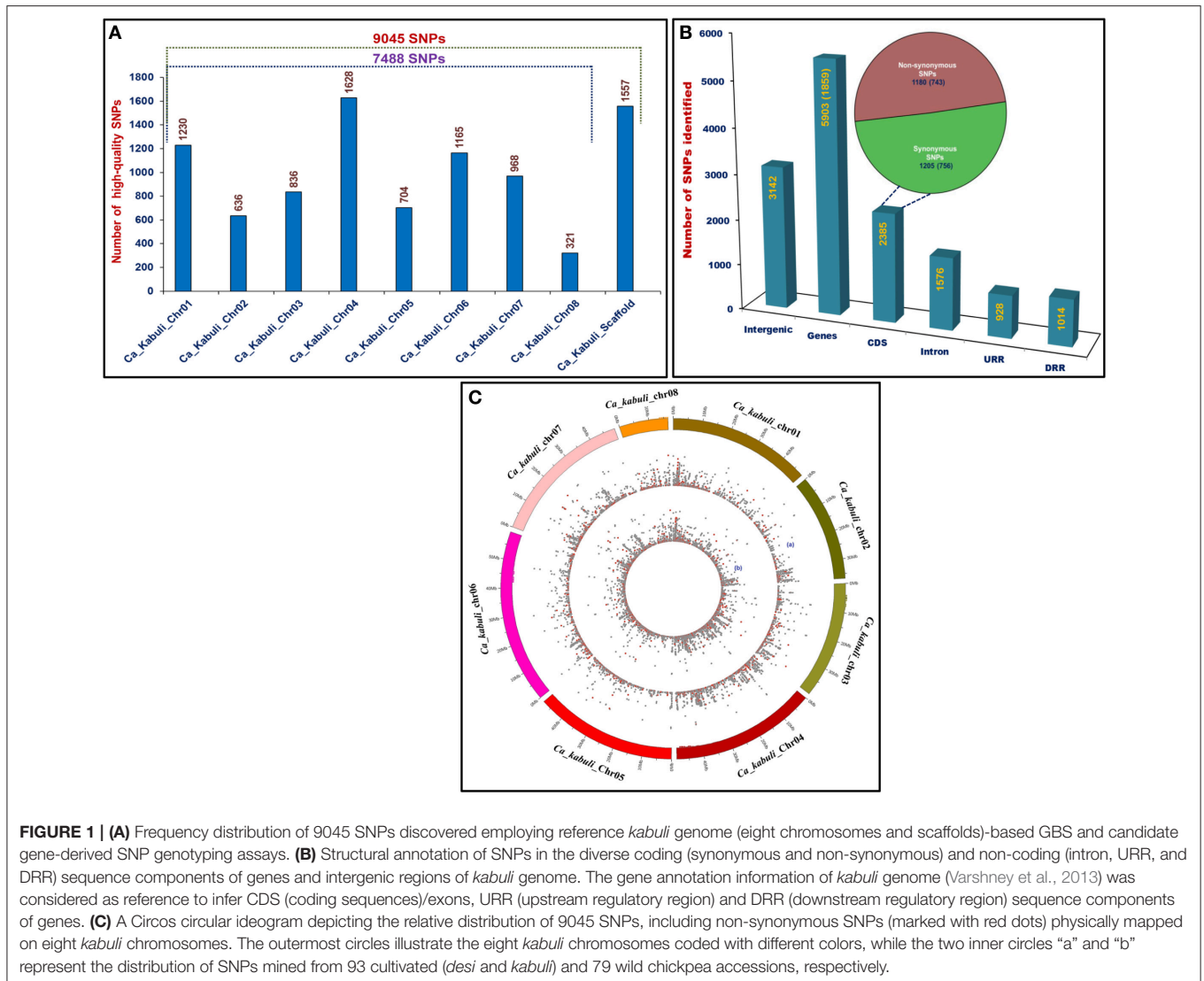


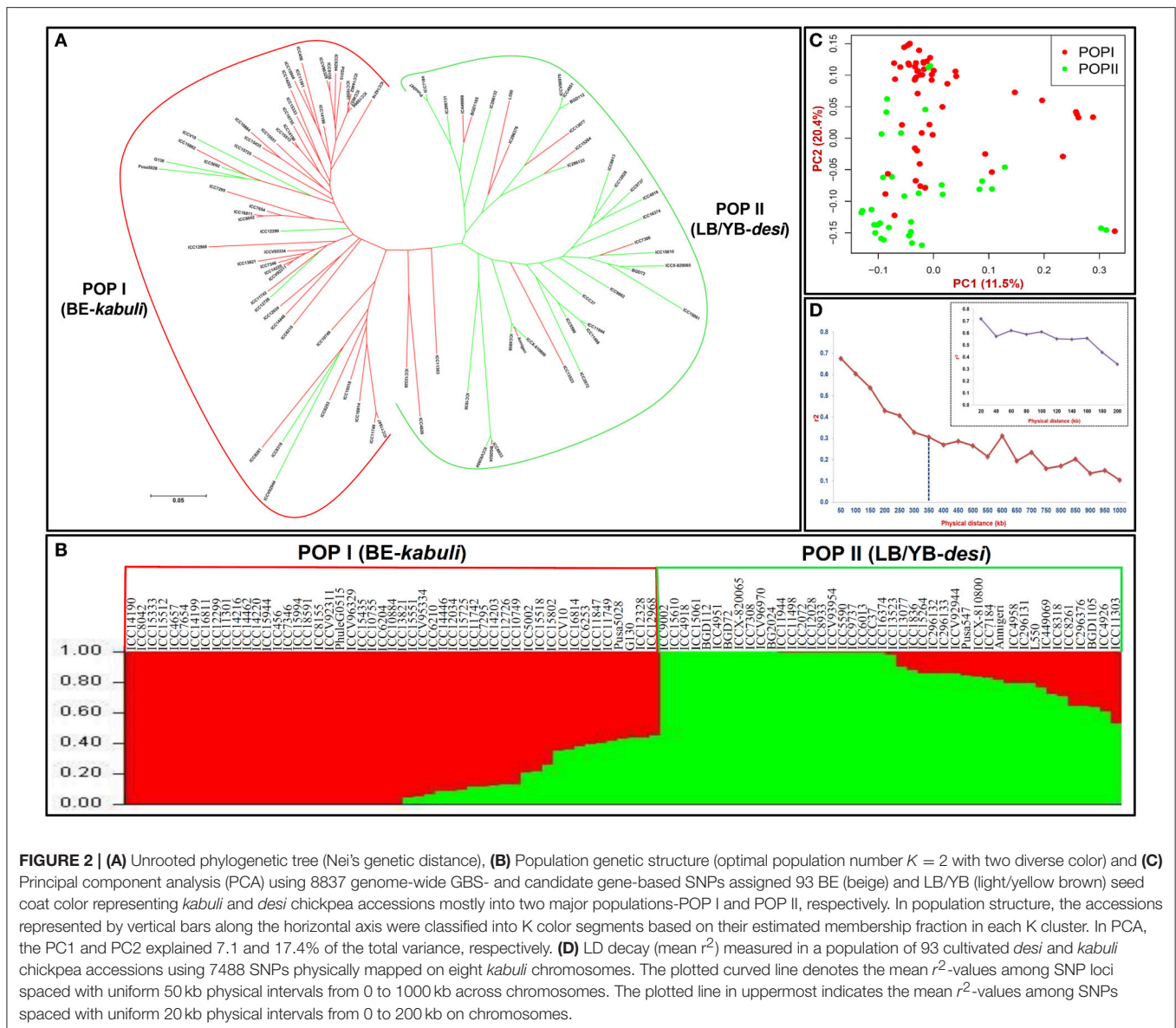
FIGURE 1 | (A) Frequency distribution of 9045 SNPs discovered employing reference *kabuli* genome (eight chromosomes and scaffolds)-based GBS and candidate gene-derived SNP genotyping assays. **(B)** Structural annotation of SNPs in the diverse coding (synonymous and non-synonymous) and non-coding (intron, URR, and DRR) sequence components of genes and intergenic regions of *kabuli* genome. The gene annotation information of *kabuli* genome (Varshney et al., 2013) was considered as reference to infer CDS (coding sequences)/exons, URR (upstream regulatory region) and DRR (downstream regulatory region) sequence components of genes. **(C)** A Circos circular ideogram depicting the relative distribution of 9045 SNPs, including non-synonymous SNPs (marked with red dots) physically mapped on eight *kabuli* chromosomes. The outermost circles illustrate the eight *kabuli* chromosomes coded with different colors, while the two inner circles “a” and “b” represent the distribution of SNPs mined from 93 cultivated (*desi* and *kabuli*) and 79 wild chickpea accessions, respectively.

and candidate gene-based association mapping would strengthen the possibility of identifying numerous high-resolution trait-associated valid genomic loci at a genome-wide scale in chickpea.

Keeping that in view, the GWAS and candidate gene-based association analysis was performed by integrating 8837 SNP genotyping data of 93 accessions with their replicated multi-location/years seed coat color field phenotyping information. The normal frequency distribution along with broader seed color phenotypic variation (Figure S3A) and 75% broad-sense heritability (H^2) among 93 accessions belonging to a population was evident. The accessions exclusively exhibiting consistent phenotypic expression of seed coat color across two geographical locations/years (supported by high H^2) were utilized for subsequent SNP marker-trait association. The integration of outcomes from four model-based approaches [TASSEL (GLM and MLM) and GAPIT (EMMA and CMLM)] with a quantile-quantile plot of the expected and observed $-\log_{10} P$ -values at FDR cut-off ≤ 0.05 in GWAS and candidate gene-based association analysis identified 15 genomic loci

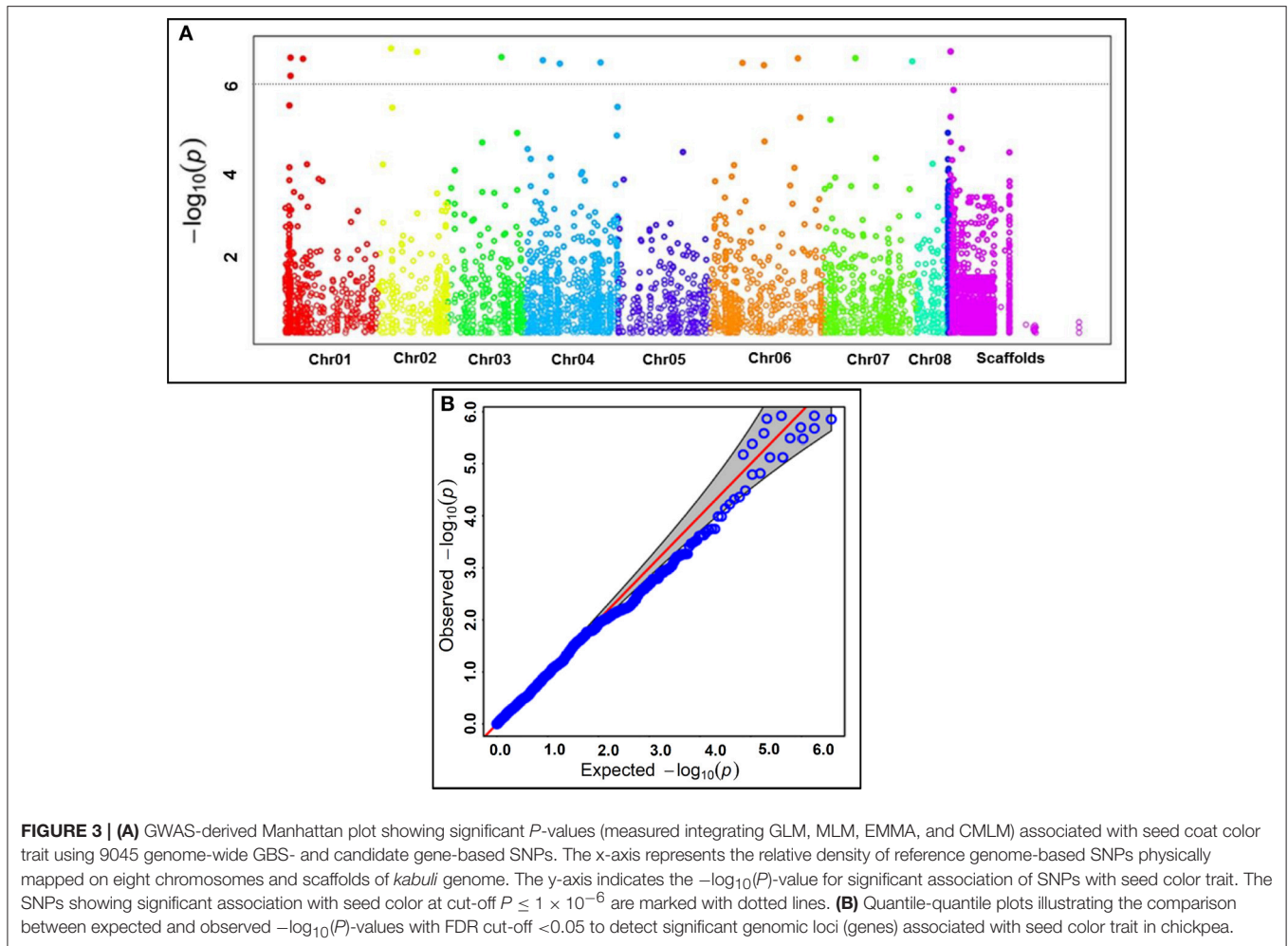
showing significant association with seed coat color at a $P \leq 10^{-6}$ (Figures 3A,B, Table 2). This includes eight genome-wide GBS-based and seven candidate gene-derived SNPs. Fourteen seed color-associated genomic loci were physically mapped on seven *kabuli* chromosomes. The rest one SNP locus was represented from scaffold region of *kabuli* genome. One of 15 seed color-associated genomic SNP loci was derived from the intergenic region, whereas rest 14 SNP loci were represented from diverse coding (seven, including six non-synonymous SNPs) as well as non-coding intronic (two SNPs) and URR (five) sequence components of 13 *kabuli* genes (Table 2).

The identified 15 significant SNPs explained (R^2) an average of 28% (ranging 20–43%) seed coat color phenotypic variation in a population of 93 chickpea accessions. The total phenotypic variation for seed coat color explained by the 15 SNPs was 61%. The significant association of 15 multiple SNP loci in more than one genomic region (genes) with seed coat color gave clues for complex genetic inheritance pattern/regulation of this target quantitative trait in chickpea. This association was evident from



high chromosomal and population-specific LD estimates with preferable LD decay of trait-associated linked/unlinked multiple SNP loci and further by the observed genetic heterogeneity of seed color traits across two populations. We observed no significant differences concerning the association potential of SNP loci with seed color in different populations despite diverse genetic architecture of seed coat color traits in two populations and along entire population. Therefore, seed coat color-associated genomic loci (genes) identified by us employing both GWAS and candidate gene-based association mapping could have potential for establishing rapid marker-trait linkages and identifying genes/QTLs regulating seed color in chickpea. Seven (three non-synonymous coding and four URR SNPs) SNP loci in the six different known cloned and candidate seed color genes (*tt12*, *tt4*, *tt3*, *tt19*, *ban-ANR*, and *tt18*) had significant association (P : 1.3×10^{-6} to 1.5×10^{-8} with R^2 : 20–35%) with seed coat color trait in chickpea (Table 2). Interestingly,

significant association of non-synonymous and regulatory SNPs identified in the different dormancy and stress-related known/candidate genes (encoding inositol monophosphate and disease resistance proteins) with seed coat color was observed in chickpea. This is consistent with previous studies that reported strong correlation/interactions among genes/QTLs regulating all three important agronomic traits, including seed coat pigmentation, seed dormancy/germination ability and stress tolerance in legumes (Harborne and Williams, 2000; Furukawa et al., 2006; Lepiniec et al., 2006; Kovinich et al., 2012; Saxena et al., 2013; Smýkal et al., 2014). The transcriptional regulation of numerous genes involved in anthocyanin and flavonoid biosynthetic pathways lead to accumulation of diverse major secondary metabolites in specific plant organs, which are known to play most important role in collective alteration of seed coat pigmentation/coloration, dormancy and defense (stress) response in crop plants



(Gore et al., 2002; Benitez et al., 2004; Senda et al., 2004; Isemura et al., 2007; Caldas and Blair, 2009; Diaz et al., 2010; Kongjaimun et al., 2012). Notably, one non-synonymous SNP locus (T/G) in the CDS of the MATE (multidrug and toxic compound extrusion) secondary transporter gene (Ca18123, mapped on *kabuli* chromosome 2) exhibiting strong association ($P: 1.7 \times 10^{-9}$ with $R^2: 43\%$) with seed color than that of other 14 seed color-associated SNPs was selected as a potential candidate for further analysis. The discovery of non-synonymous and regulatory SNP loci particularly in coding and URRs of known/candidate genes associated with seed color traits signifies their functional relevance in rapid seed color trait-regulatory gene identification and characterization in chickpea. The non-synonymous (amino acid substitutions) and regulatory SNPs in the genes have known to alter their transcriptional regulatory mechanism for controlling diverse traits of agronomic importance, including seed/grain size and weight in rice (Fan et al., 2006; Mao et al., 2010; Li et al., 2011; Zhang et al., 2012) and chickpea (Kujur et al., 2013, 2014; Bajaj et al., 2015; Saxena et al., 2014b).

Two candidate seed color-associated MATE secondary transporter genes [including one (Ca05557) reported earlier at

the selective sweep region of *kabuli* chromosome 4, (Varshney et al., 2013)] identified by use of both candidate gene-based association mapping and GWAS were selected (Table S4) to compare and deduce the nucleotide diversity measures as well as the direction and magnitude of natural selection acting between these genes. These genes revealed a similar trend of nucleotide diversity ($\theta\pi$ and $\theta\omega$) and Tajima's D-based selection pressure occurring on LB/YB and BE seed colored *desi* and *kabuli* chickpea accessions, respectively (Table S4). Interestingly, the degree of nucleotide diversity was reduced by 61% in the LB/YB seed colored cultivated *desi* accessions than that of *kabuli* chickpea accessions with BE seed color. The extreme reduction of nucleotide diversity level ($\theta\pi: 0.73\text{--}0.79$ and $\theta\omega: 0.77\text{--}0.82$) and strong purifying selection along with reduced Tajima's D (-2.89 to -2.63) specifically in LB/YB seed colored cultivated *desi* chickpea accessions compared with BE seed colored *kabuli* accessions ($\theta\pi: 1.93\text{--}1.96$, $\theta\omega: 1.95\text{--}1.98$ and $D: 1.82\text{--}1.90$) was evident (Table S4). These findings provide definite evidence regarding signature of strong purifying selection in LB/YB seed colored *desi* accessions in contrast to BE seed colored *kabuli* chickpea accessions, which is in line with enduring purifying selection for seed coat/pericarp color traits in crop plants,

TABLE 2 | Fifteen seed coat color-regulating genomic loci (genes) identified by GWAS and candidate gene-based association mapping.

SNP IDs ^a	Kabuli chromosomes/scaffolds	SNPs physical positions (bp)	SNPs (Reference/alternate alleles)	Genes/gene-associated targets ^b	Structural annotation	Association analysis	
						P	R ² (%)
CWSNP151	1	1858509	(C/T)	Ca00233 (Inositol monophosphate)	CDS (Syn)	2.2×10^{-7}	25
CWSNP156	1	1859914	(G/A)	Ca00233 (Inositol monophosphate)	Intron	2.0×10^{-7}	27
CWSNP208	1	2053788	(A/G)	Ca00260 (Disease resistance protein)	CDS (NSyn) Threonine (ACG)-Alanine (GCG)	1.8×10^{-7}	22
CWSNP1275	2	7701199	(T/G)	Ca18123 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Valine (GTG)-Glycine (GGG)	1.7×10^{-9}	43
gSNP1	2	33782378	(C/A)	Ca10158 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Alanine (GCT)-Aspartic acid (GAT)	1.2×10^{-7}	20
gSNP2	3	26031490	(A/G)	Ca08294 (<i>tt4</i> -Chalcone synthase)	URR	1.5×10^{-7}	33
gSNP5	4	48021943	(G/T)	Ca10786 (<i>tt3</i> -Dihydroflavonol-4-reductase)	CDS (NSyn) Glycine (GGG)-Valine (GTG)	1.7×10^{-7}	29
gSNP3	4	7931791	(T/C)	Ca03442 (<i>tt19</i> -Glutathione-S-reductase)	URR	1.4×10^{-7}	30
gSNP4	4	15386163	(T/G)	Ca05557 (<i>tt12</i> -MATE secondary transporter)	CDS (NSyn) Serine (TCA)-Alanine (GCA)	1.3×10^{-6}	24
gSNP6	6	15514534	(C/G)	Ca05336 (ban-ANR-Anthocyanin reductase)	URR	2.1×10^{-8}	35
CWSNP5434	6	28325392	(C/T)	Ca25968 (Ribonuclease T2)	CDS (NSyn) Serine (TCA)-Leucine (TTA)	2.4×10^{-7}	27
CWSNP5762	6	56793021	(G/T)	–	Intergenic	1.5×10^{-7}	21
CWSNP6261	7	14281993	(G/T)	Ca23042 (domain of unknown function DUF231)	Intron	2.1×10^{-7}	38
gSNP7	8	2062481	(G/A)	Ca02395 (<i>tt18</i> -Leucoanthocyanidin dioxygenase)	URR	2.1×10^{-6}	26
CWSNP8591	Scaffold93	645442	(A/C)	Ca23085 (Integrase)	URR	1.9×10^{-7}	23

*CWSNP (cultivated wild SNP) and gSNP (gene-derived SNPs).

CDS, coding sequence; NSyn, non-synonymous; Syn, synonymous; URR, upstream regulatory region.

^aDetails regarding SNPs are mentioned in the Table S2.

gSNP: gene-derived SNPs associated with seed coat color trait identified by candidate gene-based association mapping.

including chickpea (Sweeney et al., 2007; Varshney et al., 2013). Therefore, seed color in chickpea vis-à-vis other crop plants can be considered as a target trait for both domestication and artificial breeding (Sweeney et al., 2007; Hossain et al., 2011; Meyer et al., 2012).

Validation of Seed Color-associated Genomic Loci through QTL Mapping

We constructed a high-density intra-specific genetic linkage map (ICC 12299 × ICC 8261) by integrating 415, including 382 SNP and 33 previously reported parental polymorphic SSR markers across eight chickpea LGs (LG1 to LG8). This genetic map covered a total map length of 1065.7 cM, with a mean inter-marker distance of 2.57 cM (Table S5). The most and least saturated genetic map was LG1 (average inter-marker distance: 1.92 cM) and LG5 (4.0 cM), respectively. The mean map density (average inter-marker distance: 2.57 cM) observed in our constructed intra-specific genetic linkage map

was higher/comparable with earlier documentation (Cho et al., 2002, 2004; Cobos et al., 2005; Radhika et al., 2007; Kottapalli et al., 2009; Gaur et al., 2011; Kujur et al., 2013, 2014; Sabbavarapu et al., 2013; Stephens et al., 2014; Varshney et al., 2014). Therefore, this constructed high-density intra-specific genetic linkage map have potential to be utilized as a reference for rapid targeted mapping of potential genes/QTLs governing diverse agronomic traits, including seed coat color in chickpea. The bi-directional transgressive segregation-based normal frequency distribution (Figure S3B), including a significant variation in seed coat color trait along with 78% H² in this developed mapping population (ICC 12299 × ICC 8261) was observed. Two principal seed coat color (BE and LB/YB) groupings; primary (PC1) and secondary (PC2) components accounted for 52 and 48% of the total seed coat color variance, respectively in 190 RIL mapping individuals and parental accessions were evident (Figure S3C). This continuous color distribution inferred the quantitative genetic inheritance pattern of seed

coat color trait in the mapping population and thus the developed bi-parental mapping population could be utilized for complex seed color trait dissection through QTL mapping in chickpea.

The QTL mapping was performed by correlating the genotyping data of 415 mapped SNP and SSR markers with multilocation/years seed coat color field phenotyping data of the RIL mapping individuals and parental accessions. This analysis identified five major (LOD: 5.6–11.5) genomic regions harboring robust QTLs (validated across two locations/years) controlling seed coat color traits, which were mapped on five *kabuli* chromosomes (Table S6). The phenotypic variation explained (PVE) for seed color by individual QTL (R^2) varied from 20.1 to 38.7%. The PVE estimated for all five major QTLs was 39.4%. All these QTLs revealed additive gene effects (ranging -3.9 to -2.5), inferring the effective contribution of ICC 8261 alleles at these loci for BE seed coat color trait. The five SNPs in the genes showing tight linkage with all five major seed color-associated robust QTLs (*CaqSC2.1*, *CaqSC4.1*, *CaqSC4.2*, *CaqSC6.1*, and *CaqSC7.1*) had high seed color trait association potential based on our GWAS and candidate gene-based association analysis (Table 2; Table S6). Remarkably, strong association potential of one non-synonymous SNP (T/G) identified in the CDS of the MATE secondary transporter gene as compared to other genomic loci associated with seed coat color trait was ascertained both by QTL mapping (R^2 : 35.6–38.7%) and genetic association analysis (R^2 : 43%) (Table 2). Therefore, we selected MATE secondary transporter gene as a target candidate for seed coat color trait regulation by its further validation through differential expression profiling and molecular haplotyping in chickpea.

To determine the potential and novelty of seed coat color-regulating 15 genomic loci and five major QTLs detected by association and QTL mapping, respectively, the markers linked/flanking the seed color-associated known QTLs/genes (reported earlier in QTL mapping studies, Hossain et al., 2011) were selected for their validation in seed color-specific 93 diverse *desi* and *kabuli* chickpea accessions and mapping population under study. These comparative analyses revealed correspondence of one known QTL (*CaqSC2.1*) regulating seed coat color between past (SSR marker interval: TA194-TR19) and our present studies based on congruent flanking [CWSNP 1273 (25.2 cM)-CWSNP1277 (30.2 cM)]/linked marker (CWSNP 1275) physical/genetic positions on chromosome (LG) 2. Henceforth, 14 genomic loci and 4 QTLs regulating seed color identified by us employing high-resolution GWAS/candidate gene-based association and QTL mapping are novel and exhibited population-specific genetic inheritance pattern for seed color trait regulation in chickpea. Further validation of these functionally relevant molecular tags is required in diverse genetic backgrounds and/or through fine mapping/map-based positional cloning prior to their utilization in marker-assisted genetic improvement of chickpea for desirable seed color.

Validation of Seed Color-associated Genes through Differential Expression Profiling

To determine the regulatory pattern of genes for seed coat color, thirteen, including five seed color-associated genes validated

both by association analysis and QTL mapping were utilized for differential expression profiling. The RNA isolated from vegetative leaf tissues, seeds without seed coats and seed coats scrapped from the immature and mature seeds of 11 contrasting chickpea accessions as well as mapping parents and two homozygous RIL individuals representing two predominant seed coat color types, LB/YB (*desi*: ICCX810800, *desi*: ICC 12299, *desi*: ICCV10, *desi*: ICC 4958, *desi*: ICC 15061, *kabuli*: ICC 6204, and *kabuli*: Annigeri) and BE (*kabuli*: ICC 20268, *kabuli*: ICC 8261, *kabuli*: Phule G0515, and *desi*: ICC 4926) was amplified with these gene-specific primers using semi-quantitative and quantitative RT-PCR assays (Table S7). Five genes, including four seed color known cloned genes (*tt4*-chalcone synthase, *tt3*-dihydroflavonol-4-reductase, *ban-ANR*-anthocyanin reductase, and *tt19*-glutathione-s-reductase) showed seed coat-specific differential (>2.5 -fold up- and/down-regulation, $P \leq 0.01$) expression [compared with leaves and seeds (without seed coats)] specifically in two seed developmental stages of LB/YB and BE seed colored *desi* and *kabuli* accessions (Figure S4). This suggests that differential regulation and accumulation of transcripts encoded by these four known genes might be essential in synthesis of proanthocyanidin and anthocyanin biosynthetic enzymes in chickpea seed coats for its coloration/pigmentation (Lepiniec et al., 2006; Zhao and Dixon, 2009; Zhao et al., 2010).

Remarkably, a seed coat-specific MATE secondary transporter gene (validated by both association and QTL mapping) showing strong association with seed coat color trait exhibited its higher differential down-regulation (~ 5 – 7 -folds, $P \leq 0.001$) in the mature/immature seed coats of BE seed colored chickpea accessions (ICC 20268, ICC 8261, Phule G0515, and ICC 4926) as compared to that of mature/immature seed coats of LB/YB seed colored accessions (ICCX810800, ICC 12299, ICCV10, ICC 4958, ICC 15061, ICC 6204, and Annigeri) (Figure S4). A pronounced down regulation (~ 13 -folds, $P \leq 0.001$) of this seed coat-specific gene in the mature seed coats of all four BE seed colored chickpea accessions in contrast to immature seed coats of all seven LB/YB seed colored accessions was observed. However, differential down-regulation (~ 3 -folds, $P \leq 0.001$) of MATE gene in the mature seed coats than the immature seed coats of the accessions within an individual LB/YB and BE seed coat colored group was evident (Figure S4). The reduction of chickpea MATE gene transcript level specifically in the mature seed coat parallels with differential expression pattern that observed previously using *TT12* and *MATE1* genes of *Arabidopsis* and *Medicago*, respectively (Marinova et al., 2007; Pang et al., 2008; Zhao and Dixon, 2009; Zhao et al., 2010). These outcomes over imply the functional significance of MATE gene in possible transcriptional regulation of seed color trait in chickpea.

Molecular Haplotyping of a Strong Seed Color-regulating Gene

For molecular haplotyping of a strong seed color-associated gene (validated by association analysis, QTL mapping and expression profiling), the 6843 bp cloned amplicon covering the entire 2 kb URR, 7 exons, 1 kb DRR and 6 intronic regions of MATE secondary transporter *kabuli* gene (Ca18123)

were sequenced and compared among 93 cultivated and 79 wild chickpea accessions. These analyses identified four SNPs, including one URR and two coding SNPs as well as one coding SSR in the MATE gene (Figure 4A, Table S8). This includes one missense non-synonymous coding SNPs (T/G) showing valine (GTG) to glycine (GGG) amino acid substitution in the gene. The gene-based haplotype analysis by integrating the genotyping data of four SNPs and one SSR among 172 accessions constituted four haplotypes (Figure 4B). The haplotype-specific

LD mapping and association analysis using four marker-based haplotypes identified in a MATE gene inferred strong association potential ($P: 2.3 \times 10^{-11}$ with $R^2: 45\%$) of the gene with seed coat color trait. Moreover, we observed a significant higher degree of LD ($r^2 > 0.85$ with $P < 1.1 \times 10^{-6}$) resolution across the entire 6843 bp sequenced region of this strong seed color-associated gene (Figure 4C). Remarkably, two specific haplotypes, haplotype I [C-(GTTG)₃-T-T-A] and haplotype II [C-(GTTG)₃-G-T-A] affected by one functional

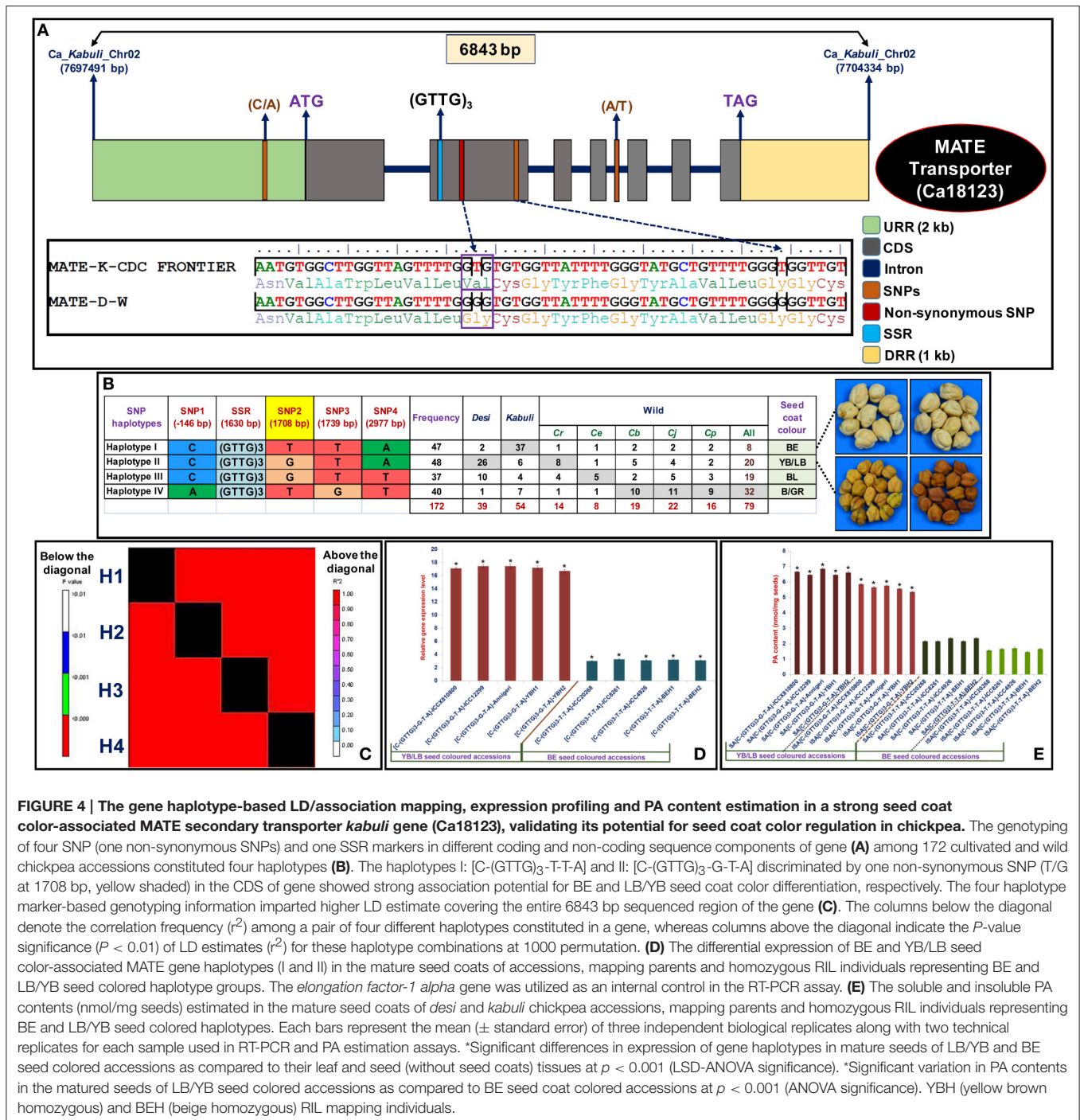


FIGURE 4 | The gene haplotype-based LD/association mapping, expression profiling and PA content estimation in a strong seed coat color-associated MATE secondary transporter kabuli gene (Ca18123), validating its potential for seed coat color regulation in chickpea. The genotyping of four SNP (one non-synonymous SNPs) and one SSR markers in different coding and non-coding sequence components of gene (A) among 172 cultivated and wild chickpea accessions constituted four haplotypes (B). The haplotypes I: [C-(GTTG)₃-T-T-A] and II: [C-(GTTG)₃-G-T-A] discriminated by one non-synonymous SNP (T/G at 1708 bp, yellow shaded) in the CDS of gene showed strong association potential for BE and LB/YB seed coat color differentiation, respectively. The four haplotype marker-based genotyping information imparted higher LD estimate covering the entire 6843 bp sequenced region of the gene (C). The columns below the diagonal denote the correlation frequency (r^2) among a pair of four different haplotypes constituted in a gene, whereas columns above the diagonal indicate the P -value significance ($P < 0.01$) of LD estimates (r^2) for these haplotype combinations at 1000 permutation. (D) The differential expression of BE and YB/LB seed color-associated MATE gene haplotypes (I and II) in the mature seed coats of accessions, mapping parents and homozygous RIL individuals representing BE and LB/YB seed colored haplotype groups. The elongation factor-1 alpha gene was utilized as an internal control in the RT-PCR assay. (E) The soluble and insoluble PA contents (nmol/mg seeds) estimated in the mature seed coats of desi and kabuli chickpea accessions, mapping parents and homozygous RIL individuals representing BE and LB/YB seed colored haplotypes. Each bars represent the mean (\pm standard error) of three independent biological replicates along with two technical replicates for each sample used in RT-PCR and PA estimation assays. *Significant differences in expression of gene haplotypes in mature seeds of LB/YB and BE seed colored accessions as compared to their leaf and seed (without seed coats) tissues at $p < 0.001$ (LSD-ANOVA significance). *Significant variation in PA contents in the matured seeds of LB/YB seed colored accessions as compared to BE seed coat colored accessions at $p < 0.001$ (ANOVA significance). YBH (yellow brown homozygous) and BEH (beige homozygous) RIL mapping individuals.

non-synonymous coding SNPs (T/G) in the *MATE* gene had strong association potential for BE and YB/LB seed coat color differentiation, respectively (**Figure 4B**). The haplotypes I and II than other two haplotypes (III and IV) identified in the gene was represented most commonly by YB/LB and BE seed colored *desi* (66.7%) and *kabuli* (68.5%) accessions, respectively. A higher down-regulated expression (~5-fold) of BE seed color-specific gene haplotype I was observed in mature seed developmental stages of BE seed colored three contrasting chickpea accessions (ICC 20268, ICC 8261, and ICC 4926) as well as mapping parents and homozygous RIL individuals compared with three YB/LB seed colored accessions (ICCX 810800, ICC 12299, and Annigeri), mapping parents and homozygous mapping individuals (**Figure 4D**). To correlate the differential expression/regulatory pattern of *MATE* gene-encoding transcript with proanthocyanidins (PA) synthesis/accumulation (Lepiniec et al., 2006; Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2010), we estimated the soluble and insoluble PA content (three replicates) in aforementioned mature seeds of six contrasting accessions as well as mapping parents and homozygous RIL individuals representing BE and YB/LB-specific haplotype groups I and II, respectively. This detected a significant reduction of both soluble and insoluble PA contents in the mature seed coats of BE seed colored accessions (soluble and insoluble PA: 2.1 and 1.6 nmol/mg seeds, respectively) than that of LB/YB (6.7 and 5.9 nmol/mg seeds) seed colored accessions (**Figure 4E**). Collectively, a strong association potential of *MATE* gene with seed color trait was ensured by integrating GWAS and candidate gene-based association analysis with QTL mapping, differential expression profiling, SNP-SSR marker-based high-resolution gene-specific LD mapping/haplotyping, haplotype-specific transcript profiling and PA estimation (soluble and insoluble) assay. These analyses identified novel natural allelic variant (T/G) showing non-synonymous amino acid substitution (valine to glycine) and potential haplotypes [C-(GTTG)₃-T-T-A] and [C-(GTTG)₃-G-T-A] in a *MATE* gene regulating BE and YB/LB seed coat color differentiation, respectively in *kabuli* and *desi* chickpea accessions. The implication of such combinatorial approach for rapid delineation of functionally relevant genes/QTLs controlling diverse agronomic traits, including starch biosynthesis, seed pericarp coloration/pigmentation, seed shattering and seed setting rate in rice as well as seed size/weight and pod number in chickpea have been well understood (Konishi et al., 2006; Sweeney et al., 2007; Kujur et al., 2013, 2014; Li et al., 2013; Saxena et al., 2014a; Bajaj et al., 2015). Therefore, natural allelic variants and haplotypes identified in a strong seed color-associated *MATE* gene by use of an integrated genomic approach can have potential to decipher its complex transcriptional regulatory function of seed coat coloration during seed development and subsequently for marker-assisted genetic enhancement of chickpea for preferred seed color.

We observed sharing of all four SNP-SSR marker-based haplotypes among 93 cultivated *desi* and *kabuli* as well as 79 wild accessions representing primary, secondary and tertiary gene pools (**Figure 4B**). The B/GR and BE seed color-associated haplotypes IV and I were represented most (40.5%) and least

(10.1%) in the wild chickpea accessions, respectively. However, the B/GR-associated haplotype IV was primarily represented by the wild chickpea accessions (38%) of secondary gene pool (*C. bijugum*, *C. judaicum*, and *C. pinnatifidum*) (**Figure 4B**). The wild accessions of *C. reticulatum* from primary gene pool had maximum YB/LB (10.1%) seed color-associated haplotype II. A maximum haplotypes sharing of BE seed color-associated haplotype I in cultivated *kabuli* accessions with YB/LB seed color-associated haplotype II in *desi* accessions following YB/LB seed color-associated haplotype II between *desi* and wild *C. reticulatum* accessions was evident (**Figure 4B**). The B/GR seed color-associated haplotype IV gave evidence for its putative recombination event during chickpea domestication. These marker haplotypes sharing analyses reflected single origin of “G” functional SNP allele primarily from *C. reticulatum*, and subsequently it was introgressed into most of the LB/YB seed colored *desi* accessions as compared to *kabuli* accessions with BE seed coat color. The nucleotide diversity ($\theta\pi$ and $\theta\omega$) and Tajima’s D estimation (Table S3) further inferred the occurrence of very strong purifying selection (with reduced D) in favor of retention of LB/YB seed coat color-associated “G” SNP locus and haplotype II in the *MATE* secondary transporter gene specifically in *desi* and wild *C. reticulatum* accessions toward assortment of more preferential LB/YB seed coat color in chickpea. These findings may be a result of domestication bottlenecks coupled with artificial selection/modern breeding efforts that are constantly practiced during the genetic improvement program of chickpea accessions for diverse seed coat color characteristics of high consumer preference and trade value. These findings infer that the natural allelic/haplotype variation discovered in the *MATE* gene are possibly associated with seed coat color trait evolution and therefore, the seed color is expected to represent an important component of domestication trait in chickpea.

The chickpea *MATE* gene is a homologs of *Arabidopsis TT12* (transparent testa 12) and *Medicago MATE1* genes encodes a membrane protein of multidrug and toxic compound extrusion (*MATE*) secondary transporter (Debeaujon et al., 2001). The detailed phenotypic characterization of *tt12* and *mate1* mutants and complementation analysis of their corresponding genes in *Arabidopsis* and *Medicago*, respectively inferred that these tonoplast late biosynthetic genes mediate transportation of precursors for proanthocyanidin (PA) biosynthesis into the vacuoles of seed coat endothelial cells in developing immature seeds for their coloration/pigmentation (Lepiniec et al., 2006; Marinova et al., 2007; Zhao and Dixon, 2009; Zhao et al., 2010). The use of combinatorial approach of GWAS, QTL mapping, transcript profiling, gene-based marker haplotyping/LD mapping and PA estimation assay indicate that BE seed color-associated haplotype I [C-(GTTG)₃-T-T-A] identified in a *MATE* gene might be involved in down-regulation/lower expression and accumulation of its encoding transcript, which led to reduced synthesis of both soluble and insoluble PAs in pericarps of BE seed colored mature seeds than that of YB/LB seed colored mature seeds with haplotype II [C-(GTTG)₃-G-T-A]. These results are in agreement with earlier observations on lower accumulation of PAs

(soluble/insoluble) in the vacuole endothelial cells of mature seed coat during later stages of seed development and their involvement in differential seed coat/pericarp coloration in crop plants (Abrahams et al., 2003; Kitamura et al., 2004; Liu et al., 2014). However, detailed molecular characterization of *MATE* secondary transporter gene is required to decipher its underlying regulatory mechanism, evolutionary cues and biochemical interactions governing diverse seed coat coloration in *desi*, *kabuli* and wild accessions during chickpea domestication. Therefore, the strong seed color-regulatory *MATE* gene once functionally well characterized, can be utilized in marker-assisted seed color breeding for developing varieties with improved seed coat color of consumer preference and trade value in chickpea.

REFERENCES

- Abbo, S., Berger, J., and Turner, N. C. (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30, 1081–1087. doi: 10.1071/FP03084
- Abrahams, S., Lee, E., Walker, A. R., Tanner, G. J., Larkin, P. J., and Ashton, A. R. (2003). The *Arabidopsis* *TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant J.* 35, 624–636. doi: 10.1046/j.1365-313X.2003.01834.x
- Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., 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. doi: 10.1093/jxb/eru478
- Baxter, I. R., Young, J. C., Armstrong, G., Foster, N., Bogenschutz, N., Cordova, T., et al. (2005). A plasma membrane H⁺-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2649–2654. doi: 10.1073/pnas.0406377102
- Benitez, E. R., Funatsuki, H., Kaneko, Y., Matsuzawa, Y., Bang, S. W., and Takahashi, R. (2004). Soybean maturity gene effects on seed coat pigmentation and cracking in response to low temperatures. *Crop Sci.* 44, 2038–2042. doi: 10.2135/cropsci2004.2038
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57, 289–300.
- Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the genus *Cicer* vernalisation response and low temperature pod set in chickpea (*C. arietinum* L.) and its annual wild relatives. *Aust. J. Agric. Res.* 56, 1191–1200. doi: 10.1071/AR05089
- Bowerman, P. A., Ramirez, M. V., Price, M. B., Helm, R. F., and Winkel, B. S. (2012). Analysis of T-DNA alleles of flavonoid biosynthesis genes in *Arabidopsis* ecotype Columbia. *BMC Res. Notes* 5:485. doi: 10.1186/1756-0500-5-485
- Burger, J. C., Champan, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95, 113–122. doi: 10.3732/ajb.95.2.113
- Caldas, G. V., and Blair, M. W. (2009). Inheritance of seed condensed tannins and their relationship with seed-coat color and pattern genes in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 119, 131–142. doi: 10.1007/s00122-009-1023-4
- Cassells, J., and Caddick, L. (2000). Harvest timing critical for chickpea quality. *Farming Ahead* 107, 41–42.
- Chen, M., Wang, Z., Zhu, Y., Li, Z., Hussain, N., Xuan, L., et al. (2012). The Effect of *TRANSPARENT TESTA2* on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in *Arabidopsis*. *Plant Physiol.* 160, 1023–1036. doi: 10.1104/pp.112.202945
- Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., et al. (2014). *TRANSPARENT TESTA8* inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support for this research study provided by a research grant from the Department of Biotechnology (DBT), Government of India (102/IFD/SAN/2161/2013-14). SD and RR acknowledge the DBT and CSIR (Council of Scientific and Industrial Research) for Junior/Senior Research Fellowship awards.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00979>

- Cho, S., Chen, W., and Muehlbauer, F. J. (2004). Pathotype-specific genetic factors in chickpea (*Cicer arietinum* L.) for quantitative resistance to *Ascochyta* blight. *Theor. Appl. Genet.* 109, 733–739. doi: 10.1007/s00122-004-1693-x
- Cho, S., Kumar, J., Shultz, J. F., Anupama, K., Tefera, F., and Muehlbauer, F. J. (2002). Mapping genes for double podding and other morphological traits in chickpea. *Euphytica* 125, 285–292. doi: 10.1023/A:1020872009306
- Cobos, M. J., Fernandez, M., Rubio, J., Kharrat, M., Moreno, M. T., Gil, J., et al. (2005). Linkage map of chickpea (*Cicer arietinum* L.) based on populations from *Kabuli* × *Desi* crosses: Location of genes for resistance to *Fusarium* wilt race 0. *Theor. Appl. Genet.* 110, 1347–1353. doi: 10.1007/s00122-005-1980-1
- Debeaujon, I., Peeters, A. J. M., Léon-Kloosterziel, K. M., and Koornneef, M. (2001). The *TRANSPARENT TESTA12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* 13, 853–872. doi: 10.1105/tpc.13.4.853
- Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* 15:708. doi: 10.1186/1471-2164-15-708
- Diaz, A. M., Caldas, G. V., and Blair, M. W. (2010). Concentrations of condensed tannins and anthocyanins in common bean seed coats. *Food Res. Int.* 43, 595–601. doi: 10.1016/j.foodres.2009.07.014
- Dixon, R. A., Achnine, L., Kota, P., Liu, C. J., Reddy, M. S. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant Pathol.* 3, 371–390. doi: 10.1046/j.1364-3703.2002.00131.x
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379
- Erdmann, R., Gramzow, L., Melzer, R., Theissen, G., and Becker, A. (2010). *GORDITA (AGL63)* is a young paralog of the *Arabidopsis thaliana* B (sister) *MADS* box gene *ABS (TT16)* that has undergone neofunctionalization. *Plant J.* 63, 914–924. doi: 10.1111/j.1365-313X.2010.04290.x
- Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (2006). *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* 112, 1164–1171. doi: 10.1007/s00122-006-0218-1
- Ferraro, K., Jin, A. L., Nguyen, T. D., Reinecke, D. M., Ozga, J. A., and Ro, D. K. (2014). Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biol.* 14:238. doi: 10.1186/s12870-014-0238-y
- Focks, N., Sagasser, M., Weisshaar, B., and Benning, C. (1999). Characterization of *tt15*, a novel transparent testa mutant of *Arabidopsis thaliana* (L.) Heynh. *Planta* 208, 352–357. doi: 10.1007/s004250050569
- Forkmann, G. and Martens, S. (2001). Metabolic engineering and applications of flavonoids. *Curr. Opin. Biotechnol.* 12, 155–160. doi: 10.1016/S0958-1669(00)00192-0
- Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., et al. (2006). The *Rc* and *Rd* genes are involved in proanthocyanidin

- synthesis in rice pericarp. *Plant J.* 49, 91–102. doi: 10.1111/j.1365-313X.2006.02958.x
- Gaur, R., Azam, S., Jeena, G., Khan, A. W., Choudhary, S., Jain, M., 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. doi: 10.1093/dnares/dss018
- Gaur, R., Sathy, N. K., Choudhary, S., Shokeen, B., Gupta, V., and Bhatia, S. (2011). Advancing the STMS genomic resources for defining new locations on the intraspecific genetic linkage map of chickpea (*Cicer arietinum* L.). *BMC Genomics* 12:117. doi: 10.1186/1471-2164-12-117
- Golam Masum Akond, A. S. M., Khandaker, L., Berthold, J., Gates, L., Peters, K., Delong, H., et al. (2011). Anthocyanin, total polyphenols and antioxidant activity of common bean. *Am. J. Food. Technol.* 6, 385–394. doi: 10.3923/ajft.2011.385.394
- Gore, M. A., Hayes, A. J., Jeong, S. C., Yue, Y. G., Buss, G. R., and Saghai Maroof, M. A. (2002). Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. *Genome* 45, 592–599. doi: 10.1139/g02-009
- Graham, J., Matassa, V., Panozzo, J., and Starick, N. (2001). “Genotype and environment interaction for whole grain color in chickpea,” in *4th European Conference on Grain Legumes* (France), 372–373.
- Gujaria, N., Kumar, A., Dauthal, P., Dubey, A., Hiremath, P., Bhanu Prakash, A., et al. (2011). Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 122, 1577–1589. doi: 10.1007/s00122-011-1556-1
- Harborne, J. B., and Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry* 55, 481–504. doi: 10.1016/S0031-9422(00)00235-1
- Haseneyer, G., Stracke, S., Piepho, H. P., Sauer, S., Geiger, H. H., and Graner, A. (2010). DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC Plant Biol.* 10:5. doi: 10.1186/1471-2229-10-5
- Hossain, S., Ford, R., McNeil, D., Pittock, C., and Panozzo, J. F. (2010). Inheritance of seed size in chickpea (*Cicer arietinum* L.) and identification of QTL based on 100-seed weight and seed size index. *Aust. J. Crop. Sci.* 4, 126–135.
- Hossain, S., Panozzo, J. F., Pittock, C., and Ford, R. (2011). Quantitative trait loci analysis of seed coat color components for selective breeding in chickpea (*Cicer arietinum* L.). *Can. J. Plant Sci.* 91, 49–55. doi: 10.4141/cjps10112
- Hu, J., Reddy, V. S., and Wessler, S. R. (2000). The rice *R* gene family: two distinct subfamilies containing several miniature inverted-repeat transposable elements. *Plant Mol. Biol.* 42, 667–678. doi: 10.1023/A:1006355510883
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Huang, Z., Ban, Y., Yang, L., Zhang, Y., Li, H., Xiao, E., et al. (2012). Fine mapping of the yellow seed locus in *Brassica juncea* L. *Genome* 55, 8–14. doi: 10.1139/g11-072
- Hyten, D. L., Choi, I. Y., Song, Q., Shoemaker, R. C., Nelson, R. L., Costa, J. M., et al. (2007). Highly variable patterns of linkage disequilibrium in multiple soybean populations. *Genetics* 175, 1937–1944. doi: 10.1534/genetics.106.069740
- Ichino, T., Fuji, K., Ueda, H., Takahashi, H., Koumoto, Y., Takagi, J., et al. (2014). *GFS9/TT9* contributes to intracellular membrane trafficking and flavonoid accumulation in *Arabidopsis thaliana*. *Plant J.* 80, 410–423. doi: 10.1111/tj.12637
- Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O. K., et al. (2007). Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* 100, 1053–1071. doi: 10.1093/aob/mcm155
- Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2014). Genotyping-by-sequencing based intra-specific genetic map refines a “QTL-hotspot” region for drought tolerance in chickpea. *Mol. Genet. Genomics* 2, 559–571. doi: 10.1007/s00438-014-0932-3
- Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tj.12173
- Kharabian-Masouleh, A., Waters, D. L. E., Reinke, R. F., Ward, R., and Henry, R. J. (2012). SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. *Sci. Rep.* 2:557. doi: 10.1038/srep00557
- Kitamura, S., Shikazono, N., and Tanaka, A. (2004). *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J.* 37, 104–114. doi: 10.1046/j.1365-313X.2003.01943.x
- Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinives, P. (2012). The genetics of domestication of yardlong bean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *Sesquipedalis*. *Ann. Bot.* 109, 1185–1200. doi: 10.1093/aob/mcs048
- Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410
- Kottapalli, P., Gaur, P. M., Katiyar, S. K., Crouch, J. H., Buhariwalla, H. K., Pande, S., et al. (2009). Mapping and validation of QTLs for resistance to an Indian isolate of *Ascochyta* blight pathogen in chickpea. *Euphytica* 165, 79–88. doi: 10.1007/s10681-008-9762-x
- Kovnich, N., Saleem, A., Arnason, J. T., and Miki, B. (2012). Identification of two anthocyanidin reductase genes and three red-brown soybean accessions with reduced anthocyanidin reductase 1 mRNA, activity, and seed coat proanthocyanidin amounts. *J. Agric. Food Chem.* 60, 574–584. doi: 10.1021/jf2033939
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–374. doi: 10.1093/dnares/dst015
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2014). An efficient and cost-effective approach for genic microsatellite marker-based large-scale trait association mapping: identification of candidate genes for seed weight in chickpea. *Mol. Breed.* 34, 241–265. doi: 10.1007/s11032-014-0033-3
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166
- Kumar, A., Choudhary, A. K., Solanki, R. K., and Pratap, A. (2011). Towards marker-assisted selection in pulses: a review. *Plant Breed.* 130, 297–313. doi: 10.1111/j.1439-0523.2011.01851.x
- Kumar, V., Singh, A., Amitha Mithra, S. V., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 2, 133–145. doi: 10.1093/dnares/dsu046
- Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252
- Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The cradle of agriculture. *Science* 288, 1602–1603. doi: 10.1126/science.288.5471.1602
- Li, S., Zhao, B., Yuan, D., Duan, M., Qian, Q., Tang, L., et al. (2013). Rice zinc finger protein *DST* enhances grain production through controlling *Gn1a/OsCKX2* expression. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3167–3172. doi: 10.1073/pnas.1300359110
- Li, X., Chen, L., Hong, M., Zhang, Y., Zu, F., Wen, J., et al. (2012). A large insertion in *bHLH* transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. *PLoS ONE* 7:e44145. doi: 10.1371/journal.pone.0044145
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi: 10.1038/ng.977
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPI: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, C., Jun, J. H., and Dixon, R. A. (2014). *MYB5* and *MYB14* play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol.* 165, 1424–1439. doi: 10.1104/pp.114.241877
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282

- Liu, L. Z., Meng, J. L., Lin, N., Chen, L., Tang, Z. L., Zhang, X. K., et al. (2006). QTL mapping of seed coat color for yellow seeded *Brassica napus*. *Acta Genet. Sinica* 33, 181–187. doi: 10.1016/S0379-4172(06)60037-1
- Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R. (1989). *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcription activators and contains the *myc* homology region. *Proc. Natl. Acad. Sci. U.S.A.* 86, 7092–7096. doi: 10.1073/pnas.86.18.7092
- Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107
- Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J. M., et al. (2007). The *Arabidopsis* MATE transporter *TT12* acts as a vacuolar flavonoid/H⁺-antipporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* 19, 2023–2038. doi: 10.1105/tpc.106.046029
- Mather, K. A., Caicedo, A. L., Polato, N. R., Olsen, K. M., McCouch, S., and Purugganan, M. D. (2007). The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 177, 2223–2232. doi: 10.1534/genetics.107.079616
- McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12273–12278. doi: 10.1073/pnas.0900992106
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Mol, J., Grotewold, E., and Koes, R. (1998). How genes paint flowers and seeds. *Trends Plant Sci.* 3, 212–217. doi: 10.1016/S1360-1385(98)01242-4
- Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H. K., Horres, R., 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. doi: 10.1007/s00122-010-1265-1
- Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi: 10.1016/j.tplants.2004.05.006
- Negrão, S., Almadanim, M. C., Pires, I. S., Abreu, I. A., Maroco, J., Courtois, B., et al. (2013). New allelic variants found in key rice salt-tolerance genes: an association study. *Plant Biotechnol. J.* 11, 87–100. doi: 10.1111/pbi.12010
- Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* 19, 153–170. doi: 10.1007/BF02300753
- Nguyen, H. N., Kim, J. H., Hyun, W. Y., Nguyen, N. T., Hong, S. W., and Lee, H. (2013). *TTG1*-mediated flavonols biosynthesis alleviates root growth inhibition in response to ABA. *Plant Cell Rep.* 32, 503–514. doi: 10.1007/s00299-012-1382-1
- Pang, Y., Peel, G. J., Sharma, S. B., Tang, Y., and Dixon, R. A. (2008). A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14210–14215. doi: 10.1073/pnas.0805954105
- Parida, S. K., Mukerji, M., Singh, A. K., Singh, N. K., and Mohapatra, T. (2012). SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. *BMC Genomics* 13:426. doi: 10.1186/1471-2164-13-426
- Pesch, M., Dartan, B., Birkenbihl, R., Somssich, I. E., and Hülskamp, M. (2014). *Arabidopsis* *TTG2* regulates TRY expression through enhancement of activator complex-triggered activation. *Plant Cell* 26, 4067–4083. doi: 10.1105/tpc.114.129379
- Pourcel, L., Irani, N. G., Koo, A. J., Bohorquez-Restrepo, A., Howe, G. A., and Grotewold, E. (2013). A chemical complementation approach reveals genes and interactions of flavonoids with other pathways. *Plant J.* 74, 383–397. doi: 10.1111/tpj.12129
- Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). *PHA* of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18, 1274–1291. doi: 10.1105/tpc.105.034041
- Radhika, P., Gowda, S. J. M., Kadoo, N. Y., Mhase, L. B., Jamadagni, B. M., Sainani, M. N., et al. (2007). Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L.) using two recombinant inbred line populations. *Theor. Appl. Genet.* 115, 209–216. doi: 10.1007/s00122-007-0556-7
- Rahman, M., Li, G., Schroeder, D., and McVetty, P. B. E. (2011). Inheritance of seed coat color genes in *Brassica napus* (L.) and tagging the genes using SRAP, SCAR and SNP molecular markers. *Mol. Breed.* 26, 439–453. doi: 10.1007/s11032-009-9384-6
- Roorikiwal, M., Sawargaonkar, S. L., Chitkineeni, A., Thudi, M., Saxena, R. K., Upadhyaya, H. D., et al. (2013). Single nucleotide polymorphism genotyping for breeding and genetics applications in chickpea and pigeonpea using the BeadXpress platform. *Plant Genome* 6, 1–10. doi: 10.3835/plantgenome2013.05.0017
- Routaboul, J. M., Dubos, C., Beck, G., Marquis, C., Bidzinski, P., Loudet, O., et al. (2012). Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. *J. Exp. Bot.* 63, 3749–3764. doi: 10.1093/jxb/ers067
- Sabbavarapu, M. M., Sharma, M., Chamarthi, S. K., Swapna, N., Rathore, A., Thudi, M., et al. (2013). Molecular mapping of QTLs for resistance to *Fusarium* wilt (race 1) and *Ascochyta* blight in chickpea (*Cicer arietinum* L.). *Euphytica* 193, 121–133. doi: 10.1007/s10681-013-0959-2
- Sagasser, M., Lu, G. H., Hahlbrock, K., and Weisshaar, B. (2002). *A. thaliana* *TRANSPARENT TESTA 1* is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. *Genes Dev.* 16, 138–149. doi: 10.1101/gad.212702
- Saitoh, K., Onishi, K., Mikami, I., Thidar, K., and Sano, Y. (2004). Allelic diversification at the *C* (*OsC1*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics* 168, 997–1007. doi: 10.1534/genetics.103.018390
- Sakamoto, W., Ohmori, T., Kageyama, K., Miyazaki, C., Saito, A., Murata, M., et al. (2001). The Purple leaf (*Pl*) locus of rice: the *Pl* (*w*) allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. *Plant Cell Physiol.* 42, 982–991. doi: 10.1093/pcp/pce128
- Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014a). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 6, 695–710. doi: 10.1093/dnares/dsu031
- Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014b). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484
- Saxena, S. C., Salvi, P., Kaur, H., Verma, P., Petla, B. P., Rao, V., et al. (2013). Differentially expressed myo-inositol monophosphatase gene (*CaIMP*) in chickpea (*Cicer arietinum* L.) encodes a lithium-sensitive phosphatase enzyme with broad substrate specificity and improves seed germination and seedling growth under abiotic stresses. *J. Exp. Bot.* 64, 5623–5639. doi: 10.1093/jxb/ert336
- Senda, M., Masuta, C., Ohnishi, S., Goto, K., Kasai, A., Sano, T., et al. (2004). Patterning of virus-infected soybean seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. *Plant Cell* 16, 807–818. doi: 10.1105/tpc.019885
- Smykal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5:351. doi: 10.3389/fpls.2014.00351
- Sonah, H., Bastien, M., Iqura, E., Tardivel, A., Lègaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603
- Sonah, H., O'Donoghue, L., Cober, E., Rajcan, I., and Belzile, F. (2014). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* 2, 211–221. doi: 10.1111/pbi.12249
- Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., et al. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* 126, 2699–2716. doi: 10.1007/s00122-013-2166-x
- Stephens, A., Lombardi, M., Cogan, N. O. I., Foster, J. W., Hobson, K., Materne, M., et al. (2014). Genetic marker discovery, intraspecific linkage map construction and quantitative trait locus analysis of *Ascochyta* blight resistance in chickpea (*Cicer arietinum* L.). *Mol. Breed.* 33, 297–313. doi: 10.1007/s11032-013-9950-9
- Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single

- mutation conferring white pericarp in rice. *PLoS Genetics* 3:e133. doi: 10.1371/journal.pgen.0030133
- Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. (2006). *Caught red-handed: Rc* encodes a basic helix–loop–helix protein conditioning red pericarp in rice. *Plant Cell* 18, 283–294. doi: 10.1105/tpc.105.038430
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 12, 2725–2729. doi: 10.1093/molbev/mst197
- Thudi, M., Bohra, A., Nayak, S. N., Varghese, N., Shah, T. M., Penmetsa, R. V., et al. (2011). Novel SSR markers from BAC-end sequences, DARt arrays and a comprehensive genetic map with 1,291 marker loci for chickpea (*Cicer arietinum* L.). *PLoS ONE* 6:e27275. doi: 10.1371/journal.pone.0027275
- Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS ONE* 9:e96758. doi: 10.1371/journal.pone.0096758
- Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21760–21765. doi: 10.1073/pnas.0912396106
- Toker, C. (2009). A note on the evolution of *kabuli* chickpeas as shown by induced mutations in *Cicer reticulatum* Ladizinsky. *Genet. Resour. Crop. Evol.* 56, 7–12. doi: 10.1007/s10722-008-9336-8
- Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41, 206–210. doi: 10.2135/cropsci2001.411206x
- Upadhyaya, H. D., Dwivedi, S. L., Baum, M., Varshney, R. K., Udupa, S. M., Gowda, C. L. L., et al. (2008). Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8:106. doi: 10.1186/1471-2229-8-106
- Upadhyaya, H. D., Furman, B. J., Dwivedi, S. L., Udupa, S. M., Gowda, C. L. L., Baum, M., et al. (2006). Development of a composite collection for mining germplasm possessing allelic variation for beneficial traits in chickpea. *Plant Genet. Resour.* 4, 13–19. doi: 10.1079/PGR2005101
- Upadhyaya, H. D., and Ortiz, R. (2001). A mini-core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y
- Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2002). Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. *Euphytica* 123, 333–342. doi: 10.1023/A:1015088417487
- Van Ooijen, J. W. (2009). *MapQTL 6: Software for the Mapping Of Quantitative Trait Loci in Experimental Populations of Diploid Species*. Wageningen: Kyazma BV.
- Varshney, R. K., Paulo, M. J., Grando, S., van Eeuwijk, F. A., Keizer, L. C. P., Guo, P., et al. (2012). Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.). *Field Crop Res.* 126, 171–180. doi: 10.1016/j.fcr.2011.10.008
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6
- Wang, H., Fan, W., Li, H., Yang, J., Huang, J., and Zhang, P. (2013). Functional characterization of dihydroflavonol-4-reductase in anthocyanin biosynthesis of purple sweet potato underlies the direct evidence of anthocyanins function against abiotic stresses. *PLoS ONE* 8:e78484. doi: 10.1371/journal.pone.0078484
- Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485
- Winter, P., Benko-Iseppon, A. M., Hüttel, B., Ratnaparkhe, M., Tullu, A., Sonnante, G., et al. (2000). A linkage map of chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for *Fusarium* wilt races 4 and 5. *Theor. Appl. Genet.* 101, 1155–1163. doi: 10.1007/s001220051592
- Winter, P., Pfaff, T., Udupa, S. M., Hüttel, B., Sharma, P. C., Sahi, S., et al. (1999). Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol. Gen. Genet.* 262, 90–101. doi: 10.1007/s004380051063
- Wroblewski, T., Matvienko, M., Piskurewicz, U., Xu, H., Martineau, B., Wong, J., et al. (2014). Distinctive profiles of small RNA couple inverted repeat-induced post-transcriptional gene silencing with endogenous RNA silencing pathways in *Arabidopsis*. *RNA* 20, 1987–1999. doi: 10.1261/rna.046532.114
- Xie, D. Y., and Dixon, R. A. (2005). Proanthocyanidin biosynthesis - still more questions than answers? *Phytochemistry* 66, 2127–2144. doi: 10.1016/j.phytochem.2005.01.008
- Xu, W., Grain, D., Bobet, S., Le Gourrierc, J., Thévenin, J., Kelemen, Z., et al. (2014). Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 202, 132–144. doi: 10.1111/nph.12620
- Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050
- Yang, K., Jeong, N., Moon, J. K., Lee, Y. H., Lee, S. H., Kim, H. M., et al. (2010). Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *J. Hered.* 101, 757–768. doi: 10.1093/jhered/esq078
- Zhang, K., Lu, K., Qu, C., Liang, Y., Wang, R., Chai, Y., et al. (2013). Gene silencing of *BnTT10* family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PLoS ONE* 8:e61247. doi: 10.1371/journal.pone.0061247
- Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare allele of *OsPPKL1* associated with grain length causes extralarge grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21534–21539. doi: 10.1073/pnas.1219776110
- Zhao, J., and Dixon, R. A. (2009). MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *Plant Cell* 21, 2323–2340. doi: 10.1105/tpc.109.067819
- Zhao, J., Pang, Y., and Dixon, R. A. (2010). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153, 437–443. doi: 10.1104/pp.110.155432
- Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* 2:467. doi: 10.1038/ncomms1467
- Zuo, J., and Li, J. (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* 48, 99–118. doi: 10.1146/annurev-genet-120213-092138

Conflict of Interest Statement: The reviewer Bharadwaj Chellapilla declares that, despite being affiliated with the same institute and having previously collaborated with the author Shailesh Tripathi, the review process was handled objectively. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Bajaj, Das, Upadhyaya, Ranjan, Badoni, Kumar, Tripathi, Gowda, Sharma, Singh, Tyagi and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.