Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea

Alice Kujur1*, Hari D. Upadhyaya2*, Tanima Shree1, Deepak Bajaj1, Shouvik Das1, Maneesha S. Saxena1, Saurabh Badoni1, Vinod Kumar3, Shailesh Tripathi4, C. L. L. Gowda2, Shivali Sharma2, Sube Singh2, Akhilesh K. Tyagi1 & Swarup K. Parida1

1National Institute of Plant Genome Research (NIPGR), Aruna Asaf Ali Marg, New Delhi 110067, India, 2International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Telangana, India, 3National Research Centre on Plant Biotechnology (NRCPB), New Delhi 110012, India, 4Division of Genetics, Indian Agricultural Research Institute (IARI), New Delhi 110012, India.

We discovered 26785 and 16573 high-quality SNPs differentiating two parental genotypes of a RIL mapping population using reference desi and kabuli genome-based GBS assay. Of these, 3625 and 2177 SNPs have been integrated into eight desi and kabuli chromosomes, respectively in order to construct ultra-high density (0.20–0.37 cM) intra-specific chickpea genetic linkage maps. One of these constructed high-resolution genetic map has potential to identify 33 major genomic regions harbouring 35 robust QTLs (PVE: 17.9–39.7%) associated with three agronomic traits, which were mapped within <1 cM mean marker intervals on desi chromosomes. The extended LD (linkage disequilibrium) decay (~15 cM) in chromosomes of genetic maps have encouraged us to use a rapid integrated approach (comparative QTL mapping, QTL-region specific haplotype/LD-based trait association analysis, expression profiling and gene haplotype-based association mapping) rather than a traditional QTL map-based cloning method to narrow-down one major seed weight (SW) robust QTL region. It delineated favourable natural allelic variants and superior haplotype-containing one seed-specific candidate embryo defective gene regulating SW in chickpea. The ultra-high-resolution genetic maps, QTLs/genes and alleles/haplotypes-related genomic information generated and integrated strategy for rapid QTL/gene identification developed have potential to expedite genomics-assisted breeding applications in crop plants, including chickpea for their genetic enhancement.

Chickpea (Cicer arietinum L.), the second most grown legume crop world-wide is an annual self-pollinated and diploid (2n = 2x = 16) pulse species with a genome size of ~740 Mb1,2. A member of Fabaceae family, this diploid food legume is a major source of human dietary protein packed with essential amino acids3. Chickpea is broadly categorized into two cultivars- Kabuli and Desi- types basing upon their plant characteristics and diverse gene pools-based geographical distribution. They are known to enrich the soil nutritional status and fertility by symbiotic nitrogen fixation. The seed and pod traits are considered the most prominent characteristics of chickpea defining its economic value as a diet for human race. Consequently, they draw major interest of researchers for yield and product quality improvement towards generation of high-yielding genetically tailored chickpea cultivars via genomics-assisted breeding4.

A large genome with a narrow genetic base, chickpea requires numerous informative and genome-wide well-distributed SNP (single nucleotide polymorphism) markers for construction of ultra-high density genetic linkage map (~1 cM average map-density5) in order to identify and fine-map the useful trait-influencing candidate genes/ QTLs (quantitative trait loci) for marker-assisted genetic enhancement. A number of high-resolution inter-specific genetic linkage maps (with mean map-density: 0.59–1.7 cM) have been generated in chickpea by genotyping of
large-scale SNP markers in its diverse mapping populations. Few recent efforts have also been made to construct SNP marker-based high-density genetic linkage maps (with map density: 1.74–3.68 cM) by utilizing diverse advanced generation desi and kabuli intra-specific mapping populations. Indeed, all these studies have utilized high-throughput genotyping assays like Illumina GoldenGate/Infinium and Competitive Allele Specific PCR (KASPar) for large-scale validation and genotyping of SNP markers (~1000 SNPs still now) in intra- and inter-specific chickpea mapping populations to construct high-density genetic linkage maps. Aside chickpea, similar efforts have also been undertaken to construct comprehensive high-density genetic linkage map (average inter-marker distance ≤1 cM) by deploying aforesaid strategies in diverse legume species (soybean, Medicago, pigeonpea, common bean and mung bean). However, the need of prior information regarding SNP loci and their high-quality flanking sequence regions for genotyping of SNPs largely limits the use of these assays in both discovery and validation of SNPs at genome-wide scale for constructing efficient high-resolution genetic linkage maps in chickpea. It thus implicates the necessity of a suitable modern and fast advanced approach/assay for simultaneous large-scale genome-wide discovery and high-throughput genotyping of SNPs in diverse mapping populations to generate required ultra-high resolution genetic linkage maps in chickpea. Numerous past reports have established the essentiality of high-density linkage maps as references for more effective assembly and orientation of the high-quality sequenced scaffolds into the pseudomolecules corresponding the chromosomes of crop plant genomes. Therefore, it is indispensable to generate ultra-high density genetic linkage maps enriched with a greater number of SNP markers for efficient anchoring of assembled scaffolds into the chickpea chromosomes. This is expected to overall greatly facilitate the chickpea genome assembly and sequencing efforts in near future.

Genotyping-By-Sequencing (GBS) is a simple, rapid and cost-effective next-generation sequencing (NGS)-based assay having potential for simultaneous large-scale discovery and genotyping of SNPs in diverse crop plants at a genome-wide scale. The high-barcoded multiplexing capacity (pooling up to 384 barcoded accessions in a single sequencing lane) at methylation-sensitive RE (restriction endonuclease)-sites (like ApeKI) have enhanced the efficiency of GBS assay in fast discovery and genotyping of genome-wide SNPs in larger mapping populations. This in turn made the GBS assay much more advantageous over other available high-throughput traditional genotyping assays for constructing ultra-high density genetic linkage maps and genome-wide high-resolution (major and minor effects QTLs) gene/QTL mapping in small diploid as well as large genome crop species, including rice, wheat, barley, sorghum, Medicago and soybean. Considering its utility in many high-throughput marker-based genotyping applications in genomics-assisted breeding, the use of GBS assay for generating ultra-high resolution genome maps and molecular mapping of robust major as well as minor genes/QTLs with sub-optimal use of resources in a crop species like chickpea assumes significance. Nevertheless, very limited reports are available with regard to the use of GBS approach in discovery and genotyping of genome-wide SNPs in intra- and inter-specific mapping populations for constructing ultra-high resolution genetic linkage maps and fine mapping of QTLs regulating important agronomic traits in chickpea vis-à-vis other crop plants.

Over the past few years, a number of QTLs associated with important agronomic traits, including yield component and stress tolerance traits have been identified and mapped on low/high-resolution intra- as well as inter-specific genetic linkage maps of chickpea. However, very limited number of these QTLs could either be successfully validated in diverse genetic backgrounds of accessions across geographical locations/years or fine-mapped through map-based cloning for their utilization in marker-assisted genetic improvement of chickpea. To expedite such process of identification and mapping of high-resolution and robust major and minor effect QTLs at genome-wide scale, the construction of SNP marker-based ultra-high density intra- and inter-specific genetic linkage maps using GBS assay at present is anticipated as an attractive approach for chickpea genetic enhancement.

Keeping all above in view, we discovered 26785 and 16573 genome-wide SNPs showing differentiation between parental genotypes [ICC 12299 (desi) and ICC 8261 (kabuli)] by their genotyping in a 275 F2 RIL (recombinant inbred line) mapping population (ICC 12299 × ICC 8261) using reference desi and kabuli genome-based GBS assay, respectively. A selected 3625 and 2177 high-quality SNPs of these, were integrated into eight desi and kabuli chromosomes, respectively to construct ultra-high density intra-specific genetic linkage maps in chickpea. These high resolution genetic maps were further used as a reference to identify and map the novel major genomic regions underlying robust QTLs associated with three important agronomic traits in chickpea. To delineate functionally relevant candidate gene(s) regulating seed weight, one of the strong trait-associated major genomic region harbouring robust QT has been narrowed down by integrating comparative desi and kabuli genome-based QTL mapping with QTL region-specific haplotype (linkage disequilibrium)-based high-resolution trait association analysis, differential gene expression profiling and gene haplotype-based association mapping in chickpea.

Results

Discovery and large-scale validation of GBS-based genome-wide SNPs. We generated an average of 207.3 million high-quality sequence reads (ranged from 1.9 to 8.3 with a mean of 3.4 million reads per individual) (Fig. S1) by sequencing of each 3 × 96-plex (288-plex) ApeKI libraries made from 275 RIL mapping individuals and two parental genotypes (ICC 12299 × ICC 8261) using GBS assay. However, 84.5 and 82.6% of sequence reads of these, were mapped to unique physical location of desi and kabuli reference genomes, respectively. It identified 43358, including 26785 and 16573 high-quality SNPs [with read-depth ≥10, SNP base quality ≥20, <10% missing data in each individual, 5% minor allele frequency (MAF) and 100% reproducibility] differentiating two parental genotypes using the reference desi and kabuli genome-based GBS assay, respectively (Fig. 1A). Of these, 22331, including 7968 and 14363 SNPs were physically mapped on eight chromosomes of desi and kabuli genomes with average map-densities of 15.6 and 24.2 kb, respectively (Table S1, Fig. 1A, B, C). A higher proportion of SNPs were mapped on desi and kabuli chromosomes 3 (19.4%, 1548 SNPs) and 4 (19.9%, 2854), respectively. The des1 chromosomes 7 and 8 (13.9 kb each) and kabuli chromosome 4 (17.2 kb) had maximum mean map-density (Table S1). The remaining 21027 SNPs, including 18817 and 2210 SNPs were physically mapped on scaffolds of desi and kabuli genomes, respectively (Fig. 1A). The A/G and C/T transitions (62.7 and 55.1% of total SNPs identified in desi and kabuli, respectively) were more abundant compared with that of transversions (Fig. 1D). Detailed characteristics of 43358 SNPs discovered in our study are provided in the Table S2. Of these, 16573 SNPs discovered by kabuli genome-based GBS assay have been submitted to NCBI dbSNP (http://www.ncbi.nlm.nih.gov/SNP) with SNP submission (SS) accession numbers 1067289484 to 1067306056 for unrestricted use (NCBI db SNP Build (B142) year 2014 (http://www.ncbi.nlm.nih.gov/SNP/snp_view/Table.cgi?handle=NIPGR)).

Four hundred-eighty SNPs mined by reference desi and kabuli genome-based GBS assay were targeted for their validation through resequencing of PCR amplons (384 SNPs) and MALDI-TOF mass array (96 SNPs). It validated 437 (91%) SNPs successfully in parental genotypes and selected mapping individuals of a RIL population (ICC 12299 × ICC 8261) (Fig. S2A, B). Remarkably, all these validated SNPs revealed expected homozygous and heterozygous alleles discrimination in RILs as detected through our GBS assay.
Construction of ultra-high density intra-specific genetic linkage maps. The genotyping information of 4448 and 2689 high-quality SNPs (MAF ≥ 20%) discovered by reference desi and kabuli genome-based GBS assay, respectively (Table S3) in parental genotypes and 275 individuals of a RIL mapping population (ICC 12299 × ICC 8261) were used to construct ultra-high density intra-specific genetic linkage maps in chickpea. The linkage analysis of SNPs showing significant Mendelian segregation ratio (1:5:1) mapped 3625 and 2177 SNPs on eight desi and kabuli linkage groups (LGs) (LG1 to LG8) of intra-specific genetic maps, respectively according to their physical positions (bp) on corresponding chromosomes (Table 1, Fig. S3, S4, S5). Considering significant variations in the genome assembly and differences in sequence coverage and length (bp) of chromosomal pseudomolecules between desi and kabuli genomes78, we preferred to construct two genetic maps separately by using desi and kabuli genome-based high-quality GBS-SNPs and their corresponding physical positions on eight chromosomes as reference. The constructed desi and kabuli SNP-based ultra-high density intra-specific genetic maps comprising of eight LGs covered a total map length of 714.089 and 798.467 cM with mean inter-marker distances of 0.20 and 0.37 cM, respectively (Table 1). Maximum numbers of SNPs were mapped on desi LG1 (736 SNPs) and kabuli LG3 (463), while desi LG8 (281) and kabuli LG5 (104) contained minimum numbers of SNPs (Table 1). The map length spanned by each LG of desi ranged from 85.662 cM (386 SNPs) in LG2 to 97.782 cM (574) in LG4, while for kabuli, it varied from 81.444 cM (292 SNPs) in LG5 to 120.359 cM (386) in LG2 (Table 1). In desi, most saturated genetic map was LG1 (average inter-SNP marker distance: 0.12 cM), whereas LG5 and LG8 were least saturated (0.32 cM). In kabuli, LG3 (mean inter-SNP marker distance: 0.18 cM) and LG5 (0.78 cM) had most and least saturated genetic maps, respectively (Table 1).

Estimation of genome-wide LD patterns in chickpea. The genome-wide LD estimates (r²) and extent of LD decay using all possible pair-combinations of 3625 and 2177 SNPs genetically mapped on eight desi and kabuli LGs of intra-specific genetic maps, respectively were determined. The average LD estimates in eight LGs of desi genome

Figure 1 | (A) Relative distribution and frequency of 43358 SNPs discovered in eight chromosomes and scaffolds of desi and kabuli chickpea using reference genome-based GBS assay. Distribution of 22331 SNPs physically mapped on eight desi (B) and kabuli (C) chromosomes are depicted in the Circos circular ideogram. The innermost circles display the distribution of SNPs identified through GBS assay, while outermost circles represent the eight chickpea chromosomes coded with different colours. (D) Proportionate distribution of transition and transversion SNPs identified using the reference desi and kabuli genomes-based GBS assays.
coefficient of variation (CV) ranged from 0.14 for PN to 0.33 in SW (Table S5). We observed bi-directional trans-

significant LD decay (r^2) sustained a significant level of LD up to a genetic distance of 10 cM. A decreasing trend of LD decay (r^2 < 0.3) was observed with increase in the genetic distance (cM) of SNP markers mapped on the desi and kabuli LGs (Fig. S6). Remarkably, a rapid LD decay was observed at the genetic distance of 5 cM in both desi and kabuli genomes. The desi and kabuli LGs overall sustained a significant level of LD up to a genetic distance of 10 cM. A significant LD decay (r^2 < 0.1) was observed near about 15 cM genetic distance (Fig. S6) in LGs of these two genomes.

Identification and mapping of high-resolution trait-governing QTLs. A significant difference of three quantitative agronomic traits, namely pod number/plant (PN) (51.1–141.2 with 79–80% heritability), seed number/plant (SN) (60.6–231.6 with 77–81%) and 100-seed weight (7.1–43.7 g with 88–90%, Fig. S7) in 275 RIL mapping individuals and parental genotypes across three years based on ANOVA was observed (Table S5). We observed bi-directional transgressive segregation of these traits beyond that of parental genotypes in RILs (Fig. S8). The normal frequency distribution of all three quantitative traits in RILs and parental genotypes was evident. The coefficient of variation (CV) ranged from 0.14 for PN to 0.33 in SW (Table S5). The Pearson’s correlation coefficient estimated among three pair-wise combinations of quantitative traits indicated significant (P < 10^-4) negative correlation of PN and SN with SW (r = -0.42 to −0.51) as well as positive correlation between PN and SN (0.95) (Fig. S9).

For genetic/QTL mapping, the genotyping data of 3625 parental polymorphic desi-genome based SNPs mapped on an ultra-high density intra-specific genetic linkage map was integrated with phenotyping data of 275 RIL mapping population. It identified and mapped 33 major genomic regions underlying 34 robust QTLs showing consistent allelic contributions from ICC 8261. Collectively, all these QTLs showed 36.4% PVE. The PN QTL regions covered with 57 SNP markers on LGs (0.015 cm on LG4 to 1.978 cm on LG1) showed mostly the positive additive gene effects indicating the effective contributions of ICC 12299 alleles at these loci for increasing pod number in chickpea. For SN, eight major genomic regions harbouring eight robust QTLs explaining 21.7–38.1% PVE (LOD: 7.44–12.6) were identified. These SN QTLs covered with 24 SNPs (0.072 cm on LG2 to 1.083 cm on LG5) were mapped on four LGs (Table 2 and Fig. S10). The combined PVE estimated for all these QTLs was 35.8%. The QTLs showed positive additive gene effects for increasing seed number with large allelic contribution from ICC 12299. For SW, nine major genomic regions harbouring 10 robust QTLs covered with 27 SNP markers (0.115 cm on LG1 to 1.07 cm on LG5) were mapped on six LGs explaining 23.2–39.7% PVE (LOD: 7.4–14.5) (Table 2 and Fig. S10). The combined PVE for all these QTLs was 39.4%. These QTLs showed positive additive gene effects for increasing seed weight with allelic contributions from ICC 8261.

Structural and functional annotation of SNPs mapped on intra-specific genetic linkage maps. The structural annotation of 4448 and 2689, including 3625 and 2177 high-quality desi and kabuli GBS-SNPs mapped on intra-specific genetic maps revealed the presence of 3017 and 1349 SNPs in the intergenic regions, while 1431 and 1340 SNPs in the different sequence components of 635 and 760 genes, respectively (Table S6A and B). The detailed annotation of SNPs within genes identified maximum frequency (35.4%, 507 SNPs in desi and 49.4%, 662 SNPs in kabuli) of SNPs in the exons and minimum in the 1000-bp downstream regulatory regions (DRR) (27.9%, 399 and 17.7%, 238) (Fig. 2A and S12A). Remarkably, 278 and 229 coding SNPs-containing desi genes, while 352 and 328 coding SNPs-carrying kabuli genes showed synonymous and non-synonymous (missense and nonsense) substitutions, respectively (Fig. 2B and S12B). The non-synonymous SNPs identified in desi and kabuli comprised of 233 and 323 missense SNPs in the 149 and 234 genes showing amino acid substitutions as well as five and nonsense SNPs in the six and five genes, respectively culminating into premature termination codons introduced by nucleotide replacements (Fig. 2B and S12B). Maximum (5.3 SNPs/gene) and minimum (1.6 SNPs/gene) average frequency of SNPs was observed in the non-coding intronic and coding sequence components of desi genes, respectively. In contrast, we observed comparable average SNP frequency between intronic (1.7 SNPs/gene) and coding (1.9) sequence components of kabuli genes.

The functional annotation of 635 desi and 760 kabuli genes with SNPs detected a higher proportion of SNPs belonging to growth, development and metabolism-related proteins (58 and 63% in desi and kabuli, respectively) followed by transcription factors (16 and 18%) (Fig. S13A). The KOG-based determination of putative functional categories (excluding unknown and general functions) for SNPs-carrying genes revealed their maximum correspondence to transcription (K,

| Linkage groups (LGs) | SNPs mapped | Map length covered (cM) | Average inter-marker distance (cM) |
|----------------------|-------------|-------------------------|-----------------------------------|
|                      | Desi      | Kabuli     | Desi     | Kabuli     | Desi     | Kabuli     |
| CalLG01              | 736       | 343        | 86.076   | 91.704     | 0.12     | 0.27       |
| CalLG02              | 386       | 199        | 85.662   | 120.359    | 0.22     | 0.60       |
| CalLG03              | 587       | 463        | 85.993   | 82.124     | 0.15     | 0.18       |
| CalLG04              | 574       | 399        | 97.782   | 105.127    | 0.17     | 0.26       |
| CalLG05              | 292       | 104        | 94.069   | 81.444     | 0.32     | 0.78       |
| CalLG06              | 430       | 285        | 87.601   | 110.009    | 0.20     | 0.38       |
| CalLG07              | 339       | 194        | 87.821   | 105.184    | 0.26     | 0.54       |
| CalLG08              | 281       | 180        | 89.085   | 102.516    | 0.29     | 0.54       |
| **Total**            | **3625**  | **2177**   | **714.089** | **798.467** | **0.20** | **0.37**   |
| Linkage groups [LGs] | QTL identity | LGs | Marker intervals with genetic positions (cM) | Peak LOD score | Phenotypic variance explained (PVE R²%) by QTLs | Traits associated | Additive effect |
|---------------------|--------------|-----|---------------------------------------------|----------------|----------------------------------------------|------------------|----------------|
| Ca_Desi_LG01        | qSW1.1       | LG_1 | Ca_Desi_SNP_1 (0) to Ca_Desi_SNP_2 (0.146)  | 13.5           | 29.8                                         | SW               | 4.5            |
|                     | qPN1.1       | LG_1 | Ca_Desi_SNP_1.5 (5.54) to Ca_Desi_SNP_1.7 (7.518) | 8.2            | 22.4                                         | PN               | 3.2            |
|                     | qPN1.2       | LG_1 | Ca_Desi_SNP_40 (13.733) to Ca_Desi_SNP_42 (14.48) | 8.31           | 22.8                                         | PN               | 2.8            |
|                     | qPN1.3       | LG_1 | Ca_Desi_SNP_250 (34.293) to Ca_Desi_SNP_252 (34.453) | 8.11           | 22                                           | PN               | 1.9            |
|                     | qSN1.1       | LG_1 | Ca_Desi_SNP_486 (54.514) to Ca_Desi_SNP_488 (54.597) | 10.95          | 29.6                                         | SN               | 2.7            |
|                     | qSN1.2       | LG_1 | Ca_Desi_SNP_513 (56.043) to Ca_Desi_SNP_515 (56.158) | 13             | 28.2                                         | SW               | 3.2            |
|                     | qPN1.4       | LG_1 | Ca_Desi_SNP_591 (61.419) to Ca_Desi_SNP_595 (61.738) | 14             | 35                                           | PN               | 5.1            |
|                     | qSW1.2       | LG_1 | Ca_Desi_SNP_636 (64.659) to Ca_Desi_SNP_638 (64.817) | 7.39           | 23.2                                         | SW               | 4.5            |
|                     | qPN1.5       | LG_1 | Ca_Desi_SNP_655 (67.104) to Ca_Desi_SNP_657 (67.22) | 7.25           | 22.6                                         | PN               | 3.8            |
| Ca_Desi_LG02        | qSN2.1       | LG_2 | Ca_Desi_SNP_752 (9.224) to Ca_Desi_SNP_754 (10.101) | 8.91           | 25.4                                         | SN               | 2.5            |
|                     | qSN2.2       | LG_2 | Ca_Desi_SNP_902 (39.18) to Ca_Desi_SNP_904 (39.252) | 8.42           | 23.3                                         | SN               | 2.8            |
|                     | qPN2.1       | LG_2 | Ca_Desi_SNP_946 (48.4) to Ca_Desi_SNP_948 (48.982) | 10.56          | 28                                           | PN               | 3.4            |
|                     | qPN2.2       | LG_2 | Ca_Desi_SNP_1_1012 (58.324) to Ca_Desi_SNP_1_1014 (58.377) | 8.12           | 22                                           | PN               | 4.7            |
| Ca_Desi_LG04        | qSN4.1       | LG_4 | Ca_Desi_SNP_1_1113 (80.057) to Ca_Desi_SNP_1_1115 (81.06) | 10.72          | 32.5                                         | SN               | 4.2            |
|                     | qPN4.2       | LG_4 | Ca_Desi_SNP_2011 (51.285) to Ca_Desi_SNP_2013 (51.308) | 8.32           | 22                                           | PN               | 3.4            |
|                     | qPN4.3       | LG_4 | Ca_Desi_SNP_2023 (51.702) to Ca_Desi_SNP_2025 (51.717) | 9.74           | 24.7                                         | PN               | 4.3            |
| Ca_Desi_LG07        | qSN4.1       | LG_7 | Ca_Desi_SNP_2088 (66.304) to Ca_Desi_SNP_2090 (66.442) | 9.02           | 21.6                                         | SW               | 2.6            |
|                     | qSN5.1       | LG_5 | Ca_Desi_SNP_2336 (23.792) to Ca_Desi_SNP_2338 (24.875) | 12.6           | 38.1                                         | SN               | 6.7            |
|                     | qSN5.2       | LG_5 | Ca_Desi_SNP_2557 (76.802) to Ca_Desi_SNP_2559 (77.188) | 10.5           | 24.5                                         | SN               | 1.9            |
| Ca_Desi_LG08        | qSN6.1       | LG_6 | Ca_Desi_SNP_2608 (10.981) to Ca_Desi_SNP_2610 (11.552) | 7.8            | 17.9                                         | SW               | 2.6            |
|                     | qSN6.2       | LG_6 | Ca_Desi_SNP_2902 (56.733) to Ca_Desi_SNP_2904 (57.25) | 8.37           | 21.8                                         | SW               | 3.5            |
|                     | qPN6.1       | LG_6 | Ca_Desi_SNP_2903 (56.755) to Ca_Desi_SNP_2907 (57.422) | 7.86           | 20.7                                         | PN               | 3.8            |
|                     | qPN6.2       | LG_6 | Ca_Desi_SNP_2931 (62.038) to Ca_Desi_SNP_2933 (62.381) | 11.62          | 25.7                                         | SW               | 4.6            |
| Ca_Desi_LG06        | qPN6.1       | LG_6 | Ca_Desi_SNP_2993 (77.157) to Ca_Desi_SNP_2995 (78.929) | 9.22           | 26.4                                         | PN               | 5.2            |
|                     | qSN7.1       | LG_7 | Ca_Desi_SNP_3006 (0) to Ca_Desi_SNP_3007 (1.014) | 8.28           | 22.7                                         | SW               | 3.6            |
|                     | qPN7.1       | LG_7 | Ca_Desi_SNP_3089 (24.832) to Ca_Desi_SNP_3091 (24.848) | 8.35           | 22.5                                         | PN               | 4.3            |
|                     | qPN7.2       | LG_7 | Ca_Desi_SNP_3157 (34.458) to Ca_Desi_SNP_3159 (34.858) | 8.28           | 22.1                                         | PN               | 3.9            |
|                     | qPN7.3       | LG_7 | Ca_Desi_SNP_3230 (44.792) to Ca_Desi_SNP_3232 (45.282) | 9.65           | 24.3                                         | PN               | 4.5            |
|                     | qPN7.4       | LG_7 | Ca_Desi_SNP_3269 (54.679) to Ca_Desi_SNP_3271 (54.792) | 10.45          | 27.6                                         | PN               | 5.1            |
| Ca_Desi_LG08        | qSN8.1       | LG_8 | Ca_Desi_SNP_3387 (19.508) to Ca_Desi_SNP_3389 (19.65) | 9.06           | 21.7                                         | SW               | 3.9            |
|                     | qSN8.2       | LG_8 | Ca_Desi_SNP_3394 (20.582) to Ca_Desi_SNP_3396 (21.005) | 7.44           | 21.7                                         | SN               | 2.8            |
|                     | qPN8.1       | LG_8 | Ca_Desi_SNP_3356 (54.055) to Ca_Desi_SNP_3358 (54.218) | 8.45           | 27.6                                         | SN               | 4.6            |
25% in desi and 31% in kabuli (Fig. S13B). The GO enrichment analysis of SNPs-carrying desi and kabuli genes showed a significant (P ≤ 5.9 × 10^{-3}) overrepresentation/enrichment of GO terms in the genes associated with transport (15.4%) belonging to biological process (Fig. S13C).

**QTL region-specific haplotype (LD)-based trait association mapping.** One major and robust (LOD: 14.5 and R²: 39.7%) QTL (qSW5.1) region [Ca_Des1_SNP2337 (23.805 cM)-Ca_Des1_SNP2338 (24.875 cM)] governing SW mapped on desi LG5 of intra-specific genetic map was selected for fine mapping through QTL target-specific haplotype (LD)-based high-resolution trait association mapping (Fig. S10 and Fig. 3A). The genomic region underlying such a major QTL was defined by integrating the genetic linkage map information with that of physical map of desi genome. The target QTL interval genetically mapped on desi LG5 was compared and correlated with that mapped on kabuli LGs of a genetic map (based on QTL mapping). In kabuli, the SW trait-influencing QTL region spanned with 1.6 cM genetic interval between the markers Ca_Kabuli_SNP1459 (34.6 cM) and Ca_Kabuli_SNP1462 (36.2 cM) was mapped on LG5 (Fig. 3B). The integration of genetic linkage map information with physical map of kabuli genome, the 1.6 cM QTL interval corresponded to 713 kb genomic region (40020.4–40733.4 kb) on chromosome 5 (Fig. 3C).

The targeted resequencing of this 713 kb qSW5.1 QTL region in parental genotypes (ICC 12299 and ICC 8261) and 10 of each homozygous low and high seed weight RIL mapping individuals detected 804 high-quality SNPs (mean SNP density: 1/886.8 bp). It includes 435 intergenic SNPs and 369 SNPs derived from the diverse coding and non-coding sequence components of 64 genes. The decay of LD over longer genetic distance (~15 cM, Fig. S6) particularly at the target SW QTL interval mapped on kabuli LG5 gave us clue to narrow-down this QTL region of interest using SNP haplotype (LD)-based high-resolution trait association mapping. The genotyping of 804 SNPs identified and mapped on the 713 kb qSW5.1 QTL region in 244 accessions belonging to an SW-specific association panel (Table S7) enabled to constitute 12 haplotypes (Fig. 3D). The SNP haplotype-based genotyping information were further correlated with their SW-specific robust field phenotyping data (SW: 5.9–57.6 g) for QTL region-specific trait association analysis. The integration of GLM and MLM analysis with EMMA and false discovery rate (FDR) correction based on multiple-comparisons (min-
imizing the confounding effect of population structure) identified one best haplotype (H1) between two SNPs (CaSNP1 and CaSNP2) in two genes (Ca07590 and Ca07594) showing strong association (P: 2.4 × 10⁻⁴ and R²: 37.6–40.3%) with SW in contrast to any other SNP combinations. This haplotype region covered a maximum of 22.5 kb (40710.6–40733.1 kb) physical distance between CaSNP1 and CaSNP2 at qSW5.1 QTL interval. The structural and functional annotation of this 22.5 kb sequenced short QTL region with kabuli genome annotation database identified five protein-coding candidate genes (Fig. 3D). The comprehensive analysis of strong SW-associated haplotypes-constituting two genic SNPs detected one regulatory SNP (CaSNP2: C/T) in the URR (upstream regulatory region) of an embryo defective protein-coding gene showing strong association (P: 2.0 × 10⁻⁵ and R²: 39.8%) with SW.

Expression profiling of SW-associated genes. To infer the differential regulatory gene expression patterns, the expression profiling of five protein-coding candidate genes annotated in the 22.5 kb delineated SW-specific robust QTL region with kabuli genome annotation database identified five protein-coding candidate genes (Fig. 3D). The comprehensive analysis of strong SW-associated haplotypes-constituting two genic SNPs detected one regulatory SNP (CaSNP2: C/T) in the URR (upstream regulatory region) of an embryo defective protein-coding gene showing strong association (P: 2.0 × 10⁻⁵ and R²: 39.8%) with SW.

Haplotype-based association mapping in a strong SW-associated gene. The sequencing of 9286 bp cloned amplicon covering the entire coding and non-coding regions (including 2 kb URR) of a strong SW-influencing embryo defective gene (Fig. 4A) (validated by QTL mapping, QTL region-specific association analysis and differential expression profiling) among 244 accessions belonging to a SW-specific association panel identified 12 SNP loci (Fig. 4B). It includes four regulatory SNPs in the URR of this gene. The gene-based haplotype analysis combining the genotyping data of 12 SNPs constituted a maximum of four haplotypes among accessions (Fig. 4B).

The association analysis using the four SNP marker-based haplotypes in an embryo defective gene with SW (5.9 to 57.6 g)-specific multi-location replicated field phenotyping data of 244 accessions (association panel) revealed its strong association with SW (P: 2.5 × 10⁻⁷ and R²: 44.7%). The haplotype-pair-based LD estimation produced a significant high degree of LD (r² > 0.70 and P < 0.0001) across entire 9286 bp sequenced gene (Fig. 4C), which increased its overall potential for trait association. The 94 high SW (40–57.6 g) accessions represented by single haplotype group 1 (TAC) in the gene differentiated distinctly from another haplotype group 2 (>7-fold) in parental genotypes during seed development (Fig. S14). Henceforth, this gene localized in a major SW-influencing QTL interval (qSW5.1) was selected as target candidate for understanding its significance in seed weight regulation through high-resolution gene haplotype-specific association/LD mapping in chickpea.
SCIENTIFIC REPORTS

of studies in the past have accessed the potential of diverse NGS-MALDI-TOF mass array genotyping assay was evident. A number total of 43358 GBS-based SNPs by amplicon resequencing and validation success rate (91%) of 480 randomly selected SNPs from a densities of 1/15.6 and 1/24.2 kb, respectively. A high experimental coverage across eight desi including 7968 and 14363 SNPs of these, showed extensive genome population with 100% reproducibility, respectively. Notably, 22331, two parental genotypes (ICC 12299 and ICC 8261) of a RIL mapping 26785 and 16573 high-quality SNPs (5% MAF) differentiating the desi The reference

Discussion

We utilized NGS-based 3 × 96-plex GBS assay for large-scale validation and high-throughput genotyping of genome-wide SNPs in a 275 RIL mapping population (ICC 12299 × ICC 8261) to construct ultra-high density intra-specific genetic linkage maps in chickpea. The reference desi and kabuli genome-based GBS assay identified 26785 and 16573 high-quality SNPs (5% MAF) differentiating the two parental genotypes (ICC 12299 and ICC 8261) of a RIL mapping population with 100% reproducibility, respectively. Notably, 22331, including 7968 and 14363 SNPs of these, showed extensive genome coverage across eight desi and kabuli chromosomes with mean densities of 1/15.6 and 1/24.2 kb, respectively. A high experimental validation success rate (91%) of 480 randomly selected SNPs from a total of 43358 GBS-based SNPs by amplicon resequencing and MALDI-TOF mass array genotyping assay was evident. A number of studies in the past have accessed the potential of diverse NGS-based assay for efficient mining and genotyping of non-erroneous SNPs through experimental validation of randomly chosen smaller set (~1–2%) of SNPs from a whole genome SNP dataset in diverse crop plants, including chickpea\textsuperscript{22,40–50}. The correspondence of on an average of 39% (16910 SNPs) desi and kabuli genome-derived GBS-SNPs with that of SNP allelic information available in various chickpea genotypes\textsuperscript{22,40–42} in silico based on their congruent physical positions was clearly evident. These findings overall suggest the robustness and utility of GBS assay for rapid large-scale mining and high-throughput genotyping of valid and high-quality SNPs at a genome-wide scale in chickpea. The advantages of GBS-assay in simultaneous genome-wide discovery and genotyping of SNPs and their ability to expedite various large-scale genotyping applications have been demonstrated in many crop plants\textsuperscript{22,40–50}. Collectively, a large number of 43358 high-quality reference desi and kabuli genome-based GBS-SNPs (61% novel SNPs) developed by us (submitted to NCBI SNPdb for unrestricted use) could serve as a valuable genomic resource for their immense use in genomics-assisted breeding applications of chickpea.

An ultra-high density 3625 and 2177 SNP marker-based intra-specific genetic linkage maps (ICC 12299 × ICC 8261) with eight desi (mean inter-marker distance 0.20 cM) and kabuli (0.37 cM) LGs constructed in our study are highly saturated compared to the previously reported intra- and inter-specific genetic linkage maps (0.50–8.01 cM) in chickpea\textsuperscript{11,12,22,23,31}. Remarkably, the currently generated genetic linkage maps had much lower marker map-density and thus highly saturated in contrast to latest available high-density intra-specific genetic linkage maps (0.50–3.68 cM) in chickpea\textsuperscript{11,14,22,23,31}. Beside chickpea, the efficacy of different high-throughput genome-wide SNP discovery and genotyping approaches, including GBS assay
for construction of ultra-high density genetic linkage maps have been well demonstrated in several other crop plants\textsuperscript{13–25,45}. The ultra-high density genetic linkage maps have been successfully deployed as a reference for effective assembly, precise orientation and anchoring of high-quality scaffolds into pseudomolecules of the chromosomes arising out of whole genome sequencing efforts (finished and/or in progress) in many plant species\textsuperscript{13,23,25,35–38}. Therefore, ultra-high density intra-specific genetic linkage maps constructed in our study can serve as a reference for efficient generation of integrated genetic, physical and genome map, including identification and fine-mapping of major as well as minor QTLs regulating important agronomic traits in chickpea.

A significant LD estimates across eight desi (0.62) and kabuli (0.59) LGs of constructed intra-specific genetic maps using mapped SNP markers reflected direct correlation of LD patterns with the density of SNPs required to cover the genomes. In spite of large differences in SNP marker map density between desi (0.20 cM) and kabuli (0.37) LGs of genetic maps, we observed higher LD estimates and extended LD decay (~15 cM) in LGs of both desi and kabuli genomes. This is much higher than the LD decay estimated in other self- and cross-pollinated crop plants\textsuperscript{83,90}. It could be due to sequential bottlenecks during chickpea domestication resulting in its narrow genetic base in contrast to other crop plants\textsuperscript{36,37,74,91}.

Overall, the genome-wide LD patterns and constructed LD maps gave an approximate perception regarding the SNP density required to identify the functionally relevant trait-regulatory robust genes/ QTLs in a large chickpea genome with narrow genetic base using genetic (fine-mapping and map-based cloning) and association mapping.

Our study utilized ‘\(z\)-lattice’ designs in multi-location experimental field trials to efficiently grow and comprehensively phenotype the 275 RIL mapping individuals (ICC 12299 × ICC 8261) along with their parental genotypes for three important yield component traits (PN, SN and SW) in chickpea. The ‘\(z\)-lattice’ designs are more preferred than the commonly adopted ‘randomized complete block design (RCBD)’, since they are less prone to experimental errors, cost-effective and provide more precise (~35% higher) phenotyping information in field experiments aside high reproducibility.

The ‘\(z\)-lattice’ designs are also associated with smaller coefficient of variation and error mean square, which collectively gives them an edge over the traditionally used RCBD designs specifically when phenotyping of yield component trait variables was conducted for very large numbers of diverse crop genotypes in field experiments with a smaller plot size\textsuperscript{94,96}.

Thirty-three major genomic regions harbouring 35 robust QTLs (LOD: 7.25–14.5 and PVE: 17.9–39.7%) associated with three important agronomic traits (SW, PN and SN) were identified and mapped on seven desi LGs of an ultra-high density genetic map, which were well validated in multiple geographical locations/environments across years in field. The reliability and validity of the identified PN, SN and SW QTLs were determined by correlating/comparing their underlying genomic regions with that of earlier QTL mapping studies\textsuperscript{9,52,69,90,97–104} based on correspondence of physical positions of markers covering the target QTL intervals. This analysis identified one of each PN (qPN8.1) and SW (qSW4.1) QTLs in line with previous QTL mapping studies using diverse intra- and inter-specific mapping populations of chickpea\textsuperscript{14,20,22,51,72,76}. This analysis indicated novelty and population-specific characteristics of our identified 35 robust QTLs for SW, PN and SN traits in chickpea. Several QTL mapping studies in crop plants, including legumes have effectively utilized ultra-high density genetic linkage maps derived from different intra- and inter-specific mapping populations as foundation for rapid molecular mapping as well as fine mapping/map-based cloning of major genes harbouring robust QTLs governing multiple yield contributing traits (like pod and seed number and seed weight) to accelerate genomics-assisted crop improvement\textsuperscript{20,21,48,102–105}. The use of ultra-high density genetic linkage maps as a reference in the present study, significantly narrowed down the marker intervals in the QTL regions to 0.015–1.978 cM (mean < 1 cM), which is lower than that reported (1–3 cM) so far in diverse QTL mapping studies of chickpea\textsuperscript{4,12–14,20,22,51,72–74}. It further authenticates the use of SNP markers that are tightly linked to the QTLs for subsequent validation in diverse genetic backgrounds along with fine mapping/map-based cloning of QTLs controlling three agronomic traits in chickpea. Moreover, SNPs localized in the QTL region mapped on eight desi and kabuli LGs, including SW, PN and SN QTL intervals of ultra-high density intra-specific genetic maps will enable us to select functional gene-based SNPs (like non-synonymous coding and regulatory SNPs) for establishing marker-trait linkages and identifying genes/QTLs associated with important qualitative and quantitative traits through genetic and association mapping in chickpea.

One strong SW-associated major genomic region ( LOD: 14.5, PVE: 38.7%) harbouring a robust QTL (qSW3.1) mapped on desi LG3 of an intra-specific genetic map was compared/correlated with that of integrated genetic and physical maps of kabuli to fine map the major QTL region. The extended LD decay (~15 cM) in desi and kabuli LGs of genetic maps suggested that high-resolution haplotype (LD)-based trait association mapping at target QTL interval could be an attractive alternative approach for narrowing-down the QTL regions to candidate gene(s) regulating three said agronomic traits in chickpea. Henceforth, we combined haplotype (LD)-based high-resolution trait association mapping at target SW-specific QTL region with differential gene expression profiling (during seed development in low and high seed weight parental genotypes), which delineated one regulatory SNP-containing embryo defective protein-coding gene governing seed weight in chickpea. Such haplotype-specific association analysis in presence of high-resolution significant LD at major QTL region have utility to scale-down the QTLs to specific candidate gene(s) regulating important agronomic traits in crop plants\textsuperscript{96–109}.

The strong trait association potential of an embryo defective gene with SW has been ascertained by higher contribution of a significant superior haplotype (TAC) identified in the URR of this gene with 44.7% SW phenotypic variation in a constituted 244 association panel. This was further supported by the seed-specific expression and significant down-regulated expression of gene haplotype transcript during seed development in contrasting SW parental genotypes and RIL mapping individuals. The favourable natural allelic variants and optimal superior haplotype identified in an embryo defective gene will be helpful in dissection of gene regulatory networks underlying this complex quantitative SW trait in chickpea. This identified gene could eventually be used as a potential candidate in marker-assisted genetic enhancement of chickpea for increasing seed weight and higher yield. The embryo defective chickpea gene ortholog in Arabidopsis is known to control the normal embryo development, which implies the role of this vital gene in natural growth and seed development of crop plants\textsuperscript{109}. Considering the availability of very limited information regarding embryo defective gene in plants, a further comprehensive molecular characterization and functional validation of this gene in over-expression/knock-down transgenics is required to ascertain its definite involvement in regulating seed development and seed weight of chickpea.
Methods

Generation of an intra-specific chickpea mapping population. An intra-specific F1 RIL mapping population (ICC 12299 × ICC 8261) comprising of 275 RILs derived from the crosses between two parental chickpea genotypes ICC 12299 and ICC 8261 was generated using a single seed descent method. The chickpea genotype ICC 12299 originated from Nepal is a low 100-seed weight (9 g) and high pod (103) and seed (158) number-containing desi (C. arietinum) landrace. In contrast, the genotype ICC 8261 originated from Turkey gives a high 100-seed weight (30 g) and low pod (64) and seed (73) number-containing kabuli (C. arietinum) landrace.

Phenotyping of mapping population for agronomic traits. The RIL mapping population and their parental genotypes were grown in the field according to a split-plot design with each RIL replicated three times. The agronomic traits, namely PN, SN and SW. The PN and SN was estimated by counting the average number of fully formed pods and seeds per plant (from 10–15 representative plants from each RILs at maturity), respectively. The SW was measured by taking the average weight (g) of 100-matured seeds at 10% moisture content by selecting 10–15 representative plants from each RILs. The mean, standard deviation, coefficient of variation, broad-sense heritability, frequency distribution, correlation coefficient and analysis of variance (ANOVA) of three agronomic traits in RIL mapping population were analyzed using SPSS v17.0 and following the methods of Kujur et al.14 and Saxena et al.17.

Genome-wide discovery and genotyping of SNPs using GBS assay. The genomic DNA isolated from 275 RILs, including two parental accessions (used as biological replicates) were digested with Apel1 and the digestion products were pooled together following the detailed procedures of Elshire et al.18 and Spindel et al.19. All these libraries were sequenced by Illumina TrueSeq V3.0 single end sequencing chemistry with read lengths of 100-bp using HiSeq2000 (Illumina Inc., San Diego, CA, USA) NGS platform. The FASTQ sequence reads were prepared for high-quality sequence filtering and mapping through TASSEL-GBS. The sequence reads with a minimum phred Q-score of 10 across the first 72-bp nucleotide sequences were considered as high-quality, which were further rechecked for their quality using FASTQC v0.10.1. The retained good quality sequences were de-multiplexed using unique barcodes attached to each mapping individual. The sequence reads of each individuals were aligned and mapped to reference draft desi (ICC 495820) and kabuli (CDC Frontier)21 chickpea genome sequences separately using Burrows-Wheeler alignment tool (BWA) with default parameters. The missing SNP allele data in individuals were imputed using the haplotype probability method of fastPHASE. A maximum likelihood statistical model and a catalog was used to identify and genotype valid and high-quality SNPs (minimum sequence read depth: 10 with SNP base quality ≥20) in 275 RIL mapping individuals along with two parental genotypes.

Large-scale validation of SNPs. To validate the genome-wide SNPs identified through GBS assay, the primers designed from the 200-bp sequences flanking either side of 384 selected SNPs were PCR amplified using the genomic DNA of two parental chickpea genotypes and selected RIL mapping individuals. The amplified PCR products were resequenced, aligned the high-quality sequences and the SNPs were detected among individuals following the methods22,23. A selected 96 SNPs identified by reference genome-based GBS assay was PCR amplified in parental genotypes and representative mapping individuals using MALDI-TOF (matrix-assisted laser desorption ionization-time of flight) mass array SNP genotyping assay (http://www.sequenom.com) following the methods of Pandit et al.24 and Saxena et al.25.

Construction of intra-specific genetic linkage maps. The genotyping data of reference desi and kabuli genome-based high-quality GBS-SNP genotyping data (MAF ≥ 0.2) showing differentiation between parental genotypes (ICC 12299 and ICC 8261) were analysed in 275 RIL mapping individuals using the χ2 test (p ≤ 0.05). The SNP genotyping data showing goodness-of-fit to the expected Mendelian 1:1 segregation ratio were further used for linkage analysis using MAPMAKER/EXP3.0 and JoinMap 4.1 at higher LOD (logarithm of odds) threshold (>10.0) with Kosambi mapping function. The SNPs were integrated into eight desi and kabuli LGs of intra-specific genetic maps based on their centimorgan (cM) genetic distance using the methods of Kujur et al.14,15 and Saxena et al.17. The marker ordering was performed using the RECORD algorithm implemented in RECORD WIN. The best marker ordered genetic linkage map with shortest map distance (cM) was preferred. The LGs with genetically mapped SNPs were designated and numbered (LG1 to LG8) based on their corresponding physical positions (bp) on the chromosomes of chickpea genomes as determined in our study.

Determination of LD patterns. To determine genome-wide LD patterns in chickpea, the LD estimates (significant P-value ≤ 10−8) as average squared-allele frequency correlations (r2) among a pairs of SNP loci that are genetically mapped on eight desi and kabuli LGs of intra-specific genetic maps were analysed using the sliding window approach of TASSEL v5.0 to assess the decay of LD across the genome. The clonality LD was calculated by combining the r2 values of SNP-pairs mapped in a uniform genetic intervals of 5 cM (maximum up to 120 cM) across eight desi and kabuli LGs of genetic maps. The graph was plotted between pooled r2 and genetic distance (cM) based on nonlinear regression model considering the r2 value = 1 at marker genetic distance of 0 cM74,94,101,105 to determine the trend of LD decay in desi and kabuli genomes.

Structural and functional annotation of genome-wide SNPs. The reference desi and kabuli genome-based SNPs mapped on the eight LGs of intra-specific genetic maps were annotated in diverse intragenic and intergenic sequence components of genomes (chromosomes/pseudomolecules) based on gene functional information (GFF) of desi26 and kabuli27. The structural and functional annotation of SNPs, including synonymous and non-synonymous coding and regulatory SNPs were performed using the customized perl scripts and single-nucleotide polymorphism effect predictor (SNPEff v3.1h). The SNPs were plotted individually based on their unique physical positions (bp) on either desi or kabuli (pseudomolecules) of desi and kabuli genomes using Circos v0.67-pre3. The functional annotation of SNP-containing genes was performed according to desi and kabuli genome annotation and PFAM database v27.0. The genes with SNPs were BLAST searched against the KOGnter NCBI database. Gene Ontology (GO) enrichment analysis of genes with SNPs was performed using the BioNGO plugin of Cytoscape v2.6 based on Benjamini and Hochberg false discovery rate correction at 5% significance level.

QTL mapping. For identification and mapping of QTLs, the reference desi genome-based high-quality GBS SNP genotyping data and genetic linkage map information of SNPs mapped on eight desi LGs of intra-specific genetic map were correlated with multi-location/years replicated field phenotyping data (PN, SN and SW) of RIL mapping individuals and parental genotypes using single marker analysis, interval mapping and composite interval mapping functions of QTL Cartographer v2.5 and MapQTL v6.0. To identify and map the novel genomic regions harbouring the major effect QTLs (PN, SN, and SW) containing SNP or LD threshold score of ≥5.0 at 1000 permutations was considered significant (p < 0.05). The significant major trait-influencing QTLs, which were validated across multiple environments (locations/years) were considered as “robust QTLs” controlling agronomic traits (as per Ref. 77). The positional genetic effects, phenotypic variation explained (PVE) % by QTLs and their additive effect (evaluated by parental origin of favourable alleles) on traits were evaluated at significant LOD (p ≤ 0.05).

QTL-region targeted resequencing and association analysis. A selected strong SW major genomic region underlying robust QTL was sequenced in parental genotypes (ICC 12299 and ICC 8261) along with 10 of each homogenous low and high SW RIL mapping individuals using the multiplexed amplicons sequencing protocol (following manufacturer’s instructions) of TrueSeq Custom Amplicon v1.5 in Illumina MiSeq next-generation sequencer. The high-quality amplicon sequence reads (> 90% bases covered at 0.5× mean coverage) were mapped to reference desi and kabuli genomes and SNPs were detected among parental genotypes and mapping individuals adopting the methods of Agarwal et al.28, Jain et al.29 and Saxena et al.30. The sequenced genomic region underlying SW-specific QTL was structurally and functionally annotated (following aforementioned methods). The SNPs showing differentiation between low and high SW parental genotypes and mapping individuals at QTL region of interest were genotyped in 244 chickpea accessions (211 mini-core) belonging to an SW-specific association panel (Table S7) using the MALDI-TOF SNP genotyping assay (following Refs. 77, 111). The SNP genotyping data among accessions were used to constitute haplotypes at target QTL interval. The replicated multi-location/years SW field phenotyping data, population structure (K = 5) and kinship coefficients were calculated using STRUCTURE v2.3.4. The association analysis was performed using general linear model (GLM) and mixed linear model (MLM at optimum level of compression with P3D method) of TASSEL v5.0 and mixed model approach of EMMA adopting the detailed procedures of Kujur et al.31, Saxena et al.32 and Thudi et al.33. To eliminate the confounding effect of population structure and correct the false discovery rate (FDR) based on multiple comparisons, the Bonferroni correction of P-value was performed for each trait-associated SNPs at 5% significance level. By integrating the results of GLM and MLM with EMMA and FDR correction, the SNPs showing strong association (R2 = correlation potential of significant SNPs with traits) with SW at significant cut-off P ≤ 10−4 were selected.

Differential gene expression profiling. The SNPs-carrying genes annotated at the short delineated major genomic region underlying the robust SW-specific QTL were selected to infer their differential regulation during seed development through expression profiling. The RNA isolated from three different vegetative and reproductive tissues (leaf, root and flower bud) and two seed developmental stages14,36,38 of high (ICC 8261) and low (ICC 12299) SW parental genotypes was amplified using the gene-specific primers through quantitative RT-PCR assay (following Refs. 74, 100).

Gene haplotype-based LD mapping. To perform high-resolution haplotype-based association/LD mapping, the fragments covering the entire coding and non-coding sequences, including 2 kb upstream regulatory region (URR) of one strong SW-associated gene (validated by QTL mapping, association mapping and differential expression profiling) were amplified using the genomic DNA of 224 SW-specific mapping individuals and the sequencing of amplicons. SNP mining, haplotype constitution and LD pattern determination of a gene including, its association potential with SW were estimated using the methods of Kujur et al.74,94 Saxena et al.32 and Bajaj et al.33. To determine the potential of diverse haplotypes constituted in the
gene for regulating SW, the differential expression profiling in two seed development stages of parental genotypes and homologous RIL mapping individuals representing the low and high SW haplotype groups was performed using the haplotype-specific primers.

1. Jukanti, A. K. et al. Nutritional quality and health benefits of chickpea (Cicer arietinum L.): a review. Br. J. Nutr. 108, S11–S26 (2012).

2. Arumuganathan, K. & Earle, E. D. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9, 208–219 (1991).

3. Kumar, A., Choudhary, A. K., Solanki, R. K. & Pratap, A. Towards marker-assisted selection in pulses: a review. Plant Breed. 130, 297–313 (2011).

4. Yu, H. et al. 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 (2011).

5. Truco, M. J. et al. An ultra-high density, transcript-based, genetic map of lettuce. G3 (Bethesda) 3, 617–631 (2013).

6. Nayak, S. N. et al. Integration of novel SSR 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 (2010).

7. Gujaria, N. et al. 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 (2011).

8. Gaur, R. et al. High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (Cicer arietinum L.). DNA Res. 19, 357–373 (2012).

9. Hiremath, P. J. et al. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in leguminous plants. Biotechnol. J. 3, e2253 (2013).

10. Roorkiwal, M. et al. Single nucleotide polymorphism genotyping for breeding genetics and applications in chickpea and pigeonpea using the BeadXpress platform. Plant Genome 6, 1–10 (2013).

11. Choudhary, S., Gaur, R. & Gupta, S. EST-derived genic molecular markers, development and utilization for generating an advanced transcript map of chickpea. Theor. Appl. Genet. 124, 1449–1462 (2012).

12. Stephens, A. et al. 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 (2014).

13. Varshney, R. K. et al. Genetic dissection of disease resistance in chickpea (Cicer arietinum L.). Theor. Appl. Genet. 127, 445–462 (2014a).

14. Varshney, R. K. et al. Marker-assisted backcrossing to introgress resistance to Fusarium wilt race 1 and Ascochyta blight in C. C. 214, an elite cultivar of chickpea. Plant Genome 7, 1–11 (2014b).

15. Hyten, D. L. et al. High density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping in legumes. Plant Biotechnol. J. 3, 716–732 (2013).

16. Huo, N. et al. A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. PLoS One 7, e92453 (2012).

17. Varshney, R. K. et al. Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. 30, 83–89 (2012).

18. Varshney, R. K. et al. Draft genome sequence of chickpea (Cicer arietinum) an orphan crop for trait improvement. Nat. Biotechnol. 31, 240–246 (2013).

19. Jain, M. et al. A draft genome sequence of the pulse crop chickpea (Cicer arietinum L.). Plant J. 74, 715–729 (2013).

20. Michael, T. P. & Jackson, S. The first 50 plant genomes. Plant Genome 6, doi:10.3835/plantgenome2013.03.0011 (2013).

21. Schnable, P. S. et al. The B73 maize genome: complexity, diversity and dynamics. Science 326, 1112–1115 (2009).

22. Elshire, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379 (2011).

23. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J. L. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7, e32253 (2012).

24. Sonah, H. et al. An improved genotyping-by-sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS One 8, e46003 (2013).

25. Thurber, C. S., Ma, J. M., Higgins, R. H. & Brown, P. J. Retrogressive genomic analysis of sorghum adaptation to temperate-zone grain production. Genome Biol. 14, R88 (2013).

26. Morris, G. P. et al. Population genomic and genome-wide association studies of agromictic traits in sorghum. Proc. Natl. Acad. Sci. USA 110, 453–458 (2013).

27. Spindel, J. et al. 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 (2013).

28. Varala, K., Swaminathan, K., Li, Y. & Hudson, M. E. Rapid genotyping of soybean cultivars using high throughpout sequencing. PLoS One 6, e24811 (2011).

29. Bastien, M., Sonah, H. & Belzile, F. Genome-wide association mapping of Sclerotinia sclerotiorum resistance in soybean with a genotyping-by-sequencing approach. Plant Genet. 7, 1–13 (2014).

30. He, J. et al. Genotyping-by-sequencing (GBS), an ultimate marker assisted selection (MAS) tool to accelerate plant breeding. Front. Plant. Sci. 5, 484 (2014).

31. Jarquin, D. et al. Genotyping-by-sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15, 740 (2014).

32. Sonah, H., O’Donoughue, L., Cober, E., Rajcan, I. & Belzile, F. 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 (2014).

33. Jaganathan, D. et al. Genotyping-by-sequencing-based intra-specific genetic map reveals a ”QTL-hotspot” region for drought tolerance in chickpea. Mol. Genomics Genomics doi:10.1097/00438-014-0932-3 (2014).

34. Cho, S. et al. Mapping genes for double podding and other morphological traits in chickpea. Euphytica 125, 285–292 (2002).

35. Rakhsh, S. et al. DAF marker tightly linked to a major locus for Ascochyta blight resistance in chickpea (Cicer arietinum L.). Euphytica 132, 23–30 (2003).

36. Chandra, S. et al. Identifying QTL-linked markers in marker-depent crops. Proceedings of the 4th International Crop Science Congress Brisbane, Australia, 26 September–1 October 2004, Fisher, T. (ed.), The Regional Institute Ltd (Publisher), Gosford New South Wales, Australia, (2004).

37. Sharma, K. D., Winter, P. J. & Muehlbauer, F. J. Molecular mapping of Fusarium oxysporum f. sp. ciceris race 3 resistance gene in chickpea. Theor. Appl. Genet. 108, 1234–1248 (2004).

38. Sharma, K. D., Chen, W. & Muehlbauer, F. J. Genetics of chickpea resistance to five races of Fusarium wilt and a concise set of race differentials for Fusarium oxysporum L. sp. ciceris. Plant Dis. 89, 385–390 (2005).

39. Cobos, M. J. et al. Linkage map of chickpea (Cicer arietinum L.) based on populations from kabuli x desi crosses: Location of genes for resistance to Fusarium wilt race 0. Theor. Appl. Genet. 110, 1347–1353 (2005).

40. Cobos, M. J. et al. A new QTL for Ascochyta blight resistance in an RI population of four inter-specific crosses in chickpea. Euphytica 149, 105–111 (2006).

41. Cobos, M. J. et al. Genetic analysis of agronomic traits in a wide cross of chickpeas. Field Crops Res. 111, 130–136 (2009).

42. Iruela, M. et al. Detection of two quantitative trait loci for resistance to Ascochyta blight in an intra-specific cross of chickpea (Cicer arietinum L.): development of new GBS markers associated with resistance. Theor. Appl. Genet. 112, 278–287 (2006).

43. Iruela, M. et al. Validation of a QTL for resistance to Ascochyta blight linked to resistance to Fusarium wilt race 5 in chickpea (Cicer arietinum L.). Eur. J. Plant Pathol. 119, 29–37 (2007).
68. Gowda, S. J. M., Radhika, P., Kadoo, N. Y., Mhase, L. B. & Gupta, V. S. Molecular mapping of wilt resistance genes in chickpea (Cicer arrietinum L.) using two recombining inbred line populations. Theor. Appl. Genet. 115, 209–216 (2007).

69. Tara, S. C. SCIENTIFIC.

70. Saxena, M. S.

71. Varshney, R. K.

72. Kujur, A.

73. Hiremath, P. J.

74. Agarwal, G.

75. Hiremath, P. J.

76. Jain, M., Moharana, K. C., Shankar, R., Kumari, R. & Garg, R. Genome-wide analyses and conservation of QTL for resistance in a chickpea interspecific cross (Cicer arrietinum L.). Theor. Appl. Genet. 119, 757–765 (2009).

77. Gowda, S. J. M., Radhika, P., Kadoo, N. Y., Mhase, L. B. & Gupta, V. S. Molecular mapping of wilt resistance genes in chickpea. Mol. Breed. 24, 177–183 (2009).

78. Gowda, C. L. L., Upadhyaya, H. D., Dronavalli, N. & Singh. Identification of large-seeded high-yielding stable kabuli chickpea germplasm lines for use in crop improvement. Crop Sci. 5, 198–209 (2011).

79. A. Sakiroglu, M.

80. Branca, A.

81. Zhao, K.

82. Bajaj, D. et al. Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. J. Exp. Bot. 66:1271–1290 (2015).

83. Thudi, M. et al. Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. PLoS One 9, e96758 (2014).

84. Pathan, S. M. et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 53, 765–774 (2012).

85. Kato, S. et al. A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. Theor. Appl. Genet. 127, 1365–1374 (2014).

86. Liu, Y.-L. et al. Identification of quantitative trait loci underlying plant height and seed weight in soybean. Plant Genome 6, 1–11 (2014).

87. Yuste-Lisbona, F. J. et al. Genetic analysis of single-locus and epistatic QTLs for seed traits in an adapted × nuria RIL population of common bean (Phaseolus vulgaris L.). Theor. Appl. Genet. 127, 897–912 (2014).

88. Lu, Y. et al. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Proc. Natl. Acad. Sci. USA 107, 19585–19590 (2010).

89. Famoso, A. N. et al. Genetic architecture of aluminum tolerance in rice (Oryza sativa L.) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7, e1002221 (2011).

90. Li, S. et al. Natural variation in PBT1 regulates rice seed setting rate by controlling pollen tube growth. Nat. Commun. 4, 2793 (2013).

91. Zhang, Z. et al. QTL analysis of kernel-related traits in maize using an inbred line-population F2 population. PLoS One 9, e96965 (2014).

92. Trafier, L. et al. Identification of genes required for embryo development in Arabidopsis. Plant Physiol. 135, 1206–1220 (2004).

93. Saxena, M. S. et al. Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. PLoS One 9, e107484 (2014b).

94. Pandit, A. et al. Combining QTL mapping and transcriptome profiling of bulked RILs for identification of functional polymorphism for salt tolerance genes in rice (Oryza sativa L.). Mol. Genet. Genomics 284, 121–136 (2010).

95. Yan, W. G. et al. Association mapping of stigma and sileptae characteristics in rice (Oryza sativa L.). Mol. Breed. 24, 277–292 (2009).

Acknowledgments
The authors gratefully acknowledge the financial support for this research study provided by a research grant from the Department of Biotechnology (DBT), Government of India (No. BT/PR13114/DBT/23/1716/2013-14). Alice Kujur and Shouvik Das acknowledge the CSR (Council of Scientific and Industrial Research) and DBT for Junior/Senior Research Fellowship awards. Authors declare no conflict of interest.

Author contributions
A.K., H.D.U. and T.S. conducted all experiments and drafted the manuscript. T.S., D.B., S. M. S., S.B. and V.K. involved in genotyping, sequencing and data analysis. H.D.U., S.T., C.G., S.H.S. and S.S. helped in generation of mapping population, and performed their phenotyping. A.K., N.P. and A.K.T. conceived and designed the study, guided data analysis and interpretation, participated in drafting and correcting the manuscript critically and gave the final approval of the version to be published. All authors have read and approved the final manuscript.
Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kujur, A. et al. Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. Sci. Rep. 5, 9468; DOI:10.1038/srep09468 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/