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

Alice Kujur · Deepak Bajaj · Maneesha S. Saxena · Shailesh Tripathi ·
Hari D. Upadhyaya · C. L. L. Gowda · Sube Singh · Akhilesh K. Tyagi ·
Mukesh Jain · Swarup K. Parida

Abstract The large-scale validation and high-throughput genotyping of numerous informative genic microsatellite markers are required for association mapping to identify candidate genes for complex quantitative traits in chickpea. However, the screening and genotyping of such informative markers in individual genotypes/whole association panels for trait association mapping involves massive costs in terms of resources, time and labour due to low genetic polymorphism in chickpea. We have developed an alternative time-saving and cost-effective pool-based trait association mapping approach by combining pooled DNA analysis (with 616 genic microsatellite markers) and individual genotype (large structured association panel) genotyping. Using this approach we have identified seven seed weight-associated transcription factor gene-derived microsatellite markers (with minor allele frequency >15 %) in *desi* and *kabuli* chickpea. Strong marker allele effects of these five transcription factors with increasing seed weight in the contrasting *desi* and *kabuli* genotypes were evident. Bi-parental linkage mapping using 241 of the informative gene-based microsatellite markers resulted in the identification and mapping of nine such markers linked with three major quantitative trait loci (explaining a total phenotypic variance of 23.5–34.7 %) on chromosomes 1 (CaqSW1.1: 73.5–74.5 cM and CaqSW1.2: 79.3–81.3 cM) and 2 (CaqSW2.1: 65.7–67.5 cM) controlling 100-seed weight in chickpea. The integration of pool-based trait association mapping with differential expression profiling, traditional bi-parental linkage mapping and high-resolution microsatellite-single nucleotide polymorphism marker-based haplotyping/linkage disequilibrium mapping delineated four transcription factor genes (DUF3594, bZIP, DUF1635 and SBP) controlling seed weight in *desi* and *kabuli* chickpea. The strategies implemented in our study can be used in large-scale trait association mapping for the rapid identification of candidate genes and in the development of functional markers for traits of agricultural importance in crop species including chickpea.

Keywords Association mapping · *Desi* · *Kabuli* · Microsatellite · Seed weight · SNP

A. Kujur · D. Bajaj · M. S. Saxena · A. K. Tyagi ·
M. Jain · S. 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

S. Tripathi
Division of Genetics, Indian Agricultural Research
Institute (IARI), New Delhi 110012, India

H. D. Upadhyaya · C. L. L. Gowda · S. Singh
International Crops Research Institute for the Semi-Arid
Tropics (ICRISAT), Patancheru 502324, Andhra Pradesh,
India

Introduction

The genome-wide and candidate gene-based association analysis/linkage disequilibrium (LD) mapping strategy involving the use of informative microsatellite and single nucleotide polymorphism (SNPs) markers has proven to be an effective approach for dissecting the simple qualitative and complex quantitative traits in many plant species, including legumes (Jun et al. 2008; Zhu et al. 2008; Yan et al. 2009; Atwell et al. 2010; Cockram et al. 2010; Hall et al. 2010; Huang et al. 2010; Shi et al. 2010; Hou et al. 2011; Li et al. 2011; Zhao et al. 2011; Jia et al. 2012; Riedelsheimer et al. 2012; Sakiroglu et al. 2012; Upadhyaya et al. 2012). In chickpea, as has been found in other self-pollinated crop species, extended chromosomal LD decay along with low-resolution genome-wide LD mapping is anticipated. Consequently, candidate gene-based association mapping could be an attractive, rapid, and inexpensive approach for identifying markers associated with genes/quantitative trait loci (QTLs) controlling the agronomic traits in chickpea (Haseneyer et al. 2010; Huang et al. 2010; Zhao et al. 2011; Varshney et al. 2012; Kujur et al. 2013). However, the success of any systematic analysis and of such a candidate gene-based association mapping/high-resolution gene-specific LD mapping approach specifically for identifying complex quantitative traits in crops like chickpea requires large-scale validation and high-throughput genotyping of numerous genic microsatellite and SNP markers derived from different coding and non-coding (upstream regulatory regions) sequence components of the genes among a phenotypically and genotypically diverse contrasting germplasm sets/association panel. In recent years, the availability of highly multiplexed, efficient, robust and high-throughput next-generation sequencing and genotyping platforms, such as the Roche 454 Pyrosequncer, ABI SOLiD, Illumina Solexa Genome Analyser, ABI Automated DNA fragment Analyser (Applied Biosystems, Verson Hills IL; http://www.appliedbiosystems.com), Illumina GoldenGate/Infinium assay (Illumina, San Diego, CA; http://www.illumina.com), Competitive Allele Specific PCR (KASPar) (KBiosciences, Hertfordshire, UK; http://www.kbioscience.co.uk) and restriction site-associated (RAD)-based genotyping by sequencing (GBS) assay have specifically expedited large-scale genome-wide microsatellite and SNP marker validation and genotyping in many useful genes encoding expressed sequence tags (ESTs)/transcripts of chickpea (Nayak et al. 2010; Gujaria et al. 2011; Hiremath et al. 2011, 2012; Choudhary et al. 2012; Jain et al. 2013; Varshney et al. 2013). Among the assays currently available, microsatellite marker-based automated fragment analysis usually involves relatively high costs due to the fluorescent dye labeling of numerous primers and the need for specific commercial size standards for large-scale genotyping. In addition, the deployment of such a high-throughput microsatellite marker genotyping assay requires laboratories with sophisticated infrastructural facilities which are generally not affordable for most research groups. Further, due to the narrow genetic base in chickpea, large-scale genotyping of genic microsatellite markers which are derived from the conserved transcribed component of the genome usually reveals low (20–30 %) intra- and inter-specific polymorphism even in a high-resolution automated fragment analyser (Choudhary et al. 2009, 2012; Nayak et al. 2010; Gujaria et al. 2011; Hiremath et al. 2011). Therefore, the high costs involved in high-throughput genotyping of a large proportion (70–80 %) of monomorphic genic microsatellite markers, which remain further unutilized for genetic studies, including trait association mapping in chickpea, appear to be a suboptimal use of resources, time and labour. Therefore, an efficient strategy is needed to reduce the costs of genotyping, particularly the costs of genotyping larger monomorphic microsatellite markers, and to screen a smaller set of appropriate and informative polymorphic markers which can be further utilized in large-scale validation and high-throughput genotyping for efficient trait association mapping in chickpea.

An integrated approach of DNA pooling and individual genotyping is a less time-consuming, non-labourious and cost-effective strategy which has been implemented successfully in large-scale genome-wide association studies for identifying potential loci associated with qualitative and complex quantitative traits, including diseases such as schizophrenia, rheumatoid arthritis and bipolar disorder in humans (Craig et al. 2005; Steer et al. 2007; Pearson et al. 2007; Abraham et al. 2008; Baum et al. 2008; Shifman et al. 2008; Kirov et al. 2009; Zhao and Wang 2009; Diergaarde et al. 2010; Earp et al. 2011) and plant height and saccharification yield in sorghum (Wang et al. 2011, 2012). The utility of such integrated DNA pooling and individual genotyping studies has been well understood in trait association analysis: firstly, as a means to

efficiently screen polymorphic and informative markers by genotyping the larger marker sets in at least two pooled DNA, with each pool representing diverse genotypes of two different contrasting phenotypic traits, and secondly, through the genotyping of individuals constituting the trait-specific pools with selected polymorphic markers (Bansal et al. 2002; Sham et al. 2002; Wang et al. 2011, 2012). However, the advantages of pool-based association studies are partially offset due to a major concern on complications/ambiguity associated with the estimation of microsatellite allele frequencies based on genotyping information from pooled DNA which contains numerous multiple and variable sized alleles in contrast to genotyping data on individual genotypes. In diploid and self-pollinated crop species like chickpea, with low intra- and inter-specific allelic variation (ranging from two to four alleles), particularly at genic microsatellite marker loci (Choudhary et al. 2009, 2012; Nayak et al. 2010; Gujaria et al. 2011; Hiremath et al. 2011), the estimation of microsatellite allele frequencies from pooled DNA genotyping information would be relatively easier and simpler, even in a low-resolution agarose gel, than in other diploid/polyploid crop species showing higher genetic polymorphism by amplifying a large number of alleles at single loci and/multiple loci. Further, confirmation of the results of pooled DNA analysis through the genotyping of individual genotypes that constituted the pools in the agarose gel following automated fragment analysis would certainly provide more robust genotyping and precise allelic information, which could be utilized for trait association mapping in chickpea. Using this DNA pooling strategy in chickpea, a large proportion of monomorphic genic microsatellite markers that are ultimately of no use in trait association mapping could be screened effectively, rapidly, and easily in an agarose gel, thereby reducing the time, cost and labour involved in individual sample-by-sample genotyping of all the markers. Consequently, the candidate gene-based association mapping strategy, by integrating the pooled DNA analysis with genotyping of individual genotypes/association panel using informative genic microsatellite markers, would be of great relevance and could expedite the large-scale marker validation and high-throughput genotyping, including trait association mapping, to pin down the causal genes for simple qualitative and complex quantitative traits in *desi* and *kabuli* chickpea.

The aim of the study reported here was, therefore, undertaken to evaluate the potential of integrated pooled DNA analysis and the individual genotyping approach in large-scale candidate gene-based association mapping using a freshly constituted association panel for the rapid identification of gene-derived microsatellite markers associated with 100-seed weight in chickpea. The screened seed weight-associated genic markers were further validated through traditional bi-parental linkage mapping, differential expression profiling and high-resolution gene-specific molecular haplotyping/LD mapping to identify the most informative functionally relevant candidate genes controlling seed weight in *desi* and *kabuli* chickpea.

Materials and methods

Constitution of the association panel

We previously utilized 96 germplasm lines belonging to a seed- and pod trait-specific association panel to identify transcription factor genes associated with seed size/weight, pod number per plant and seed number per plant in chickpea (Kujur et al. 2013). In the present study, for more precise identification of the transcription factor genes specifically associated with 100-seed weight, we constructed a new association panel by selecting a large number of genotypically and phenotypically diverse *desi* and *kabuli* germplasm lines from available global germplasm collections of chickpea. This set ultimately consisted of 242 germplasm lines, including 211 minicore germplasm lines (representing >85 % diversity of total germplasm lines evaluated), which were selected from 16,991 accessions (1,956 chickpea core germplasm lines representing diverse eco-geographical regions of 58 countries of the world). The selection criteria were based on multi-locational replicated field evaluation phenotypic data on 22 morphological and seed/pod yield contributing to agronomic traits, including 100-seed weight (Upadhyaya and Ortiz 2001; Upadhyaya et al. 2001, 2002). In addition, in order to validate the seed weight/size-associated genes further through traditional bi-parental linkage mapping and differential expression profiling, we preferentially included two contrasting small and large seed size/weight chickpea genotypes namely, ICCX-810800

(*desi* small, 100-seed weight 11 g) and ICC 20268 (*kabuli* large, 47 g), in the association panel; the F_3 mapping population (ICCX-810800 × ICC 20268) and tissues of two different seed developmental stages, respectively, are available for these two genotypes. The phenotypic evaluation data specifically for 100-seed weight in the 244 genotypes and genotyping information on 96 genomic microsatellite markers distributed over eight chromosomes/linkage groups (LGs) of chickpea (Winter et al. 1999) were analyzed in PowerCore (Kim et al. 2007). To ascertain the phenotypic and genotypic diversity and homogeneity of the constituted association panel, we compared the level of genetic diversity estimated among genotypes with that of the original reference core and minicore collections (Upadhyaya et al. 2008) based on the standard genetic diversity coefficient (Nei 1972) using PowerMarker V3.51 (Liu and Muse 2005). Based on the results of these analyses, 244 germplasm lines, including 67 contrasting small and large seed size/ weight *desi* and *kabuli* genotypes (with 100-seed weight varying from 5.9 to 57.6 g), representing diverse eco-geographical regions of 29 countries of the world were selected to constitute a seed weight trait-specific association panel in chickpea [Electronic Supplementary Material (ESM) Table S1]. Higher phenotypic diversity for 100-seed weight in the association panel comprising 167 *desi* (5.9–35.4 g) and 77 *kabuli* (14.4–57.6 g) genotypes was evident and is thus significant for seed weight-specific genetic association analysis in chickpea.

Genic microsatellite marker-based pooled DNA analysis and individual genotyping

Total genomic DNA isolated from the 244 genotypes constituting the association panel was purified and the quality (intactness, A_{260}/A_{280}) and quantity (concentration) assessed using both spectrophotometric and fluorospectrophotometric assays followed by agarose gel-based visual estimation. After quantification, the high-quality ($A_{260}/A_{280} = 1.8$–1.9) and intact genomic DNA isolated from genotypes was diluted to equal concentration of 50 ng/μl for further construction of pools. To construct the DNA pools, the genotypes were first classified into three distinct groups (small, medium and large) separately for each *desi* and *kabuli* germplasm based on their 100-seed weight-specific traits. Using this criterion, the equal amount and equal

concentration of high-quality genomic DNA isolated from individual genotypes was bulked to constitute the two pools in *desi* as "*desi* small" and "*desi* large" and another two pools in *kabuli* as "*kabuli* small" and "*kabuli* large" for initial screening of polymorphic genic microsatellite markers. The four DNA pools of *desi* and *kabuli* were quantified and diluted to an equal concentration (50 ng/μl) following the methods described above and used further for the genotyping of genic microsatellite markers.

A selected set of informative 616 genic microsatellite markers was designed from the available gene-encoding transcript sequences (Garg et al. 2011) of chickpea *desi* genotype ICC 4958 (ESM Table S2). This set includes 321 informative genic microsatellite markers (TFGMS and TFFDMS) developed from the transcription factor (TF) gene-encoding transcripts of ICC 4958 (Kujur et al. 2013), as well as 295 gene-encoding transcript sequence-derived (other than TF genes) microsatellite markers showing *in silico* polymorphism between ICC 4958 (*desi Cicer arietinum*) and PI 489777 (wild *C. reticulatum*) based on variation in the microsatellite repeats. The genomic distribution of these 616 genic microsatellite markers on the chickpea genome was determined (BLAST search at $E = 0$) based on their physical positions (basepairs, bp) on the corresponding protein-coding genes annotated on the eight chromosomes as per the latest released draft genome pseudomolecule V1.1 (Varshney et al. 2013). The developed genic markers were amplified by PCR (Jhanwar et al. 2012) in the genomic DNA of four small and large seed size *desi* and *kabuli* DNA pools along with the two chickpea genotypes ICC 4958 (*Cicer arietinum*) and PI 489777 (wild *C. reticulatum*) as controls. The standard PCR constituents and optimum cyclic conditions (Jhanwar et al. 2012), except for annealing temperature (depending on the primers used, see ESM Table S2) were used for PCR amplification. The PCR products of each genic microsatellite marker amplified from the four DNA pools were resolved in a 3.5 % metaphor agarose gel-based assay, and their allele-sizing was performed against 50-bp DNA ladder following the protocol described by Wang et al. (2012) and Sham et al. (2002). The fragment size of alleles amplified by genic microsatellite markers in four DNA pools was confirmed by analysis in an automated fragment analyser using the fluorescent dye-labeled primers following the procedure of Kujur et al. (2013). A set of

genic microsatellite markers showing polymorphism among the four small and large *desi* and *kabuli* DNA pools was further validated in individual genotypes that constituted the pools using both a 3.5 % metaphor agarose gel-based assay and the automated fragment analyser. A selected set of well-validated informative fluorescent dye-labeled genic microsatellite markers were genotyped in the association panel using the automated 96 capillary ABI 3730xl DNA Analyser (Applied Biosystems). The actual allele size (bp) and fragment length polymorphism (bp) detected by the genic microsatellite markers among genotypes was determined. The genotyping data on the informative genic microsatellite markers validated in both the gel-based assay and automated fragment analysis was used to estimate the total number of alleles amplified, average polymorphic alleles per marker loci, major and minor allele frequency and polymorphism information content in the association panel employing the PowerMarker ver. 3.51 (Liu and Muse 2005).

Genetic association analysis

The population structure among the 244 germplasm lines, including 67 contrasting seed weight/size genotypes (constituted association panel), was determined using a selected set of 158 markers, including 62 genic and 96 previously reported genomic microsatellite markers (Winter et al. 1999) distributed over eight LGs/chromosomes of chickpea. For determining the population structure, the marker genotyping information among genotypes (association panel) were analyzed in STRUCTURE (Pritchard et al. 2000) using the methods described by Kujur et al. (2013). The population structure model representing better relationships among the genotypes using the optimum K (number of populations) was constructed.

For the trait association analysis, the genotyping and population structure data of informative genic and genomic microsatellite markers [minor allele frequency (MAF) >0.05] distributed over eight LGs/chromosomes of chickpea and 100-seed weight-specific phenotypic information of association panel were analyzed in TASSEL (Trait Analysis by Association, Evolution and Linkage; http://www.maizegenetics.net; Bradbury et al. 2007) based on the general linear model (GLM) and mixed linear model (MLM) following the detailed methods described by Kujur et al. (2013). The genic microsatellite markers

showing an association with 100-seed weight in the *desi* and *kabuli* genotypes at a significant cut-off P_adjusted value of ≤0.05 (with 95 % confidence interval) and R^2 value ≥0.10 were ultimately selected using the combined results of GLM and MLM analyses.

Differential gene expression profiling

The expression profiling of seed weight-specific microsatellite marker-associated genes was performed to correlate the trait association potential of genic microsatellite markers with differential expression profiling. These marker-associated genes along with internal control elongation factor 1-alpha (*EF1α*) were amplified in the RNA isolated from two different seed developmental stages [early cell division phase during the 10–20 days after podding (DAP) and late maturation phase during the 21–30 days after DAP] of four contrasting small and large seed size/weight chickpea genotypes, namely, ICCX-810800 (*desi*, 100-seed weight 11 g), ICC 5590 (*desi*, 5.9 g), ICC 20268 (*kabuli*, 47 g) and ICC 8155 (*kabuli*, 57.3 g) using both semi-quantitative and SYBR green chemistry-based real time quantitative reverse transcription (RT)-PCR assays. The expression levels of the seed weight-associated genic microsatellite markers obtained in diverse seed developmental stages of the four contrasting chickpea genotypes were correlated with each other and with the control (leaf at vegetative stage of the respective genotypes). Based on the results of these analyses, we were able to identify those seed weight-specific microsatellite marker-associated genes showing differential expression during seed development in the contrasting chickpea genotypes relative to the leaf. The cDNA and genomic DNA fragments amplified from seed weight-specific microsatellite marker-associated genes in the four contrasting genotypes (ICCX-810800, ICC 5590, ICC 20268 and ICC 8155) were cloned and sequenced in both the forward and reverse directions twice on a capillary-based Automated DNA Sequencer (ABI 3730xl DNA Analyser; Applied Biosystems) using the BigDye Terminator v3.1 Sequencing kit and M13 forward and reverse primers. The high-quality consensus sequences thus obtained for each genic marker were aligned and compared among the four chickpea genotypes to infer the presence of predicted microsatellite repeat-motifs and the correspondence of

differential expression profiling with fragment length polymorphisms based on microsatellite repeat length expansion/contraction in these genotypes.

Validation of trait-associated genic microsatellite markers in the bi-parental mapping population

The genic microsatellite markers showing parental polymorphism between *desi* (ICCX-810800) and *kabuli* (ICC 20268) were genotyped using the genomic DNA isolated from 94 segregating individuals of a F$_3$ mapping population (ICCX-810800 × ICC 20268) to construct a transcript map (genetic linkage map) of chickpea employing the procedures of Kujur et al. (2013). The genetic map constructed in this study using 113 additional novel transcript sequence-derived (other than TF genes) genic microsatellite markers was integrated with our previously constructed transcript map of a similar bi-parental mapping population that included 32 genomic and 96 TF gene-derived microsatellite markers (TFGMS and TFFDMS) (Kujur et al. 2013). For performing genetic/QTL mapping, we correlated the integrated genetic linkage map information and genotyping data of 241 genic and genomic microsatellite markers mapped on the eight LGs/chromosomes of chickpea and 100-seed weight/size trait-specific phenotyping information of segregating individuals and parental genotypes using the methods described by Kujur et al. (2013) to identify and map the significant seed weight-specific microsatellite marker-associated genes underlying QTL regions on the chickpea chromosomes. The genetic positions [centiMorgans (cM)] of the seed weight-associated genic microsatellite markers on the eight LGs/chromosomes of chickpea along with their genotyping information in ten each of small (5–15 g) and large (41–50 g) seed size/weight homozygous F$_3$ individuals and two parental genotypes were visualized in Graphical GenoTypes (Van Berloo 1999) ver. 2.0 to further validate these seed weight-specific microsatellite marker-associated genes by traditional bi-parental linkage mapping.

Gene-specific haplotyping and LD mapping

To obtain a better understanding of the association potential of candidate genes for seed weight, we cloned and sequenced the seed weight-specific

Fig. 1 Eight major steps followed in the integrated approach of ▶ pool-based trait association mapping, differential expression profiling, traditional bi-parental linkage mapping and high-resolution marker haplotyping/linkage disequilibrium (*LD*) mapping for the identification of functionally relevant microsatellite markers in the genes controlling seed weight in *desi* and *kabuli* chickpea. *SNP* single nucleotide polymorphism, *M* 50-bp DNA size standard

microsatellite marker-associated genes of the 244 genotypes belonging to the association panel using the methods described above. The high-quality sequences generated for each gene were aligned among genotypes using the CLUSTALW multiple sequence alignment tool in MEGA 4.0 (Tamura et al. 2011) and the SNP loci were mined. SNP and microsatellite marker genotyping information (MAF > 0.05) generated in the genes was used to constitute haplotypes and determine haplotype diversity and LD patterns among the *desi* and *kabuli* chickpea genotypes. The genic SNP and microsatellite marker-based haplotype genotyping information was further correlated with 100-seed weight-specific phenotypic information of the 244 genotypes for trait association analysis and high-resolution LD mapping using TASSEL as described earlier (Kujur et al. 2013). The overall integrated strategies combining the pool-based trait association mapping with differential gene expression profiling, traditional bi-parental linkage mapping and high-resolution gene-specific molecular haplotyping/ LD mapping followed in our study to identify functionally relevant microsatellite markers in the genes controlling seed weight of *desi* and *kabuli* chickpea are diagrammatically summarized in a stepwise manner in Fig. 1.

Results

Constitution and screening of DNA pools with genic microsatellite markers

A new association panel containing 244 germplasm lines, including 67 contrasting small and large seed size/weight *desi* and *kabuli* genotypes (with 100-seed weight varying from 5.9 to 57.6 g), was constituted (ESM Table S1). These genotypes were phenotyped precisely for 100-seed weight and different seed and pod yield contributing to agronomic traits, including plant height, days to flowering and maturity, pod
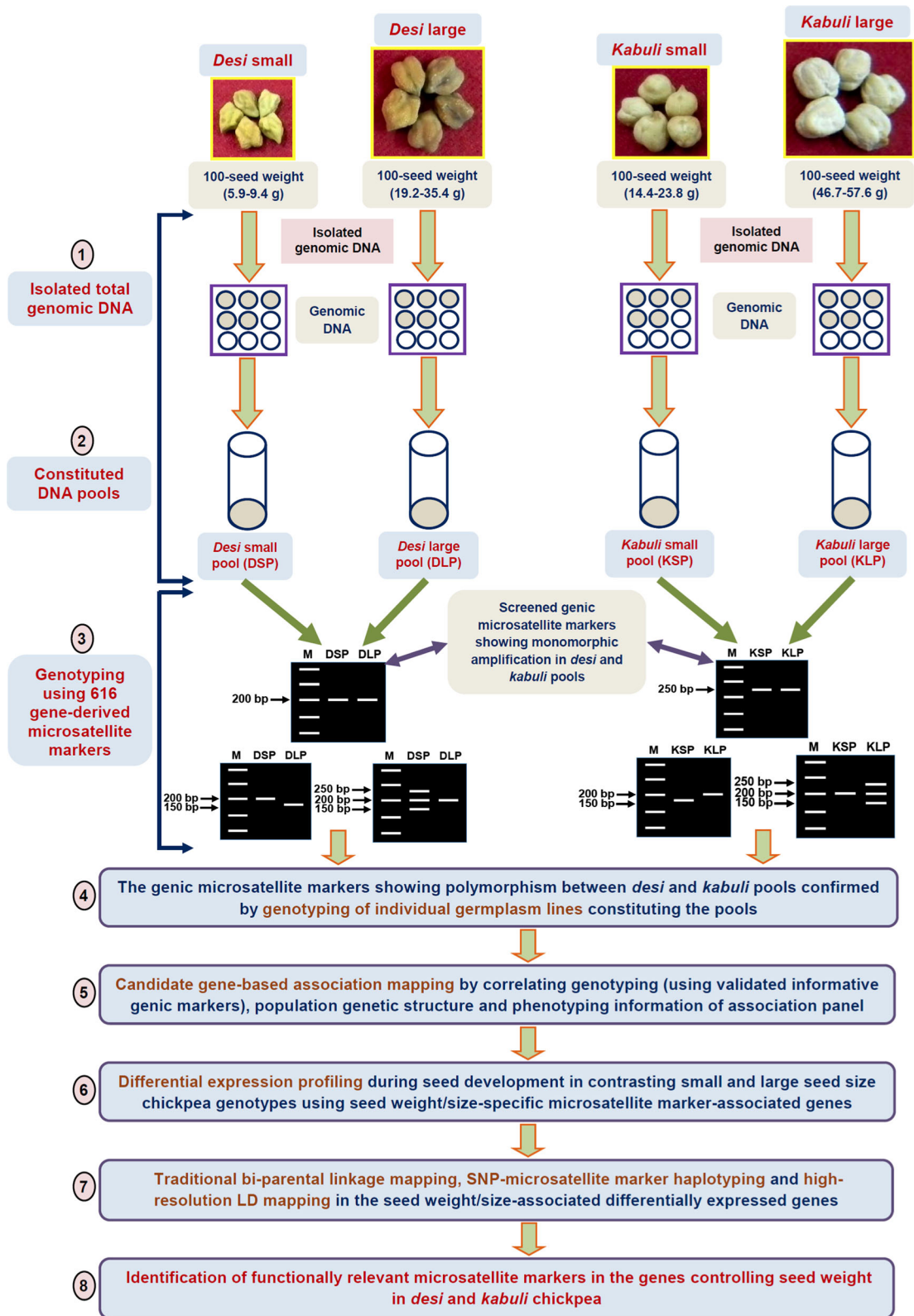
number per plant and seed number per pod. The detailed phenotypic evaluation based on Pearson correlation analysis demonstrated a wider level of both positive and negative correlation (-0.18 to 0.94, mean 0.41) among the seed and pod yield component traits in the 244 *desi* and *kabuli* genotypes belonging to the seed weight-specific association panel (ESM Fig. S1). Specifically, the 100-seed weight in the 167 *desi* and 77 *kabuli* genotypes of the association panel varied from 5.9 to 35.4 g and from 14.4 to 57.6 g, respectively. The germplams lines of the seed weight-specific association panel were further classified and categorized based on the 100-seed weight (g) of their phenotypes. Five genotypes of each contrasting small and large seed size/seed weight *desi* and *kabuli* genotypes were selected from the association panel and grouped as *desi* small (5.9–9.4 g), *desi* large (19.2–35.4 g), *kabuli* small (14.4–23.8 g) and *kabuli* large (46.7–57.6 g) pools (Fig. 1). The mean seed weight estimated for the small and large *desi* pools was 7.2 and 24.7 g, respectively; in comparison, the mean seed weight was 14.2 g in the entire 167 *desi* association mapping panel. For the 77 *kabuli* genotypes included in the association mapping panel, the average seed weight was 33.7 g; in comparison, in the small and large *kabuli* pools, it was 22.1 and 52.7 g, respectively.

The genomic DNA isolated from the four small and large pools of *desi* and *kabuli* were genotyped using 616 genic markers, including 321 TF gene-derived microsatellite markers of chickpea (ESM Table S2). Of these, 574 (93.2 %) were physically mapped on eight chickpea chromosomes with an average map density of 604.9 kb (ESM Table S3; ESM Fig. S2). The average marker density varied from 457.8 kb in chromosome 8 (36 markers) to 789.7 kb in chromosome 5 (61). Of the 321 microsatellite markers in the TF genes that produced clear reproducible amplicons with fragments of the expected product size in ICC4958 and PI489777, 94 revealed polymorphism either between the small and large *desi* pools and/or small and large *kabuli* pools (at least two combinations of DNA pools) using both the agarose gel-based assay and automated fragment analyser. Similarly, 143 (48.5 %) of the remaining 295 microsatellite markers designed from the gene-encoding transcripts other than TFs showed polymorphism among the four small and large *desi* and *kabuli* DNA pools. Remarkably, 227 (70.7 %) microsatellite markers in the TF genes

and 152 (51.5 %) markers in the transcripts other than TF genes amplified amplicons of a similar size (bp) and unique single alleles among all the four small and large *desi* and *kabuli* DNA pools using both metaphor agarose gel analysis and the automated fragment analyser (Fig. 2a, b); these were therefore screened as monomorphic genic microsatellite markers having no potential for seed weight-specific association mapping. Based on the diverse nature of the amplification and polymorphism patterns of the microsatellite markers in the four *desi* and *kabuli* pools, we selected 237 (38.5 %) informative genic microsatellite markers. In most of the cases, the screened genic microsatellite markers produced multiple amplicons with more than one allele but they showed clear differentiation either between small and large *desi* and/or small and large *kabuli* pools based on fragment length polymorphism in the gel analysis (Fig. 2c–h). These selected markers were fluorescent dye-labeled and validated further through automated fragment analysis. The actual allele size amplified by the fluorescent dye-labeled genic microsatellite markers and their fragment length polymorphism among the four *desi* and *kabuli* pools were estimated precisely (Fig. 2c–h). Using the above screening methods for the DNA pools, the selected 237 informative genic microsatellite markers were subsequently used to validate individual genotype genotyping followed by seed weight-specific association analysis.

Individual genotyping and genetic association mapping

The 237 informative genic microsatellite markers identified using the pool-based assay were genotyped in the genomic DNA isolated from 20 individual genotypes that constituted the four small and large *desi* and *kabuli* pools using both the gel-based assay and the automated fragment analyser. The detailed data thus obtained enabled validation of the 88 markers that showed allelic polymorphism between at least two combinations of contrasting small and large seed weight *desi* and *kabuli* genotypes based on fragment length polymorphism (bp) in the gel-based assay and precise allele sizing in the automated fragment analyser (Fig. 3; ESM Fig. S3). For example, a microsatellite marker in the DUF3594 TF gene amplifying two differently sized alleles showed a distinct fragment length polymorphism between the
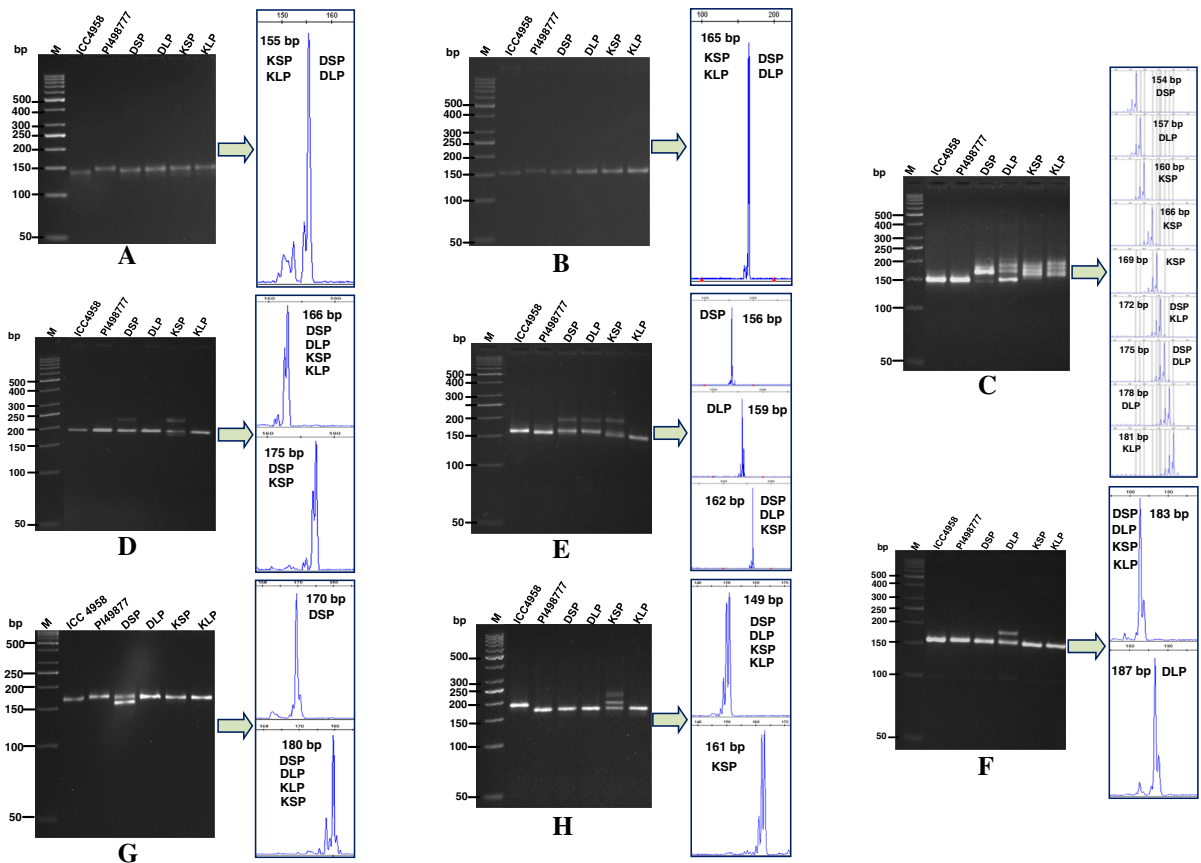
Fig. 2 Different amplification and polymorphism patterns and distribution of various allele types detected through pooled DNA analysis by the genotyping of four contrasting *desi* and *kabuli* DNA pools in terms of seed weight/size with the genic microsatellite markers using both gel-based assay and auto-mated fragment analyser. **a**, **b** Microsatellite markers in the AP2 and MADS TF genes showing monomorphic amplification, **c**, **d** markers in the bZIP transcription factor (TF) and Med15 genes showing fragment length polymorphism between DSP versus DLP and between KSP versus KLP, respectively, **e** marker in the GRAS TF gene showing fragment length polymorphism and differentiating DSP and DLP from KSP and KLP, **f**, **g** markers in the Myb and DUF3594 TF genes showing fragment length polymorphism between DSP and DLP, respectively, **h** marker in the bHLH TF gene showing fragment length polymorphism between KSP and KLP. The two chickpea genotypes ICC4958 and PI489777 from which transcript sequences the primers were designed were used as controls for evaluating the amplification efficiency of genic microsatellite markers. *DSP Desi* small pool, *DLP Desi* large pool, *KSP Kabuli* small pool, *KLP Kabuli* large pool, *M* 50-bp DNA size standard. The fragment sizes (bp) of all the polymorphic alleles detected by automated fragment analyser are indicated

desi small (170 bp) and *desi* large (180 bp) pools in the agarose gel-based assay and in the automated fragment analyser (Figs. 2g, 3a). The validation of this marker further by the genotyping of individual geno-types in the agarose gel-based assay and the automated fragment analyser revealed the presence of 170-bp alleles specific to the *desi* small pool in three (60 %) of the five individual genotypes that constituted the *desi* small pool (Fig. 3a). Of the remaining two individual genotypes from the *desi* small pool, one contained 180-bp alleles specific to the *desi* large pool and

another amplified both 170- and 180-bp heterozygous alleles at the marker loci. All five individual genotypes constituting the *desi* large pool contained the 180-bp alleles specific to the *desi* large pool as expected (Fig. 3a). The microsatellite marker in the bZIP TF gene amplifying multiple alleles in the four DNA pools, as shown in the gel-based assay, confirmed the presence of the nine distinct alleles of various fragment sizes in the individual *desi* and *kabuli* genotypes constituting the pools as shown by the automated fragment analyser (Fig. 3b). The

microsatellite marker in the GRAS TF gene amplified three different polymorphic alleles (156, 159 and 162 bp) that differentiated the *desi* small and large pools from the *kabuli* small and large DNA pools, showing a 40 % correlation with the alleles amplified from 20 individual germplam lines constituting the pools (ESM Fig. S3A). Interestingly, the microsatellite marker in the Med15 gene that amplified 166- and 175-bp polymorphic alleles in the four *desi* and *kabuli* DNA pools (Fig. 2d) did not show any correspondence with the fragment size of alleles amplified in the individual germplam lines constituting the pools; the later were therefore not considered further for trait association analysis (ESM Fig. S3B). Another microsatellite marker in the bHLH TF gene amplifying two different polymorphic alleles differentiated the *kabuli* small pool (161 bp) from the *kabuli* large pool (149 bp) (Fig. 2h). Two (40 %) of the five individual genotypes constituting the *kabuli* small pool amplified the 161-bp alleles specific to *kabuli* small, while the remaining three genotypes produced 149-bp alleles specific to *kabuli* large (ESM Fig. S4). As per expectation, all of the five individual genotypes constituting the *kabuli* large pool contained the 149-bp alleles specific to the *kabuli* large types. Based on the efficiency of the markers to discriminate the contrasting small and large seed weight/size genotypes belonging to each of the *desi* and *kabuli* genotypes, the 62 most informative genic microsatellite markers were selected and genotyped in the 244 genotypes of the association panel using the automated fragment analyser for seed weight-specific association analysis. The 62 polymorphic genic microsatellite markers overall detected two to nine alleles (average 4.2 alleles) per locus, with a total of 259 alleles (average PIC 0.76) among the 244 genotypes of the association panel. The MAF per microsatellite marker locus varied from 0.03 to 0.29, with an average of 0.13.

For genetic association mapping, the population genetic structure among the 244 genotypes, including 67 contrasting seed weight genotypes of the association panel, based on 62 selected informative genic and 96 genomic microsatellite markers (distributed over eight LGs/chromosomes of chickpea) was determined using the STRUCTURE with varying levels of K ($K = 2$–3) with 20 replications. The optimization of K from 2 to 3 revealed that the average estimate of Ln P(D) across 20 independent replication plateaus and also the best

replicate giving maximum log likelihood values with a sharp peak was obtained at $K = 2$. At $K = 2$, all of the 244 genotypes, including the 67 contrasting genotypes, were classified into two distinct populations, *desi* and *kabuli*, with a high-resolution population structure (ESM Fig. S5A and B) that corresponded well with their clustering patterns and genetic relationships as obtained by the neighbor-joining tree analysis. The estimation of molecular genetic variation among and within the two populations based on 158 informative markers revealed a wider level of quantitative genetic differentiation (F_{ST} varied from 0.16 to 0.94 with an average of 0.79) among two population groups. The entire association panel of 244 genotypes, including the 67 contrasting genotypes, clearly belonged to a single population in which about 88.3 % of the inferred ancestry was derived from one of the model-based populations and the remaining approximately 11.7 % contained admixed ancestry.

The genetic association analysis was performed by correlating the marker genotyping data of the 62 most informative genic microsatellite markers with 100-seed weight-specific phenotypic information of the 244 phenotypically and genotypically diverse genotypes/ association panel (based on population genetic structure). The combined results of GLM and MLM identified seven microsatellite markers in the TF genes, namely, bZIP (basic leucine zipper), DUF3594 (domain of unknown function), bHLH (basic helix-loop-helix), GRAS [gibberellic acid insensitive (GAI)-repressor of GAI (RGA)-SCARECROW (SCR)], SBP (squamosa promoter-binding protein), DUF1635 and NAC (no apical meristem-arabidopsis transcription activation factor-cup shaped cotyledon), that showed a significant association with 100-seed weight at $P \leq 0.00001$ (R^2 0.19–0.37 (Table 1) in the *desi* and *kabuli* chickpea genotypes. In particular, the genic microsatellite markers in the 5′untranslated region (UTR) sequence components of the bZIP and DUF3594 TF genes showed strong seed weight-specific association potential (significant at $P \leq 0.00001$, R^2 0.33–0.37) as compared to that obtained by two microsatellite markers ($P \leq 0.001$, R^2 0.19–0.26) in the coding DNA sequence (CDS) of the TF genes (bHLH and GRAS). The remaining three seed weight-specific microsatellite markers in the TF genes (SBP, DUF1635 and NAC), which were identified in our previous study using an association panel of 96 genotypes (Kujur et al. 2013), showed a higher association potential ($P \leq 0.0001$, R^2
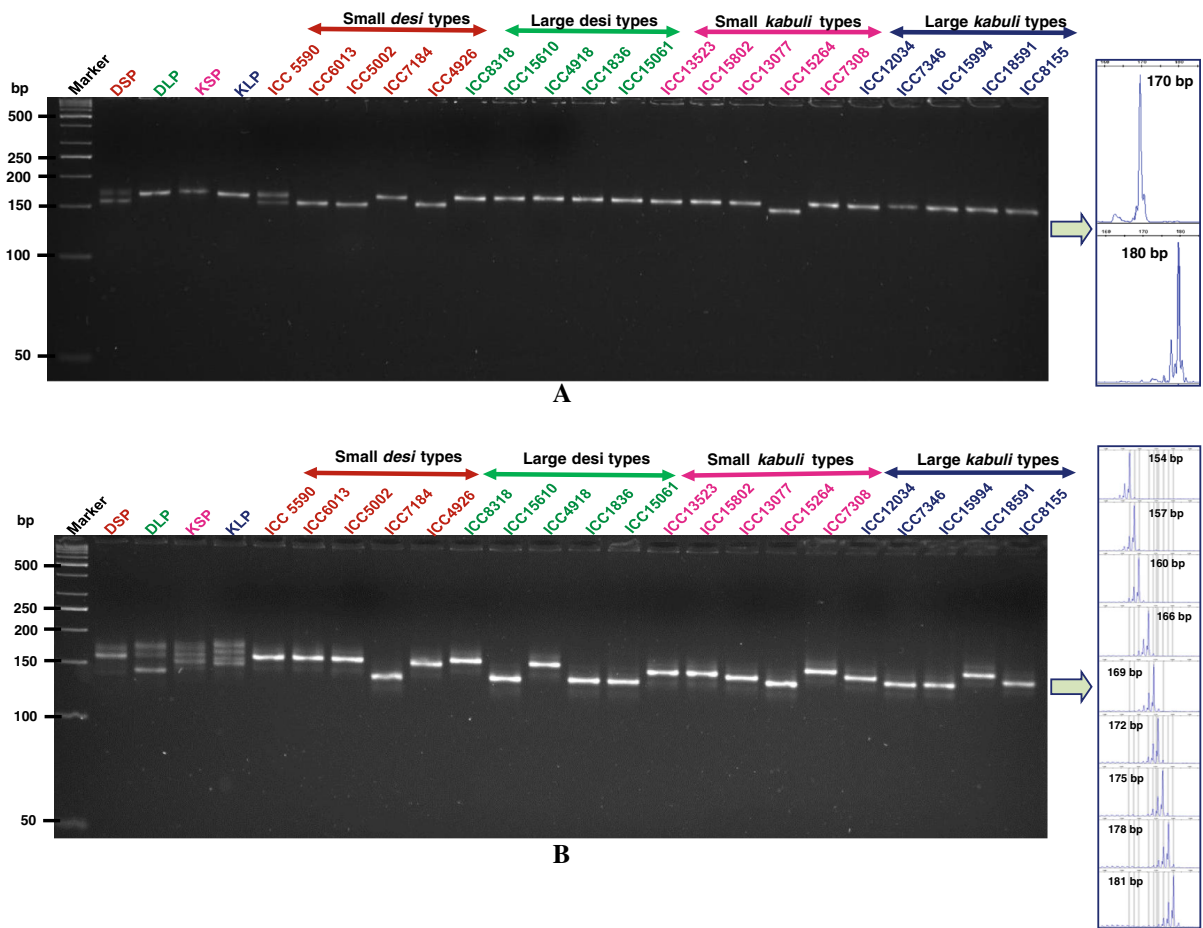
Fig. 3 Validation of informative and polymorphic microsatellite markers in the DUF3594 (**a**) and bZIP (**b**) TF genes screened from the pooled DNA analysis by their genotyping in 20 individual genotypes constituting the four small and large *desi* and *kabuli* DNA pools using both the gel-based assay and automated fragment analyser. *Lanes* are as labeled in Fig. 2. The fragment sizes (bp) of all the polymorphic alleles detected by automated fragment analyser are indicated

0.32–0.35) for seed weight than the two microsatellite marker-associated TF genes (bHLH and GRAS) in the present study. The favorable (abundant) common/rare allelic variants detected by the seven trait-associated markers in our constituted association panel was measured based on MAF that varied from 0.16 to 0.23 with a mean of 0.18.

The estimation of marker allele effects in the four trait-associated TF genes revealed the overall strong effect of these marker alleles on increasing seed weight in both *desi* and *kabuli* chickpea (ESM Fig. S6). The microsatellite marker in a strong trait-associated DUF3594 TF gene that amplified 180-bp alleles increased seed weight by 2.6-fold in 116 (69.5 %) of the 167 *desi* genotypes analysed. Similarly, a strong marker allele effect was evident for a

trait-associated bZIP TF gene that amplified 157-bp alleles in 109 *desi* genotypes and 172- and 181-bp alleles in 24 and 37 *kabuli* genotypes, respectively, with increased seed weight (2.8–2.9-fold) (ESM Fig. S6). Higher marker allele effects were also observed for 149- and 156-bp alleles amplified by the two microsatellite markers in the trait-associated bHLH and GRAS TF genes, with observed increased seed weight of 1.4- and 1.5-fold in 55 (71.4 %) and 58 (75.3 %) of the 77 *kabuli* genotypes, respectively. The remaining seed weight-specific microsatellite markers in three TF genes (SBP, DUF1635 and NAC) also showed similar strongly biased marker allele effects on increasing (two-fold) seed weight in *kabuli* genotypes (ESM Fig. S6). The results overall indicate that the strong and differential effects of the microsatellite

Table 1 Functional relevance of the seven seed weight-specific microsatellite marker-associated transcription factor (TF) genes identified through pool-based trait association mapping

Genes	NCBI probe IDs (Marker IDs*)	Microsatellite repeat-motifs	Number and size (bp) of alleles amplified in the association panel	Minor allele frequency	Location at the sequence components of genes	Microsatellite marker-based association		Seed weight-associated marker alleles (bp) showing maximum effects in *desi* and *kabuli* chickpea
						P value	R^2 value	
bZIP (Basic leucine zipper)[a]	PUID 16588123 (TFGMS9)	$(GAA)_{18}$	9 alleles 154, 157, 160, 166, 169, 172, 175, 178 and 181	0.17	5'UTR	1.1×10^{-5}	0.37	157, 172 and 181 *Desi* and *kabuli*
DUF3594 (Domain of unknown function)[a]	PUID 16588036 (TFGMS844)	$(AG)_{10}$	2 alleles 170 and 180	0.20	5'UTR	1.6×10^{-5}	0.33	180 *Desi*
bHLH (Helix-loop-Helix)	PUID 16587897 (TFGMS8)	$(AATAAC)_5$	2 alleles 149 and 161	0.16	CDS	1.2×10^{-3}	0.26	149 *Kabuli*
GRAS [Gibberellic acid insensitive (GAI)- Repressor of GAI (RGA)- SCARECROW (SCR)]	PUID 16587989 (TFGMS747)	$(TAG)_7$	3 alleles 156, 159 and 162	0.22	CDS	3×10^{-3}	0.19	156 *Kabuli*
SBP (Squamosa promoter binding protein)[a]	PUID 16586189 (TFFDMS21-TYPE-I)	$(CAC)_8$	4 alleles 153, 156, 162 and 165	0.19	CDS	1.2×10^{-3}	0.34	162 *Kabuli*
DUF1635 (Domain of unknown function)[a]	PUID 16584991 (TFGMS18)	$(CTA)_6$	3 alleles 151, 154 and 160	0.23	5'UTR	1.0×10^{-3}	0.35	154 *Kabuli*
NAC (No apical meristem- arabidopsis transcription activation factor-cup shaped cotyledon)[a]	PUID 16586186 (TFFDMS18-TYPE-I)	$(GAT)_6$	3 alleles 153, 159 and 162	0.17	CDS	1.4×10^{-3}	0.32	159 *Kabuli*

Table 1 continued

Genes	Differentially expressed (DE)/not differentially expressed (NDE) during seed development	SNP loci and marker haplotypes discovered	Sequenced gene amplicon size (bp)	Frequency (SNPs/bp)	Microsatellite-SNP marker haplotype-based association		Genetic position (cM) on the LGs/chromosomes
					P value	R^2 value	
bZIP (Basic leucine zipper)[a]	DE	C/T, A/G, G/C, C/T 4 loci 24 haplotypes	1,200	1/300	0.8×10^{-6}	0.40	79.3 CaLG (Chr1)
DUF3594 (Domain of unknown function)[a]	DE	A/G, C/T, A/T, G/A 4 loci 13 haplotypes	1,000	1/250	1.2×10^{-6}	0.45	80.6 CaLG (Chr1)
bHLH (Helix-loop-Helix)	NDE	A/G, C/T, G/C, C/A, C/T, A/G, C/A, C/T, G/A 9 loci 15 haplotypes	2,500	1/312.5	0.9×10^{-4}	0.29	75.6 CaLG (Chr2)
GRAS [Gibberellic acid insensitive (GAI)- Repressor of GAI (RGA)-SCARECROW (SCR)]	NDE	C/T, A/T, A/G, A/C, T/C 5 loci 14 haplotypes	1,500	1/300	5.8×10^{-4}	0.22	81.2 CaLG (Chr2)
SBP (Squamosa promoter binding protein)[a]	NDE	C/T, T/A, C/T, G/C, C/T, A/T, G/A, G/A, C/T, A/C, G/A, C/T, G/A, A/C 14 loci 23 haplotypes	3,000	1/214.3	1.6×10^{-5}	0.39	74.5 CaLG (Chr1)
DUF1635 (Domain of unknown function)[a]	DE	A/G, C/T, C/G, A/T, C/T, T/G, A/G 7 loci 16 haplotypes	2,000	1/285.7	1.2×10^{-5}	0.43	73.5 CaLG (Chr1)
NAC (No apical meristem- arabidopsis transcription activation factor-cup shaped cotyledon)[a]	NDE	A/G, T/C, G/C, T/A, G/A, G/T 6 loci 13 haplotypes	1,400	1/233.3	1.3×10^{-4}	0.26	73.9 CaLG (Chr1)

* TFGMS (transcription factor gene-derived microsatellite) and TFFDMS (transcription factor functional domain-associated microsatellite)

[a] Validated by traditional bi-parental linkage mapping

SNP single nucleotide polymorphism, Chr chromosome

marker alleles of seven TF genes on seed weight contributed to their higher and varied seed weight-specific genetic association potential in *desi* and *kabuli* chickpea.
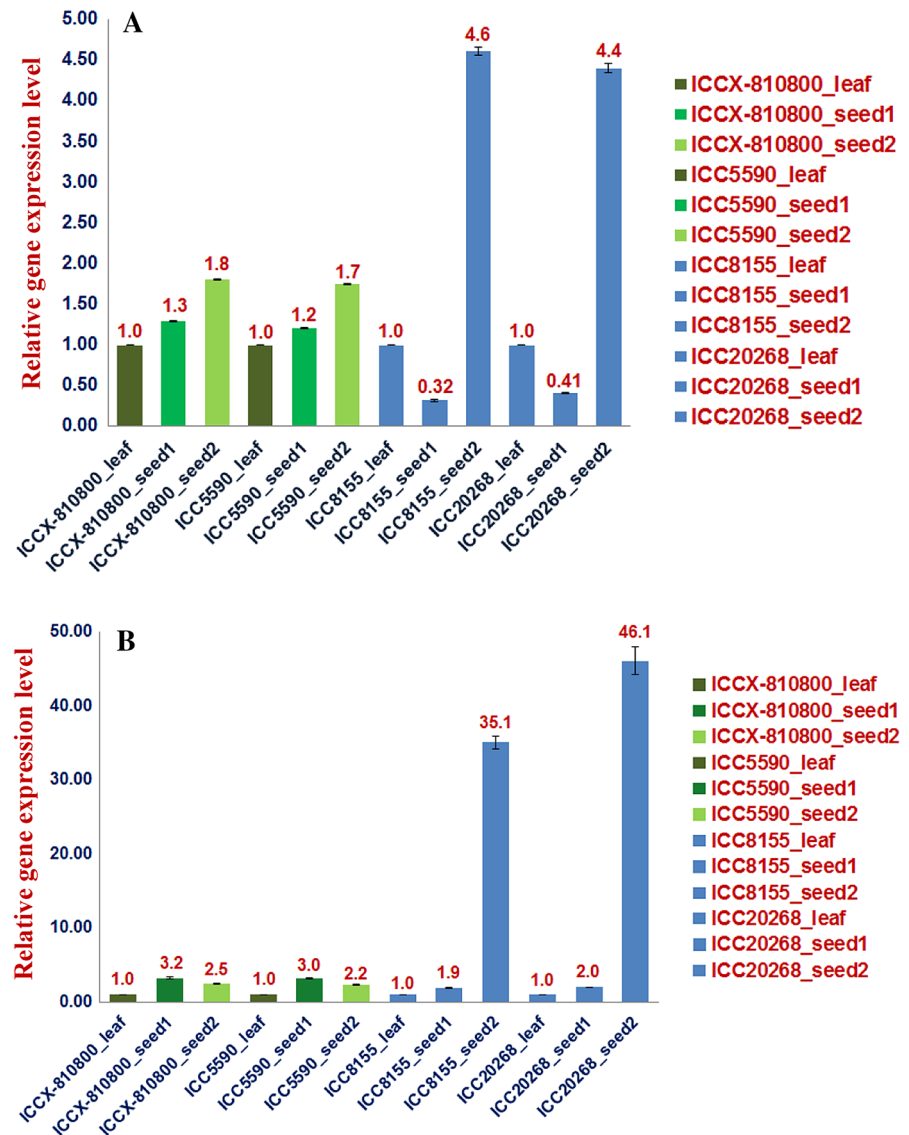
Differential expression profiling of seed weight-specific marker-associated genes

To validate the seven TF genes showing an association with seed weight more comprehensively through differential expression profiling, we amplified the four microsatellite marker-associated genes in two different seed developmental stages of the four contrasting small and large seed size/weight *desi* and *kabuli* genotypes (ICCX-810800, ICC 20268, ICC 5590 and ICC 8155) by semi-quantitative and quantitative RT-PCR assays. Three of these microsatellite marker-associated TF genes (bZIP, DUF3594 and DUF1635) showed differential expression specifically in seeds in contrast to vegetative leaf tissues of four genotypes. The microsatellite marker containing the $(AG)_{10}$ repeat-motif in the 5′UTR of a strong seed weight-associated ($R^2 = 0.33$, $P < 1.6 \times 10^{-5}$) DUF3594 TF gene revealed an upregulated pattern of expression (10.7–14.4-fold) specifically in the seed developmental stages of two contrasting large seed size/weight *kabuli* genotypes (ICC 20268 and ICC 8155) (Fig. 4a). Similarly, a microsatellite marker-carrying gene showed upregulation (at least 1.4-fold) in the seed developmental stages of two small seed size/weight *desi* genotypes (ICCX-810800 and ICC 5590). In contrast, we observed a differential upregulation (18.5–23-fold) and downregulation (1.3–1.4-fold) pattern of expression of a microsatellite marker containing the $(GAA)_{18}$ repeat-motif in the 5′UTR of a strong seed weight-associated ($R^2 = 0.37$, and $P < 1.1 \times 10^{-5}$) bZIP TF gene in large (ICC 20268 and ICC 8155) and small (ICCX-810800 and ICC 5590) seed size/weight genotypes during seed development, respectively (Fig. 4b). The microsatellite marker carrying the $(CTA)_6$ repeat-motif in the 5′UTR of the seed weight-associated DUF1635 TF gene was downregulated (3.6-fold) during seed development in the contrasting large seed size *kabuli* genotypes ICC 20268 and ICC 8155 (ESM Fig. S7A). The two microsatellite markers containing the $(AATAAC)_5$ and $(TAG)_7$ repeat-motifs in the CDS of the seed weight-associated bHLH and GRAS TF genes, respectively, did not show differential up- and

down-regulation in the seed developmental stages of the four small and large seed size/weight *desi* and *kabuli* chickpea genotypes (ESM Fig. S7B, C).

The sequencing of cloned cDNA and genomic DNA amplicons from the seven seed weight-specific microsatellite marker-associated TF genes in ICCX-810800, ICC 20268, ICC 5590 and ICC 8155 revealed the presence of the expected microsatellite repeat-motif sequences, but a variable number of microsatellite repeat-units, particularly in the 5′UTRs of three genes (DUF3594, bZIP and DUF1635) between the small seed size *desi* (ICCX-810800 and ICC 20268) and large seed size *kabuli* (ICC 5590 and ICC 8155) genotypes (ESM Fig. S8A, B). Interestingly, the two microsatellite markers in the CDS of the TF genes (bHLH and GRAS) which did not show differential expression during seed development contained similar microsatellite repeats between the small seed size *desi* (ICCX-810800 and ICC 20268) and large seed size *kabuli* (ICC 5590 and ICC 8155) genotypes (ESM Fig. S9A, B). The remaining seed weight-specific microsatellite marker-associated SBP and NAC TF genes, including the differentially expressed DUF1635 gene showing fragment length polymorphism between small and large seed size/weight *desi* and *kabuli* genotypes (as demonstrated in our previous study; Kujur et al. 2013), contained variable microsatellite repeats in the 5′UTR and CDS sequence components of these genes. The detailed comparison of sequences from the strong seed weight-associated and differentially expressed microsatellite marker carrying the DUF3594 TF gene revealed the occurrences of a variable number of microsatellite repeats $(AG)_{10}$ and $(AG)_{15}$ in the 5′UTR of this gene amplifying 170- and 180-bp alleles between the small seed size/weight *desi* (ICCX-810800 and ICC20268) and large seed size *kabuli* (ICC5590 and ICC8155) genotypes, respectively (ESM Fig. S8A). Likewise, we observed the presence of $(GAA)_{18}$ and $(GAA)_{23}$ microsatellite repeat-motifs in the 5′UTR of a strong seed weight/size-associated and differentially expressed microsatellite marker-carrying bZIP TF gene amplifying 157- and 172-bp alleles between small seed size/weight *desi* (ICCX-810800 and ICC20268) and large seed size/weight *kabuli* (ICC5590 and ICC8155) genotypes (ESM Fig. S8B). These observations implied a significant correlation between the expansion/contraction of repeat-units in the 5′UTR sequence components of seed size/weight-specific microsatellite

Fig. 4 Differential expression profiling of seed weight-specific microsatellite marker-associated TF genes using the quantitative reverse transcription (RT)-PCR assay. The markers carrying $(AG)_{10}$ and $(GAA)_{18}$ repeat-motifs in the 5′-untranslated regions (*UTRs*) of the strong seed weight-associated DUF3594 (**a**) and bZIP (**b**) TF genes revealed an up- and down-regulated pattern of expression in two seed developmental stages (*seed 1* and *seed 2*) of contrasting small seed size *desi* (ICCX-810800 and ICC5590) and large seed size *kabuli* (ICC20268 and ICC8155) chickpea by quantitative RT-PCR assay. Elongation factor-1 alpha was used as internal control in the RT-PCR reaction

marker-associated TF genes and differential expression profiling in the genes during seed development in contrasting chickpea genotypes. As a result, the transcriptional activity of the target TF gene might be altered during seed development in contrasting small and large seed size/weight *desi* and *kabuli* chickpea genotypes.

Validation of seed weight-associated genes by bi-parental linkage mapping

An integrated intra-specific genetic linkage map using genotyping information on 32 parental polymorphic genomic (Winter et al. 1999) and 209 genic microsatellite markers [including 96 TF gene-derived microsatellite markers (Kujur et al. 2013) and 113 new transcript sequence-derived microsatellite markers in this study] assayed in 94 segregating individuals of a F_3 mapping population (ICCX-810800 × ICC 20268) was constructed. This map contained a total of 241 microsatellite markers, including 209 genic microsatellite markers across eight LGs/chromosomes of chickpea (ESM Fig. S10) based on their corresponding physical positions on the chromosomes and common parental polymorphic genomic microsatellite marker positions and groupings shared between the

corresponding eight LGs as documented in previous studies (Winter et al. 2000; Nayak et al. 2010; Gaur et al. 2011; Gujaria et al. 2011; Kujur et al. 2013). The integrated chickpea transcript map comprising the eight LGs/chromosomes covered a total map length of 843.5 cM with an average inter-marker distance of 3.5 cM (ESM Table S4).

The genetic mapping using 100-seed weight-specific phenotyping information (100-seed weight varied from 5.4 to 49.6 g and showed a normal frequency distribution with a quantitative genetic inheritance pattern) of segregating individuals of a F_3 mapping population identified and mapped two major significant QTLs (CaqSW1.1 and CaqSW1.2) on chromosome 1 (LOD threshold >4.0, $P < 0.05$) (explained total phenotypic variance: $R^2 = 28.6–34.7$ %) and one QTL (CaqSW2.1) on chromosome 2 ($R^2 = 23.5$ %) based on composite interval mapping (ESM Fig. S11, ESM Table S5). The identified QTLs exhibited dominance as well as additive effects to increased seed weight (2.1–2.8-fold), particularly *kabuli* (ICC20268)-specific alleles (ESM Table S5). The target seed weight-specific QTL regions identified on chromosomes 1 (73.5–74.5 and 79.3–81.3 cM) and 2 (65.7–67.5 cM) spanned nine gene (DUF1635, NAC, SBP, bZIP, DUF3594, ARF, EF, DUF1635 and helicase)-based microsatellite markers. Three microsatellite markers in the TF genes SBP, DUF1635 and NAC correspond to the QTL region (73.5–74.5 cM) on chromosome 1 reported by Kujur et al. (2013), and two TF gene (bZIP and DUF3594)-derived microsatellite markers present in another novel QTL region (79.3–82.7 cM) on chromosome 1 (identified in this study); all had strong seed weight-specific association potential in chickpea. To further validate the trait-associated TF genes in bi-parental mapping population, the 11 polymorphic microsatellite markers in the genes (seven seed weight-associated TF genes and four other genes identified exclusively through QTL mapping; ESM Fig. S12) were genotyped in ten of each small and large seed size/weight homozygous F_3 individuals along with parental *desi* ICCX-810800 and *kabuli* ICC0268 genotypes. The marker allele effects of nine of these gene (DUF1635, NAC, SBP, bZIP, DUF3594, ARF, EF, DUF1635 and helicase)-derived microsatellite markers specific to small seed size/weight *desi* and large seed size/weight *kabuli* corresponded with the seed weight phenotypes of 18 of the 20 small and large seed size/weight homozygous F_3

individuals, showing an overall 90 % validation success rate of trait-associated genes in the bi-parental mapping population (ESM Fig. S12). For example, the 170- and 180-bp alleles specific to the small and large seed size/weight ICCX-810800 and ICC20268 amplified by the seed weight-associated DUF3594 TF gene, respectively, were present in 18 of the 20 small (100-seed weight 5.4–13.6 g) and large (41.2–49.6 g) seed weight homozygous F_3 mapping individuals. Similarly, the seed weight-associated bZIP TF gene amplifying 157- and 172-bp alleles specific to small and large seed size/weight genotypes, respectively, were matched with phenotypes of 18 of the 20 small and large seed size/weight homozygous F_3 mapping individuals (ESM Fig. S12). In contrast, alleles specific to small and large seed size/weight genotypes amplified by two other seed weight-associated TF genes (bHLH and GRAS)-derived microsatellite markers revealed a biased distribution and did not correspond to the phenotypes of the 20 small and large seed size/weight homozygous mapping individuals. Therefore, the five seed weight-specific potential microsatellite marker-associated TF genes (bZIP, DUF3594, DUF1635, SBP and NAC) validated by association analysis, traditional bi-parental linkage mapping and differential expression profiling were selected as target candidates for controlling seed weight in *desi* and *kabuli* chickpea.

High-resolution gene-specific haplotyping and LD mapping

For determining the seed weight-specific association potential of the seven identified genes, including the five most informative TF genes more precisely, we performed high-resolution molecular haplotyping and LD mapping using different combinations of amplified microsatellite marker alleles and SNPs mined in the corresponding genes among the 244 chickpea genotypes constituting the association panel. The cloned amplicon sequencing of seven seed weight-associated TF genes among these 244 genotypes mined a total of 50 SNP loci in 12,600-bp sequences with an average SNP frequency of 1/252 bp (Table 1; Fig. 5). The average nucleotide diversity and mean PIC value estimated for the 50 SNP loci carrying the seven TF genes among the association panel was $\pi = 3.8 \times 10^{-3}$ bp and 0.43, respectively. The haplotype analysis using the microsatellite (26 alleles) and SNP (50 SNP loci) marker genotyping information among the association panel constituted
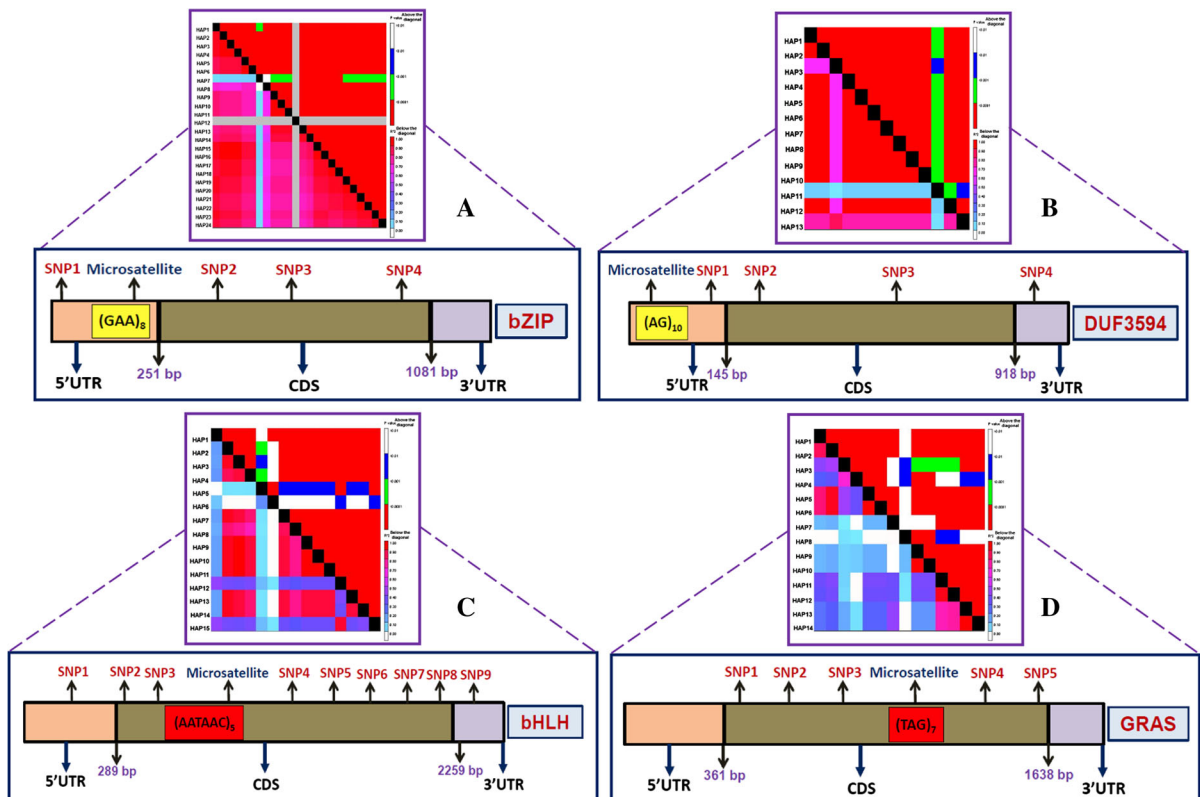
Fig. 5 Genotyping of 38 polymorphic marker alleles (22 SNPs and 16 microsatellite alleles) identified in different core DNA sequences (*CDS*) and 5'UTRs of four seed-weight associated TF genes [bZIP (**a**), DUF3594 (**b**), bHLH (**c**) and GRAS (**d**)] among the 244 genotypes in the association panel of which 13–24 haplotypes are in *desi* and *kabuli*. The haplotype marker-based genotyping information on 13 haplotypes produced significant and extensive LD estimates ($r^2 = 0.25$–0.96, $P < 0.001$) and extended LD decay covering the complete 1,000-bp sequenced region of the DUF3594 TF gene (**b**) and thus enhanced its overall LD resolution and association potential for seed weight in *desi* and *kabuli* chickpea

13–24 haplotypes in each TF gene amplicon (Table 1; Fig. 5) with a higher haplotype-based nucleotide diversity (mean $\pi = 3.9 \times 10^{-3}$) and PIC values (varied from 0.47 to 0.83 with an average of 0.64). The use of SNP–microsatellite marker haplotypes identified in the seven TF genes in the genetic association analysis revealed a strong association of the DUF3594 gene, followed by the bZIP, DUF1635, SBP, NAC, bHLH and GRAS genes, with 100-seed weight in *desi* and *kabuli* genotypes ($P \leq 5.8 \times 10^{-4}$, $R^2 = 0.22$–0.45 (Table 1). However, strong marker allele/haplotype effects in the seven trait-associated TF genes on increasing seed weight implies that these genes have a high trait association potential for seed weight/size in *desi* and *kabuli* chickpea.

The determination of LD estimates (r^2) and extent of LD decay using all possible pair-combinations of microsatellite–SNP marker haplotypes within the seven seed weight-associated TF genes revealed higher LD estimates and extended LD decay ($r^2 \geq 0.1$) of up to 1,000-bp of the sequenced region of these genes in the *desi* and *kabuli* association panel. However, a decreasing trend of LD decay overall in the genes with increases in the sequenced region (bp) was observed. The strong trait-associated DUF3594 and bZIP TF genes (R^2 0.40–0.45 and $P = 0.8$–1.2×10^{-6}), with their increased seed weight-specific marker allele/haplotype effects, particularly in *desi and kabuli* genotypes, displayed a significant LD with extended LD decay which persisted at a high level ($P < 0.001$; r^2 0.25–0.96) across the entire 1,000-bp sequenced region of these genes based on 13–24 microsatellite–SNP haplotype-pairs (Fig. 5a, b). The marker haplotype-based LD

mapping in three TF genes (bHLH, GRAS and NAC) revealed low LD estimates that uncovered their whole sequenced regions (Fig. 5c, d) and also decayed LD faster than that of strong trait-associated DUF3594 TF gene, resulting in a relatively low association potential of these three genes for seed weight in the contrasting *desi* and *kabuli* chickpea genotypes. The microsatellite–SNP haplotype-pairs constituted in the remaining two seed weight-associated DUF1635 and SBP TF genes (association potential $P < 1.6 \times 10^{-5}$, $R^2 = 0.39$–0.43) showed an extended high level of LD decay that covered the entire sequenced regions of these genes. The overall results indicated that high-resolution molecular haplotyping and LD mapping in the TF genes which combined both informative microsatellite and SNP markers explained a larger proportion of the phenotypic variance than did that of individual markers; thus, the former have a significant impact on the high seed weight trait-specific association potential in chickpea. Henceforth, molecular haplotyping/LD mapping in the TF genes could be an attractive, rapid and expedient approach for quantitative trait association mapping and genetic dissection of complex traits such as seed weight/size in chickpea.

Discussion

Whole genome and candidate gene-based association mapping is a rapid, efficient and well-established approach for identifying genes/QTLs, as well as for controlling complex quantitative traits in many crop species. However, microsatellite marker-based trait association mapping requires large-scale genotyping and phenotyping of individual genotypes/association panel by high-throughput platforms and consequently, is commonly associated with high costs in terms of resources and time. The "pool-based trait association mapping" approach has been proposed as an alternative rapid and cost-effective means to circumvent these difficulties. This approach has been extensively utilized in humans and, more recently, in plants for whole genome association mapping to identify markers associated with genes/QTLs for complex quantitative traits, such as diseases in humans and plant height and saccharification yield in sorghum (Bansal et al. 2002; Sham et al. 2002; Craig et al. 2005; Pearson et al. 2007; Steer et al. 2007; Abraham et al. 2008; Baum et al. 2008; Shifman et al. 2008; Kirov

et al. 2009; Diergaarde et al. 2010; Earp et al. 2011; Wang et al. 2011, 2012). In our study, we have applied this strategy to screen four DNA pools of 20 individual genotypes with contrasting small and large seed weight/size phenotypes in *desi* and *kabuli* chickpea with gene-derived microsatellite markers instead of genotyping all of the 244 genotypes of an association panel. Further, we have validated the polymorphic informative markers identified based on pooled DNA analysis by performing both a genotyping assay in the 20 individual genotypes constituting the pools and the candidate gene-based association analysis using the entire 244 seed weight trait-specific association panel. The genotyping of four DNA pools [small *desi* (5.9–9.4 g), large *desi* (19.2–35.4 g), *kabuli* small (14.4–23.8 g) and *kabuli* large (46.7–57.6 g)] with 616 genic microsatellite markers, including 321 TF gene-derived microsatellite markers, using the gel-based assay and automated fragment analyser enabled us to screen a large proportion (379, 61.5 %) of monomorphic microsatellite markers which did not differentiate any of these four contrasting trait-specific pools from each other. The correspondence of the monomorphic amplification pattern obtained for genic microsatellite markers in the DNA pools between the gel-based assay and automated fragment analyser suggests that the 3.5 % metaphor agarose gel assay is sufficiently efficient to resolve monomorphic amplicons from the pooled DNA and for rapid initial screening of the microsatellite markers, producing monomorphic amplification in the DNA pools of chickpea. Such a contrasting trait-specific DNA pool screening strategy in gel-based assays using genic microsatellite markers could be useful particularly in crop species showing low intra- and inter-specific polymorphism and, thereby, avoid the extensive costs in terms of time, resources and labour involved in genotyping monomorphic microsatellite markers which are ultimately of no or minimal use for trait association mapping. The remaining 237 (38.5 %) microsatellite markers in the genes which were identified to be informative based on their diverse amplification and fragment length polymorphism patterns among the four small and large *desi* and *kabuli* DNA pools were validated further in 20 contrasting individual genotypes constituting the pools using both the gel-based assay and the automated fragment analyser. Based on the correlation between marker genotyping information for an

individual genotype and the potential of the marker for discriminating the contrasting small and large seed size/weight *desi* and *kabuli* genotypes in the gel-based assay and automated fragment analyser, we selected the 62 (10.1 %) most informative gene-derived microsatellite markers for their further utilization in candidate gene-based association mapping using the 244 genotypes of the association panel. By integrating the two approaches, namely, screening of DNA pools with contrasting seed weight traits using the genic microsatellite markers and further genotyping of individual germplam lines constituting the pools with screened polymorphic markers in *desi* and *kabuli* chickpea, we were able to screen 554 (89.9 %) markers of a total of 616 markers that had very little significance in seed weight-specific association mapping. This result indicates the efficiency of such integrated DNA pooling and individual germplasm line genotyping strategies in reducing the costs, labour and time involved in high-throughput genotyping of a larger set of microsatellite markers. This approach has further utility in the rapid screening of informative and polymorphic microsatellite markers to expedite both qualitative and quantitative trait association mapping not only in chickpea but also in many other crop species.

The determination of population genetic structure for an association panel is essential to minimize the false positive rate in trait association mapping (Flint-Garcia et al. 2003). Our analysis of population structure using 158 microsatellit markers, including 62 informative genic and 96 genomic microsatellite markers, distributed over eight LGs/chromosomes classified the 244 chickpea genotypes (association panel) into two major genetically distinct *desi* and *kabuli* population groups. The results indicated that the association panel used in the study is a structured population and thus has utility in seed weight-specific association mapping. Higher population differentiation among the two chickpea population groups based on pairwise F_{ST} revealed a broad range of F_{ST} value, from 0.16 to 0.94, with an average of 0.79, indicating the utility of genic microsatellite markers in establishing the distinctness and providing an understanding of the functional molecular diversity pattern in the expressed sequence component of the genome specifically with regard to seed weight characteristics in *desi* and *kabuli* chickpea. The approximately 11.7 % admix ancestry among the two population groups reflected their complex breeding history involving

inter-crossing and introgression among contrasting genotypes coupled with different strong adaptive selection pressure during domestication. The close evolutionary relationships between the small seed size *desi* and large seed size *kabuli* genotypes was expected based on earlier documented morphological (Robertson et al. 1997), cytological (Ahmad 2000), biochemical (Toker et al. 2006) and molecular studies (Kazan and Muehlbauer 1991; Sethy et al. 2006; Redden and Berger 2007; Shan et al. 2007; van der Maesen et al. 2007; Upadhyaya et al. 2008; Choudhary et al. 2009; Toker 2009; Bharadwaj et al. 2011).

The candidate gene-based association analysis based on integration of the genotyping and population structure data of 158 microsatellite markers with the seed weight-specific phenotyping information of 244 chickpea genotypes (association mapping panel) enabled us to identify microsatellite markers in seven TF genes (including three earlier reported TF genes; DUF3594, bZIP, bHLH, GRAS, SBP, DUF1635 and NAC) that were associated ($P \leq 0.00001$, $R^2 = 0.19$–0.37) with 100-seed weight in the *desi* and *kabuli* genotypes. The seven markers with a MAF of >15 % identified in the TF genes were validated both by GLM and MLM methods in TASSEL and thus suggest the robustness and reliability of identified genic microsatellite markers for seed weight-specific association analysis in *desi* and *kabuli* genotypes. The strong marker allele effect of four trait-associated TF gene-derived microsatellite markers on increasing seed weight (1.7–2.9-fold) in contrasting *desi* and *kabuli* genotypes was evident. The one microsatellite marker in the DUF3594 TF gene and three microsatellite markers in the SBP, DUF1635 and NAC TF genes had a strong marker allele effect on increasing seed weight in *desi* and *kabuli*, respectively. The one microsatellite marker in the bZIP TF gene revealed strong allele effects on increasing seed weight in both *desi* and *kabuli* chickpea. The differential effects of marker alleles on seed weight in contrasting *desi* and *kabuli* germpalsm lines contributed to varied seed weight-specific association potential of microsatellite markers in the TF genes. For example, the stronger *kabuli*-specific microsatellite marker allele effect of the DUF3594 TF gene and both *desi* and *kabuli*-specific allele effects of the bZIP gene on increasing seed weight resulted in a higher seed weight-specific association potential ($P \leq 0.00001$, R^2 0.33–0.37) of these two genes in contrast to the other five TF genes

($P \leq 0.001$, R^2 0.19–0.26). Taking into account the effect/dependency of MAF on marker-trait association (Ehrenreich et al. 2009; Bergelson and Roux 2010; Huang and Han 2014), we observed a >18 % MAF for seven trait-associated TF gene-derived microsatellite markers in our constituted association panel. This result implies the presence of an appreciable frequency of minor alleles and favourable common/rare allelic variants detected by these markers in the association panel. Therefore, seed weight/size-specific association potential (explaining $\geq$19 % of phenotypic variance) estimated for seven TF gene-derived microsatellite markers in our study is non-spurious and reliable.

These observations reflected that using the pool-based seed weight trait-specific association mapping, we could screen a large proportion of monomorphic and less informative genic microsatellite markers (609, 98.9 %) rapidly and efficiently to delineate informative and polymorphic markers (1.1 %, 7) in those TF genes showing a strong genetic association potential with seed weight in *desi* and *kabuli* genotypes. Application of the genic microsatellite marker-based trait association mapping approach in chickpea has been limited due to its low intra-specific polymorphism, specifically among *desi* and *kabuli* genotypes, which ultimately leads to high costs in terms of resources, labour and time when screening sufficient informative markers and carrying out high-throughput genotyping of numerous polymorphic markers among a larger set of genotypes. Therefore, the development of the pool-based trait association mapping strategy demonstrated successfully in our present study for identifying microsatellite markers in those TF genes associated with seed weight in *desi* and *kabuli* genotypes is significant. In addition, the use of a cost-effective pooling strategy means that a large number of microsatellite markers can be genotyped initially in pools and subsequently in individual genotypes/whole association panel with less time and labour, which will increase association mapping resolution, resulting in an enhanced possibility of rapidly identifying genes/QTLs for important agronomic traits in chickpea. Therefore, the pool-based strategy will not only expedite the application of candidate gene-based association mapping but also whole genome association genetics in a wide range of crop plants, including legumes, using both genic and genomic microsatellite markers. However, the DNA pooling strategy has certain limitation in terms of the

constitution of individual trait-specific pools for each trait targeted in the association study. Consequently, this approach is likely to be less resource efficient with increasing number of target traits for marker-trait association study in a well-characterized association panel, which is generally used for unravelling the genetic association potential of markers for a number of agronomic traits.

Several recent reports have correlated the trait association potential of functional marker loci present in different coding and regulatory sequence components of the genes with their differential expression profiling in starch biosynthesis (Kharabian-Masouleh et al. 2012), grain size (*GS3*, Mao et al. 2010; *GS5*, Li et al. 2011) and heading date (*Ghd7*; Lu et al. 2012) genes in rice and in genes encoding chlorophyll *A/B*-binding protein in barley (Xia et al. 2012). In addition, the expansion and contraction of microsatellite repeats in the 5′UTR of genes have a significant effect in regulating gene expression and thus affecting the process of transcription and/or translation for many traits, including cancers and neuronal diseases in humans (Tidow et al. 2003; Li et al. 2004), amylose content in rice (Bao et al. 2002), quality protein in maize (Dresselhaus et al. 1999) and light and salicylic responses in *Brassica* (Zhang et al. 2006). In the present study, the strong association potential of microsatellite markers in the 5′UTRs of three TF genes (DUF3594, bZIP and DUF1635) and the correlation of these markers with a differential up- and down-regulated pattern of gene expression during seed development and fragment length polymorphism based on variable number repeats between contrasting small *desi* (ICCX-810800 and ICC5590) and large *kabuli* (ICC20268 and ICC8155) seed size/weight chickpea genotypes suggested the functional relevance of such identified microsatellite markers in the TF genes for seed weight trait association in chickpea. The differential transcriptional activity and regulation of the bZIP TF gene during the seed developmental process and the interaction of this gene with the main regulators of seed development B3/HAP (heme associated protein) domains has been documented previously in many plant species (Nijhawan et al. 2008; Agarwal et al. 2011), including legumes (Udvardi et al. 2007). However, the functionally uncharacterized DUF3594 TF gene, which shows a higher homology with another TF PHD (plant homeo domain) finger at the protein level, has a potential

role during seed development in crop plants, including legumes (Müssig et al. 2000; Udvardi et al. 2007; Agarwal et al. 2011). The remaining two TF genes (bHLH and GRAS) with seed weight-associated microsatellite markers in the CDS did not show differential expression during seed development and fragment length polymorphism based on microsatellite repeat-unit variations between small *desi* and large *kabuli* genotypes studied. Further, the validation of five seed weight-associated TF gene (DUF3594, bZIP, SBP, DUF1635 and NAC)-derived microsatellite markers by traditional bi-parental linkage mapping/ QTL mapping, as well as in 90 % homozygous F_3 mapping individuals and two parental *desi* ICCX-810800 and *kabuli* ICC 20268 genotypes (with contrasting seed size/weight phenotypes), based on their specific marker allele effects in small and large seed size/weight *desi* and *kabuli* re-emphasizes the efficacy and broader applicability of these five identified genic microsatellite markers in governing seed weight in chickpea. It also suggests the utility of using integrated approaches of candidate gene-based association mapping, differential expression profiling and traditional bi-parental linkage mapping to delineate the most suitable functionally relevant microsatellite markers in the TF genes controlling the seed weight in *desi* and *kabuli* chickpea.

The extensive LD estimate and extended LD decay expected in chickpea, in contrast to other domesticated self-pollinated crop species, are due to its low inter- and intra-specific polymorphism and insufficient and non-uniform marker coverage on the chickpea genome. Consequently, the candidate gene-based association analysis will be of great significance in trait association mapping in chickpea (Haseneyer et al. 2010; Huang et al. 2010; Zhao et al. 2011; Varshney et al. 2012). The candidate gene-based association analysis approach using a higher number of microsatellite–SNP marker haplotype pairs (13–24) identified in seven seed weight-associated TFs enhanced the overall LD resolution and trait association potential of these genes for seed weight in *desi* and *kabuli* chickpea genotypes. This result suggests that there is a correlation between LD patterning and trait association potential with the marker density required in these genes. Despite our identification of a higher number of SNP–microsatellite marker haplotypes (24) in the bZIP TF gene, the analysis showed a lower seed weight-specific association potential in this gene than

in the DUF3594 gene carrying only 13 haplotypes of the 244 genotypes of the association panel. This variation could be due to the effect of factors other than marker density, such as population genetic structure and adaptive selection pressure of the association panel, on shaping the overall LD resolution and trait association potential in these genes. Among the seven TF genes, the three differentially expressed TF genes, namely, DUF3594, bZIP and DUF1635, and one seed weight-specific SBP TF gene (validated by QTL mapping) demonstrated extended LD coverage throughout its 1,000 bp sequenced region, thus enhancing its overall association potential ($P < 1.2 \times 10^{-6}$ and $R^2 = 0.45$) for seed weight in *desi* and *kabuli* chickpea. The utility of microsatellite–SNP marker haplotype-based trait association analysis and the integration of this analysis with transcriptomics for the identification of the most appropriate genes/QTLs controlling diverse phenotypic traits have already been demonstrated in humans (Amos et al. 2010) and also in diploid and polyploid crop species, including rice (Konishi et al. 2006) and *Brassica* (Harper et al. 2012). However, the large-scale validation of such strong seed weight-associated TF genes in a natural and/or mapping population through association genetics and/or traditional bi-parental linkage (QTL) mapping, respectively, is of relevance in *desi* and *kabuli* chickpea that needs further systematic experimentation by assaying high-throughput microsatellite and SNP marker genotyping information.

The cost-effective and less time-consuming pool-based trait association mapping strategy implemented in the present study by integrating DNA pooling with individual germplasm line genotyping has been shown to be efficacious for the rapid identification of seven microsatellite markers in those TF genes (bZIP, DUF3594, bHLH, GRAS, SBP, DUF1635 and NAC) associated with seed weight/size in *desi* and *kabuli* chickpea. The further validation of these genic markers identified through large-scale candidate gene-based association mapping by traditional bi-parental linkage mapping, differential expression profiling and high-resolution gene-specific microsatellite–SNP marker haplotyping/LD mapping enabled us to identify the most informative and functionally relevant microsatellite marker in four TF genes (DUF3594, bZIP, DUF1635 and SBP) controlling seed weight in *desi* and *kabuli* chickpea. The integrated pool-based trait association mapping, differential gene expression

profiling, bi-parental linkage mapping and molecular haplotyping/LD mapping approach developed in this study, therefore, is an efficient method for screening functional microsatellite markers for rapidly establishing marker-trait association and identifying genes for many qualitative and quantitative traits of agricultural importance in crop species, including chickpea.

References

Abraham R, Moskvina V, Sims R, Hollingworth P, Morgan A, Georgieva L, Dowzell K, Cichon S, Hillmer AM, O'Donovan MC, Williams J, Owen MJ, Kirov G (2008) A genome-wide association study for late-onset Alzheimer's disease using DNA pooling. BMC Med Genomics 1:44

Agarwal P, Kapoor S, Tyagi AK (2011) Transcription factors regulating the progression of monocot and dicot seed development. BioEssays 33:189–202

Ahmad F (2000) A comparative study of chromosome morphology among the nine annual species of *Cicer* L. Cytobios 101:37–53

Amos W, Driscoll E, Hoffman JI (2010) Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? Proc R Soc B 278:1183–1188

Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Muliyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature 465:627–631

Bansal A, van den Boom D, Kammerer S, Honisch C, Adam G, Cantor CR, Kleyn P, Braun A (2002) Association testing by DNA pooling: an effective initial screen. Proc Natl Acad Sci USA 99:16871–16874

Bao JS, Corke H, Sun M (2002) Microsatellites in starch-synthesizing genes in relation to starch physicochemical properties in waxy rice (*Oryza sativa* L.). Theor Appl Genet 105:898–905

Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13:197–207

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. Nat Rev Genet 11:867–879

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

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

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

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

Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WT, Ramsay L, Mackay I, Balding DJ, AGOUEB Consortium, Waugh R, O'Sullivan DM (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci USA 107:21611–21616

Craig DW, Huentelman MJ, Hu-Lince D, Zismann VL, Kruer MC, Lee AM, Puffenberger EG, Pearson JM, Stephan DA (2005) Identification of disease causing loci using an array-based genotyping approach on pooled DNA. BMC Genomics 6:138

Diergaarde B, Brand R, Lamb J, Cheong SY, Stello K, Barmada MM, Feingold E, Whitcomb DC (2010) Pooling-based genome-wide association study implicates gamma-glutamyltransferase 1 (GGT1) gene in pancreatic carcinogenesis. Pancreatology 10:194–200

Dresselhaus T, Cordts S, Heuer S, Sauter M, Lorz H, Kranz E (1999) Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. Mol Gen Genet 261:416–427

Earp MA, Rahmani M, Chew K, Brooks-Wilson A (2011) Estimates of array and pool-construction variance for planning efficient DNA-pooling genome wide association studies. BMC Med Genomics 4:81

Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate gene association mapping of Arabidopsis flowering time. Genetics 183:325–335

Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Garg R, Patel RK, Jhanwar S, Priya P, Bhattacharjee A, Yadav G, Bhatia S, Chattopadhyay D, Tyagi AK, Jain M (2011) Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol 156:1661–1678

Gaur R, Sethy NK, Choudhary S, Shokeen B, Gupta V, Bhatia S (2011) Advancing the STMS genomic resources for defining new locations on the intra-specific genetic linkage map of chickpea (*Cicer arietinum* L.). BMC Genomics 12:117

Gujaria N, Kumar A, Dauthal P, Dubey A, Hiremath P, Bhanu Prakash A, Farmer A, Bhide M, Shah T, Gaur PM, Upadhyaya HD, Bhatia S, Cook DR, May GD, Varshney RK (2011) Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). Theor Appl Genet 122:1577–1589

Hall D, Tegstrom C, Ingvarsson PK (2010) Using association mapping to dissect the genetic basis of complex traits in plants. Brief Funct Genomics 9:157–165

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nat Biotechnol 30:798–802

Haseneyer G, Stracke S, Piepho HP, Sauer S, Geiger HH, Graner A (2010) DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. BMC Plant Biol 10:5

Hiremath PJ, Farmer A, Cannon SB, Woodward J, Kudapa H, Tuteja R, Kumar A, Bhanu Prakash A, Mulaosmanovic B, Gujaria N, Krishnamurthy L, Gaur PM, Kavi Kishor PB, Shah T, Srinivasan R, Lohse M, Xiao Y, Town CD, Cook DR, May GD, Varshney RK (2011) Large-scale transcriptome analysis in chickpea (*Cicer arietinum* L.), an orphan legume crop of the semi-arid tropics of Asia and Africa. Plant Biotechnol J 9:922–931

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, Kavi Kishor PB, Shah TM, Cook DR, Varshney RK (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J 10:716–732

Hou J, Wang C, Hong X, Zhao J, Xue C, Guo N, Gai J, Xing H (2011) Association analysis of vegetable soybean quality traits with SSR markers. Plant Breed 113:413–506

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:4.1–4.21. doi:10.1146/annurev-arplant-050213-035715

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, Shah N, Singh VK, Garg R, Jeena G, Yadav M, Kant C, Sharma P, Yadav G, Bhatia S, Tyagi AK, Chattopadhyay D (2013) A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). Plant J 74:715–729

Jhanwar S, Priya P, Garg R, Parida SK, Tyagi AK, Jain M (2012) Transcriptome sequencing of wild chickpea as a rich resource for marker development. Plant Biotechnol J 10:690–702

Jia L, Yan W, Zhu C, Agrama HA, Jackson A, Yeater K, Li X, Huang B, Hu B, McClung A, Wu D (2012) Allelic analysis of sheath blight resistance with association mapping in rice. PLoS ONE 7:e32703

Jun TH, Van K, Kim MY, Lee S-H, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. Euphytica 162:179–191

Kazan K, Muehlbauer FJ (1991) Allozyme variation and phylogeny in annual species of *Cicer* (Leguminosae). Plant Syst Evol 175:11–21

Kharabian-Masouleh A, Waters DLE, Reinke RF, Ward R, Henry RJ (2012) SNP in starch biosynthesis genes associated with nutritional and functional properties of rice. Sci Rep 2:557

Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG, Park YJ (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. Bioinformatics 23:2155–2162

Kirov G, Zaharieva I, Georgieva L, Moskvina V, Nikolov I, Cichon S, Hillmer A, Toncheva D, Owen MJ, O'Donovan MC (2009) A genome-wide association study in 574 schizophrenia trios using DNA pooling. Mol Psychiaty 14:796–803

Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M (2006) An SNP caused loss of seed shattering during rice domestication. Science 312:1392–1396

Kujur A, Bajaj D, Saxena MS, Tripathi S, Upadhyaya HD, Gowda CL, Singh S, Jain M, Tyagi AK, Parida SK (2013) Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. DNA Res 20:355–373

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991–1007

Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. Nat Genet 43:1266–1270

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

Lu L, Yan W, Xue W, Shao D, Xing Y (2012) Evolution and association analysis of *Ghd7* in rice. PLoS ONE 7:e34021

Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. Proc Natl Acad Sci USA 107:19579–19584

Müssig C, Kauschmann A, Clouse SD, Altmann T (2000) The *Arabidopsis* PHD-finger protein *SHL* is required for proper development and fertility. Mol Gen Genet 264:363–370

Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R, Jüngling R, Singh J, Kishor PB, Sivaramakrishnan S, Hoisington DA, Kahl G, Winter P, Cook DR, Varshney RK (2010) Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. Theor Appl Genet 120:1415–1441

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine

zipper transcription factor family in rice. Plant Physiol 146:333–350

Pearson JV, Huentelman MJ, Halperin RF, Tembe WD, Melquist S, Homer N, Brun M, Szelinger S, Coon KD, Zismann VL, Webster JA, Beach T, Sando SB, Aasly JO, Heun R, Jessen F, Kolsch H, Tsolaki M, Daniilidou M, Reiman EM, Papassotiropoulos A, Hutton ML, Stephan DA, Craig DW (2007) Identification of the genetic basis for complex disorders by use of pooling-based genome-wide single nucleotide polymorphism association studies. Am J Hum Genet 80:126–139

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Redden B, Berger J (2007) History and origin of chickpea. In: Yadav SS, Redden B, Chen W, Sharma B (eds) Chickpea breeding and management. CAB International, Wallingford, pp 1–13

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci USA 109:8872–8877

Robertson LD, Ocampo B, Singh KB (1997) Morphological variation in wild annual Cicer species in comparison to the cultigen. Euphytica 95:309–319

Sakiroglu M, Sherman-Broyles S, Story A, Moore KJ, Doyle JJ, Charles Brummer E (2012) Patterns of linkage disequilibrium and association mapping in diploid alfalfa (M. sativa L.). Theor Appl Genet 125:577–590

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

Sham P, Bader JS, Craig I, O'Donovan M, Owen M (2002) DNA pooling: a tool for large scale association studies. Nat Rev Genet 3:862–871

Shan F, Clarke HC, Yan G, Plummer JA, Siddique KHM (2007) Identification of duplicates and fingerprinting of primary and secondary wild annual Cicer gene pools using AFLP markers. Genet Resour Crop Evol 54:519–527

Shi A, Chen P, Zhang B, Hou A (2010) Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. Plant Breed 129:233–348

Shifman S, Bhomra A, Smiley S, Wray NR, James MR, Martin NG, Hettema JM, An SS, Neale MC, van den Oord EJ, Kendler KS, Chen X, Boomsma DI, Middeldorp CM, Hottenga JJ, Slagboom PE, Flint J (2008) A whole genome association study of neuroticism using DNA pooling. Mol Psychiatry 13:302–312

Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR (2007) Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. Genes Immun 8:57–68

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tidow N, Boecker A, Schmidt H, Agelopoulos K, Boecker W, Buerger H, Brandt B (2003) Distinct amplification of an untranslated regulatory sequence in the egfr gene contributes to early steps in breast cancer development. Cancer Res 63:1172–1178

Toker C (2009) A note on the evolution of kabuli chickpeas as shown by induced mutations in Cicer reticulatum Ladizinsky. Genet Resour Crop Evol 56:7–12

Toker C, Canci H, Ceylan FO (2006) Estimation of out-crossing rate in chickpea (Cicer arietinum L.) sown in autumn. Euphytica 151:201–205

Udvardi MK, Kakar K, Wandrey M, Montanari O, Murray J, Andriankaja A, Zhang JY, Benedito V, Hofer JM, Chueng F, Town CD (2007) Legume transcription factors: global regulators of plant development and response to the environment. Plant Physiol 144:538–549

Upadhyaya HD, Ortiz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet 102:1292–1298

Upadhyaya HD, Bramel PJ, Singh S (2001) Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci 41:206–210

Upadhyaya HD, Ortiz R, Bramel PJ, Singh S (2002) Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. Euphytica 123:333–342

Upadhyaya HD, Dwivedi SL, Baum M, Varshney RK, Udupa SM, Gowda CL, Hoisington D, Singh S (2008) Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (Cicer arietinum L.). BMC Plant Biol 8:106

Upadhyaya HD, Wang Y-H, Sharma S, Singh S, Hasenstein KH (2012) SSR markers linked to kernel weight and tiller number in sorghum identified by association mapping. Euphytica 187:401–410

Van Berloo R (1999) GGT: software for the display of graphical genotypes. J Hered 90:328–329

van der Maesen LJG, Maxted N, Javadi F, Coles S, Davies AMR (2007) Taxonomy of the genus Cicer revisited. In: Yadav SS, Redden B, Chen W, Sharma B (eds) Chickpea breeding and management. CAB International, Wallingford, pp 14–46

Varshney RK, Paulo MJ, Grando S, van Eeuwijk FA, Keizer LCP, Ceccarelli S, Kilian A, Baum M, Graner A (2012) Genome wide association analyses for drought tolerance related traits in barley (Hordeum vulgare L.). Field Crops Res 126:171–180

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CL, Singh NP, Lichtenzveig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR (2013) Draft genome sequence of chickpea (Cicer arietinum) provides a resource for trait improvement. Nat Biotechnol 31:240–246

Wang Y-H, Poudel DD, Hasenstein KH (2011) Identification of SSR markers associated with saccharification yield using pool-based genome-wide association mapping in sorghum. Genome 54:883–889

Wang YH, Bible P, Loganantharaj R, Upadhyaya HD (2012) Identification of SSR markers associated with height using pool-based genome-wide association mapping in sorghum. Mol Breed 30:281–292

Winter P, Pfaff T, Udupa SM, Hüttel B, Sharma PC, Sahi S, Arreguin-Espinoza R, Weigand F, Muehlbauer FJ, Kahl G (1999) Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. Mol Gen Genet 262:90–101

Winter P, Benko-Iseppon AM, Hüttel B, Ratnaparkhe M, Tullu A, Sonnante G, Pfaff T, Tekeoglu M, Santra D, Sant VJ, Rajesh PN, Kahl G, Muehlbauer FJ (2000) A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *Cicer arietinum* × *Cicer reticulatum* cross: localization of resistance genes for *Fusarium* wilt races 4 and 5. Theor Appl Genet 101:1155–1163

Xia Y, Ning Z, Bai G, Li R, Yan G, Siddique KH, Baum M, Guo P (2012) Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (*LHCB1*) associated with agronomic traits in barley. PLoS ONE 7:e37573

Yan WG, Li Y, Agrama HA, Luo D, Gao F, Lu X, Ren G (2009) Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). Mol Breed 24:277–292

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

Zhao Y, Wang S (2009) Optimal DNA pooling-based two-stage designs in case-control association studies. Hum Hered 67:46–56

Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20