

Accepted Manuscript

Title: Identification of candidate genes for dissecting complex branch number trait in chickpea

Author: Deepak Bajaj Hari D. Upadhyaya Shouvik Das
Vinod Kumar C.L.L. Gowda Shivali Sharma Akhilesh K.
Tyagi Swarup K. Parida



PII: S0168-9452(16)30004-8
DOI: <http://dx.doi.org/doi:10.1016/j.plantsci.2016.01.004>
Reference: PSL 9347

To appear in: *Plant Science*

Received date: 4-11-2015
Accepted date: 15-1-2016

Please cite this article as: Deepak Bajaj, Hari D. Upadhyaya, Shouvik Das, Vinod Kumar, C.L.L. Gowda, Shivali Sharma, Akhilesh K. Tyagi, Swarup K. Parida, Identification of candidate genes for dissecting complex branch number trait in chickpea, *Plant Science* <http://dx.doi.org/10.1016/j.plantsci.2016.01.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Identification of candidate genes for dissecting complex branch number trait in chickpea

Deepak Bajaj^{1a}, Hari D. Upadhyaya^{2a}, Shouvik Das^{1a}, Vinod Kumar³, C.L.L. Gowda², Shivali Sharma², Akhilesh K. Tyagi¹, Swarup K. Parida^{1*}

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

^aThese authors contributed equally to this work.

*Corresponding author

Swarup K. Parida

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

E-mail: swarup@nipgr.ac.in; swarupdbt@gmail.com

Phone: 91-11-26735228; Fax: 91-11-26741658

Highlights

- GWAS identified 11 SNP loci-containing known/candidate genes associated with branch number in chickpea.
- Integration of association analysis with mapping population validation and expression profiling delineated seven branch number-regulating genes.
- Implicates efficacy of integrated genomics-assisted breeding strategy for quantitative dissection of complex branch number trait in chickpea.

Abstract

The present study exploited integrated genomics-assisted breeding strategy for genetic dissection of complex branch number quantitative trait in chickpea. Candidate gene-based association analysis in a branch number association panel was performed by utilizing the genotyping data of 401 SNP allelic variants mined from 27 known cloned branch number gene orthologs of chickpea. The genome-wide association study (GWAS) integrating both genome-wide GBS- (4556 SNPs) and candidate gene-based genotyping information of 4957 SNPs in a structured population of 60 sequenced *desi* and *kabuli* accessions (with 350-400 kb LD decay), detected 11 significant genomic loci (genes) associated (41% combined PVE) with branch number in chickpea. Of these, seven branch number-associated genes were further validated successfully in two inter (ICC 4958 x ICC 17160)- and intra (ICC 12299 x ICC 8261)-specific mapping populations. The axillary meristem and shoot apical meristem-specific expression, including differential up- and down-regulation (4-5 folds) of the validated seven branch number-associated genes especially in high branch number was apparent as compared to the low branch number-containing parental accessions and homozygous individuals of two aforesaid mapping populations. Collectively, this combinatorial genomic approach delineated diverse naturally occurring novel functional SNP allelic variants in seven potential known/candidate genes [*PINI* (PIN-FORMED protein 1), *TBI* (teosinte branched 1), *BA1/LAX1* (*BARREN STALK1/LIKE AUXIN1*), *GRAS8* (gibberellic acid insensitive/GAI, Repressor of gal3/RGA and Scarecrow8/SCR8), *ERF* (ethylene-responsive element-binding factor), *MAX2* (more axillary growth 2) and lipase] governing chickpea branch number. The useful information generated from this study have potential to expedite marker-assisted genetic enhancement by developing high-yielding cultivars with more number of productive (pods and seeds) branches in chickpea.

Key Words: branch number; chickpea; *Cicer*; *desi*; GBS; GWAS; *kabuli*; QTL; SNP

1. Introduction

Plant architecture is a complex dynamic characteristic that determines the shape, size, geometry and external structure of a plant [1]. Shoot branching, a key determinant of plant architecture undermines the ability of a plant to produce side shoots from its axillary buds [2]. Branching density and pattern is a key yield-component trait, as increasing shoot branching can be translated into increase biomass and seed/pod production. Shoot branching plays a key role in adaptation of plant to their local environment by changing the shape of plant. The number and size of the branches formed determine the total area of the plant and the spatial distribution of leaf area in the canopy. The amount of light absorbed by leaves of the plants make them compete with the neighbouring plants in terms of light interception and capturing of other resources, thereby reducing the chances of weed growth. Branching exerts its impact on dry matter accumulation and assimilates partitioning into the vegetative compartment and the reproductive growth [3]. Branching also affects developmental phenotypes, including flowering time and reproductive success in plant [4]. The process of axillary shoot branch formation is controlled by a complex interaction between genetically regulated developmental process and the environment [5] and [6]. Branching pattern and its density determined by number of branches per plant, is a complex trait and governed by many genes/QTLs (quantitative trait loci). Moreover, this trait was most probably targeted for domestication and selective breeding [7]. Considering the importance of branch number in yield improvement and adaption to the environment, it is imperative to identify the underlying heritable forces and potential genes/QTLs regulating this complex trait. Molecular dissection of branch number trait will also pave the way to uncover the regulatory and domestication pattern of genes/QTLs governing this trait. Significant efforts have been made towards deciphering the complex genetic architecture of branching pattern and density in several crop species, including *Arabidopsis*, pea, soybean, rice, maize and barley [8], [9], [10], [11], [12], [13], [14] and [15]. In legumes, a number of loci, including *RAMOSUS* (*ram*, *rms1*, *rms2*, *rms3*, *rms4* and *rms5*) as well as *Branching* (*Br1* and *Br2*) have been identified as regulatory factors controlling branching patterns and density in pea and soybean [3] and [12].

Chickpea [*Cicer arietinum* (L.)] is an economically important food legume with a genome size of ~720 Mbp. It is a self-pollinated annual diploid crop species having indeterminate and branched growth habit. The cultivated chickpea plant is generally erect or semi-erect with primary, secondary and tertiary branching, resembling a small bush. Unlike cultivated species, the wild species of chickpea has prostrate growth habit [16]. The effect of branching on seed and pod yield

as well as water-use efficiency is extensively studied and well documented in chickpea [17]. Only limited number of QTLs/genes regulating branch number have been identified utilizing QTL mapping and trait association analysis [18], [19], [20], [21], [22] and [23]. However, these identified QTLs/genes are yet to be deployed in marker-assisted selection for developing cultivars with high branch number in chickpea.

Considering the significance of integrated genomics-assisted breeding approach for quantitative dissection of complex yield component traits in crop plants, it would be interesting to employ this strategy in natural and mapping populations to identify functionally relevant molecular tags (markers, genes/QTLs and alleles) regulating branch number in chickpea [20], [24], [25], [26], [27], [28] and [29]. This will provide much needed inputs for genetic dissection of complex branch number trait, which can be subsequently utilized in marker-assisted breeding programme leading to development of superior cultivar with a desirable high attribute of branch number/density in chickpea. The increase in branch density as determined by number of productive (pod and seed-bearing branches) branches per plant can also enhance the pod and seed yield in chickpea, which is indispensable to feed the fast growing population for sustaining food security.

Keeping above in view, an integrated genomics-assisted breeding approach involving SNP (single nucleotide polymorphism) marker-based genome-wide association study (GWAS) and candidate gene-based association analysis, QTL mapping and differential gene expression profiling was utilized in the present study to delineate novel allelic variants in the genes associated with branch number in chickpea.

2. Materials and methods

2.1 Mining and genotyping of genome-wide SNPs

We selected 60 phenotypically and genotypically diverse *desi* (17 accessions) and *kabuli* (43) *Cicer* accessions (representing various eco-geographical regions of 21 countries of the world) (**Supplemental file 1**) from the chickpea germplasm collections (16991, including 211 minicore germplasm lines) available at ICRISAT Genebank following the methods of Kujur et al. [25] and [26]. For large-scale mining and high-throughput genotyping of genome-wide SNPs, the genomic DNA was isolated from *Cicer* accessions were used to constitute a 96-plex GBS library. The library was sequenced (100-bp single end) employing Illumina HiSeq2000 NGS platform as per the modified procedures of Elshire et al. [30] and Kujur et al. [31]. The Bowtie v2.1.0 [32] and reference-based GBS pipeline/genotyping approach of STACKS v1.0 (<http://creskolab.uoregon.edu/stacks>) were employed for de-multiplexing and mapping of high-quality FASTQ sequence reads onto the reference *kabuli* draft chickpea genome [32] and detection

of high-quality SNPs from 60 accessions as per Kujur et al. [31]. The structural and functional annotation of GBS-derived SNPs detected in different coding and non-coding sequence components of genes and genomes (chromosomes/pseudomolecules and scaffolds) were performed according to *kabuli* genome annotation [33].

2.2 Discovery and genotyping of candidate gene-derived SNPs

A selected 27 cloned genes/QTLs known to be involved in regulation of branch/tiller/panicle numbers in diverse crop plants, including *Arabidopsis thaliana*, tomato, *Medicago*, soybean, maize, rice and barley [13] were acquired for large-scale discovery and high-throughput genotyping of gene-based SNPs. The coding sequences (CDS) of these known genes were sequence homology (BLAST) searched against the CDS of *kabuli* genes to identify the best true chickpea gene orthologs. All the CDS and 2000-bp upstream and 1000-bp downstream regulatory regions of identified true *kabuli* chickpea gene orthologs (*E*-value: 0 and bit score ≥ 500) were targeted to design (Batch Primer3, <http://probes.pw.usda.gov/batchprimer3>) the multiple overlapping forward and reverse primers with expected amplification product size of 500 to 700 bp. The gene-derived primers were PCR amplified using the genomic DNA of 60 *Cicer* accessions (selected for genome-wide GBS-based SNP genotyping). The amplified PCR fragments were purified, cloned and sequenced by an automated DNA sequencer (Applied Biosystems, ABI 3730xl DNA Analyzer, Vernon Hills, Illinois, USA). The high-quality sequences were aligned and compared among accessions, and SNPs were detected in diverse sequence components of orthologous chickpea genes following Kujur et al. [24] and Saxena et al. [34].

2.3 Phylogenetic tree construction, population structure and LD patterns

The high-quality SNP genotyping data among 60 *Cicer* accessions were analysed with PowerMarker v3.51 [35] and MEGA v6.0 [36] to construct an unrooted neighbour-joining (NJ)-based phylogenetic tree (with 1000 bootstrap replicates) among accessions. The population genetic structure among accessions was determined by a model-based program STRUCTURE v2.3.4 following Kujur et al. [24] and [25] and Saxena et al. [34]. The genome-wide and candidate gene-based genotyping data of SNPs physically mapped on eight *kabuli* chromosomes were analysed by PLINK and the full-matrix approach of TASSELv5.0 [34] and [37]. Based on these, the genome-wide LD patterns (r^2 , frequency correlation among pair of alleles across a pair of SNP loci) and LD decay (by plotting average r^2 against 50 and 20 kb uniform physical intervals across chromosomes) in population was determined.

2.4 Phenotyping for branch number

Sixty *Cicer* accessions were planted in a single row with a plant density of 35 x 10 cm and raised in the experimental field following RCBD (randomised complete block design) with at least two replications. Following the above criteria, these accessions were grown and phenotyped for two consecutive years (2011 and 2012) during crop growing season at two diverse geographical locations (New Delhi; latitude/longitude: 28.4⁰N/77.1⁰E and Hyderabad; 17.1⁰N/78.9⁰E) of India. The branch number (BN) was measured by counting the average number (selecting 10-15 representative plants from each accession) of productive branches with pods/seeds emerging per plant at time of harvest. The broad-sense heritability (H^2), coefficient of variation (CV), analysis of variance (ANOVA) and frequency distribution of BN among 60 *Cicer* accessions were estimated following Kujur et al. [24] and [25] and Saxena et al. [20]. The effects/interactions among accessions/genotypes (G), environments (E) (i.e., experimental years and geographical locations) and replications/block replications was measured based on ANOVA to determine the inheritance pattern of BN trait among accessions utilized. The broad-sense heritability [$H^2 = \sigma^2g/(\sigma^2g + \sigma^2ge/n + \sigma^2e/nr)$] was measured based on σ^2g (genetic), σ^2ge (G x E) and σ^2e (error) variance with n (number of experimental years/environments) = 2 and r (number of replicates) = 2. For association mapping, the accessions exhibiting high-heritability (H^2) for BN trait across two environments (geographical locations) and experimental years were selected. The environment and/replication-wise BN phenotypic data obtained from each of the selected accessions was used to estimate the mean value of BN in individual accession to be employed in genetic association analysis.

2.5 Trait association mapping

For genetic association mapping, the genome-wide and candidate gene-derived SNP genotyping and BN phenotyping data of 60 *Cicer* accessions was integrated with their ancestry coefficient (Q matrix of population structure), relative kinship matrix (K) and PCA (principal component analysis) information based on P3D/compressed mixed linear model (CMLM) interfaces of GAPIT [37], [38] and [39]. The quantile-quantile plot of GAPIT was employed to compare individually the relative distribution of observed and expected $-\log_{10}(P)$ -value for each SNP marker-trait association. The adjusted P-value threshold of significance was corrected for multiple comparisons based on false discovery rate (FDR cut-off ≤ 0.05) [40]. Based on these analyses, the potential SNPs in the target genomic (gene) regions associated significantly with BN at higher R^2 (degree of SNP marker-trait association) and lower FDR adjusted P-values (threshold $P < 1 \times 10^{-6}$) were selected.

2.6 Validation of trait-associated SNPs in bi-parental mapping populations

To ensure the potential of identified genomic loci for BN trait association, these SNPs were targeted to validate in two traditional bi-parental mapping populations. For this, 20 of each low and high branch number-containing homozygous individuals along with parental accessions derived from two diverse inter- [*C. arietinum desi* ICC 4958 (BN: 25) x *C. reticulatum* wild ICC 17160 (BN: 17)] and intra- [*C. arietinum desi* ICC 12299 (BN: 29) x *C. arietinum kabuli* ICC 8261 (BN: 18)] specific F₇ RIL mapping populations were selected for DNA isolation. The BN-associated SNPs showing polymorphism between the parents of two mapping populations were genotyped in these homozygous mapping individuals using MALDI-TOF mass array SNP genotyping assay following Saxena et al. [20] and [34]. Furthermore, the physical positions of BN-associated SNPs mapped on chromosomes were correlated with that of multiple BN QTLs identified/mapped previously on the inter-/intra-specific genetic linkage maps to validate these major genomic loci underlying BN QTLs in chickpea.

2.7 Differential expression profiling

To determine the differential regulatory pattern of BN-associated genes validated by GWAS/candidate gene-based association analysis and in bi-parental mapping populations, the *in silico* digital expression analysis was performed. The global transcript profiling data (generated from whole-genome microarray and transcriptome sequencing assays) available for BN-associated genes in different vegetative and reproductive tissues (leaf, root and shoot apical meristem) of chickpea (Chickpea Transcriptome Database; <http://www.nipgr.res.in/ctdb.html>) was retrieved and correlated. For further validation of *in silico* expression data, the differential expression profiling of BN-associated genes was performed using the semi-quantitative and quantitative RT-PCR assays. The RNA was isolated from vegetative and reproductive tissues [leaf, root, axillary meristem (AM) at primary and secondary branches differentiation stage during vegetative growth and shoot apical meristem (SAM) of short floret bearing branches during reproductive growth] of low and high branch number-containing parental accessions and one of each homozygous individual derived from aforementioned two mapping populations. The isolated RNA was amplified with the gene-specific primers (**Supplemental file 2**) and differential expression analysis was performed following Bajaj et al. [27].

3. Results and discussion

3.1 Discovery and genotyping of genome-wide GBS- and candidate gene-based SNPs

A total of 201.3 million sequence reads (with an average of 2.1 million reads per accession) were generated through GBS-based sequencing of 96-plex *Ape*KI libraries constructed from 60 diverse *desi* and *kabuli Cicer* accessions (**Supplemental file 1**). Of these, 130.7 million high-quality

sequence reads (uniformly distributed across 60 accessions) were mapped to unique physical location on *kabuli* reference genome. The sequencing data generated in the present investigation was submitted to NCBI-sequence read archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>) under BioProject ID: PRJNA300922 (submission ID: SUB1161505). Altogether, 4,556 high-quality SNPs were discovered from 60 *Cicer* accessions using *kabuli* reference genome-based GBS assay (**Supplemental file 3**). Of these, 3,656 and 900 SNPs were physically mapped across eight chromosomes and unanchored scaffolds of *kabuli* genome, respectively. This emphasizes the potential utility of GBS assay in quick large-scale discovery and high-throughput genotyping of reliable and high-quality genome-wide SNPs simultaneously in chickpea. The GBS assay has been successfully deployed in high-resolution genetic and association mapping to identify potential genes/QTLs governing important agronomic traits in chickpea [23], [26], [31], [41] and [42]. Therefore, GBS-based genome-wide SNPs differentiating cultivated *desi* and *kabuli* accessions identified in our study have significance for their immense utilization in genomics-assisted breeding applications of chickpea.

A diverse array of known cloned genes regulating branch/tiller/panicle numbers have been identified and characterized in *Arabidopsis thaliana*, tomato, *Medicago*, soybean, maize, rice and barley [13]. To evaluate the potential of these known/cloned genes for branch number trait regulation in chickpea, the trait association mapping was performed by large-scale genotyping of novel SNP allelic variants discovered from diverse coding and regulatory sequence components of genes in 60 *desi* and *kabuli Cicer* accessions. Primarily, for mining candidate gene-derived SNPs, the cloned PCR amplicons-based sequencing of 27 chickpea gene orthologs known to be involved in regulation of branch/tiller/panicle numbers in diverse crop plants was performed. The comparison of high-quality gene amplicon sequences among a constituted branch number association panel (60 *desi* and *kabuli Cicer* accessions) identified 401 SNPs in 27 genes (1,23,306 bp) with a mean frequency of one SNP per 307.5 kb (**Table 1**). A total of 4,957, including 4,556 genome-wide GBS- and 401 candidate gene-based SNPs were mapped on eight chromosomes (4,027 SNPs with an average map-density 86.2 kb) and unanchored scaffolds (930 SNPs) of *kabuli* genome (**Fig. 1A, B**). Highest (932, 18.8%) and lowest (202, 4.1%) number of SNPs were mapped on chromosomes 4 and 5 with the mean densities of 52.8 and 143.8 kb, respectively.

The detailed structural annotation of 4,957 SNPs, including 4,556 *kabuli* reference genome- and 401 gene-based SNPs revealed the occurrence of 3,131 (63%) and 1,826 (37%) SNPs in the 1,404 genes and intergenic regions, respectively (**Fig. 1C**). The gene-derived SNPs included the highest and lowest frequency of 1,764 (56%) and 107 (3%) SNPs in the exons/CDS (coding sequences) and DRRs (downstream regulatory regions), respectively. The coding SNPs comprised of 1,005 (57%) synonymous and 759 (43%) non-synonymous (missense and nonsense) SNPs,

respectively (**Fig. 1C**). The structurally and functionally annotated genome-wide and gene-derived SNPs and novel SNP allelic variants mined from diverse branch number-governing known cloned and candidate genes can be employed for various high-throughput genetic analysis especially in quantitative dissection of complex branch number trait chickpea.

3.2 Association mapping of branch number

For GWAS and candidate gene-based association mapping, the genome-wide GBS- (4,556 SNPs) and candidate gene-based SNP genotyping (401 SNPs) data of 4,957 SNPs assayed in 60 *desi* and *kabuli Cicer* accessions (branch number association panel) were utilized. The use of 4,957 SNPs in neighbour-joining phylogenetic tree construction (**Fig. 2A**), and high-resolution population genetic structure (**Fig. 2B**) and PCA (**Fig. 2C**) determination, differentiated 60 accessions from each other and clustered into two distinct populations (POP I and POP II). The determination of LD patterns in a population of 60 accessions using 4,027 SNPs (3,656 genome-wide and 371 gene-based SNPs) (physically mapped on eight chromosomes) revealed LD estimates (r^2) of 0.20-0.66 and LD decay (r^2 decreased half of its maximum value) nearly at 350-400 kb physical distance of *kabuli* chromosomes (**Fig. 2D**). An extended LD decay in a self-pollinating crop plant like chickpea due to its narrow genetic base (low intra-specific polymorphism) along with extensive contribution of four sequential bottlenecks during domestication is quite anticipated [26] and [31]. Henceforth, GWAS integrated with candidate gene-based association analysis will be of much relevance for efficient genetic dissection of diverse complex quantitative traits in chickpea [26], [42], [43] and [44]. In this context, a combinatorial strategy of GWAS and candidate gene-based association analysis has been utilized in the present study to delineate numerous high-resolution potential genomic loci regulating quantitative trait of branch number at a genome-wide scale in chickpea.

A normal frequency distribution along with broader phenotypic variation (ranged 12.4-27.5, mean \pm SD: 19.0 ± 2.2 , CV: 11.5% and H^2 : 82.3%) of BN trait in 60 *Cicer* accessions (branch number association panel) across two diverse geographical locations and years (environments) based on ANOVA was observed (**Supplemental files 4 and 5**). All the accessions belonging to an association panel exhibited a higher broad-sense heritability (H^2) for BN trait across environments (two diverse geographical locations and years) under study. We observed a highly significant difference ($P < 0.001$) among accessions for BN trait based on ANOVA despite significant environmental effects (locations and years) on the target trait (**Supplemental files 4, 5 and 6**). The replication block effects were also significant ($P < 0.001$) for BN trait measured at two diverse locations across two years (**Supplemental file 6**). A significant interaction between accessions and environments (locations and years) for BN trait in 60 accessions was apparent.

The CMLM-based association analysis at FDR cut-off ≤ 0.05 identified 11 genomic loci (gene-derived SNPs) exhibiting significant association with BN at a $P \leq 10^{-8}$ (**Fig. 3A, B, Table 2**). This includes four genome-wide and seven known branch number-regulating gene-based SNPs. All BN-associated genomic loci were physically mapped on seven chromosomes (except chromosome 4) of *kabuli* genome (**Fig. 3A, Table 2**). A highest number of four BN-associated SNPs was mapped on *kabuli* chromosome 7. Seven and four of 11 BN-associated genomic loci derived from diverse coding (five non-synonymous and two synonymous SNPs) and non-coding (one intronic and three URR-SNPs) sequence components of 11 genes, respectively were identified (**Table 1**). The proportion of BN phenotypic variation explained (PVE) by 11 SNP loci in seven known BN-regulating genes [*PINI* (PIN-FORMED protein 1), *TBI* (teosinte branched 1), *BAI/LAX1* (*BARREN STALK1/LIKE AUXINI*), *GRAS8* (gibberellic acid insensitive/GAI, Repressor of ga13/RGA and Scarecrow8/SCR8), *HUB1* (histone monoubiquitination 1), *ERF* (ethylene-responsive element-binding factor) and *MAX2* (more axillary growth 2) and four candidate genes [unknown expressed gene, *DUF827* (domain of unknown function 827), *SNF2* (Sucrose Non-Fermenting 2) and lipase] among 60 *Cicer* accessions varied from R^2 : 4 to 12%. All significant 11 SNP loci explained 41% BN phenotypic variation. Remarkably, seven non-synonymous coding, intronic and regulatory SNPs containing seven known cloned genes revealed strong association ($P \leq 10^{-8}$ with R^2 : 10-12% and 47% combined PVE) with BN as compared to four BN-associated candidate genes-derived SNPs ($P \leq 10^{-8}$ with R^2 : 4-11% and 38% combined PVE) in chickpea. Strong association of one regulatory SNP in a *BAI/LAX1* (*BARREN STALK1/LIKE AUXINI*) known cloned gene ($P \leq 10^{-9}$ with R^2 : 12%) and another synonymous SNP in a candidate lipase gene ($P \leq 10^{-9}$ with R^2 : 11%) with BN was evident in chickpea. The functionally relevant non-synonymous and regulatory SNP loci-containing known/candidate genes associated with branch number trait identified by combining both GWAS and candidate gene-based association mapping could essentially be utilized for ascertaining the rapid marker-trait linkages and efficient detection of potential genes/QTLs governing branch number trait in chickpea [20], [24], [25], [26], [27], [31] and [42].

3.3 Validation of BN-associated genes in bi-parental mapping populations and by expression profiling

To validate the BN-associated known/candidate gene-derived SNP loci, the genic SNPs revealing parental polymorphism were genotyped in 20 of each low and high branch number-containing homozygous individuals derived from two diverse inter- (ICC 4958 x ICC 17160) and intra- (ICC 12299 x ICC 8261) specific RIL mapping populations. Six known BN-regulating genes (*PINI*, *TB*, *LAX1*, *GRAS8*, *MAX2* and *ERF*) and one candidate gene (lipase) revealing significant association with BN (based on our trait association analysis), were validated in both inter- and intra-specific

mapping populations (**Table 2**). All high (BN: 25-29) and low (BN: 15-18) branch number-containing parental accessions and homozygous individuals of two mapping populations contained the identical high and low BN-associated SNP alleles identified from seven known/candidate genes. Thus, stronger marker allele effects of these seven genes with high and low branch number differentiation were observed in chickpea. In contrast, SNP alleles of one known gene (*HUB1*) and three candidate genes (unknown expressed gene, *DUF827* and *SNF2*) specific to high and low BN, could not correspond to the phenotypes of the low and high branch number-containing parents and homozygous individuals of two mapping populations and thus exhibited a biased distribution. Therefore, seven non-synonymous and regulatory SNPs-containing known/candidate genes validated by trait association analysis and in bi-parental mapping populations, were selected as target candidates for branch number trait regulation by their further validation through differential expression profiling in chickpea.

In order to ascertain the accuracy and novelty of seven genomic loci-carrying genes for branch number regulation in chickpea (identified by their validation in natural and mapping populations), the outcomes of the present study was compared/correlated with that of previous reports [18], [19], [20], [21], [22] and [23]. For this, markers linked/flanking the branch number known QTLs/genes (documented in previous QTL mapping studies), were selected for their further validation in the branch number-specific natural (60 diverse *desi* and *kabuli Cicer* accessions) and mapping populations [(ICC 4958 x ICC 17160) and (ICC 12299 x ICC 8261)] constituted in the present study. This revealed correspondence of three SNPs-containing genes [*bHLH* (basic helix-loop-helix), *LBD41* (lateral organ boundaries 41) and *NAC*: no apical meristem (*NAM*), *Arabidopsis* transcription activation factor (*ATAF*), cup-shaped cotyledon (*CUC*)] as well as five SNP loci-carrying known/candidate genes (*GRAS8*, *LAX1*, *TB*, *ERF* and lipase) harbouring the multiple branch number QTLs between past and present study based on congruent flanking/linked marker physical positions on the chromosomes of chickpea. Therefore, five newly and three previously identified SNPs-containing known/candidate genes regulating branch number delineated in this study could be the potential candidate for their utilization in marker-assisted genetic improvement of chickpea.

The BN-associated seven SNP-containing known/candidate genes (validated by association analysis and in mapping populations) were analysed for differential expression profiling to determine the functional regulatory patterns of these genes for governing branch number in chickpea. The *in silico* global transcript profiling and experimental differential expression analysis (semi-quantitative and quantitative RT-PCR assays) of BN-associated genes in diverse vegetative and reproductive tissues (leaf, root and axillary meristem (AM) at primary and secondary branches differentiation stage during vegetative growth and shoot apical meristem (SAM) of short floret

bearing branches during reproductive growth) of low and high branch number-containing parental accessions and one of each homozygous individual derived from two mapping populations [(ICC 4958 x ICC 17160) and (ICC 12299 x ICC 8261)] was performed. All seven known cloned/candidate genes exhibiting significant association with BN (validated by association analysis and in mapping populations) revealed AM/SAM-specific expression (>10-folds) compared with leaf and root tissues of low and high branch number-containing parental accessions and homozygous mapping individuals (**Fig. 4**). Of these, six known/cloned genes (*GRAS8*, *LAX1*, *ERF*, *TB*, *MAX2* and *PINI*) and one candidate gene (lipase) exhibited pronounced differential up (> 5-folds, $P \leq 0.001$)- and down (> 4-folds, $P \leq 0.001$)-regulation, respectively in AM of primary and secondary branches and SAM of high branch number-containing parental accessions/mapping individuals as compared to that of low branch number-containing parental accessions/individuals (**Fig. 4**). Remarkably, these seven differentially expressed AM/SAM-specific genes showing significant association with BN (validated by association mapping and in mapping populations) mostly contained upstream regulatory and non-synonymous coding SNPs (**Table 2**). Therefore, it would be interesting to correlate these regulatory and non-synonymous novel SNP allelic variants mined from different sequence components of the genes with BN trait association potential and differential expression profiles underlying branch number regulation in chickpea.

In our study, the integration of GWAS/candidate gene-based trait association analysis with traditional bi-parental mapping population validation and differential gene expression profiling, delineated six known/cloned genes (*GRAS8*, *LAX1*, *ERF*, *TB*, *MAX2* and *PINI*) and one candidate gene (lipase) regulating branch number in chickpea. The transcriptional up- and down-regulation, including differential accumulation of transcripts of these known/candidate genes in the AM/SAM tissues during vegetative and reproductive growth and their correlation with multiple auxin biosynthesis genes towards governing branch/panicle/tiller numbers have been well-understood in diverse crop plants, including legumes [13], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55] and [56]. The functional novel allelic variants from some of these selected genes like *MAX2* and *TBI* known to be involved in regulation of branch/panicle/tiller numbers, have been identified and functionally characterized by utilizing an integrated approach of candidate gene-based association analysis, fine mapping/map-based cloning and differential transcript profiling in crop plants [16], [57], [58] and [59]. Notably, diverse naturally occurring novel allelic variants mined from a well-known BN-regulating gene, *TBI* for instance, have been selected for branch/tiller number differentiation during domestication in diverse cultivated and wild crop accessions [13], [58], [60], [61] and [62]. Therefore, it would be interesting to target this potential *TBI* gene ortholog for mining novel functional allelic variants towards understanding the adaptation-based branch number trait domestication pattern in cultivated and wild chickpea. The functionally

relevant BN-governing molecular tags (SNPs, genes and novel allelic variants) scaled-down by deploying an integrated genomic approach can be utilized for quantitative dissection of complex branch number trait. This will eventually accelerate the marker-assisted genetic enhancement for developing genetically tailored chickpea cultivars with increased productive (pods and seeds) branch numbers and yield.

Acknowledgments

The authors gratefully acknowledge the financial support by the core grant of National Institute of Plant Genome Research (NIPGR), New Delhi, India. Shouvik Das acknowledges the Department of Biotechnology (DBT), Government of India for Junior Research Fellowship award. We are thankful to the Editor and reviewer for critically evaluating the manuscript and providing constructive comments.

Appendix A. Supplemental data

Supplemental file 1. Sixty *Cicer* accessions selected for discovery and high-throughput genotyping of genome-wide GBS- and candidate gene-based SNPs for branch number trait association mapping.

Supplemental file 2. Primers used for differential expression analysis of branch number (BN)-associated genes in chickpea.

Supplemental file 3. Structural and functional annotation of 4556 GBS-based SNPs mined from *kabuli* chickpea genome.

Supplemental file 4. Frequency distribution of branch number (BN) trait in 60 cultivated *desi* and *kabuli* *Cicer* accessions belonging to an association panel depicted a goodness of fit to the normal distribution.

Supplemental file 5. Phenotypic variation of BN trait determined in a constituted association panel across two diverse geographical locations and years using various statistical measures.

Supplemental file 6. ANOVA-based summary effects of BN trait measured in a constituted association panel across two geographical locations and experimental years.

References

- [1] J. Ross, The radiation regime and architecture of plant stands, W. Junk Publishers, The Hague-Boston-London, (1981) 391-395.
- [2] J.B. Evers, A.R. van der Krol, J. Vos, P.C. Struik, Understanding shoot branching by modelling form and function, *Trends Plant Sci.* 16 (2011) 464-467.
- [3] C. Huyghe, Genetics and genetic modifications of plant architecture in grain legumes: a review, *Agronomie* 18 (1998) 383-411.
- [4] S.P. Ward, O. Leyser, Shoot branching, *Curr. Opin. Plant Biol.* 7 (2004) 73-78.
- [5] P. McSteen, O. Leyser, Shoot branching, *Annu. Rev. Plant Biol.* 56 (2005) 353-374.
- [6] M.W.F. Yaish, D.R. Guevara, A. El-Kereamy, S.J. Rothstein, Axillary shoot branching in plants, *Plant Developmental Biology Biotechnological Perspectives*, Springer Berlin Heidelberg 1 (2010) 37-52.
- [7] D.A. Vaughan, E. Balazs, J.S. Heslop-Harrison, From crop domestication to super-domestication, *Annal. Bot.* 100 (2007) 893-901.
- [8] K. Sorefan, et al., *MAX4* and *RMS1* are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea, *Genes Dev.* 17 (2003) 1469-1474.
- [9] P. Stirnberg, K.V.D. Sande, H.M. Leyser, *MAX1* and *MAX2* control shoot lateral branching in *Arabidopsis*, *Development* 129 (2002) 1131-1141.
- [10] J.A. Aguilar-Martínez, C. Poza-Carrión, P. Cubas, *Arabidopsis* BRANCHED1 acts as an integrator of branching signals within axillary buds, *Plant Cell* 19 (2007) 458-472.
- [11] C.A. Beveridge, J.J. Ross, I.C. Murfet, Branching Mutant *rms-2* in *Pisum sativum* (grafting studies and endogenous indole-3-acetic acid levels), *Plant Physiol.* 104 (1994) 953-959.
- [12] R. Nelson, The inheritance of a branching type in soybean, *Crop Sci.* 36 (1996) 1150-1152.
- [13] A. Hussien, et al., Genetics of tillering in rice and barley, *Plant Genome* 7 (2014) 1-20.
- [14] T. Dabbert, et al., The genetics of barley low-tillering mutants: low number of tillers-1 (*lnt1*), *Theor. Appl. Genet.* 121 (2010) 705-715.
- [15] C.J. Whipple, et al., Grassy tillers1 promotes apical dominance in maize and responds to shade signals in the grasses, *Proc. Natl. Acad. Sci. U.S.A.* 33 (2011) E506-E512.
- [16] L. Ali, et al., Detection of a new QTL/gene for growth habit in chickpea CaLG1 using wide and narrow crosses, *Euphytica* 204 (2015) 473-485.
- [17] K.H.M. Siddique, R.H. Sedgley, The effect of reduced branching on yield and water use of chickpea (*Cicer arietinum* L.) in a Mediterranean type environment, *Field Crops Res.* 12 (1985) 251-269.

- [18] J. Lichtenzveig, D.J. Bonfil, H.B. Zhang, D. Shtienberg, S. Abbo, Mapping quantitative trait loci in chickpea associated with time to flowering and resistance to *Didymella rabiei* the causal agent of *Ascochyta* blight, *Theor. Appl. Genet.* 113 (2006) 1357-1369.
- [19] S.J.M. Gowda, et al., Mapping of QTLs governing agronomic and field traits in chickpea, *J. Appl. Genet.* 52 (2011) 9-21.
- [20] M.S. Saxena, et al., An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea, *DNA Res.* 21 (2014) 695-710.
- [21] M. Thudi, et al., Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches, *PLoS ONE* 9 (2014) e96758.
- [22] R.K. Varshney, et al., Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.), *Theor. Appl. Genet.* 127 (2014) 445-462.
- [23] D. Jaganathan, et al., Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea, *Mol. Genet. Genomics* 290 (2015) 559-571.
- [24] A. Kujur, et al., Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping, *DNA Res.* 20 (2013) 355-374.
- [25] A. Kujur, et al., 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 (2014) 241-265.
- [26] A. Kujur, et al., A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea, *Sci. Rep.* 5 (2015) 11166.
- [27] D. Bajaj, 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 (2015) 1271-1290.
- [28] D. Bajaj, et al., A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea, *Sci. Rep.* 5 (2015) 9264.
- [29] J. Zuo, J. Li, Molecular genetic dissection of quantitative trait loci regulating rice grain size, *Annu. Rev. Genet.* 48 (2014) 99-118.
- [30] R.J. Elshire, et al., A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, *PLoS ONE* 6 (2011) e19379.

- [31] A. Kujur, et al., Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea, *Front. Plant Sci.* 6 (2015a) 162.
- [32] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357-359.
- [33] R.K. Varshney, et al., Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement, *Nat. Biotechnol.* 31 (2013) 240-246.
- [34] M.S. Saxena, et al., Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea, *PLoS ONE* 9 (2014) e107484.
- [35] K. Liu, S.V. Muse, PowerMarker: an integrated analysis environment for genetic marker analysis, *Bioinformatics*, 21 (2005) 2128-2129.
- [36] K. Tamura, G. Stecher, D. Peterson, A. FilipSKI, S. Kumar, MEGA6: Molecular evolutionary genetics analysis version 6.0, *Mol. Biol. Evol.* 12 (2013) 2725-2729.
- [37] V. Kumar, et al., Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*), *DNA Res.* 22 (2015) 133-145.
- [38] A.E. Lipka, et al., GAPIT: genome association and prediction integrated tool, *Bioinformatics* 28 (2012) 2397-2399.
- [39] H.D. Upadhyaya, et al., A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea, *Plant Mol. Biol.* (2015) DOI:10.1007/s11103-015-0377-z.
- [40] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. Roy. Statist. Soc. Ser. B* 57 (1995) 289-300.
- [41] A.A. Deokar, et al., 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 (2014) 708.
- [42] A. Kujur, 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 (2015) 9468.
- [43] D.B. Neale, O. Savolainen, Association genetics of complex traits in conifers, *Trends Plant Sci.* (2004) 325-330.
- [44] R.K. Varshney, et al., Genome wide association analyses for drought tolerance related traits in barley (*Hordeum vulgare* L.), *Field Crop Res.* 126 (2012) 171-180.
- [45] J. Stuurman, F. Jäggi, C. Kuhlemeier, Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells, *Genes Dev.* (2002) 16 2213-2218.

- [46] K. Komatsu, et al., *LAX* and *SPA*: Major regulators of shoot branching in rice, *Proc. Natl. Acad. Sci. U.S.A.* 100 (2003) 765-770.
- [47] T. Takeda, et al., The *OsTBI* gene negatively regulates lateral branching in rice, *Plant J.* 33 (2003) 513-520.
- [48] M. Xu, L. Zhu, H. Shou, P. Wu, A PIN1 family gene, *OsPIN1*, involved in auxin-dependent adventitious root emergence and tillering in rice, *Plant Cell Physiol.* 46 (2005) 1674-1681.
- [49] J. Mach, Rice axillary meristem formation requires directional movement of LAX PANICLE1 protein, *Plant Cell* 21(2009) 1027.
- [50] T. Oikawa, J. Kyojuka, Two-step regulation of LAX PANICLE1 protein accumulation in axillary meristem formation in rice, *Plant Cell* 21 (2009) 1095-1108.
- [51] W. Qi, et al., Rice ethylene-response AP2/ERF factor *OsEATB* restricts internode elongation by down-regulating a gibberellin biosynthetic gene, *Plant Physiol.* 157 (2011) 216-228.
- [52] H. Tong, et al., DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice, *Plant J.* 58 (2009) 803-816.
- [53] H. Tong, et al., DWARF AND LOW-TILLERING acts as a direct downstream target of a *GSK3/SHAGGY*-like kinase to mediate brassinosteroid responses in rice, *Plant Cell* 24 (2012) 2562-2577.
- [54] S. Guo, et al., The interaction between *OsMADS57* and *OsTBI* modulates rice tillering via *DWARF14*, *Nat. Commun.* 4 (2013) 1566.
- [55] W. Liu, et al., *THIS1* is a putative lipase that regulates tillering, plant height, and spikelet fertility in rice, *J. Exp. Bot.* 64 (2013) 4389-4402.
- [56] G.M. Deshpande, K. Ramakrishna, G.L. Chongloi, U. Vijayraghavan, Functions for rice *RFL* in vegetative axillary meristem specification and outgrowth, *J. Exp. Bot.* 66 (2015) 2773-2784.
- [57] I.M. Ehrenreich, P.A. Stafford, M.D. Purugganan, The genetic architecture of shoot branching in *Arabidopsis thaliana*: A comparative assessment of candidate gene associations vs. quantitative trait locus mapping, *Genetics* 176 (2007) 1223-1236.
- [58] M.S. Remigereau, et al., Cereal domestication and evolution of branching: evidence for soft selection in the *Tb1* orthologue of pearl millet (*Pennisetum glaucum* [L.] R. Br.), *PLoS ONE* 6 (2011) e22404.
- [59] L. Vann, T. Kono, T. Pyhäjärvi, M.B. Hufford, J. Ross-Ibarra, Natural variation in teosinte at the domestication locus teosinte branched1 (*tb1*), *Peer J* 3 (2015) e900.
- [60] R.M. Clark, E. Linton, J. Messing, J.F. Doebley, Pattern of diversity in the genomic region near the maize domestication gene *tb1*, *Proc. Natl. Acad. Sci. U.S.A.* 101 (2004) 700-707.

- [61] R.S. Meyer, A.E. DuVal, H.R. Jensen, Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops, *New Phytol.* 196 (2012) 29-48.
- [62] A.N. Doust, et al., Beyond the single gene: How epistasis and gene-by-environment effects influence crop domestication, *Proc. Natl. Acad. Sci. U.S.A.* 111 (2014) 6178-6183.

Figure legends

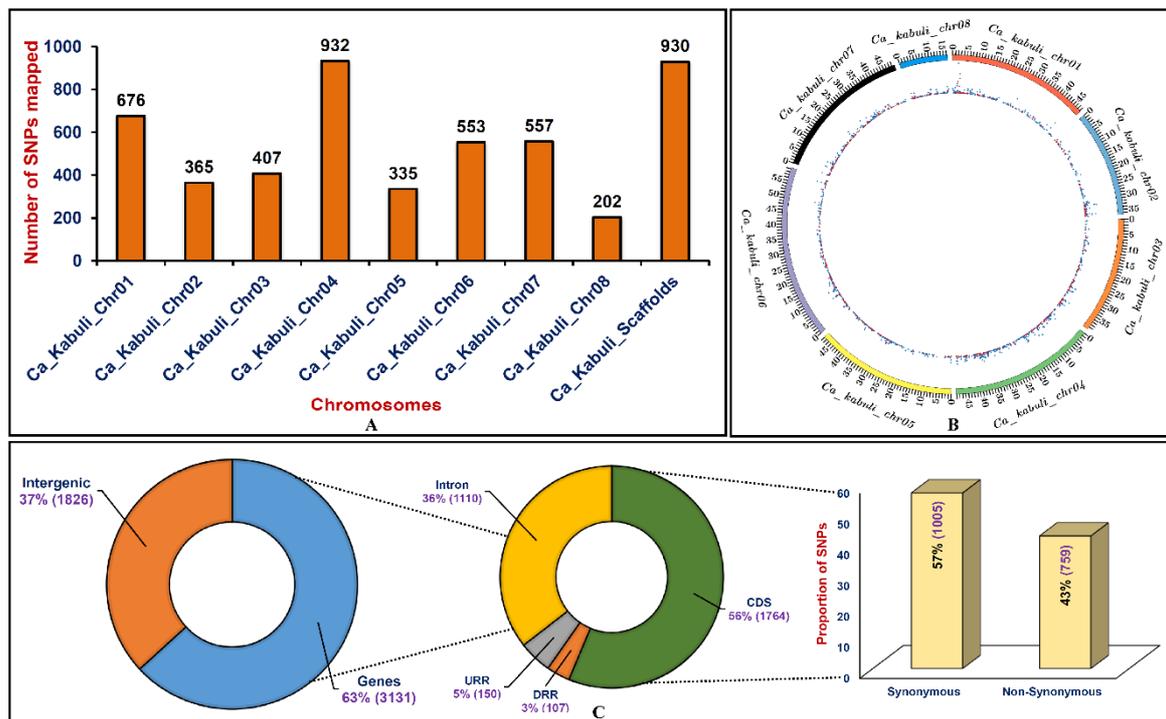


Fig. 1

Fig. 1. (A) Frequency distribution of 4957 SNPs mined utilizing reference *kabuli* genome (eight chromosomes and scaffolds)-based GBS and known/candidate gene-derived SNP genotyping assays. (B) A Circos circular ideogram illustrating the genomic distribution of 4957 SNPs, including non-synonymous SNPs (marked with red dots) physically mapped on eight *kabuli* chromosomes. The outermost circles depict the eight *kabuli* chromosomes coded with different colours, while the inner circle represents the relative distribution of SNPs mined from 60 cultivated (*desi* and *kabuli*) accessions. (C) Annotation of SNPs in the diverse coding (synonymous and non-synonymous) and non-coding (intron, URR and DRR) sequence components of genes and intergenic regions of *kabuli* genome. The gene annotation of *kabuli* genome [42] was used as a reference to deduce the CDS (coding sequences)/exons, introns, URR (upstream regulatory region) and DRR (downstream regulatory region) sequence components of genes.

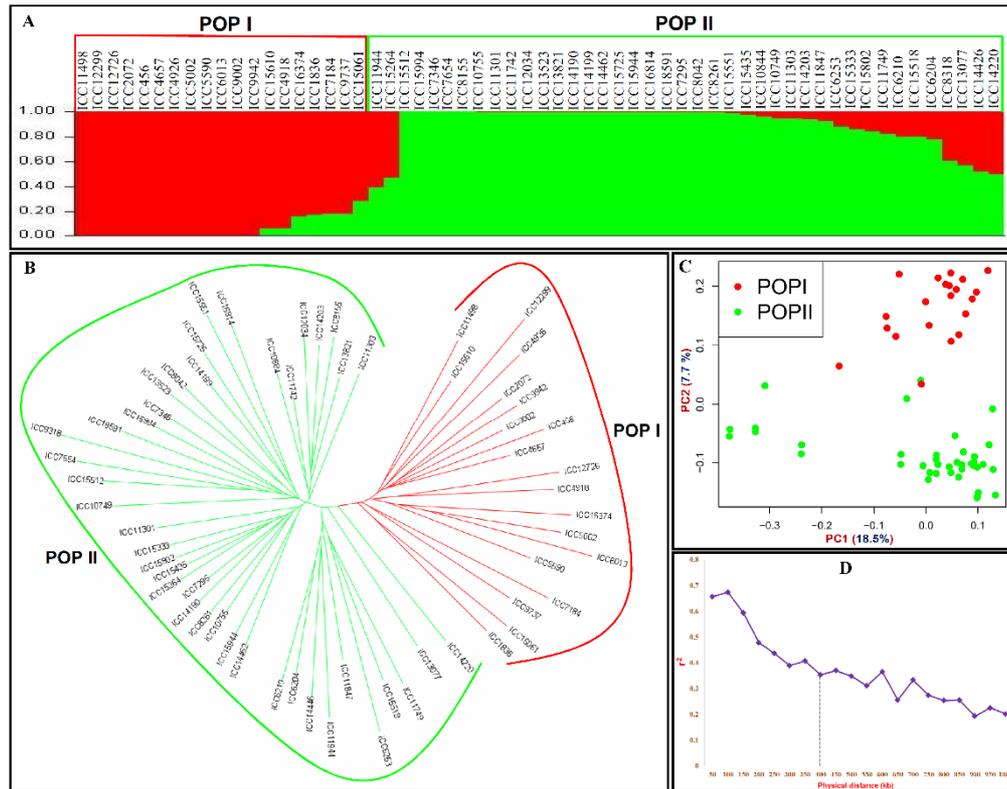


Fig. 2

Fig. 2. The determination of (A) Population structure (with optimal population number $K = 2$ indicated by two different colours: red and blue), (B) unrooted phylogenetic tree construction and (C) Principal component analysis (PCA) employing 4957 genome-wide GBS- and candidate gene-derived SNPs classified 60 *desi* and *kabuli* *Cicer* accessions into two major populations (POP I and POP II). In the population genetic structure, the accessions represented by vertical bars along the horizontal axis were categorized into K colour segments according to their estimated membership fraction in each K cluster. In PCA, the PC1 and PC2 explained 18.5 and 7.7% of the total variance, respectively. (D) LD decay (mean r^2) estimated in a population of 60 *Cicer* accessions using 4027 SNPs physically mapped on eight *kabuli* chromosomes. The plotted curved line represents the mean r^2 values among SNP loci spaced with uniform 50 kb physical intervals from 0 to 1000 kb across chromosomes.

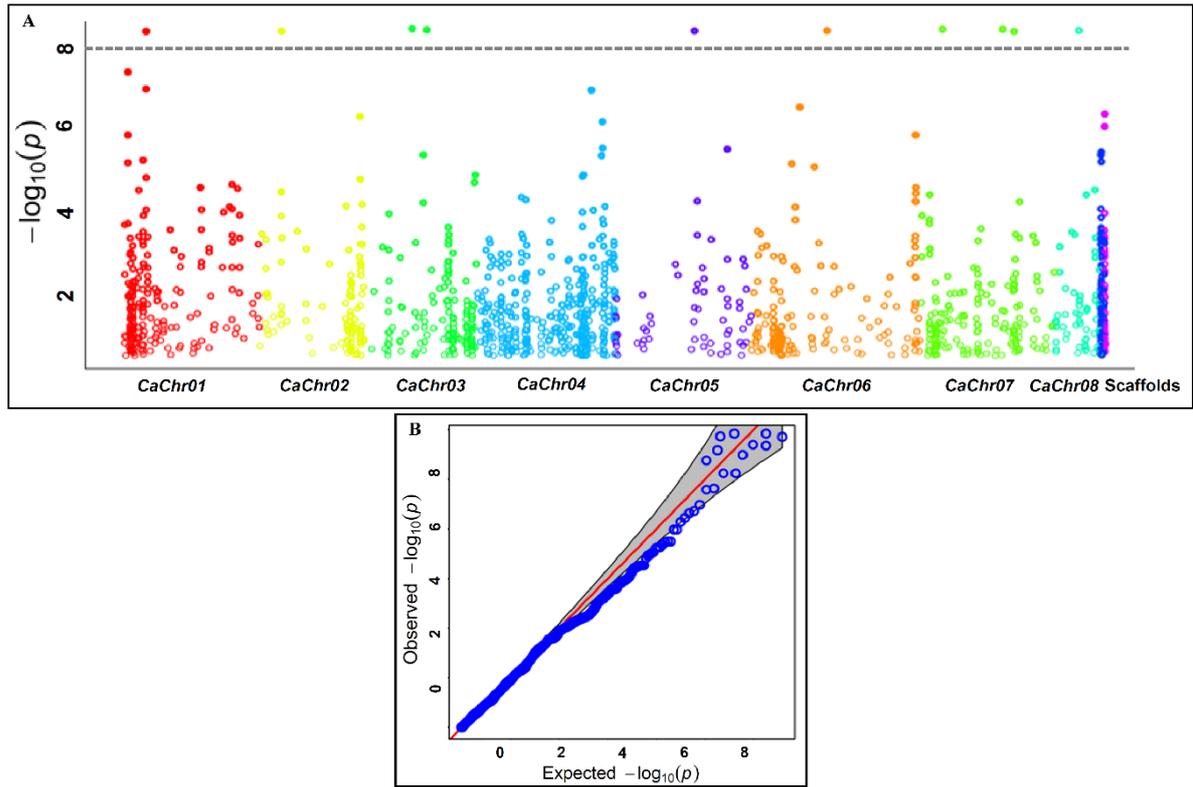


Fig. 3

Fig. 3. (A) GWAS-based Manhattan plot depicting significant P-values (CMLM approach) associated with branch number trait using 4957 genome-wide GBS- and candidate gene-based SNPs. The x-axis signifies the genomic distribution of reference genome-based SNPs physically mapped on eight chromosomes and scaffolds of *kabuli* genome. The y-axis represents the $-\log_{10}(P)$ -value for significant association of SNP loci with branch number trait. The SNPs revealing significant association with branch number trait at cut-off P value $\leq 1 \times 10^{-8}$ are indicated by dotted lines. (B) Quantile-quantile plots depicting the comparison between expected and observed $-\log_{10}(P)$ -values with FDR (false discovery rate) cut-off < 0.05 to scan the significant genomic loci (genes) associated with branch number trait in chickpea.

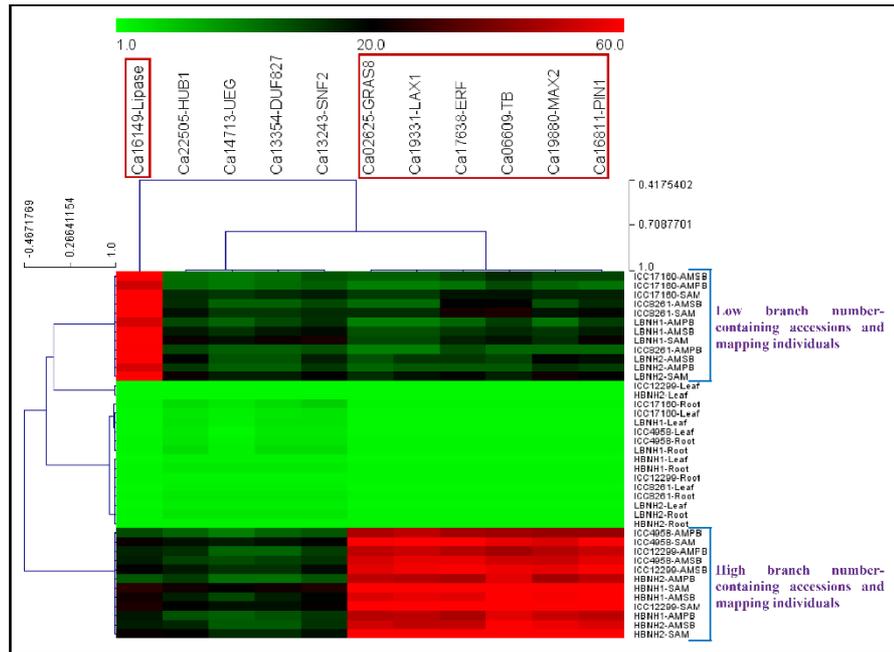


Fig. 4

Fig. 4. Hierarchical cluster display illustrating the differential expression profiles of 11, including seven branch number-associated genes (validated by association analysis and in bi-parental mapping population) in vegetative and reproductive tissues [leaf, root and axillary meristem (AM) at primary and secondary branches differentiation stage during vegetative growth and shoot apical meristem (SAM) of short floret bearing branches during reproductive growth] of low and high branch number-containing parental accessions and one of each homozygous individual derived from two mapping populations [(ICC 4958 x ICC 17160) and (ICC 12299 x ICC 8261) using quantitative RT-PCR assay. The colour scale at the top signifies the average log signal expression values of genes in different tissues; in which green, black and red color denote low, medium and high level of expression, respectively. Genes exhibiting high pronounced differential up-/down-regulation especially in AM and SAM of high branch number-containing accessions/individuals are indicated with red boxes. The details regarding genes are mentioned in the Tables 1 and 2. The tissues and genes used for expression profiling are represented on the right and top side of expression map, respectively. To normalize the expression values across diverse tissues of accessions/individuals, an endogenous control *elongation factor-1 alpha* was utilized in quantitative RT-PCR assay. AMPB: axillary meristem at primary branch and AMSB: axillary meristem at secondary branch differentiation stage, and LBNH: low branch number homozygous and HBNH: high branch number homozygous mapping individuals.

Table 1: Twenty-seven branch/tiller/panicle numbers-regulating known cloned gene orthologs of chickpea selected for trait association mapping

Characterized known cloned genes			<i>Kabuli</i> chickpea gene homologs				
Crop plants	Gene identities		Gene accession IDs	Gene accession IDs	Chromosomes	Sequenced gene amplicon size (bp)	Coding and upstream regulatory SNPs mined
<i>Arabidopsis</i>	<i>ERFs</i>	ethylene-responsive element-binding factors	AT1G01250	Ca17638	<i>Ca_Kabuli_Chr07</i>	3552	12
<i>Arabidopsis</i>	<i>GA2oxs</i>	gibberellin-2 oxidases	AT1G02400	Ca03051	<i>Ca_Kabuli_Chr07</i>	3981	14
<i>Arabidopsis</i>	<i>GSK3</i>	Glycogen synthase kinase 2	AT1G06390	Ca06987	<i>Ca_Kabuli_Chr01</i>	4263	15
<i>Arabidopsis</i>	<i>DRIP1</i>	Dehydration-responsive element binding protein 2A (DREB2A)-Interacting protein 1	AT1G06770	Ca20466	<i>Ca_Kabuli_Chr06</i>	3954	13
<i>Arabidopsis</i>	<i>GRAS8</i>	Gibberellic acid insensitive (GAI), Repressor of ga1-3 (RGA) and Scarecrow (SCR) 8	AT1G07520	Ca02625	<i>Ca_Kabuli_Chr01</i>	5262	18
<i>Arabidopsis</i>	<i>miR393</i>	microRNA 393	AT1G12820	Ca05981	<i>Ca_Kabuli_Chr03</i>	4716	15
<i>Arabidopsis</i>	<i>TPL/REL2</i>	TOPLESS/RAMOSIA 1-enhancer locus 2	AT1G15750	Ca02899	<i>Ca_Kabuli_Chr01</i>	6450	22
<i>Arabidopsis</i>	<i>APO1-F-box</i>	Aberrant panicle organization 1-F- Box	AT1G30950	Ca25518	<i>Ca_Kabuli_Scaffold1369</i>	4329	14
<i>Arabidopsis</i>	<i>MOC1-GRAS</i>	Monoculm 1-Gibberellic acid insensitive (GAI), Repressor of ga1-3 (RGA) and Scarecrow (SCR)	AT1G55580	Ca17862	<i>Ca_Kabuli_Chr08</i>	5658	19
<i>Arabidopsis</i>	<i>PIN1</i>	PIN-FORMED protein 1	AT1G73590	Ca16811	<i>Ca_Kabuli_Chr08</i>	4830	16
<i>Arabidopsis</i>	<i>NSP1/NSP2</i>	Nodulation signalling pathway 1/Nodulation signalling pathway 2	AT2G33070	Ca13569	<i>Ca_Kabuli_Chr01</i>	3978	13
<i>Arabidopsis</i>	<i>MAX2</i>	More axillary growth 2	AT2G42620	Ca19880	<i>Ca_Kabuli_Chr06</i>	5133	17
<i>Arabidopsis</i>	<i>HUB1</i>	Histone monoubiquitination 1	AT2G44950	Ca22505	<i>Ca_Kabuli_Chr03</i>	5637	19
<i>Arabidopsis</i>	<i>MAX3</i>	More axillary growth 3	AT2G44990	Ca10867	<i>Ca_Kabuli_Chr04</i>	4866	16
<i>Arabidopsis</i>	<i>D14</i>	DWARF14	AT3G03990	Ca18371	<i>Ca_Kabuli_Chr04</i>	3804	13
<i>Arabidopsis</i>	<i>BRC1/TBL1</i>	Branched 1/Teosinte branched 1 like-1	AT3G12060	Ca15881	<i>Ca_Kabuli_Chr07</i>	4656	15
<i>Arabidopsis</i>	<i>TB1</i>	Teosinte branched1	AT3G18550	Ca06609	<i>Ca_Kabuli_Chr07</i>	4044	13
<i>Arabidopsis</i>	<i>GLDH</i>	L-galactono-1,4-lactone dehydrogenase	AT3G47930	Ca26185	<i>Ca_Kabuli_Scaffold1196</i>	4800	16
<i>Arabidopsis</i>	<i>CCS52A</i>	Cell cycle switch 52 A	AT4G11920	Ca08728	<i>Ca_Kabuli_Chr01</i>	4428	14
<i>Arabidopsis</i>	<i>WES1</i>	Indol-3-acetic acid amido synthetase GH3.5	AT4G27260	Ca06937	<i>Ca_Kabuli_Chr01</i>	4845	16
<i>Arabidopsis</i>	<i>SPLs</i>	Squamosa promoter binding likes	AT4G27330	Ca23431	<i>Ca_Kabuli_Chr06</i>	4233	13
<i>Arabidopsis</i>	<i>MAX4</i>	More axillary growth 4	AT4G32810	Ca01909	<i>Ca_Kabuli_Chr05</i>	4623	15
<i>Arabidopsis</i>	<i>BA1/LAX1</i>	Barren stalk 1/ Like Auxin 1	AT5G01240	Ca19331	<i>Ca_Kabuli_Chr03</i>	4470	14
<i>Arabidopsis</i>	<i>TFL1</i>	Terminal flower 1	AT5G03840	Ca00790	<i>Ca_Kabuli_Chr03</i>	3525	10
<i>Arabidopsis</i>	<i>PRLIP8</i>	Pathogenesis-related lipase 8	AT5G50890	Ca22813	<i>Ca_Kabuli_Chr05</i>	4200	16
<i>Arabidopsis</i>	<i>PIN2</i>	PIN-FORMED protein 2	AT5G57090	Ca03207	<i>Ca_Kabuli_Chr07</i>	4881	12
<i>Arabidopsis</i>	<i>LFY</i>	LEAFY	AT5G61850	Ca01386	<i>Ca_Kabuli_Chr05</i>	4188	11

Table 2: Eleven branch number-regulating SNP loci-containing genes identified by association mapping and in traditional bi-parental mapping populations

SNP IDs	Kabuli chromosomes	SNP physical positions (bp)	SNPs	Gene accession IDs	Structural annotation	Known/putative functions	Association analysis	
							P	R2 (%)
*BN_SNP344	<i>Ca_Kabuli_Chr01</i>	11285291	(C/A)	Ca02625	CDS (NSyn)	<i>GRAS8</i> (Gibberellic acid insensitive (GAI), Repressor of <i>ga13</i> (RGA) and Scarecrow (SCR)8	2.1×10^{-9}	10
BN_SNP606	<i>Ca_Kabuli_Chr02</i>	4981666	(G/C)	Ca14713	CDS (NSyn)	Unknown expressed gene	1.7×10^{-8}	5
BN_SNP1024	<i>Ca_Kabuli_Chr03</i>	14653193	(T/G)	Ca22505	Intron	<i>HUB1</i> (Histone monoubiquitination 1)/Zinc finger	1.5×10^{-8}	10
*BN_SNP1	<i>Ca_Kabuli_Chr03</i>	19270907	(A/G)	Ca19331	URR	<i>BA1/LAX1</i> (Barren stalk 1/Like Auxin 1)	1.3×10^{-9}	12
BN_SNP2352	<i>Ca_Kabuli_Chr05</i>	28757467	(A/T)	Ca13354	CDS (NSyn)	<i>DUF827</i> (Domain of unknown function 827)	2.0×10^{-8}	4
*BN_SNP2763	<i>Ca_Kabuli_Chr06</i>	21238711	(C/T)	Ca19880	CDS (NSyn)	<i>MAX2</i> (More axillary growth 2)/F-box leucine-rich repeat domain	1.6×10^{-9}	11
*BN_SNP2	<i>Ca_Kabuli_Chr07</i>	7508884	(T/C)	Ca06609	URR	<i>TBI</i> (Teosinte branched 1)	1.0×10^{-9}	11
BN_SNP3136	<i>Ca_Kabuli_Chr07</i>	7756455	(C/T)	Ca13243	CDS (NSyn)	<i>SNF2</i> (Sucrose Non-Fermenting 2)	2.0×10^{-8}	5
*BN_SNP3347	<i>Ca_Kabuli_Chr07</i>	33182932	(G/A)	Ca16149	CDS (Syn)	Lipase	1.0×10^{-9}	11
*BN_SNP3391	<i>Ca_Kabuli_Chr07</i>	35134863	(T/C)	Ca17638	CDS (Syn)	<i>ERF</i> (ethylene-responsive element-binding factor)	1.4×10^{-9}	10
*BN_SNP3	<i>Ca_Kabuli_Chr08</i>	11544119	(G/A)	Ca16811	URR	<i>PIN1</i> (PIN-FORMED protein 1)	1.1×10^{-9}	10

*Branch number-regulating SNPs-carrying genes validated in two inter (ICC 4958 x ICC 17160)- and intra (ICC 12299 x ICC 8261)-specific mapping population