![frontiers in Plant Science]

Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea

Alice Kujur[1†], Deepak Bajaj[1†], Hari D. Upadhyaya[2†], Shouvik Das[1], Rajeev Ranjan[1], Tanima Shree[1], Maneesha S. Saxena[1], Saurabh Badoni[1], Vinod Kumar[3], Shailesh Tripathi[4], C. L. L. Gowda[2], Shivali Sharma[2], Sube Singh[2], Akhilesh K. Tyagi[1] and Swarup K. Parida[1*]

[1] National Institute of Plant Genome Research (NIPGR), New Delhi, India, [2] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Telangana, India, [3] National Research Centre on Plant Biotechnology (NRCPB), New Delhi, India, [4] Division of Genetics, Indian Agricultural Research Institute (IARI), New Delhi, India

The genome-wide discovery and high-throughput genotyping of SNPs in chickpea natural germplasm lines is indispensable to extrapolate their natural allelic diversity, domestication, and linkage disequilibrium (LD) patterns leading to the genetic enhancement of this vital legume crop. We discovered 44,844 high-quality SNPs by sequencing of 93 diverse cultivated desi, kabuli, and wild chickpea accessions using reference genome- and de novo-based GBS (genotyping-by-sequencing) assays that were physically mapped across eight chromosomes of desi and kabuli. Of these, 22,542 SNPs were structurally annotated in different coding and non-coding sequence components of genes. Genes with 3296 non-synonymous and 269 regulatory SNPs could functionally differentiate accessions based on their contrasting agronomic traits. A high experimental validation success rate (92%) and reproducibility (100%) along with strong sensitivity (93–96%) and specificity (99%) of GBS-based SNPs was observed. This infers the robustness of GBS as a high-throughput assay for rapid large-scale mining and genotyping of genome-wide SNPs in chickpea with sub-optimal use of resources. With 23,798 genome-wide SNPs, a relatively high intra-specific polymorphic potential (49.5%) and broader molecular diversity (13–89%)/functional allelic diversity (18–77%) was apparent among 93 chickpea accessions, suggesting their tremendous applicability in rapid selection of desirable diverse accessions/inter-specific hybrids in chickpea crossbred varietal improvement program. The genome-wide SNPs revealed complex admixed domestication pattern, extensive LD estimates (0.54–0.68) and extended LD decay (400–500 kb) in a structured population inclusive of 93 accessions. These findings reflect the utility of our identified SNPs for subsequent genome-wide association study (GWAS) and selective sweep-based domestication trait dissection analysis to identify potential genomic loci (gene-associated targets) specifically regulating important complex quantitative agronomic traits in chickpea. The numerous informative

genome-wide SNPs, natural allelic diversity-led domestication pattern, and LD-based information generated in our study have got multidimensional applicability with respect to chickpea genomics-assisted breeding.

Keywords: *Cicer*, chickpea, *desi*, GBS, *kabuli*, linkage disequilibrium, SNP

Introduction

The second most abundantly grown food legume chickpea (*Cicer arietinum* L.) is a self-pollinated and diploid (2n = 2x = 16) crop species belonging to Fabaceae (Kumar et al., 2011). It is a vital source of human dietary protein enriched with essential amino acids. Chickpea is believed to have diverged from a wild progenitor *C. reticulatum* by single domestication event at a Fertile Crescent (South Eastern Turkey and adjacent Syria) near about 10,000 years ago (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., 2013a; Saxena et al., 2014a). A series of four sequential evolutionary bottlenecks coupled with strong adaptation-based selection pressure/selective sweeps during chickpea domestication might have significantly narrowed-down the genetic base of the presently cultivated varieties (Abbo et al., 2003; Berger et al., 2005; Burger et al., 2008; Toker, 2009). The draft genomes of two major chickpea cultivars-*kabuli* (large seeded) and *desi* (small seeded) representing diverse gene pools have been successfully sequenced (Jain et al., 2013; Varshney et al., 2013a). These genome sequencing efforts signified that ∼70% of their total sequenced draft genomes are represented by low-complexity regions, which could serve as reference for subsequent resequencing of diverse *desi* and *kabuli* accessions in order to discover and validate informative sequence-based markers at a genome-wide scale by deploying suitable high-throughput genotyping assay.

Single nucleotide polymorphisms (SNPs) are highly preferred in plant genetic and genome analyses because of their excellent genetic attributes, such as wide genomic distribution, co-dominant inheritance, high reproducibility, and chromosome-specific location (Brookes, 1999; Cho et al., 1999; Gupta et al., 2001; Sachidanandam et al., 2001; Rafalski, 2002; Gupta and Rustgi, 2004). For a large chickpea genome (∼740 Mb) with a narrow genetic base, these informative SNPs aids in identifying trait-regulatory genes/QTLs (quantitative trait loci) for marker-assisted genetic enhancement. Tremendous efforts have been made in chickpea toward mining a vast number of SNPs *in silico* from the ESTs (expressed sequence tag), transcripts, genes and genome sequences of diverse *desi, kabuli,* and wild accessions using traditional Sanger and high-throughput next-generation sequencing (NGS) approaches (Varshney et al., 2007, 2013a; Rajesh and Muehlbauer, 2008; Nayak et al., 2010; Garg et al., 2011; Gujaria et al., 2011; Hiremath et al., 2011; Agarwal et al., 2012; Azam et al., 2012; Jhanwar et al., 2012; Jain et al., 2013; Kujur et al., 2013). Also a smaller subset of selected SNPs have been successfully validated and genotyped in diverse chickpea accessions by low-throughput allele-specific resequencing, allele-specific PCR (polymerase chain reaction) and CAPS (cleaved amplified polymorphic sequence) assays to be utilized for various marker-based genotyping applications. More recently, the

assistance of several array-based SNP genotyping approaches, like Illumina GoldenGate/Infinium (Bead Xpress array) and Competitive Allele Specific PCR (KASPar) assays have pushed ahead the process of large-scale validation and high-throughput genotyping of SNPs in different chickpea accessions. In chickpea, so far about 3000 such validated SNPs have been exploited, specifically in genetic diversity and evolutionary studies (Gaur et al., 2012; Hiremath et al., 2012; Thudi et al., 2012; Roorkiwal et al., 2013; Stephens et al., 2014).

With advent of barcoded multiplexing and restriction enzyme (RE)-based NGS approaches, the simultaneous integration of large-scale discovery, validation and genotyping of SNPs is now possible in numerous diverse crop accessions. These NGS-based methods have evolved plenty of high-throughput and cost-effective SNP genotyping assays and applications for many small diploid and larger polyploid crop genomes. Some of these assays include reduced representation libraries (RRLs) (van Tassell et al., 2008), complexity reduction of polymorphic sequences (CRoPs) (van Orsouw et al., 2007), restriction-site-associated DNA sequencing (RAD-seq) (Miller et al., 2007; Baird et al., 2008), and more recently, genotyping-by-sequencing (GBS) (Elshire et al., 2011) assays. In addition to simultaneous high-throughput genome-wide SNP discovery and genotyping in a larger set of diverse crop accessions, these assays are also particularly helpful for ultra-high density genetic and physical linkage map construction and genome-wide high-resolution trait association and genetic mapping (Huang et al., 2009, 2010; Barchi et al., 2011; Chutimanitsakun et al., 2011; Davey et al., 2011; Pfender et al., 2011; Yu et al., 2011). Quite recently, the significance of RAD-seq has been clearly demonstrated in mining and genotyping of ∼4.4 million genome-wide SNPs for understanding the molecular diversity and phylogenies among 90 diverse chickpea accessions (Varshney et al., 2013a).

GBS is a simple, yet fast and reproducible latest RE (restriction enzyme)-based improved NGS method that has gained considerable attention of crop researchers for large-scale discovery and genotyping of SNPs simultaneously in diverse crop germplasm accessions and mapping populations at a genome-wide scale (Schnable et al., 2009; Elshire et al., 2011). The maximal efficiency of GBS in mining and genotyping of SNPs with wider genomic distribution is due to its highest barcoded multiplexing capability (pooling up to 384 barcoded accessions in a single sequencing lane) at methylation-sensitive RE (restriction endonuclease)-sites (*APe*KI) and also complexity reduction (targeting lower copy number of genome) ability, especially in larger complex plant genomes (Schnable et al., 2009; Elshire et al., 2011; Poland et al., 2012a). This assay is now the most preferred and well-established SNP discovery and genotyping approach, thus has got multiple utilities in high-throughput marker-based applications of plant genomics and breeding with optimal use of

resources, high genotype multiplexing and rapid SNP genotyping potential.

The GBS method has also been extensively applied in diverse germplasm lines of barley (Mayer et al., 2012), wheat (Poland et al., 2012b), soybean (Sonah et al., 2013), sorghum (Morris et al., 2013; Thurber et al., 2013), and rice (Spindel et al., 2013) for large-scale mining and high-throughput genotyping of SNPs at a genome-wide scale. The significant outcomes of these studies have proven the efficacy of GBS for understanding the molecular diversity and phylogeny in crop plants (Byrne et al., 2013; Crossa et al., 2013; Mascher et al., 2013; Uitdewilligen et al., 2013; Bastien et al., 2014; Huang et al., 2014; Jarquín et al., 2014; Liu et al., 2014; Tardivel et al., 2014; Sonah et al., 2014). Notably, the above assay was recently employed for constructing high-resolution intra- and inter-specific genetic linkage maps and fine mapping of QTLs/genes associated with drought tolerance in chickpea (Deokar et al., 2014; Jaganathan et al., 2014). Nonetheless, as far as molecular diversity analysis in core/mini-core germplasm lines is concerned, the GBS assay is yet to be utilized in chickpea. Thus, considering the successful application of GBS assay in high-throughput genetic analyses in other crop plants (He et al., 2014), its use for mining and validating of genome-wide SNPs in diverse natural populations of chickpea assumes significance. The information acquired from diverse germplasm lines by GBS assay could enable the realistic estimation of natural allelic diversity and domestication pattern in chickpea. Moreover, this will accelerate various large-scale genotyping applications, including construction of high-density integrated genetic, physical and genome maps, comparative genome mapping, genome-wide association study (GWAS), selective sweep-based trait dissection analysis and fine mapping/map-based cloning of QTLs/genes controlling traits of agronomic importance and eventually marker-assisted genetic improvement of chickpea.

Keeping that in view, the present study has utilized GBS approach for the first time to discover, validate and genotype SNPs in 93 diverse [with significant seed yield (g) per plant variability] cultivated (*desi* and *kabuli*) and wild chickpea accessions at a genome-wide scale. The genotyping information of validated SNPs was further utilized to understand their functional significance and determine the molecular diversity pattern, population genetic structure and LD patterns among chickpea accessions.

Materials and Methods

Chickpea Accessions used for Genomic DNA Extraction

Ninty-three accessions belonging to 92 cultivated *C. arietinum desi* (39 accessions) and *kabuli* (53), and one *C. reticulatum* wild accession (ICC 17160) were utilized (Supplementary Table S1) for genome-wide discovery and genotyping of SNPs employing GBS assay. Ninety-two *desi* and *kabuli* accessions of these, with significant phenotypic [seed yield (g)/plant] and genotypic diversity (>80% diversity of total germplasm lines evaluated) were selected from available chickpea germplasm collections (16,991, including 211 minicore and 300 reference core germplasm lines) (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2001, 2008) following the methods of Kujur et al. (2014). An additional wild

accession (ICC 17160) was included in GBS assay for understanding its molecular diversity and phylogenies with cultivated *desi* and *kabuli* chickpea. The genomic DNA was isolated from the young leaf samples of 93 accessions using a QIAGEN DNeasy 96 Plant Kit (QIAGEN, CA, USA) following the manufacturer's instructions.

Library Preparation, Sequencing, and Sequence Read Mapping

The isolated genomic DNA of 96 chickpea accessions (93 accessions along with three accessions as biological replicates) was digested with *Ape*KI and ligated to adapters containing one of 96 unique barcodes to construct 96-plex GBS libraries. These libraries were pooled together (following Elshire et al., 2011; Spindel et al., 2013) and sequenced (100-bp single end) using Illumina HiSeq 2000. The reproducibility of the GBS assay was evaluated using three *desi* and *kabuli* accessions (ICCX-810800, ICCV10, and ICCV95334) as biological replicates (Supplementary Table S1). The FASTQ sequence reads generated from accessions were processed for quality assessment. For this assessment, a sliding window approach implemented in the STACKS v1.0 (Catchen et al., 2013; http://creskolab.uoregon.edu/stacks) was utilized to examine the quality distribution of each sequence read. The reads with an average quality below 90% confidence (a *phred* score of 10; Ewing and Green, 1998) within a sliding window were discarded. Additionally, the sequence reads showing prolonged drops in quality were discarded. The rest of the good quality sequence reads were further analyzed for their quality using FASTQC v0.10.1. The high-quality reads were de-multiplexed based upon their unique barcodes to extract sequence reads of individual accessions. The individual sequence reads of 96 accessions were separately aligned and mapped to reference drafts of *desi* (ICC 4958; Jain et al., 2013) and *kabuli* (CDC Frontier; Varshney et al., 2013a) chickpea genome sequences using Bowtie v2.1.0 with default parameters (Langmead and Salzberg, 2012). The unaligned sequence reads of *desi* and *kabuli* reference genomes were further processed individually using the *de novo* genotyping approach of STACKS.

Discovery and Genotyping of SNPs

The sequence alignment map (SAM) files generated from each *desi* and *kabuli* genome were processed using reference-based GBS pipeline of STACKS to identify accurate SNPs in 96 accessions. The SNP identification process in pstacks is efficient enough for handling gapped alignments properly to accommodate InDels and soft-masked alignment fragments. This prevents the SNP model from wrongly calling polymorphisms due to InDel frameshifts. Stacks algorithm also filters paralog and high-copy loci by read coverage, assuming random coverage across loci. We removed the putative loci with more than twice the standard diversion of coverage depth to filter out repetitive elements and stacks of paralogs loci. A maximum likelihood statistical model was used in STACKS to identify valid and high-quality SNPs (with no sequencing errors and SNP base quality = 20), supported by at least 10 sequence reads. A catalog was built to record all SNP loci identified in 96 accessions by matching each accession to that catalog and detecting accurate alleles (minimum

sequence read depth: 10) that are present at every SNP locus in each accession.

Using the *de novo* genotyping approach of STACKS, the sampled SNP loci from each accession were reassembled from the sequenced reads (reads unaligned with reference genome) for creating stacks. A *de novo* method was employed to group all exactly matching sequences into stacks, forming putative loci from high-confidence unique stacks. To detect polymorphism and infer SNP alleles, each putative locus of one-nucleotide position at a time was examined using a maximum likelihood framework (Hohenlohe et al., 2010) implemented in STACKS. We set the parameters in a way that for each stack in the locus there was another stack in the locus, which is at most one nucleotide divergent (−M 1). A minimum of three raw reads that requires to form a stack (−m 3) were specified. The stack information was further used to create a catalog against which the other accessions matched, resulting in a map of all the alleles (minimum sequence read depth: 10 with SNP base quality = 20) in 96 accessions for a set of SNP loci.

Structural and Functional Annotation of SNPs

The physical positions (bp) of SNPs identified by reference-based GBS approaches were correlated with the GFF file containing the genome annotation of *desi* (CGAP v1.0; Jain et al., 2013) and *kabuli* (Varshney et al., 2013a) chickpea to determine the distribution of SNPs in various coding and non-coding sequence components of genes and genomes (chromosomes/pseudomolecules and unanchored scaffolds). The customized Perl scripts and a single-nucleotide polymorphism effect predictor (SnpEff v3.1h; http://snpeff.sourceforge.net/) were used to deduce accurate genomic distribution, including the structural and functional annotation of SNPs (synonymous and non-synonymous SNPs). To determine the genomic distribution of identified SNPs across chickpea chromosomes, the SNPs with synonymous and non-synonymous substitutions were plotted individually based on their unique physical positions (bp) on eight chromosomes (pseudomolecules) of *desi* and *kabuli* genomes and visualized using Circos visualization tool (Krzywinski et al., 2009).

The putative functions, including conserved domains in the proteins encoded by SNP-carrying genes were predicted based on available *desi* and *kabuli* genome annotations (Jain et al., 2013; Varshney et al., 2013a) and PFAM database v27.0 (http://pfam.sanger.ac.uk). The SNP-carrying genes were BLAST searched (cut-off *E*-value = 1e-41) against the KOGnitor NCBI database (ftp://ftp.ncbi.nih.gov/pub/COG/KOG) for detailed functional classification. Gene Ontology (GO) enrichment analysis of SNPs-containing genes was performed using BiNGO plugin of Cytoscape V2.6 (Maere et al., 2005). For the GO enrichment significance test (*P* = 0.05), the Benjamini and Hochberg false discovery rate correction was employed.

For estimation of Ka (number of non-synonymous substitutions per non-synonymous site)/Ks (number of synonymous substitutions per synonymous site) substitution ratio, the SNP-carrying coding sequences of genes annotated from the reference *desi* and *kabuli* genomes were compared among accessions and analyzed following the methods of Li et al. (2011) and Devisetty et al. (2014). To further deduce the direction as well

as magnitude of natural selection acting on protein-coding *desi* and *kabuli* genes, the significant departure from standard neutral model Ka/Ks = 1 was evaluated using maximum-likelihood ratio method implemented in the CODEML program of PAMLv4.8a (http://abacus.gene.ucl.ac.uk/software/paml.html).

Large-Scale Validation of SNPs

For validating SNPs identified through GBS assay, the SNP genotyping information among four chickpea accessions (ICC 4958, ICC 4951, ICC 12968, and ICC 17160) was compared/correlated with an available *in silico* SNP database of same accessions according to their nature/types and bp on the *desi* (ICC 4958) genome (Gaur et al., 2012; Jain et al., 2013). For experimental validation, 384 non-synonymous and upstream regulatory SNPs were selected based on their presence in the transcription factor (TF) genes, potential to reveal polymorphism between a representative set of stress tolerant and sensitive chickpea accessions and localization in the genes underlying the known QTLs for stress tolerance in chickpea. Moreover, 70 *de novo*-based GBS-SNPs screened as per their uniform physical distribution (within ~1 Mb physical distance between SNPs) on *desi* and *kabuli* genomes, were included in our validation study. For these analyses, the primers designed from 300-bp sequences flanking either side of 214 selected SNPs were PCR amplified using the genomic DNA of 93 chickpea accessions. The amplified PCR products were purified and sequenced. The aligned high-quality sequences were used to detect SNPs among accessions following Kujur et al. (2013) and Saxena et al. (2014a,b). Additionally, to validate GBS-based SNPs, a representative set of 240 SNPs were genotyped in the genomic DNA of 93 accessions using 34-plex Sequenom MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) MassARRAY (http://www.sequenom.com) following Saxena et al. (2014a,b). The SNP alleles scanned among accessions were cataloged according to allele-specific differences in mass between extension products. The sensitivity [TP/(TP + FN) × 100] and specificity [TN/(TN + FP) × 100] rate (%) of validated SNPs were determined by estimating the number of true positives (TP), false negatives (FN), true negatives (TN), and false positives (FP).

Estimation of Diversity Statistics, Population Genetic Structure, and LD Patterns

The SNP genotyping data (minor allele frequency/MAF ≥ 5% with missing data < 10%, Zhao et al., 2011) were analyzed with PowerMarker v3.51 (Liu and Muse, 2005), MEGA v5.0 (Tamura et al., 2011), and TASSEL v5.0 (Bradbury et al., 2007; http://www.maizegenetics.net) to estimate the PIC (polymorphism information content) and for constructing an unrooted neighbor-joining (NJ)-based phylogenetic tree (Nei's genetic distance; Nei et al., 1983 with 1000 bootstrap replicates) among 93 chickpea accessions. Various nucleotide diversity parameters, including θπ (average number of nucleotide substitutions per site between any two DNA sequences selected randomly from chickpea population under study), θω (average estimates of mutation rate derived from a total number of segregating sites that corrected for population size) and Tajima's D (statistical test to infer the evidence of neutral evolution for each defined region

within a selected population) were measured using a 100-kb non-overlapping sliding window approach of TASSEL v5.0 and following the methods of Xu et al. (2012) and Varshney et al. (2013a).

For determining the population structure among 93 cultivated and wild chickpea accessions, the SNP genotyping data were analyzed in a model-based program STRUCTURE v2.3.4 (Pritchard et al., 2000) adopting the methods of Kujur et al. (2013, 2014). The optimal value of population number (K) was determined in accordance with *ad-hoc* method of Pritchard et al. (2000) and *delta* K procedure of Evanno et al. (2005). To determine the genome-wide LD patterns (r^2; frequency correlation among pair of alleles across a pair of SNP loci), including LD decay (by plotting average r^2 against 10 and 100 kb uniform physical intervals across eight *desi* and *kabuli* chromosomes) in each *desi* and *kabuli* genomes, the genotyping data of SNPs that were physically mapped on eight chromosomes were analyzed using a command (−r2 −ld-window 99999 −ld-window-r2 0) line interface of PLINK (Purcell et al., 2007) and the full-matrix approach of TASSEL (following Zhao et al., 2011; Kujur et al., 2013, 2014). The ANOVA (analysis of variance)-based significance test of LD estimates by comparing their r^2-values across population groups and chromosomes was performed using SPSS v17.0 (http://www.spss.com/statsistics) and following Saxena et al. (2014a).

Results

Genome-Wide SNPs Discovery using a GBS Assay

The sequencing of 96-plex *ApeKI* libraries constructed from 96 diverse chickpea accessions (Supplementary Table S1) using a GBS assay generated 246.5 million reads (∼30-fold sequencing depth of coverage), with an average of 2.5 million reads (varied from 1.5 to 9.9 million reads) per accession (Supplementary Figure S1). After quality filtering, 207.9 (84.3%) million processed high-quality reads uniformly distributed across 96 accessions were obtained. Of these high-quality sequence reads, 81.2 and 72.6% were mapped to unique physical locations on *desi* and *kabuli* reference genomes, respectively (Supplementary Table S2). The sequencing data generated in our study have been submitted to a free publicly-accessible NCBI-short read archive (SRA) database (www.ncbi.nlm.nih.gov/sra) with accession number SRX845396 (Release date: 31st March 2015). Using both reference and *de novo*-based GBS approaches in 96 accessions, 44,844, including 20,439 and 24,405 high-quality SNPs (with read-depth = 10, SNP base quality = 20, < 10% missing data and ∼2% heterozygosity in each accession) were identified from *desi* and *kabuli* genomes, respectively (**Table 1**, **Figures 1A–C**, Supplementary Table S3). A high degree of reproducibility (100%) of SNPs identified by GBS assay in three *desi* and *kabuli* chickpea accessions (ICCX-810800, ICCV10, and ICCV 95334) used as biological replicates was observed. The transitions were more frequent than the transversions that composed slightly more than a-half (57.9%; 11,831 SNPs in *desi* and 57.5%; 14,022 SNPs in *kabuli*) of the identified SNPs (**Figure 1D**). A higher frequency of A/G transitions was evident in both *desi* (29.3%) and *kabuli* (29.7%) chickpea compared to

A/C (11.8% in *desi*) and G/T (12.1% in *kabuli*) transversions (**Figure 1D**).

A total of 20,439 SNPs identified in the *desi* genome includes 13,593 SNPs from reference genome as well as 6846 *de novo* discovered SNPs from its unaligned sequence reads (**Figure 1A**, Supplementary Table S3). The mined 6063 reference genome-based SNPs were physically mapped across eight *desi* chromosomes with an average map density of 20.5 kb (**Table 1**, **Figures 1B,C**). A genome-wide SNP density plot revealed that a maximum proportion of SNPs were physically mapped on *desi* chromosome 3 (21.6%, 1311 SNPs) (**Figures 1A, 2A**). The average marker density was maximum on chromosome 3 (17.8 kb), whereas it was minimum on chromosome 7 (**Table 1, Figure 2A**). The remaining 7530 SNPs were physically mapped on unanchored scaffolds of *desi* genome.

A total of 24,405 SNPs discovered from the *kabuli* genome includes 16,376 and 8029 reference genome and *de novo*-based SNPs, respectively (**Figure 1A**, Supplementary Table S3). The reference genome-based 14,115 SNPs were physically mapped across eight *kabuli* chromosomes with a mean map density of 24.6 kb (**Table 1**, **Figures 1B,C**). A genome-wide SNP density plot indicated that a higher proportion (18.4%, 2598 SNPs) of SNPs were physically mapped on *kabuli* chromosome 4 (**Figures 1A, 2B**). The *kabuli* chromosomes 4 and 7 had maximum (18.9 kb) and minimum (29.8 kb) average marker density, respectively (**Table 1**, **Figure 2B**). The rest 2261 SNPs were physically mapped on unanchored scaffolds of *kabuli* genome. In total, 16,736 SNPs identified from the reference *kabuli* genome have been submitted to NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP), with SNP submission (SS) accession numbers 974751673–974768048, which will be publicly available in the next NCBI db SNP Build (B142), year 2015 (http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=NIPGR).

Structural and Functional Annotation of SNPs

The structural annotation of 13,593 *desi* reference genome-based SNPs showed the presence of 9567 SNPs in 4010 genes and 4026 SNPs in the intergenic regions (Supplementary Table S2). Likewise, the annotation of 16,372 *kabuli* reference genome-based SNPs identified 10,656 and 5720 SNPs in 4643 genes and intergenic regions, respectively. The SNPs-carrying gene sequences, including SNPs with their 100-bp flanking sequences were compared between *desi* and *kabuli* genomes based on sequence homology to filter out the duplicated genes and SNPs observed commonly between the two genomes. This analysis identified 6170 non-redundant SNPs (including 2117 SNPs in 945 genes and 4053 intergenic SNPs) specifically from *desi* genome (Supplementary Table S2). Combining these unique *desi*-specific SNPs data with 16,376 SNPs identified from *kabuli*, the detailed structural annotation of 22,542 SNPs (Supplementary Table S4) in different sequence components of chickpea genes/genomes and the functional annotation of SNP-carrying genes were performed. The maximum percentage (56.7%, 12,773 SNPs) of SNPs were identified in 5588 chickpea genes, whereas the remaining 43.3% (9773) of SNPs were detected in the intergenic regions (**Figure 3**).

TABLE 1 | Genomic distribution and nucleotide diversity potential of SNPs physically mapped on eight *desi* and *kabuli* chickpea chromosomes.

Chromosomes	Size (Mb) of chromosomes (pseudomolecules)		Number (%) of SNPs mapped		Average map density (kb)		Nucleotide diversity					
							Desi			Kabuli		
	Desi	*Kabuli*	*Desi*	*Kabuli*	*Desi*	*Kabuli*	Theta Pi $(\theta\pi)$ kb^{-1}	Theta w $(\theta\omega)$ kb^{-1}	Tajima's D	Theta Pi $(\theta\pi)$ kb^{-1}	Theta w $(\theta\omega)$ kb^{-1}	Tajima's D
Ca_Chr01	14.79	48.36	753 (12.4)	2117 (15.0)	19.6	22.8	0.89	1.56	−2.54	0.87	1.62	−2.71
Ca_Chr02	17.30	36.63	758 (12.5)	1283 (9.1)	22.8	28.6	0.81	1.45	−2.61	0.96	1.32	−1.99
Ca_Chr03	23.38	39.99	1311 (21.6)	1664 (11.8)	17.8	24.0	0.90	1.54	−2.51	0.78	1.29	−2.43
Ca_Chr04	22.09	49.19	1005 (16.6)	2598 (18.4)	22.0	18.9	1.33	1.69	−2.78	1.31	1.72	−2.91
Ca_Chr05	16.30	48.17	792 (13.1)	1782 (12.6)	20.6	27.0	0.47	1.25	−2.27	0.43	1.01	−2.11
Ca_Chr06	11.48	59.46	636 (10.5)	2182 (15.5)	18.1	27.3	0.74	1.42	−2.73	0.70	1.51	−2.99
Ca_Chr07	8.46	48.96	331 (5.5)	1643 (11.6)	25.6	29.8	1.13	1.53	−1.96	0.95	1.67	−2.58
Ca_Chr08	10.57	16.48	477 (7.9)	846 (6.0)	22.2	19.5	0.98	1.30	−1.88	0.75	1.19	−2.35
Total	124.37	347.24	6063	14115	20.5	24.6	0.99	1.47	−2.31	0.92	1.65	−2.39
Unanchored scaffolds	NA	NA	7530	2261	NA	NA	1.06	1.93	−2.66	1.10	2.28	−2.86
De novo	NA	NA	6846	8029	NA	NA	0.93	0.66	−2.99	2.17	0.99	−2.95
Total	NA	NA	20,439	24,405	NA	NA	1.02	1.35	−2.67	1.40	1.64	−2.75

Ca_Chr: Cicer arietinum Chromosomes.

θπ per kb: Average pair-wise nucleotide diversity.

θω per kb: Watterson's estimator of segregating sites.

The number of SNPs identified in the genes varied from 1 to 34, with an average frequency of 3 SNPs/gene. The annotation of SNPs in various structural components of genes revealed the presence of a greater percentage (56.5%, 7212 SNPs) of SNPs in the exons, followed by introns (41.4%, 5292) and 2000-bp upstream regulatory regions (URRs) (1.5%, 191), and a minimum percentage (0.6%, 78 SNPs) in the 1000-bp downstream regulatory regions (DRRs) (**Figure 3**). Among the identified coding SNPs, 3916 SNPs within 2510 genes showed synonymous substitutions, whereas 3296 SNPs in 2048 genes resulted in non-synonymous substitutions (**Figure 3**). The identified non-synonymous SNPs included 3232 missense SNPs in 2025 genes showing amino acid substitutions and 64 non-sense SNPs in 23 genes culminating into premature termination codons introduced by nucleotide replacements (**Figure 3**). The estimation of average non-synonymous to synonymous substitution rates in 7212 coding SNPs across 4558 chickpea genes revealed that a larger proportion (~97%) of these genes had Ka/Ks = 0.84 (<1.0) and thus under negative/purifying selection pressure. A maximum (1.04) Ka/Ks was obtained between ICC 15264 (*kabuli*) and ICC 17160 (*C. reticulatum*) and minimum (0.47) between ICC 7346 (*kabuli*) and ICC 15512 (*kabuli*). The Tajima's D estimate further inferred that a higher proportion (95%) of SNP-carrying genes under purifying selection with reduced $D = -2.48$ ($D < -2$) and remaining 5% genes had elevated $D > 2$.

The functional annotation of 5588 genes with SNPs revealed their maximum correspondence to growth, development and

metabolism-related proteins (3073, 55%), followed by TFs (92, 16.3%) and signal transduction proteins (39, 7%), with a minimal correspondence to genes encoding hypothetical proteins (12, 2.2%) (Supplementary Figure S2A). The KOG-based determination of putative functions (excluding unknown and general functions) for SNP-carrying genes indicated their primary roles in signal transduction mechanisms (20.4%), followed by transcription (10.2%) and carbohydrate transport and metabolism (8.8%) (Supplementary Figure S2B). Furthermore, GO enrichment analysis of SNP-carrying genes depicted a significant over-representation/enrichment of GO terms in the genes associated with molecular function (nucleic acid binding, 22%, P: 1.9 × 10^{-31}), followed by biological process (metabolism, 6.6%, 3.4 × 10^{-94}) and cellular component (macromolecular complex, 1.7%, 1.4 × 10^{-50}) (Supplementary Figure S3). The detailed functional annotation of genes with SNPs detected a greater percentage of SNPs belonging to growth, development and metabolism-related protein-encoding genes (11.8%, 659 of total identified 5588 SNPs carrying genes), TF-encoding genes (3.7%, 207), and unknown expressed proteins (4.1%, 229) that result in missense and non-sense non-synonymous substitutions. Interestingly, 123 SNPs, including 13 non-synonymous SNPs and 12 upstream regulatory SNPs identified in the 73 chickpea genes showed polymorphism between abiotic stress tolerant (ICC 4958 and IC 296131) and sensitive (ICCV 93954) chickpea accessions (Supplementary Tables S4, S5). For instance, one SNP (C–G) showing missense non-synonymous amino acid substitution [aspartic acid (GAC) to glutamic acid (GAG)] in the plant homeodomain (PHD)
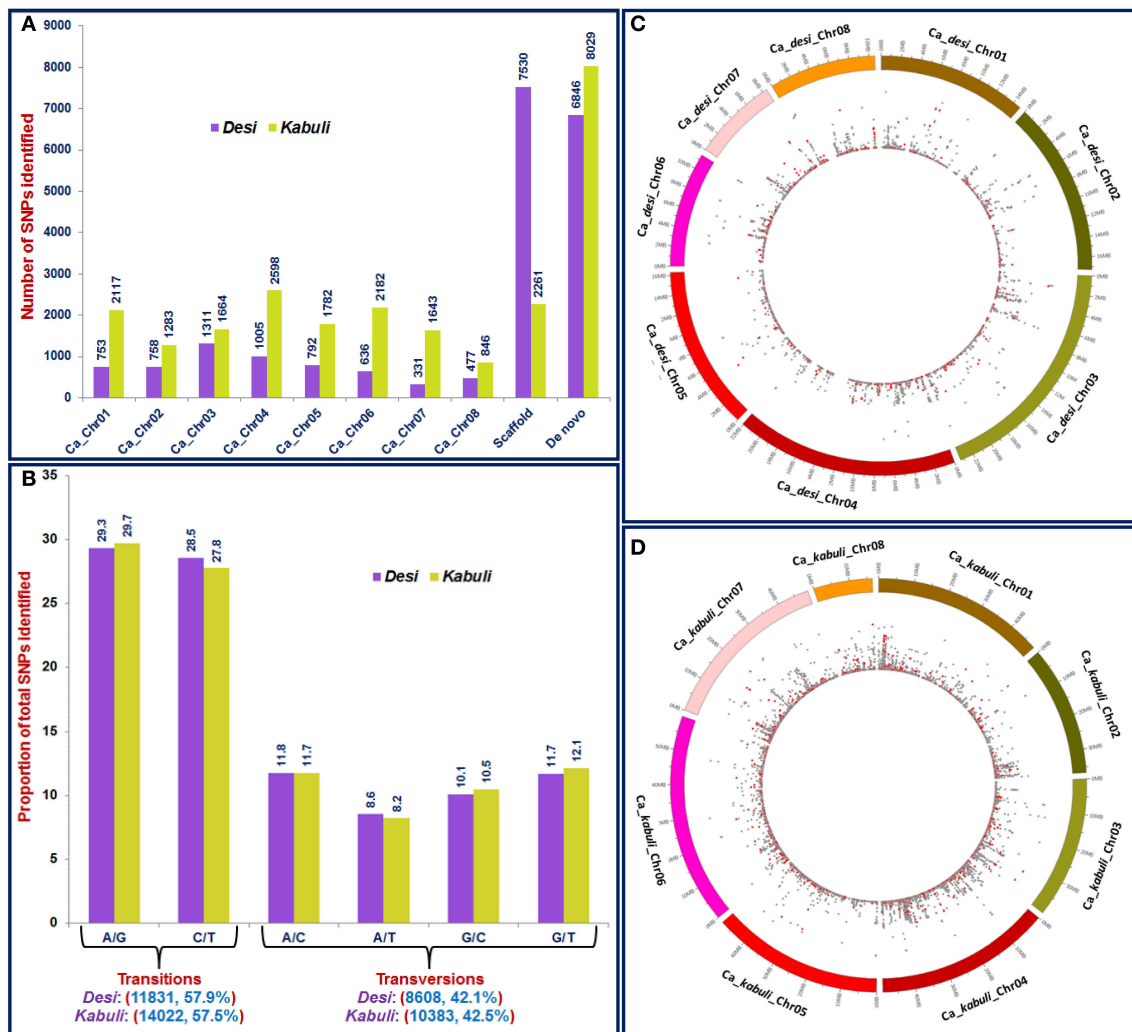
FIGURE 1 | (A) Frequency and distribution of 44,844 SNPs identified from *desi* and *kabuli* chickpea using reference genome (eight chromosomes and unanchored scaffolds)- and *de novo*-based GBS assays. **(B)** Percentage distribution of transition and transversion SNPs identified using GBS assay. The relative distribution of 20,178 SNPs physically mapped on eight chromosomes of *desi* **(C)** and *kabuli* **(D)** genomes are depicted in the Circos circular ideogram. The outermost circles represent the eight chickpea chromosomes coded with different colors, whereas the innermost circles display the distribution of SNPs, including non-synonymous SNPs (marked with red dots) identified through GBS assay.

zinc finger encoding *kabuli* transcription factor (TF) gene was found to be polymorphic between drought tolerant (ICC 4958 and ICCV 93954) and sensitive (ICCV 93954) chickpea accessions (Supplementary Figure S4). Similarly, 21 SNPs (including three non-synonymous SNPs and four upstream regulatory SNPs) detected in 16 genes showed polymorphism between *Fusarium* wilt resistant (ICC 8933, ICCV 92311, and ICC 12968) and susceptible (ICC 4951) chickpea accessions (Supplementary Tables S4, S5). A total of 242 SNPs (42 non-synonymous SNPs and 10 upstream regulatory SNPs) were detected in 156 genes showing polymorphism between the salinity stress tolerant (ICC 4951) and sensitive (ICC 12968) accessions (Supplementary Tables S4, S5). For example, SNP (A–C) showing missense non-synonymous amino acid substitution [asparagine (AAT) to histidine (CAT)] in the phosphotransferase domain encoding

methylthioribose kinase *desi* gene was found to be polymorphic between *Fusarium* wilt resistant (ICC 8933, ICCV 92311, and ICC 12968) and susceptible (ICC 4951) chickpea accessions. In addition, this non-synonymous SNP revealed polymorphism between salinity tolerant (ICC 4951) and sensitive (ICC 12968) chickpea accessions (**Figure 4**). Notably, SNPs identified in the coding regions of two stress-responsive PHD zinc finger TF and methylthioribose kinase genes had lower mean Ka/Ks = (0.39) (< 1.0) and reduced Tajima's $D = -2.66$ ($D < -2$) compared to that estimated in other genic SNPs (Ka/Ks: 0.84 and Tajima's D: −2.48).

In order to determine the significance of identified genome-wide GBS-based SNPs for subsequent fine-mapping/map-based cloning of QTLs/genes, these SNPs were annotated in the genes underlying the known QTLs reported earlier for drought
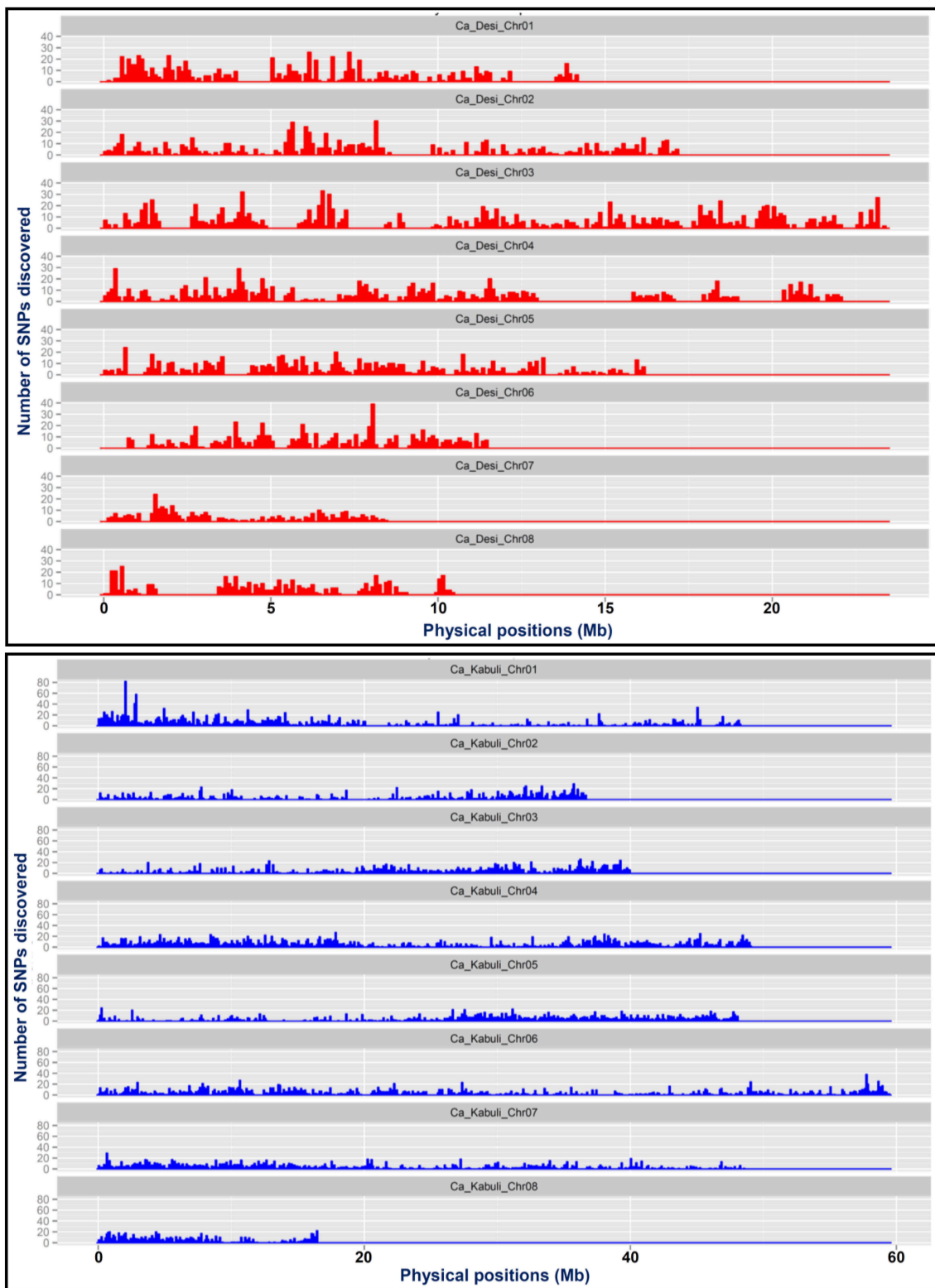
FIGURE 2 | A genome scan plot depicting the density of SNPs physically mapped across eight chromosomes of *desi* (A) and *kabuli* (B) genomes at 100-kb sliding window. The SNP density is represented by the number of SNPs mapped within 1-Mb physical intervals across chromosomes.
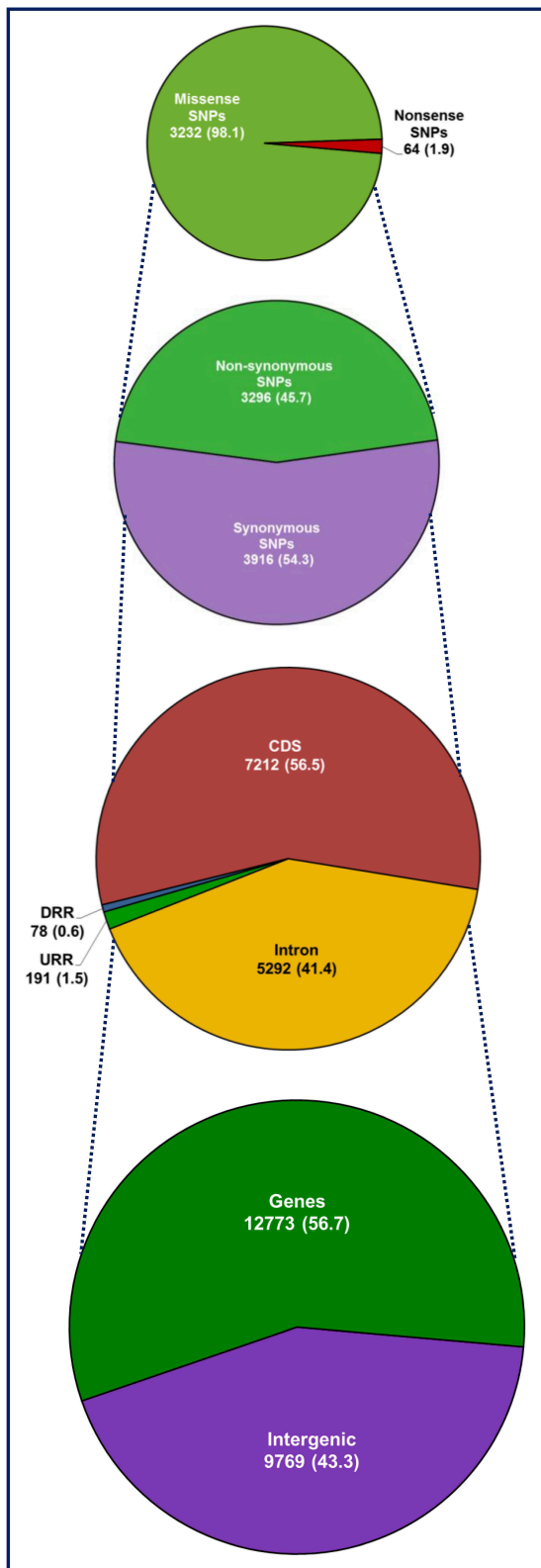
tolerance, and *Ascochyta* blight and *Fusarium* wilt resistance traits in chickpea (Udupa and Baum, 2003; Sabbavarapu et al., 2013; Varshney et al., 2013b, 2014a,b; Thudi et al., 2014). Based on these analyses, we identified 303 SNPs (78 non-synonymous SNPs) in the 140 genes localized within a 0.03 Mb (10.39–10.42 Mb) major "*QTL-hotspot*" region (mapped on *kabuli* chromosome 4) controlling drought tolerance-component root traits in chickpea (Supplementary Tables S4, S6). Interestingly, 18 SNPs-carrying 11 genes of these, showed polymorphism between drought tolerant (ICC 4958 and IC 296131) and sensitive (ICCV 93954) chickpea accessions. A total of 190 (59 non-synonymous SNPs) and 12 SNPs (1 non-synonymous SNPs) in 64 and 3 genes harboring the two major *Ascochyta* blight (*ar2aQTL*) and *Fusarium* wilt (*FW_Q_APR_6_1*) resistance QTLs mapped on *kabuli* chromosomes 2 (2.39 Mb: 31.58–33.97 Mb) and 6 (0.1 Mb: 29.47–29.57 Mb), respectively were discovered (Supplementary Tables S4, S6).

Validation and Polymorphic Potential of SNPs

For validating SNPs identified through GBS assay, 12,719 *desi* (ICC 4958) reference genome-based SNPs (with MAF $\geq$ 0.05) discovered in 93 accessions were compared and correlated with the available resequencing-based SNP database (31,019 SNPs) of four chickpea accessions (*desi*: ICC 4958, *desi*: ICC 4951, *kabuli*: ICC 12968, and wild *C. reticulatum*: ICC 17160). On the basis of congruent physical positions (bp), the correspondence of 8.8% (777 SNPs) of the SNPs present in our GBS data (8776 SNPs) was observed with that of SNPs mined by genome resequencing of the four accessions (Supplementary Table S7). Among these corresponding SNPs, the maximum percentage of SNPs (9.5%, 593 SNPs) showing polymorphism between ICC 4958 (*desi*) and ICC 17160 (wild) and minimum between ICC 4958 (*desi*) and ICC 4951 (*kabuli*) (7%, 86) were found to be common between past genome sequencing and our present GBS data (6244 SNPs) (Supplementary Table S7). Notably, the sensitivity and specificity of these *in silico* validated SNPs were 96.5% and 99.6%, respectively. Additionally, the PCR amplicon-based Sanger sequencing (214 SNPs) and MALDI-TOF mass array SNP genotyping (240 SNPs) of 454 selected non-synonymous and regulatory SNP loci discovered by reference genome- and *de novo*-based GBS approaches (Supplementary Table S8) successfully validated 418 (92.2%) SNPs in 93 *desi*, *kabuli,* and wild chickpea accessions. The sensitivity and specificity of these experimentally validated SNPs were estimated to be 93.7% and 99.8%, respectively. Multiple sequence alignment of amplicons resequenced from a representative set of *desi* and *kabuli* accessions exhibited the presence

FIGURE 3 | Relative frequency of 22,542 SNPs identified in the intergenic regions and different coding and non-coding sequence
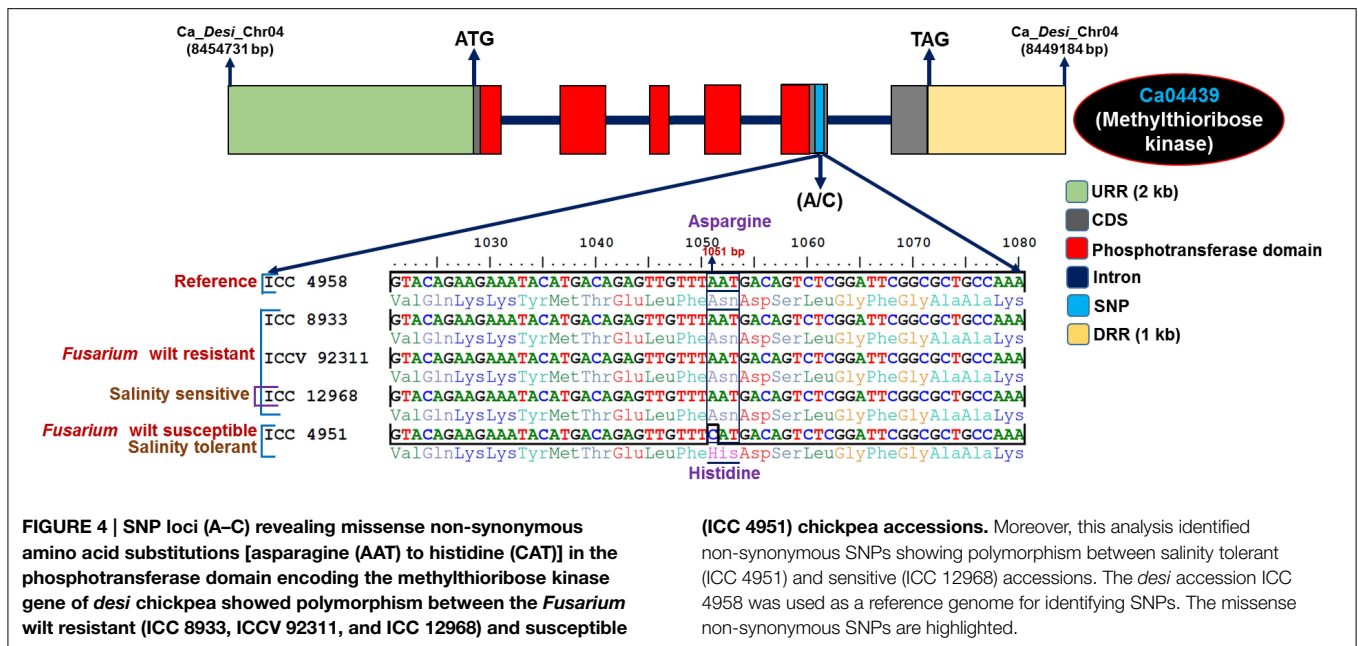
(Continued)

FIGURE 4 | SNP loci (A–C) revealing missense non-synonymous amino acid substitutions [asparagine (AAT) to histidine (CAT)] in the phosphotransferase domain encoding the methylthioribose kinase gene of *desi* chickpea showed polymorphism between the *Fusarium* wilt resistant (ICC 8933, ICCV 92311, and ICC 12968) and susceptible (ICC 4951) chickpea accessions. Moreover, this analysis identified non-synonymous SNPs showing polymorphism between salinity tolerant (ICC 4951) and sensitive (ICC 12968) accessions. The *desi* accession ICC 4958 was used as a reference genome for identifying SNPs. The missense non-synonymous SNPs are highlighted.

of similar expected SNP alleles in these accessions at specific genomic locations as discovered through our reference genome- and *de novo*-based GBS assay (Supplementary Figure S5). The MALDI-TOF mass array-based genotyping of SNPs in a selected set of *desi* and *kabuli* accessions further enabled to validate similar expected homozygous/heterozygous SNP alleles in these accessions at definite bp as detected through our GBS assay (Supplementary Figure S6).

In total, 23,798 SNPs (including 12,719 and 11,079 SNPs from *desi* and *kabuli*, respectively with MAF ≥ 0.05) showing polymorphism among 93 diverse *desi*, *kabuli*, and wild chickpea accessions had higher PIC (varied from 0.12 to 0.47, mean 0.42), MAF (mean: 0.37), and nucleotide diversity (mean θπ: 1.30 and mean θω: 1.51) (Supplementary Table S8, Supplementary Figure S7). Remarkably, 11,780 (49.5%) of these SNPs were polymorphic (PIC: 0.38, MAF: 0.34, θπ: 1.15, and θω: 1.26) between at least two accessions belonging to the cultivated *C. arietinum* species. The remaining 50.5% (12,018) of SNPs revealed polymorphism between cultivated *C. arietinum* and wild *C. reticulatum* species. The SNPs detected a higher intra-specific polymorphic potential within 39 *desi* accessions (42.1% polymorphism and mean PIC: 0.35, MAF: 0.31, θπ: 1.20, and θω: 1.33) compared with that of 53 *kabuli* accessions (37.6%, 0.30, 0.27, 0.96, and 1.07) (Supplementary Table S9). The nucleotide diversity estimated in eight *desi* chromosomes (mean θπ: 0.99, θω: 1.47, and Tajima's D: −2.31) was almost comparable to that of *kabuli* chromosomes (0.92, 1.65, and −2.39) (**Table 1**). Maximum nucleotide diversity was obtained in *desi* (mean θπ: 1.33 and θω: 1.69) and *kabuli* chromosomes 4 (1.31 and 1.72).

Molecular Diversity and Population Genetic Structure among Chickpea Accessions Revealed by Genome-Wide SNPs

The genetic distance estimation among 93 *desi*, *kabuli*, and wild chickpea accessions using 23,798 SNPs (with MAF ≥ 0.05)

revealed a wider range of genetic distances, from 0.13 (ICC 7346 and ICC 15512) to 0.89 (ICC 15264 and ICC 17160), with an average of 0.56 (Supplementary Table S9). The 11,780 SNPs-based genetic distances among 92 *desi* and *kabuli* accessions within the cultivated (*C. arietinum*) species varied from 0.12 to 0.76, with a mean of 0.45. A relatively less genetic diversity in 53 cultivated *kabuli* accessions (mean genetic distance: 0.38) as compared to the 39 *desi* accessions (0.43) was evident (Supplementary Table S9). A broader range of genetic distances (0.18–0.77, mean: 0.51) and thus, a higher functional allelic diversity level was observed among 93 accessions using 12,802 gene-derived SNPs (MAF ≥ 0.05). The unrooted neighbor-joining phylogenetic tree among 93 accessions using 23,978 genome-wide SNPs reflected a clear differentiation and clustering of accessions into two major population groups (*kabuli* and *desi*) in accordance with their pedigree relationships and parentage (**Figure 5**). One wild *C. reticulatum* accession ICC 17160 being divergent, got clearly separated and clustered distinctly within *desi* group.

The population genetic structure determination among 93 accessions with genotyping data of genome-wide (physically mapped on eight chickpea chromosomes) 96 SSR (simple sequence repeats selected from Kujur et al., 2013) and 23,798 SNP markers classified all the accessions into following three distinct populations at a K (population number) value of 3: POP I (44 *kabuli* accessions), POP II (1 wild, 4 *desi* and 4 *kabuli* accessions), and POP III (40 accessions) (**Figure 6**). The results of *K*-value with the best replicate were confirmed by accessing the average LnP(D) (log-likelihood) value (showing the greatest apparent inflection) (Supplementary Figure S8A) and using the second order statistics (Supplementary Figure S8B) of ΔK estimation (showing a sharp peak with maximum value of ΔK). The estimation of significance of SNPs to detect polymorphism among the three populations inferred their maximum polymorphic potential in POP II (PIC: 0.10–0.48 with mean: 0.40 and mean MAF: 0.36, θπ: 1.28, and θω: 1.47), followed by POP III

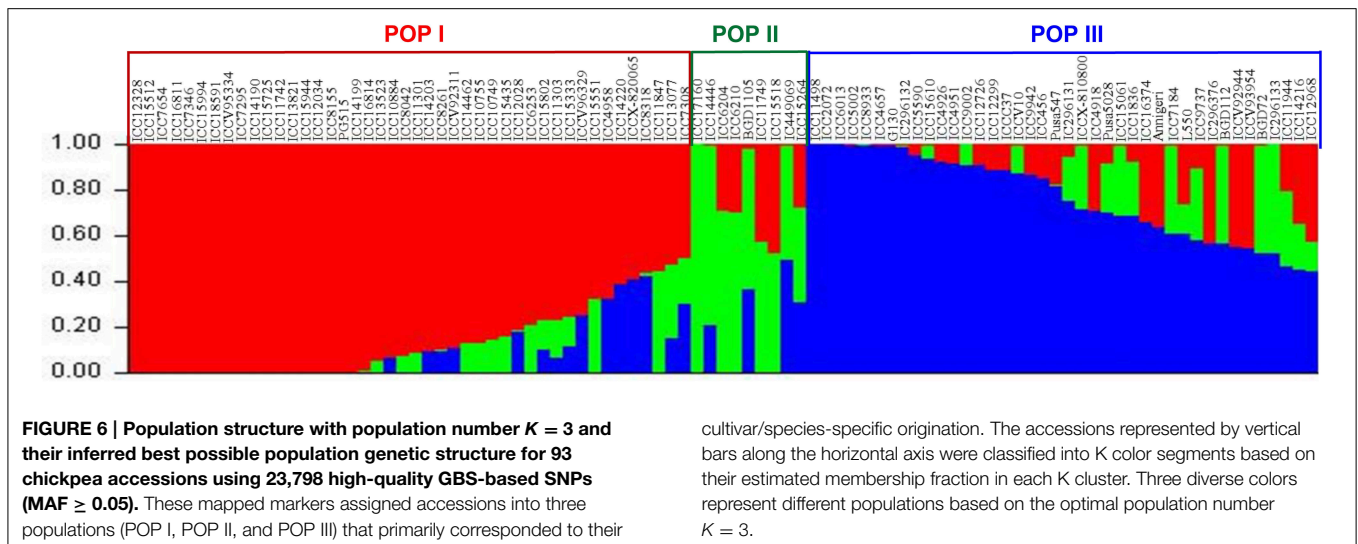**FIGURE 5 | Unrooted phylogenetic tree depicting the genetic relationships among 93 cultivated and wild chickpea accessions based on Nei's genetic distance using 23,798 high-quality GBS-based SNPs (MAF ≥ 0.05).** Molecular classification differentiated these accessions into two major clusters/population groups (*kabuli* and *desi*) as expected based on their cultivar/species-specific origination. The wild accession grouped distinctly within a POP III. Three different colors correspond to three populations (POP I, POP II, and POP III) as defined by population genetic structure.

and POP I (Supplementary Table S10). A wider level of significant quantitative genetic and population divergence based on pair-wise F_{ST} ($P < 0.001$) was observed between POP I and POP III (0.62), followed between POP II and POP III (0.43) and minimum between POP I and POP II (0.30). The F_{ST}-based population differentiation was highest within POP II (0.21), followed within POP III (0.16), whereas it was least within POP I (0.10).

All 93 accessions clearly belonged to a structured population with 78% inferred ancestry derived from one of the model-based population and the remaining 22% had admixed ancestry. Admixed ancestry was maximum in POP II (16%), followed by POP III (11%) and minimum in POP I (5%). Highest admixed ancestry was obtained between POP III and POP II (19%) compared to that between POP III and POP I (13%). The admixture was

FIGURE 6 | Population structure with population number K = 3 and their inferred best possible population genetic structure for 93 chickpea accessions using 23,798 high-quality GBS-based SNPs (MAF ≥ 0.05). These mapped markers assigned accessions into three populations (POP I, POP II, and POP III) that primarily corresponded to their cultivar/species-specific origination. The accessions represented by vertical bars along the horizontal axis were classified into K color segments based on their estimated membership fraction in each K cluster. Three diverse colors represent different populations based on the optimal population number K = 3.

TABLE 2 | LD estimates among linked, unlinked, and global SNP-pairs in three model-based individual populations and entire populations.

Populations	Global			Linked			Unlinked		
	Number of SNP-pairs (MAF ≥ 0.05) used	[a]Number (%) of significant SNP-pairs in LD	Extent of (mean r^2)	Number of SNP-pairs (MAF ≥ 0.05) used	Number (%) of significant marker-pairs in LD	Extent of (mean r^2)	Number of SNP-pairs (MAF ≥ 0.05) used	Number (%) of significant marker-pairs in LD	Extent of LD (mean r^2)
POP I	119,453	71,319 (59.7)	0.58	111,328	67,642 (60.8)	0.60	8125	3571 (44.0)	0.36
POP II	69,194	60,301 (87.1)	0.68	62,599	54,817 (87.6)	0.69	6594	5484 (83.2)	0.60
POP III	131,749	74,331 (56.4)	0.54	123,199	70,645 (57.3)	0.55	8550	3686 (43.1)	0.38
Total populations	320,396	205,951 (64.3)	0.60	297,126	193,104 (65.0)	0.63	23,269	12,741 (54.8)	0.45

[a] Percentage of SNP-pairs in significant (P < 0.0001) LD.

not detected in 9.7% (nine accessions) of chickpea accessions, whereas 37.6% (35) and 47.3% (44) of accessions showed up to 20 and 10% admixed ancestry, respectively.

Genome-Wide and Population-Specific LD Patterns in Chickpea

The LD estimates (average r^2) and extent of LD decay using all possible pair-combination of genome-wide 3321 and 8592 *desi* and *kabuli* SNPs (physically mapped on eight chromosomes) were calculated among 93 chickpea accessions constituting three populations (POP I, II, and III) and a diversity panel. In the whole population, 64.3% global marker-pairs exhibited significant LD (P < 0.0001), indicating the existence of an extreme LD level in a panel of 93 accessions (**Table 2**). The LD estimates of global markers ranged from 0.54 in POP III to 0.68 in POP II, with an average of 0.60. The proportion of significant LD and LD estimates for global, linked and unlinked markers was maximum in POP II compared to POP I and POP III (**Table 2**, **Figure 7**). The LD estimates of linked markers in all three populations was 0.63, whereas it was 0.45 for unlinked markers. Significant LD (P < 0.0001) was observed for 65 and 54.8% of the linked and unlinked marker-pairs, respectively (**Table 2**).

However, the trend of significant LD proportion and LD estimates with linked and unlinked marker-pairs was similar to that observed among three populations using global markers. The linked marker-pairs detected maximum percentage of LD estimates and significant LD proportion on *desi* chromosome 3 (67.3%) and *kabuli* chromosome 4 (58.4%) (Supplementary Table S11, Supplementary Figures S9A, B). The significant LD proportion and LD estimates of linked markers obtained in *desi* (56.9%, 0.56) and *kabuli* (54.6%, 0.52) chromosomes are comparable to each other, whereas these parameters for unlinked markers were higher in the chromosomes of *desi* (43%, 0.53) than *kabuli* (35.9%, 0.33) (Supplementary Table S11). The LD decay of genome-wide 3321 *desi* and 8592 *kabuli* reference genome-based SNPs was determined individually across three populations by pooling the r^2 estimates across eight chromosomes and plotting their average r^2 against the uniform physical distance of 0–100 kb (**Figures 8A,B**). In all three populations, a non-linear regression curve exhibiting a decreasing trend of LD decay with an increase in the physical distance (kb) was observed. All these populations sustained a significant level of LD up to a physical distance of 1000 kb. No significant decay of LD (below $r^2 = 0.1$) was observed in any population up to a 1000 kb physical distance
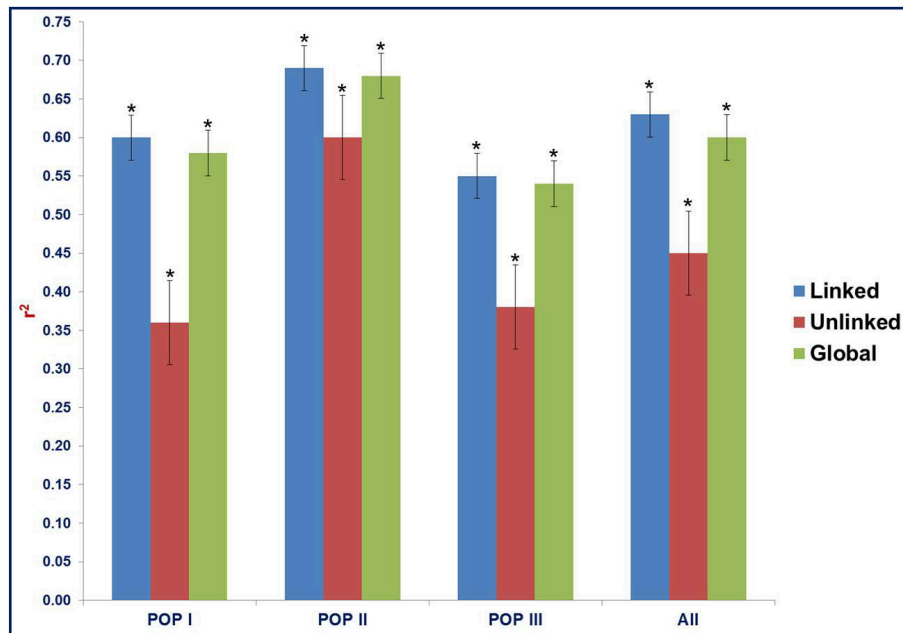
FIGURE 7 | Estimates of LD (mean r^2) for linked, unlinked, and global markers (3321 *desi* and 8592 *kabuli* SNPs) in three populations as defined by the population genetic structure. "All" includes the LD estimates and decay across three populations. The bar indicates the standard error. *ANOVA significance at $p < 0.001$.
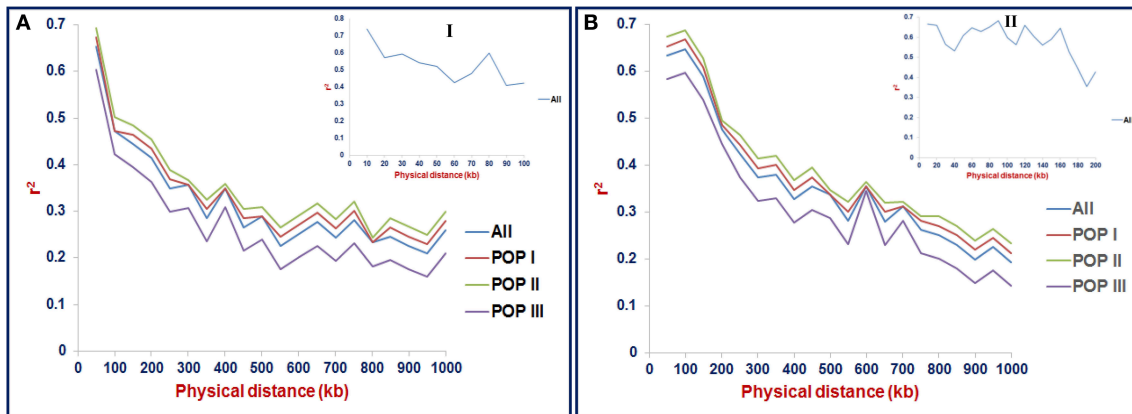

FIGURE 8 | LD decay (mean r^2) estimated in three populations (as defined by population genetic structure) using 3321 *desi* (A) and 8592 *kabuli* (B) SNPs, respectively. For LD decay, the r^2-value of the marker physical distance of 0 kb is considered 1. The plotted curved lines indicate the mean r^2-values among markers spaced with uniform 100 kb physical intervals from 0 to 1000 kb. The plotted line in uppermost panel **I** indicate the mean r^2-values among markers spaced with uniform 10 kb physical intervals from 0 to 100 kb. The marked line in panels **I, II** indicate the mean r^2-values among markers spaced with uniform 10 and 20 kb physical intervals from 0 to 100 kb and 0 to 200 kb, respectively. "All" includes the LD decay across entire three populations.

for the *desi* and *kabuli* chromosomes (**Figures 8A,B**). However, the r^2 decreased to half of its maximum value at ∼400–500 kb physical distance in both *desi* and *kabuli* chromosomes. The genetically more diverse population (POP III) showed a shorter LD decay (∼400 kb in *desi* chromosomes and ∼600 kb in *kabuli* chromosomes) than that of the less diverse populations of POP I (∼700 kb in both *desi* and *kabuli* chromosomes) and POP II (∼750 kb) (**Figures 8A,B**).

Discussion

GBS Assay Expedites Genome-Wide Discovery and High-Throughput Genotyping of SNPs in Chickpea

GBS assay is currently considered as the most convenient approach for high-throughput SNP discovery and genotyping in crop plants, including chickpea (Mayer et al., 2012; Poland et al.,

2012a,b; Byrne et al., 2013; Crossa et al., 2013; Mascher et al., 2013; Morris et al., 2013; Sonah et al., 2013; Spindel et al., 2013; Thurber et al., 2013; Uitdewilligen et al., 2013; Bastien et al., 2014; Deokar et al., 2014; He et al., 2014; Huang et al., 2014; Jaganathan et al., 2014; Jarquín et al., 2014; Liu et al., 2014; Sonah et al., 2014; Tardivel et al., 2014). Nevertheless, it also suffers from certain technical drawbacks, including non-uniform distribution of short sequence reads in different genotyped accessions and a high percentage of missing and erroneous (heterozygous) SNP genotyping data. This assay is also accompanied with difficulties in storing and handling much low quality SNP genotyping data generated from diverse crop accessions during bioinformatics analyses (Davey et al., 2011; Mir and Varshney, 2012; Poland and Rife, 2012). Recent efforts have been made to address these technical concerns involving generation of high-quality, uniform and accurate SNP genotyping information by physical mapping of short sequence reads (~64-bp tag) on available reference genomes, identifying bi-allelic tags through SNP imputation and use of suitable computational genomics softwares/pipelines (Beissinger et al., 2013). Moreover, the possibility of identifying such valid SNPs from large-scale GBS data can be enriched using the latest optimized de novo approaches, as suggested by several early reports in diverse crop plants, including chickpea (Craig et al., 2008; Cronn et al., 2008; Schnable et al., 2009; Poland et al., 2012a,b; Deokar et al., 2014; Jaganathan et al., 2014).

In the present study, an optimized GBS assay protocol was developed by sequencing of 96-plex ApeKI GBS libraries (generated from 96 diverse chickpea accessions) and integrating both reference genome- and de novo-based GBS pipelines for genome-wide discovery and genotyping of SNPs in chickpea. By implementing these novel approaches, many (44,844) non-erroneous and high-quality SNPs with wider genome coverage were identified in chickpea. In contrast to the conventional TASSEL GBS pipeline (maximum SNP detection limited to 64-bp sequence reads), the use of integrated reference genome (desi and kabuli)– and de novo-based GBS approaches would be of heightened interest for SNP discovery and genotyping. Consequently, our study for the first time identified SNPs in high-quality longer sequence reads (average sequence read length of > 75 bp after removal of barcode and adapter sequences) by involving a broader range of filtration tools (modifying stringency of SNP calling parameters), thereby enriching the chickpea genome with reduced computational time. The advantages of this integrated GBS approach for discovery and genotyping of SNPs at a genome-wide scale is well-documented in many crop plants, including soybean and rice (Sonah et al., 2013, 2014; Spindel et al., 2013). A high (100%) reproducibility of SNPs identified from the 93 accessions with our optimized integrated GBS assay was clearly evident. Their correspondence with an available SNP database in silico along with high experimental validation success rate (92%) through amplicon resequencing and MALDI-TOF mass array SNP genotyping assay based on strong sensitivity (93–96%) and specificity (99%) were also apparent. These findings reflect the major strength and reliability of GBS assay in rapid discovery and high-throughput genotyping of accurate SNPs in chickpea genome with sub-optimal use of resources. Henceforth, this highly versatile integrated GBS approach developed in our study could find

multidimensional applicability for genome-wide discovery and genotyping of SNPs in small as well as large genome plant species. The identification of a relatively higher frequency of SNPs showing transition substitutions (57%) than transversions is consistent with previous genome-wide SNP discovery studies in crop plants, including chickpea (Agarwal et al., 2012; Parida et al., 2012; Subbaiyan et al., 2012; Jain et al., 2013, 2014; Varshney et al., 2013a).

The physical mapping of a lesser number of SNPs on eight desi chromosomes (6063 SNPs) compared with kabuli chromosomes (14,115) could be attributed to the greater (~2.8 times) differences in pseudomolecule assemblies/lengths between desi (124.37 Mb) and kabuli (347.24 Mb) genomes (Jain et al., 2013; Varshney et al., 2013a). This infers a direct correlation between numbers of SNPs physically mapped on chromosomes and length of desi and kabuli chromosomes. However, the SNPs mined with GBS assay were represented from different portions of the desi and kabuli genomes, which could be the possible reason for significant differences in physical map density of SNPs between the two genomes. This in turn is due to uneven pseudomolecule assemblies with significant differences in genome coverage (kabuli: 39.4% and desi: 14.3%) resulting in sequencing of different regions of genomes between desi and kabuli chickpea (Ruperao et al., 2014). These clues indicate that large-scale mining and genotyping of SNPs at a genome-wide scale can be essentially enriched in chickpea by use of both desi and kabuli genomes as references along with comparison of their individual SNP outcomes from GBS assay. This strategy appears to be more relevant considering identification of more valid and accurate common as well as unique SNPs covering different fractions of desi and kabuli genomes, thereby accelerating the large-scale genome-wide SNP discovery in chickpea. The 20,178 SNPs-based physical maps composed of desi and kabuli chickpea chromosomes generated in our study can serve as references for a faster selection of genome-wide SNP markers for large-scale genotyping applications, including mapping of genomes and genes/QTLs controlling important agronomic traits and comparative genome mapping involving chickpea and legumes. Additionally, 24,666 high-quality and valid GBS-based SNPs (including 14,875 SNPs annotated on unanchored scaffolds and 9791 de novo SNPs) identified from desi and kabuli genomes could prove useful for various marker-based applications of chickpea genetics, genomics, and breeding.

Functional Significance of GBS-Based Genome-Wide SNPs in Chickpea

The identified SNPs were structurally annotated in different coding and non-coding (URR, intron, and DRR) sequence components of protein-coding genes individually predicted from desi (9567 SNPs in 4010 genes) and kabuli (10,656 SNPs in 4643 genes) genomes. Furthermore, the structural and functional annotation of 12,773 SNPs (submitted to NCBI dbSNP) detected in 5588 non-redundant chickpea genes and 3296 non-synonymous and 269 regulatory SNPs in 2225 genes provided a faster mode of selection of functional gene-based SNPs to establish marker-trait linkages and identify a number of genes/QTLs governing qualitative and quantitative agronomic

traits in chickpea. The estimated mean ratio of non-synonymous to synonymous SNPs (Ka/Ks: 0.84) in chickpea genes is comparable with that obtained (Ka/Ks < 1) from orthologous chickpea genes/transcripts (Agarwal et al., 2012; Jhanwar et al., 2012; Jain et al., 2013; Varshney et al., 2013a) and some selected stress-responsive rice genes (Parida et al., 2012). Consequently, this finding revealed the superior efficiency of purifying selection over non-synonymous sequence polymorphism in the scanned chickpea genes (Jhanwar et al., 2012; Parida et al., 2012; Jain et al., 2013). The purifying selection pressure on a larger proportion of SNPs-carrying genes is further evident from their reduced mean Tajima's D estimates ($D < -2$). The identification of non-synonymous and regulatory SNPs in protein-coding genes, (e.g., PHD zinc finger and methylthioribose kinase) showing polymorphism between abiotic (drought and salinity)/biotic (*Fusarium* wilt) stress tolerant/resistant and sensitive/susceptible accessions, have documented the evolutionary/adaptive advantages of these non-synonymous SNP loci (Wang et al., 1998; Parida et al., 2012) and probable functional significance of SNP-carrying genes for stress tolerance in chickpea. The adaptive significance of SNPs in two stress-responsive genes was well-supported by their estimated low mean Ka/Ks (0.39) and reduced Tajima's D ($D < -2$), and thus enduring more purifying selection compared with other genic SNPs. The differential expression and transcriptional regulation of plant homeodomain (PHD) zinc finger-encoding TF genes have confirmed its involvement in improving abiotic (drought) stress tolerance in plant species (Wei et al., 2009; Ray et al., 2011). The potential role of methylthioribose kinase-encoding genes in regulating abiotic and biotic stress responses through methionine homeostasis, ROS-mediated oxidation and polyamine biosynthesis has also been clearly demonstrated in crop plants, including rice and *Arabidopsis* (Sauter et al., 2004; Burstenbinder et al., 2007; Atkinson et al., 2013; Zagorchev et al., 2013). Henceforth, these informative SNPs identified, particularly in diverse coding and non-coding regulatory sequence components of various genes, once validated with a selected set of contrasting stress tolerant and susceptible accessions in a large-scale, can be utilized as potential markers in genetic and association mapping for identifying major trait-regulatory candidate genes/QTLs in chickpea.

Broader Functional Molecular Diversity, Admixed Domestication, and High-Resolution Complex LD Patterns in a Structured Chickpea Population

The intra (49.5% and PIC: 0.38)- and inter (50.5% and 0.40)-specific polymorphic potential detected by 23,798 SNPs among 93 cultivated and wild chickpea accessions was much higher than that estimated previously with SNP markers (Nayak et al., 2010; Gujaria et al., 2011; Roorkiwal et al., 2013; Varshney et al., 2013a). However, the nucleotide diversity measured by these SNPs in 39 *desi* (mean θ_π: 1.20 and θ_w: 1.33) and 53 *kabuli* (θ_π: 0.96 and θ_w: 1.07) accessions was comparable to that obtained using genome-wide SNP markers (Varshney et al., 2013a). A plenty of genome-wide genic SNP markers showing relatively high intra (*desi*: 42.1% and *kabuli*: 37.6%)–and inter-specific polymorphic potential and wider genetic diversity (13–89%, mean: 56%)/functional molecular diversity (18–77%, mean:

51%) among 93 accessions were identified. They could assist us in selecting desirable diverse accessions/inter-specific hybrids in cross-breeding program (introgression breeding) for chickpea varietal improvement.

The classification of 93 chickpea accessions into three major model-based genetically distinct populations; POP I, POP II, and POP III by wider population differentiation (F_{ST}: 0.30–0.62, mean: 0.42) was consistent with their expected pedigree relationships and parentage. A significant deviation from accurate population assignment based on pedigree relations was more pronounced in cultivated *desi* and *kabuli* accessions belonging to POP II. The geographical origin and adaptive environment, rather than parentage, of these accessions had a greater effect on their assignment to a specific population group. Higher (22%) admixed ancestry among accessions reflected their multiple domestication at archeological sites of south eastern Turkey ~10,000 years ago, followed by a complex breeding history involving inter-crossing/introgression coupled with strong adaptive selection pressure. The inclusion of diverse non-recurrent *desi* and *kabuli* germplasm lines as common parental lines in crop improvement breeding program to develop/breed most of the selected chickpea accessions in our study for valuable agronomic traits and higher yield might have influenced their population group assignment and domestication pattern, resulting in numerous admixtures among these accessions (Kujur et al., 2013; Varshney et al., 2013a). A more admixed ancestry between POP III and POP II compared with POP III and POP I is quite relevant because of complex domestication patterns during their evolutionary divergence from wild relatives.

A greater LD estimate (mean r^2: 0.63) in three populations of 93 accessions and furthermore, across eight *desi* (0.56) and *kabuli* (0.52) chromosomes using linked SNP-pairs compared with unlinked SNPs, reflected the direct correlation of LD patterning with the physical linkage of SNPs on chromosomes and the density required to cover genomic regions. The extensive LD estimates and extended chromosomal LD decay (~1000 kb) calculated in three populations of self-pollinated chickpea, were much higher than the cross-pollinated (Riedelsheimer et al., 2012; Sakiroglu et al., 2012; Xiao et al., 2012) and selfing (Mather et al., 2007; Atwell et al., 2010; Lam et al., 2010; Branca et al., 2011; Zhao et al., 2011) crop plants. It could be due to extensive contribution of selective sweeps and four sequential bottlenecks during chickpea domestication (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., 2013a; Saxena et al., 2014a), reducing its genetic base in contrast to other domesticated selfing plant species. Shorter LD decay (~400–600 kb) in a more diverse population (POP III) compared with POP I and POP II suggests the effect of factors other than marker density, including genetic diversity, population genetic structure, and demographic history on shaping the LD patterns in chickpea. The genome-wide and population-specific LD patterns generated in our study using 3321 *desi* and 8592 *kabuli* SNP genotyping information provided insight regarding the SNP marker density required for GWAS to identify potential genomic loci (gene-associated targets) regulating important agronomic traits in chickpea. Moreover, our studies regarding the molecular diversity,

population genetic structure and LD mapping among chickpea accessions (three population groups) using genome-wide SNPs could essentially identify trait-influencing target genomic regions (candidate genes) across chromosomes/genomes that possibly got influenced by strong selection pressure during domestication.

Conclusions

We identified 44,844 non-erroneous SNPs in chickpea at a genome-wide scale using reference genome- and *de novo*-based GBS assays. The structurally and functionally annotated SNPs in diverse coding and non-coding sequence components of genes could have potential to be utilized for various large-scale marker-based genotyping applications in chickpea (Supplementary Figure S10). High reproducibility (100%) and experimental validation success rate (92%) along with strong specificity (99%) and sensitivity (93–96%) of SNPs identified from 93 accessions using our optimized integrated GBS assay suggest the strength and reliability of GBS assay in rapid genome-wide discovery and high-throughput genotyping of accurate SNPs in chickpea. A wider natural allelic diversity, admixed domestication pattern and differential genome-wide and population-specific

LD estimates/decay assayed by mapped SNPs in a structured population of cultivated and wild accessions further implies their efficacy in genomics-assisted breeding applications of chickpea.

Acknowledgments

Supplementary Material

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2015.00162/abstract

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

Agarwal, G., Jhanwar, S., Priya, P., Singh, V. K., Saxena, M. S., Parida, S. K., et al. (2012). Comparative analysis of *kabuli* chickpea transcriptome with *desi* wild chickpea provides a rich resource for development of functional markers. *PLoS ONE* 7:e52443. doi: 10.1371/journal.pone.0052443

Atkinson, N. J., Lilley, C. J., and Urwin, P. E. (2013). Identification of genes involved in the response of *Arabidopsis* to simultaneous biotic and abiotic stresses. *Plant Physiol.* 162, 2028–2041. doi: 10.1104/pp.113.222372

Atwell, S., Huang, Y. S., Vilhjalmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800

Azam, S., Thakur, V., Ruperao, P., Shah, T., Balaji, J., Amindala, B., et al. (2012). Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results to identify SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. *Am. J. Bot.* 99, 186–192. doi: 10.3732/ajb.1100419

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376

Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Valè, G., Toppino, L., et al. (2011). Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics* 12:304. doi: 10.1186/1471-2164-12-304

Bastien, M., Sonah, H., and Belzile, F. (2014). Genome wide association mapping of *Sclerotinia sclerotiorum* resistance in soybean with a genotyping-by-sequencing approach. *Genome* 7, 1–13. doi: 10.3835/plantgenome2013.10.0030

Beissinger, T. M., Hirsch, C. N., Sekhon, R. S., Foerster, J. M., Johnson, J. M., Muttoni, G., et al. (2013). Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* 193, 1073–1081. doi: 10.1534/genetics.112.147710

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

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., et al. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 108, E864–E870. doi: 10.1073/pnas.1104032108

Brookes, A. J. (1999). The essence of SNPs. *Gene* 234, 177–186. doi: 10.1016/S0378-1119(99)00219-X

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

Burstenbinder, K., Rzewuski, G., Wirtz, M., Hell, R., and Sauter, M. (2007). The role of methionine recycling for ethylene synthesis in *Arabidopsis*. *Plant J.* 49, 238–249. doi: 10.1111/j.1365-313X.2006.02942.x

Byrne, S., Czaban, A., Studer, B., Panitz, F., Bendixen, C., and Asp, T. (2013). Genome wide allele frequency fingerprints (GWAFFs) of populations via genotyping by sequencing. *PLoS ONE* 8:e57438. doi: 10.1371/journal.pone.0057438

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

Cho, R. J., Mindrinos, M., Richards, D. R., Sapolsky, R. J., Anderson, M., Drenkard, E., et al. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* 23, 203–207.

Chutimanitsakun, Y., Nipper, R. W., Cuesta-Marcos, A., Cistue, L., Corey, A., Filichkina, T., et al. (2011). Construction and application for QTL analysis of a restriction site associated DNA (RAD) linkage map in barley. *BMC Genomics* 12:4. doi: 10.1186/1471-2164-12-4

Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., et al. (2008). Identification of genetic variants using bar-coded multiplexed sequencing. *Nat. Methods* 5, 887–893. doi: 10.1038/nmeth.1251

Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., and Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Res.* 36:e122. doi: 10.1093/nar/gkn502

Crossa, J., Beyene, Y., Kassa, S., Pérez, P., Hickey, J. M., Chen, C., et al. (2013). Genomic prediction in maize breeding populations with

genotyping-by-sequencing. *G3 (Bethesda)*. 3, 1903–1926. doi: 10.1534/g3.113.008227

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

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

Devisetty, U. K., Covington, M. F., Tat, A. V., Lekkala, S., and Maloof, J. N. (2014). Polymorphism identification and improved genome annotation of Brassica rapa through deep RNA sequencing. *G3 (Bethesda)*. 4, 2065–2078. doi: 10.1534/g3.114.012526

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

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8, 186–194. doi: 10.1101/gr.8.3.175

Garg, R., Patel, R. K., Jhanwar, S., Priya, P., Bhattacharjee, A., Yadav, G., et al. (2011). Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiol.* 156, 1661–1678. doi: 10.1104/pp.111.178616

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

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

Gupta, P. K., Roy, J. K., and Prasad, M. (2001). Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr. Sci.* 80, 524–535.

Gupta, P. K., and Rustgi, S. (2004). Molecular markers from the transcribed/expressed region of the genome in higher plants. *Funct. Integr. Genomics* 4, 139–162. doi: 10.1007/s10142-004-0107-0

He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H., and Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* 5:484. doi: 10.3389/fpls.2014.00484

Hiremath, P. J., Farmer, A., Cannon, S. B., Woodward, J., Kudapa, H., Tuteja, R., et al. (2011). Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. *Plant Biotechnol. J.* 9, 922–931. doi: 10.1111/j.1467-7652.2011.00625.x

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., and Cresko, W. A. (2010). Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genet.* 6:e1000862. doi: 10.1371/journal.pgen.1000862

Huang, B. E., Raghavan, C., Mauleon, R., Broman, K. W., and Leung, H. (2014). Efficient imputation of missing markers in low-coverage genotyping-by-sequencing data from multi-parental crosses. *Genetics* 197, 401–404. doi: 10.1534/genetics.113.158014

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108

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

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.* doi: 10.1007/s00438-014-0932-3. [Epub ahead of print].

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/tpj.12173

Jain, M., Moharana, K. C., Shankar, R., Kumari, R., and Garg, R. (2014). Genome-wide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnol. J.* 12, 253–264. doi: 10.1111/pbi.12133

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. doi: 10.1186/1471-2164-15-740

Jhanwar, S., Priya, P., Garg, R., Parida, S. K., Tyagi, A. K., and Jain, M. (2012). Transcriptome sequencing of wild chickpea as a rich resource for marker development. *Plant Biotechnol. J.* 10, 690–702. doi: 10.1111/j.1467-7652.2012.00712.x

Krzywinski, M., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

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

Kumar, J., 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

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

Li, Y., Haseneyer, G., Schön, C. C., Ankerst, D., Korzun, V., Wilde, P., et al. (2011). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biol.* 11:6. doi: 10.1186/1471-2229-11-6

Liu, H., Bayer, M., Druka, A., Russell, J. R., Hackett, C. A., Poland, J., et al. (2014). An evaluation of genotyping by sequencing (GBS) to map the *Breviaristatum-e* (*ari-e*) locus in cultivated barley. *BMC Genomics* 15:104. doi: 10.1186/1471-2164-15-104

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

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Mascher, M., Richmond, T. A., Gerhardt, D. J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J.* 76, 494–505. doi: 10.1111/tpj.12294

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

Mayer, K. F., Waugh, R., Brown, J. W., Schulman, A., Langridge, P., Platzer, M., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711–716. doi: 10.1038/nature11543

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res.* 17, 240–248. doi: 10.1101/gr.5681207

Mir, R. R., and Varshney, R. K. (2012). "Future prospects of molecular markers in plants," in *Molecular Markers in Plants*, ed R. J. Henry (Oxford, UK: Wiley-Blackwell Publishing Ltd), 169–190.

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of

agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110

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

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

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

Pfender, W. F., Saha, M. C., Johnson, E. A., and Slabaugh, M. B. (2011). Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor. Appl. Genet.* 122, 1467–1480. doi: 10.1007/s00122-011-1546-3

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012a). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253

Poland, J. A., and Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102. doi: 10.3835/plantgenome2012.05.0005

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012b). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi: 10.1016/S1369-5266(02)00240-6

Rajesh, P. N., and Muehlbauer, F. J. (2008). Discovery and detection of single nucleotide polymorphism (SNP) in coding and genomic sequences in chickpea (*Cicer arietinum* L.). *Euphytica* 162, 291–300. doi: 10.1007/s10681-008-9675-8

Ray, S., Dansana, P. K., Giri, J., Deveshwar, P., Arora, R., Agarwal, P., et al. (2011). Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. *Funct. Integr. Genomics* 11, 157–178. doi: 10.1007/s10142-010-0187-y

Riedelsheimer, C., Lisec, J., Czedik-Eysenberg, A., Sulpice, R., Flis, A., Grieder, C., et al. (2012). Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8872–8877. doi: 10.1073/pnas.1120813109

Roorkiwal, M., Sawargaonkar, S. L., Chitikineni, 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

Ruperao, P., Chan, C. K., Azam, S., Karafiátová, M., Hayashi, S., Cížková, J., et al. (2014). A chromosomal genomics approach to assess and validate the *desi* and *kabuli* draft chickpea genome assemblies. *Plant Biotechnol. J.* 12, 778–786. doi: 10.1111/pbi.12182

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

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933. doi: 10.1038/35057149

Sakiroglu, M., Sherman-Broyles, S., Story, A., Moore, K. J., Doyle, J. J., and Charles Brummer, E. (2012). Patterns of linkage disequilibrium and association mapping in diploid alfalfa (*M. sativa* L.). *Theor. Appl. Genet.* 125, 577–590. doi: 10.1007/s00122-012-1854-2

Sauter, M., Cornell, K. A., Beszteri, S., and Rzewuski, G. (2004). Functional analysis of methylthioribose kinase genes in plants. *Plant Physiol.* 136, 4061–4071. doi: 10.1104/pp.104.053587

Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014b). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031

Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014a). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Legare, 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'Donoughue, 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 soybean. *Plant Biotechnol. J.* 13, 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., Forster, 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

Subbaiyan, G. K., Waters, D. L., Katiyar, S. K., Sadananda, A. R., Vaddadi, S., and Henry, R. J. (2012). Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J.* 10, 623–634. doi: 10.1111/j.1467-7652.2011.00676.x

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

Tardivel, A., Sonah, H., Belzile, F., and O'Donoughue, L. S. (2014). Rapid identification of alleles at the soybean maturity gene E3 using genotyping by sequencing and a haplotype-based approach. *Plant Genome* 4, 106–111. doi: 10.3835/plantgenome2013.10.0034

Thudi, M., Li, Y., Jackson, S. A., May, G. D., and Varshney, R. K. (2012). Current state-of-art of sequencing technologies for plant genomics research. *Brief. Funct. Genomics* 11, 3–11. doi: 10.1093/bfgp/elr045

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

Thurber, C. S., Ma, J. M., Higgins, R. H., and Brown, P. J. (2013). Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biol.* 14:R68. doi: 10.1186/gb-2013-14-6-r68

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

Udupa, S. M., and Baum, M. (2003). Genetic dissection of pathotype-specific resistance to *Ascochyta* blight disease in chickpea (*Cicer arietinum* L.) using microsatellite markers. *Theor. Appl. Genet.* 106, 1196–1202. doi: 10.1007/s00122-002-1168-x

Uitdewilligen, J. G. M. L., Wolters, A. M. A., D'Hoop, B. B., Borm, T. J. A., Visser, R. G. F., and van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355. doi: 10.1371/journal.pone.0062355

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., 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., 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

van Orsouw, N. J., Hogers, R. C., Janssen, A., Yalcin, F., Snoeijers, S., Verstege, E., et al. (2007). Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2:e1172. doi: 10.1371/journal.pone.0001172

van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., et al. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat. Methods* 5, 247–252. doi: 10.1038/nmeth.1185

Varshney, R. K., Gaur, P. M., Chamarthi, S. K., Krishnamurthy, L., Tripathi, S., Kashiwagi, J., et al. (2013b). Fast-track introgression of "QTL-hotspot" for root traits and other drought tolerance trait in JG 11, an elite and leading variety of chickpea (*Cicer arietinum* L.). *Plant Genome* 6, 1–9. doi: 10.3835/plantgenome2013.07.0022

Varshney, R. K., Mohan, S. M., Gaur, P. M., Chamarthi, S. K., Singh, V. K., Srinivasan, S., et al. (2014b). Marker-assisted backcrossing to introgress resistance to *Fusarium* wilt race 1 and *Ascochyta* blight in C214, an elite cultivar of chickpea. *Plant Genome* 7, 1–11. doi: 10.3835/plantgenome2013.10.0035

Varshney, R. K., Nayak, S., Jayashree, B., Eshwar, K., Upadhyaya, H. D., and Hoisington, D. A. (2007). Development of cost-effective SNP assays for chickpea genome analysis and breeding. *J. Sat. Agri. Res.* 3, 29–31.

Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013a). 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. (2014a). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6

Wang, D. G., Fan, J.-B., Siao, C. J., Berno, A., Young, P., Sapolsky, R., et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280, 1077–1082. doi: 10.1126/science.280.5366.1077

Wei, W., Huang, J., Hao, Y. J., Zou, H. F., Wang, H. W., Zhao, J. Y., et al. (2009). Soybean *GmPHD*-type transcription regulators improve stress tolerance in transgenic *Arabidopsis* plants. *PLoS ONE* 4:e7209. doi: 10.1371/journal.pone.0007209

Xiao, Y., Cai, D., Yang, W., Ye, W., Younas, M., Wu, J., et al. (2012). Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. *Theor. Appl. Genet.* 125, 437–447. doi: 10.1007/s00122-012-1843-5

Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2012). 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

Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., et al. (2011). Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS ONE* 6:e17595. doi: 10.1371/annotation/f2eb75fb-ae22-4a57-b828-1506aa506c6d

Zagorchev, L., Seal, C. E., Kranner, I., and Odjakova, M. (2013). A central role for thiols in plant tolerance to abiotic stress. *Int. J. Mol. Sci.* 14, 7405–7432. doi: 10.3390/ijms14047405

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