9-1-2016

Multiple post-domestication origins of kabuli chickpea through allelic variation in a diversification-associated transcription factor

R. Varma Penmetsa
University of California, Davis

Noelia Carrasquilla-Garcia
University of California, Davis

Emily M. Bergmann
University of California, Davis

Lisa Vance
University of California, Davis

Brenna Castro
University of California, Davis

See next page for additional authors

Recommended Citation

Varma Penmetsa R, Carrasquilla-Garcia N, Bergmann EM, Vance L, Castro B, Kassa MT, Sarma BK, Datta S, Farmer AD, Baek JM, Coyne CJ. Multiple post-domestication origins of kabuli chickpea through allelic variation in a diversification-associated transcription factor. New Phytologist. 2016 Sep;211(4):1440-51.

New Phytologist

Multiple post-domestication origins of *kabuli* chickpea through allelic variation in a diversification-associated transcription factor

R. Varma Penmetsa[1], Noelia Carrasquilla-Garcia[1], Emily M. Bergmann[1], Lisa Vance[1], Brenna Castro[1], Mulualem T. Kassa[1], Birinchi K. Sarma[1,2], Subhojit Datta[1,3], Andrew D. Farmer[4], Jong-Min Baek[1], Clarice J. Coyne[5], Rajeev K. Varshney[6], Eric J. B. von Wettberg[7,8] and Douglas R. Cook[1]

[1]Department of Plant Pathology, University of California, One Shields Ave, Davis, CA 95616, USA; [2]Department of Mycology and Plant Pathology, Banaras Hindu University, Pandit Madan Mohan Malviya Road, Varanasi, Uttar Pradesh 221005, India; [3]Indian Institute of Pulses Research, Kanpur, Uttar Pradesh 208024, India; [4]National Center for Genome Resources, 2935 Rodeo Park Drive East, Santa Fe, NM 87505, USA; [5]USDA-ARS, Western Regional Plant Introduction Station, Washington State University, 59 Johnson Hall, Pullman, WA 99164-6402, USA; [6]International Crops Research Institute for the Semi-Arid Tropics, Center of Excellence in Genomics, Patancheru, Andhra Pradesh 502324, India; [7]Department of Biological Sciences, Florida International University, 11200 SW 8th Street, Miami, FL 33199, USA; [8]Kushlan Institute for Tropical Science, Fairchild Tropical Botanic Garden, 10901 Old Cutter Road, Coral Gables, FL 33156, USA

Authors for correspondence:
R. Varma Penmetsa
Tel: +1 530 752 6886
Email: rvpenmetsa@ucdavis.edu

Douglas R. Cook
Tel: +1 530 754 6561
Email: drcook@ucdavis.edu

Received: *9 November 2015*
Accepted: *24 March 2016*

New Phytologist (2016) **211:** 1440–1451
doi: 10.1111/nph.14010

Key words: basic helix–loop–helix (bHLH) transcription factor, chickpea B locus, *Cicer arietinum*, diversification, domestication, legume ortholog, Mendel's A locus.

Summary

• Chickpea (*Cicer arietinum*) is among the founder crops domesticated in the Fertile Crescent. One of two major forms of chickpea, the so-called *kabuli* type, has white flowers and light-colored seed coats, properties not known to exist in the wild progenitor. The origin of the *kabuli* form has been enigmatic.

• We genotyped a collection of wild and cultivated chickpea genotypes with 538 single nucleotide polymorphisms (SNPs) and examined patterns of molecular diversity relative to geographical sources and market types. In addition, we examined sequence and expression variation in candidate anthocyanin biosynthetic pathway genes.

• A reduction in genetic diversity and extensive genetic admixture distinguish cultivated chickpea from its wild progenitor species. Among germplasm, the *kabuli* form is polyphyletic. We identified a basic helix–loop–helix (bHLH) transcription factor at chickpea's B locus that conditions flower and seed colors, orthologous to Mendel's A gene of garden pea, whose loss of function is associated invariantly with the *kabuli* type of chickpea.

• From the polyphyletic distribution of the *kabuli* form in germplasm, an absence of nested variation within the bHLH gene and invariant association of loss of function of bHLH among the *kabuli* type, we conclude that the *kabuli* form arose multiple times during the phase of phenotypic diversification after initial domestication of cultivated chickpea.

Introduction

Crop domestication and subsequent diversification represent adaptations to human-built environments and offer insights into the evolutionary forces that shape phenotypic diversity. Chickpea (*Cicer arietinum*), a widely cultivated food legume, was domesticated in the Fertile Crescent together with several other so-called founder crops (Zohary *et al.*, 2012). This is evidenced by the Neolithic archeological record (Tanno & Wilcox, 2006) and the prevalence of crop wild relatives in the eastern Mediterranean, particularly *Cicer reticulatum*, the wild annual *Cicer* species from which the cultigen is derived.

Both wild and cultivated chickpea are predominantly selfing (e.g. Singh & Diwakar, 1995) and diploid ($2n = 16$), with a genome size of 738 Mbp. Cultivated chickpea differs from its wild relatives in its erect plant growth habit, altered phenology, seed coat texture and reduced seed dormancy. These are 'domestication syndrome' traits that were probably favored by humans during crop domestication (Hammer, 1984). Further variation exists within cultivated chickpea, notably for seed features that reflect cultural preferences (size, shape, coat thickness and color), based on which cultivated chickpea is categorized as either '*desi*' or '*kabuli*' types. Seeds of *desi* are smaller in size, darker in color and with thicker seed coats, and are typically consumed after dehulling. In South Asia and sub-Saharan Africa, cultivation of *desi* predominated historically. The *kabuli* type is associated with light-colored, larger sized seeds and with thinner seed coats, and is typically consumed as whole seeds in west Asia and Mediterranean regions where *kabuli* is common.

The greater similarity of *desi* seed to that of wild *C. reticulatum* underlies the proposal that the *desi* type represents the early domesticated form (Ladizinsky & Adler, 1976), with *kabuli* being a subsequently derived type. These differences in market class partially correspond to differences in seed size and other morphological traits that have been used to draw a distinction between proposed *microspermum* and *macrospermum* subspecies

(Moreno & Cubero, 1978). Despite the widely held view that the *desi* and *kabuli* types are distinct subgroups within cultivated chickpea, molecular marker-based analyses have been equivocal in supporting such a distinction (Iruela *et al.*, 2002; Upadhyaya *et al.*, 2008; Roorkiwal *et al.*, 2014). Thus, the origin of the *kabuli* type is currently ambiguous.

The dark-seeded *desi* and the light-seeded *kabuli* types also differ in pigmentation in vegetative tissues (stem) and in flowers. Pink to purple pigmented petals are typical of *desi* chickpea, whereas non-pigmented (white) petals are characteristic of *kabuli* chickpea. Additional variants exist for color of seed coats (green, black) and flowers (blue), but these occur at low frequencies (Pundir *et al.*, 1985). Studies on the genetics of pigmentation in chickpea go back several decades (Pundir *et al.*, 1985; and references therein). Three genetic loci, C, B and P, conditioning flower color in chickpea have been described, where recessive alleles at either C or B confer white petal color, whereas the recessive condition at P results in blue petals (Kumar *et al.*, 2000).

Human selection for lighter seed color during crop domestication is well documented in pea (Hellens *et al.*, 2010) and sorghum (Wu *et al.*, 2012). Molecular genetic analyses of seed and flower color in soybean and garden pea (Hellens *et al.*, 2010; Yang *et al.*, 2010; Moreau *et al.*, 2012) implicate genes in the phenylpropanoid pathway (Winkel-Shirley, 2001; Grotewold, 2006), in which a bifurcation typically leads either to the production of condensed tannins (proanthocyanidins) in seeds, giving them their dark appearance, or to the brightly colored anthocyanins characteristic of flowers. Both the biosynthetic enzymes and their transcriptional regulators have been well characterized in several plants, notably Arabidopsis, maize and *Medicago*, the latter of which is a close relative of chickpea (Xie *et al.*, 2003; Baudry *et al.*, 2004; Broun, 2005; Appelhagen *et al.*, 2011; Dixon *et al.*, 2013).

We performed molecular diversity analyses in a representative collection of chickpea germplasm, together with its wild annual relatives, to characterize the impact of domestication and to reveal relationships within chickpea germplasm that hitherto were masked by extensive admixture (intermating) introduced through crop breeding. In parallel, using a candidate pathway and gene approach, we identified a transcription factor whose molecular variants tightly associate with the seed coat and flower pigmentation differences characteristic of *desi* vs *kabuli* types. Our results indicate that these *kabuli* characteristics arose multiple times after domestication of the cultivated chickpea during the phase of post-domestication diversification, and that the varied and conflicting support for *kabuli* being a distinct subtype within cultivated chickpea is probably a consequence of demographic factors, such as admixture and isolation, that occurred in germplasm collection and crop breeding.

Materials and Methods

Plant genotypes

We assembled a diverse range of germplasm representing both *desi* and *kabuli* market types, the primary (*Cicer reticulatum* Ladiz.) and secondary (*C. echinospermum* P. H. Davis) gene pools, and annual *Cicer* accessions from the tertiary gene pools (see Supporting Information Table S1) to assess patterns of population genetic variation and correspondence with the *kabuli* form. Plants of the recombinant inbred population CRIL7 (Tekeoglu *et al.*, 2000) and of the chickpea minicore were obtained from the US Department of Agriculture-Agricultural Research Service (USDA-ARS) in Pullman, WA, USA. Genomic DNA for the chickpea reference set (Upadhyaya *et al.*, 2008) was obtained from ICRISAT, Hyderabad, India.

Single nucleotide polymorphism (SNP) assay development

Genotyping was based on SNPs (Table S2) discovered from a comparison of Sanger-sequenced PCR amplicons between *C. arietinum* accession ICC4958 (a *desi* type) and wild *C. reticulatum* accession PI489777. These two genotypes are the parents of an existing recombinant inbred line (RIL) population. Genomic DNA was isolated from young leaf tissue using a Qiagen plant DNA isolation kit (Valencia, CA, USA) following the manufacturer's protocols. Polymorphisms were identified by sequence alignment in Jalview among sequence pairs for a set of 1440 primarily single copy genes, with orthology inferred from legume expressed sequence tag (EST) data (i.e. the transcriptomes of *Medicago truncatula*, *Lotus japonicus* and *Glycine max*) and subsequently based on conserved genome location in a multispecies comparative genetic analysis (D. R. Cook *et al.*, unpublished; and validated in chickpea, Hiremath *et al.*, 2012). SNPs meeting the assay design criteria, determined by Illumina Inc. using their proprietary Assay Design Tool, were converted to a 768 Illumina GoldenGate genotyping assay performed at the UC Davis Genome Center DNA Technologies Core. The resulting data were curated using the Illumina Beadstudio software package (Illumina, San Diego, CA, USA). SNPs that failed in > 20 samples or when genotyping data failed to accurately identify parental genotypes among replicated samples were excluded. This filtering yielded 538 high-quality SNPs (of the 768 SNPs evaluated) representing polymorphisms in 538 genes.

Population genetic analyses

Genetic relationships among genotypes were analyzed with the program STRUCTURE using an admixture model and the default settings (Pritchard *et al.*, 2000). Genotypes were from the primary, secondary and tertiary gene pools assayed for 538 SNPs in orthologous genes. In the initial analysis, we examined $K = 2$ to $K = 10$, using 20 iterations of 25 000–50 000 samples with a 25 000–50 000 burn-in phase. To select K that best fits the data, we used the inflection and plateaus of the log probability curves on various STRUCTURE runs, together with genetic, biological and geographical coherence of grouping. This indicated $K = 8$ as the maximal number of groups beyond which further subgroupings were poorly supported. $K = 3$ to $K = 8$ were analyzed in greater detail using three iterations of 1 000 000 samples with a 50 000 burn-in phase. The Evanno method, implemented in STRUCTUREHARVESTER, was used to select K that best fits the data (following Evanno *et al.*, 2005; Earl, 2012). Principal coordinate

analysis (PCoA) in GenAlEx 6.5 (Peakall & Smouse, 2012) was used as a further step to visualize the clustering of genetic variation (Peakall & Smouse, 2006, 2012).

To further assess genetic diversity within and among groups, a set of 93 low (< 10%) admixed accessions identified from Structure over $K = 5$–9 was analyzed using DARwin5 software (Perrier & Jacquemoud-Collet, 2006) to calculate the pairwise dissimilarity by simple matching according to the method of Saitou & Nei (1987). The resulting dissimilarity matrix was used to derive a weighted neighbor-joining tree (Saitou & Nei, 1987) with 1000 bootstraps. This weighted neighbor-joining analysis employs a likelihood-based criterion that models distance between genotypes as random variables that obey a Gaussian distribution (Bruno *et al.*, 2000).

GenAlEx 6.5 (Peakall & Smouse, 2012) was used to estimate the observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F_{is}) and percentage of polymorphic loci among groups that were circumscribed by geographical origin and population genetic analyses. Accessions were clustered by geographical origin hierarchically, first considering 10 geographical source regions (*C. reticulatum*, *C. echinospermum*, Indian subcontinent 1, 2 and 3, Iran 1 and 2, Syria, USSR and Ethiopia). Comparisons among groups were made using GenoDive 2.0 (Meirmans & Van Tienderen, 2004). Analysis of molecular variance (AMOVA) was implemented in GenAlEx 6.5 with codominant genotypic data.

Characterization of basic helix–loop–helix (bHLH) candidate and transcriptional targets

Plants were surface sterilized in bleach, planted in potting soil and grown in glasshouses at the UC Davis core plant growth facility. Flower colors of plants were recorded at maturity. For gene expression analysis, flower petals and pools of seed coats from developing seeds were harvested. Young developing leaves served as the source of genomic DNA. Genomic DNA and RNA were extracted with the Qiagen DNeasy and Qiagen RNeasy total RNA extraction kits using the manufacturer's standard procedures and quantified in a NanoDrop instrument (Wilmington, DE, USA).

For real-time reverse transcription polymerase chain reaction (rtPCR), equal amounts of total RNA were used in parallel rtPCRs analyzed on an ABI7500 instrument. The expression of the target genes BANYULS (BAN) and LDOX was quantified, together with that of the control gene Elongation Factor-1 (EF1) (Garg *et al.*, 2010), and the relative transcript abundance was calculated by the cycle threshold (CT) method (Schefe *et al.*, 2006) from triplicate samples relative to EF1. The PCR conditions are listed in Table S3.

For the determination of sequence variants, genomic DNA and cDNA were used as template in PCR amplification with ExTaq polymerase, and the resulting amplicons were purified by an ExoSAP kit (Affymetrix, Santa Clara, CA, USA) before Sanger sequencing (for bHLH allele B (wild-type) and the missense and short deletion variants, b1, b2, b3 and b5), or analyzed by agarose gel electrophoresis (for the large deletion variant b4). Sequence trace files were examined in Sequencher software to curate for quality and to identify nucleotide variants. The oligonucleotides used for amplification and sequencing are given in Table S3.

Results

Patterns of molecular diversity in cultivated chickpea and its close wild annual relatives

With the goal of describing the molecular diversity of cultivated chickpea relative to wild annual *Cicer* species, we conducted multi-locus genotyping in a representative set of 322 accessions, including 224 cultivated chickpea and 98 wild annual genotypes which, together, span the primary, secondary and tertiary gene pools (Table S1). Genotyping was based on 538 high-quality, bi-allelic SNPs that distinguish alleles of low-copy conserved orthologous genes of cultivated *C. arietium* ICC4958 and wild *C. reticulatum* PI489777 (Table S2).

Genetic relationships inferred from allele frequencies (i.e. Structure: Pritchard *et al.*, 2000; Falush *et al.*, 2003; Hubisz *et al.*, 2009) revealed the separation of cultivated from wild accessions (Fig. 1). Moreover, analyses involving increased division into subgroups (i.e. $K = 2, 3, \ldots, n$) effectively resolved the wild progenitor species (i.e. *C. reticulatum*) and its nearest wild neighbor (*C. echinospermum*) from other wild species, and revealed numerous subgroups within cultivated material, including the bifurcation of cultivated material into accessions primarily from the Indian subcontinent and from West/Central Asia and the Middle East (Fig. 1a–c). The mean log probability of the data (ln P[D]) calculations (following Evanno *et al.*, 2005; Earl, 2012) indicated that cultivated material is best described as six subgroups.

Estimates of genetic diversity among all groups, cultivated and wild, revealed the impact of domestication and breeding. AMOVA partitioned diversity among hierarchical sets of individual genotypes (two groups (cultivated and wild) and 10 subgroups) that were circumscribed by Structure. In an AMOVA, the majority of variation was attributed to differences between cultivated and wild groups (93%), with low levels of differentiation among cultivated groups (4%) and low levels of variation among individuals within groups (3%) (Table S4). Cultivated germplasm had significantly lower heterozygosity ($P = 0.028$) than either *C. reticulatum* or *C. echinospermum* (Table S4). Similarly, the cultivated group had a significantly lower level of gene diversity ($H_e = 0.065$; Table S5; $P = 0.028$ for the comparison) when compared with the wild relatives, *C. reticulatum* ($H_e = 0.332$; Table S5) or *C. echinospermum* ($H_e = 0.301$; Table S5). Although relatively small sample sizes may bias gene diversity estimates (Nei, 1978), genetic diversity within and differentiation among all cultivated groups were low.

Interestingly, none of the genetic groups identified based on allele frequency data corresponded to the often-asserted dichotomy of a *desi* and *kabuli* division within cultivated chickpea. This was most apparent at $K = 3$ (Fig. 1a), but remained a feature as additional subgroups were partitioned (Fig. 1b). To further test the hypothesis that *kabuli* has a single origin, not

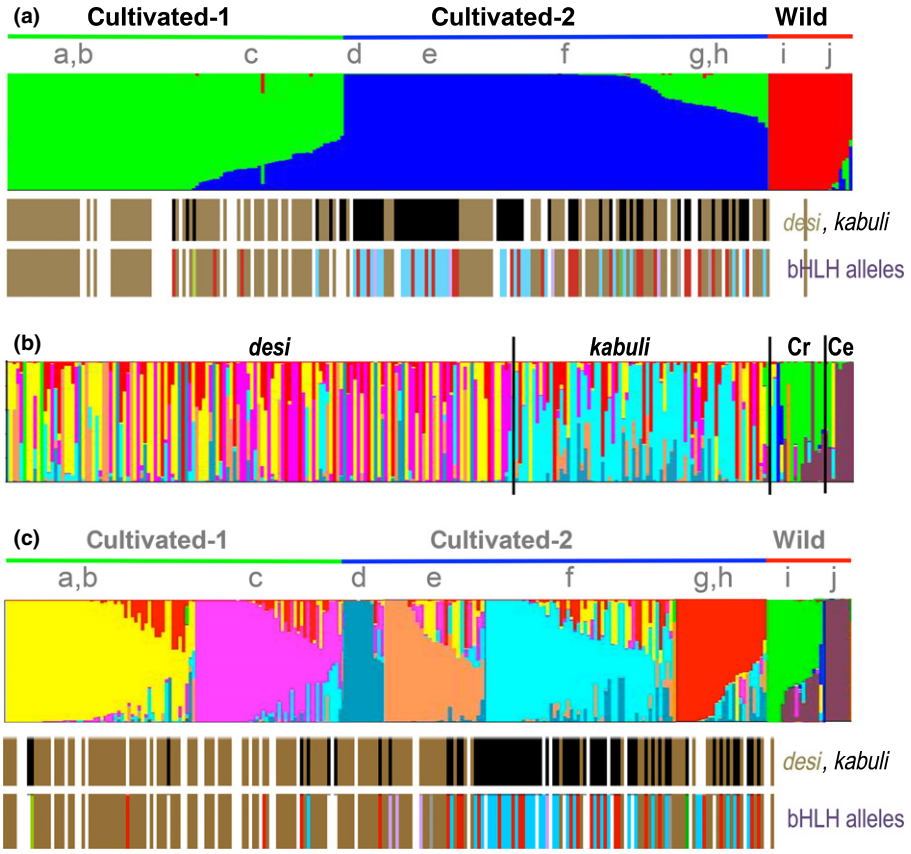

Fig. 1 Analysis of molecular variation in 247 accessions of cultivated chickpea and its wild relatives *Cicer reticulatum* and *C. echinospermum*. Multilocus single nucleotide polymorphism (SNP) data from 538 low-copy genes were analyzed. (a) Upper panel: proportional membership coefficients in STRUCTURE at $K = 3$. Each accession is represented by a single vertical line, with accessions ordered in decreasing coefficient values within each group. Middle panel: biological classification of accessions from passport data: brown, *desi*; black, *kabuli*; white, unknown. Lower panel: basic helix–loop–helix (bHLH) allelic variant of accessions: brown, wild-type *B* allele; light green, variant *b1*; turquoise, variant *b2*; dark green, variant *b3*; magenta, variant *b4*; red, variant *b5*. (b) Proportional membership coefficients in STRUCTURE at $K = 8$. Accessions are arranged by the biological category of their passport data which is indicated above the box; Cr, wild *C. reticulatum*; Ce, wild *C. echinospermum*. (c) Upper panel: accessions sorted by cluster membership at $K = 8$. Correspondence to the three clusters shown in (a) is marked by thin colored lines, which indicates subfractionation of the $K = 3$ subgroups into additional subgroups at $K = 8$; for example, the wild group (a; red) resolves into the two distinct species Cr and Ce at higher K. Letters above correspond to the groupings marked in Fig. 2. Middle panel: biological classification (brown, black and white colors represent *desi*, *kabuli* and unknown types, respectively) as shown for (a). Lower panel: allelic form of the bHLH gene, with assignments as given in (a), and congruent with those in Fig. 2.

reticulating with *desi*, a dissimilarity matrix was constructed and visualized as a neighbor-joining dendrogram (Fig. S1). At the species level, the resulting topology agreed with the prevailing taxonomic relationships among annual *Cicer* species (Javadi *et al.*, 2007; Fig. S1). At a finer scale (Fig. 2), it readily identified the misclassified nature of accession PI 593709, a *C. echinospermum* genotype that was categorized as *C. reticulatum* at its collection time and in extant germplasm repositories. Accession PI 593709 grouped within a cluster that was exclusively composed of *C. echinospermum* accessions, as has been suggested previously (Iruela *et al.*, 2002). Importantly, and in agreement with the results of STRUCTURE, among cultivated accessions the dissimilarity matrix did not yield a bipartite topology for *desi* and *kabuli* types. The same was true of analyses using the alternative minimum evolution method.

The analysis of population structure also revealed a high rate of admixture (Fig. 1). Among cultivated genotypes, such admixture is probably the consequence of breeding history. Because intermating during breeding might obscure inherent genetic stratification, we used membership coefficients obtained from STRUCTURE to identify 93 accessions with low levels (< 10%) of admixture: 81 of 222 *C. arietinum*, five of 16 *C. reticulatum* and all seven *C. echinospermum* accessions met this criterion. A dissimilarity matrix constructed from these 93 low admixed accessions separated domesticated from wild, and partitioned the 83 cultivated accessions into eight groups (Fig. 2). PCoA (Fig. 3) yielded congruent groupings, dividing clusters of genotypes along discriminant axes 1 and 2. The first axis (which accounted for 84.49% of the variation) separated cultivated from wild groups, whereas the second axis (5.75% of the variation) revealed four groups among the cultivated accessions.

In marked contrast with the result for all 322 accessions, groupings obtained with the low admixed set, either by neighbor joining or PCoA, reflected their geographical origins. Minimally,

New Phytologist

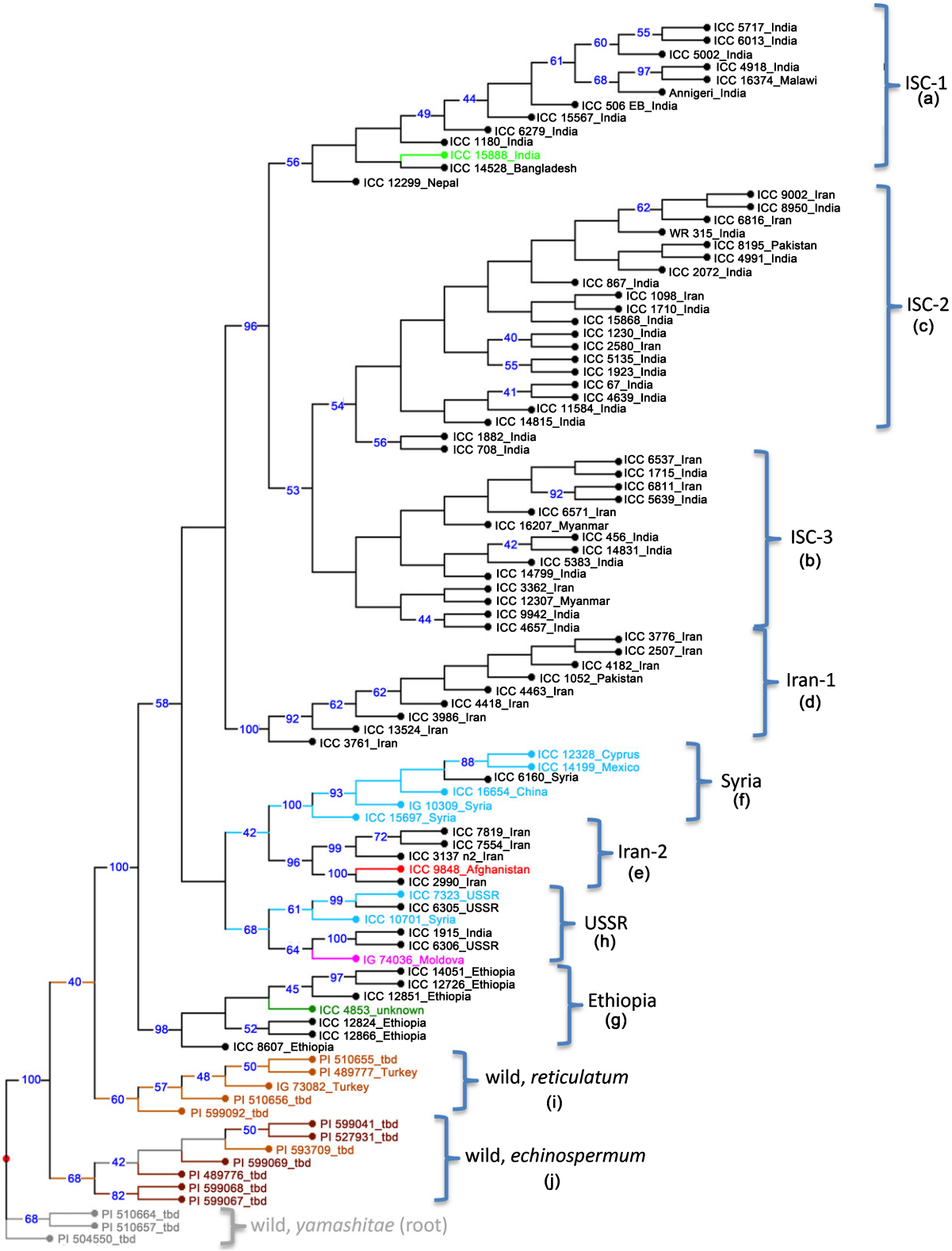

Fig. 2 Weighted neighbor-joining dendrogram of the subset of 93 accessions with low levels of admixture. Eight groups within cultivated accessions and two additional groups, one each of the wild annual relatives *Cicer reticulatum* (light brown) and *C. echinospermum* (dark brown), are discernible. The dendrogram is rooted in three accessions from the distantly related wild species *C. yamashitae* (light gray). Bootstrap values are shown when > 40 (blue numbers). Names of the groups are to the right, with lowercase letters in parentheses corresponding to those in Fig. 1(c). ISC, accessions predominantly from the Indian subcontinent (largely from India and occasionally from Pakistan, Nepal, Bangladesh, Myanmar and Iran). In the dendrogram, *desi* accessions are in black text. The colors of the *kabuli* accessions correspond to their basic helix–loop–helix (bHLH) gene variants: light green, variant *b1*; turquoise, variant *b2*; dark green, variant *b3*; magenta, variant *b4*; red, variant *b5*.

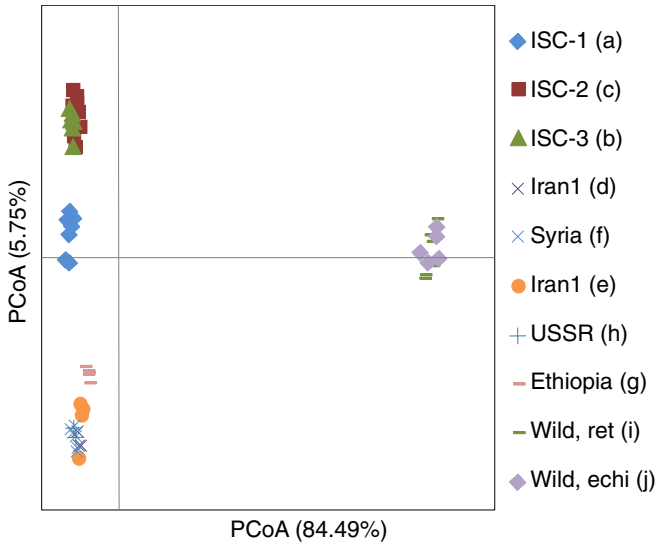

Fig. 3 Principal coordinate analysis (PCoA) of the 93 accessions with low levels of admixture. Wild relatives are separated from cultivated accessions (*x*-axis), with additional separation among the cultivated accessions into three major clusters (*y*-axis). Letters in parentheses correspond to the clades identified in Figs 1 and 2. Wild accessions: ret (i), *Cicer reticulatum*; echi (j), *C. echinospermum*; with remaining clusters comprising cultivated chickpea accessions.

we identified a basal group consisting of accessions from Ethiopia, a subtending group composed of four geographically related subgroups from West/Central Asia and the Middle East, and a coherent subgroup predominantly from the Indian subcontinent that was further divisible into two or three subgroups based on neighbor-joining analyses or PCoA, respectively (Figs 2, 3). The fact that groups that formed from low admixed genotypes reflect geographical origins of accessions is probably a consequence of local breeding practices and regional cultivar preferences. Interestingly, even among these low admixed individuals, we observed *kabuli* and *desi* flower and seed coat phenotypes distributed throughout multiple groups (Figs 1c, 2), suggestive of introgression and extensive backcross selection for these traits, and/or of the possibility of their multiple origins.

Functional variation in a transcriptional regulator assorts with the *kabuli* type

The differences in pigmentation of flowers, stems and seed coats between *desi* and *kabuli* chickpea (Fig. 4) are suggestive of differences in flavonoid biosynthesis. Because transcriptional changes often accompany crop domestication (e.g. Doebley *et al.*, 2006; Olsen & Wendel, 2013), we examined components of the ternary 'BMW' transcriptional complex, composed of a bHLH, a Myb-domain and a WD40-repeat protein, which is a key regulatory module for flavonoid biosynthesis (Baudry *et al.*, 2004; Broun, 2005; Grotewold, 2006; Appelhagen *et al.*, 2011; Hichri *et al.*, 2011). Our analysis of previous low-resolution genetic maps between chickpea (Millan *et al.*, 2010) and pea (Hellens *et al.*, 2010), using the *Medicago* genome as a bridge, determined that the phenotypically defined chickpea 'B' locus is syntenic and thus potentially orthologous to the A locus of pea. The A locus is Mendel's flower color gene A and is conditioned by the BMW bHLH protein (Hellens *et al.*, 2010).

Degenerate PCR primers, based on the pea and *Medicago* orthologs of A (bHLH), were used to amplify and then sequence cDNAs from petals and seed coats of *desi* (colored) chickpea genotype ICC4958 and wild *C. reticulatum* genotype PI599072 (also colored petals and dark seed coats), yielding identical sequence homologs of the bHLH protein. Phylogenetic analysis (Fig. 4f) positioned the chickpea candidate as orthologous to the authentic BMW bHLH proteins of *Medicago* and pea, whereas comparison with the whole-genome sequence of the reference *kabuli* accession CDC Frontier (Varshney *et al.*, 2013) placed the chickpea ortholog on CaLG04, consistent with previous linkage mapping studies that position the B locus of chickpea to the top arm of CaLG4 (Millan *et al.*, 2010). Recombinants in a RIL population derived from *C. reticulatum* PI599072 and *kabuli* FLIP94-92C (Tekeoglu *et al.*, 2000) and segregating for flower color delimited a physical interval of 741 kbp containing 61 genes, including the bHLH, in a region syntenic to *Medicago* chromosome 8 and pea LG-II, where the functionally characterized orthologs of bHLH are located. Sequencing of this bHLH revealed a *c.* 1 kbp deletion specific to the *kabuli* parent FLIP84-92C that removes 223 bp of the coding region, spanning all of exon 4 and a portion of exon 5 (Fig. 4g). This deletion is shared in common with the genome sequenced reference *kabuli* genotype 'CDC Frontier' (Varshney *et al.*, 2013) and results in coding region variation, including a premature stop codon that is predicted to eliminate protein function.

Multiple altered functional alleles of the bHLH transcription factor are correlated with absent target gene expression

Sequencing of full-length bHLH transcripts from flower and seed coat tissues in a set of 35 accessions (11 *desi* and 23 *kabuli* *C. arietinum* accessions and one accession of the wild relative *C. echinospermum*), encompassing the known seed color variation of chickpea (dark, green, dark tan and light tan), identified five coding region variants exclusive to and characteristic of *kabuli* individuals (Fig. 4g). We designated the allele of the *desi* type as 'B' and the variant alleles of *kabuli* types as 'b'1–5 to reflect their known or predicted recessive nature relative to B (Fig. 4g). Conceptual translation of the three deletion alleles 'b1', 'b3' and 'b4' indicates that these result in frameshifts and premature termination of translation (Fig. 4g). In particular, allele b4, which is a 223-bp deletion in the transcript, corresponds to the *c.* 1 kbp length variant identified in *kabuli* genomic DNAs and used for co-segregation analysis, as described earlier. The missense variant allele b2 converts glutamine to proline (Q71P) within the Myb interaction domain of the bHLH protein (Hichri *et al.*, 2011). This glutamine residue is invariant among bHLH orthologs, including proteins from basal vascular lineages (mosses), suggesting that it may be critical for protein–protein interactions between the Myb and bHLH components of the BMW regulatory complex, a prediction that is also supported by SIFT (Kumar *et al.*, 2009). Missense variant allele b5 is restricted to and characteristic of

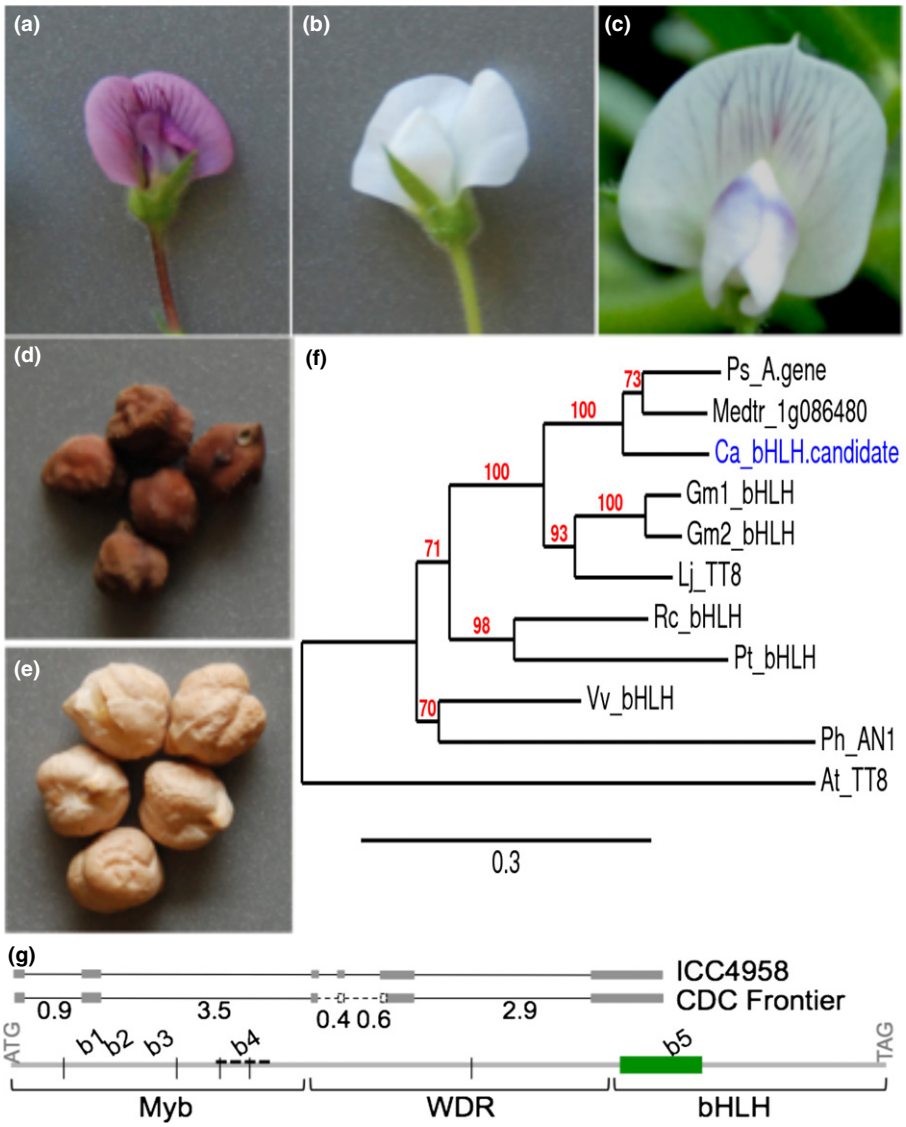

Fig. 4 Morphology of flower and seed coats of the two major types of cultivated chickpea. Flower and seeds of the (a, d) *desi* type and (b, e) *kabuli* type. (c) Pigmentation on a flower of the weak allele *b5* of the chickpea *B* locus. (f) Phylogram of the basic helix–loop–helix (bHLH) candidate gene of chickpea (blue), together with the best hit sequences from several other eudicot plants: Medicago (Medtr), garden pea (Ps_A; Mendel's A gene), two homeologs from soybean (Gm1, Gm2) and Lotus (Lj), together with those from castor (Rc), Poplar (Pt), grape (Vv) and the functionally characterized orthologs from Petunia (Ph_AN1) and Arabidopsis (At_TT8). Bootstrap values are given in red text. (g) Gene and transcript structures of the chickpea bHLH transcription factor. Gene structures of the bHLH gene in the reference genomes of the *desi* accession ICC4958 (upper row) and the *kabuli* accession CDC Frontier (middle row). Exons are denoted by shaded gray boxes and introns are represented by a thin horizontal line. Intron lengths in kbp are listed below the line. Lower panel: transcript structure of the 'wild-type' *desi* bHLH gene, from start codon ATG to stop codon TAG. The six exons are depicted by the horizontal line, with exon–exon junctions demarcated by vertical lines. The positions of the five coding region variants are listed above the exons; the coding region absent in variant b4 is marked by a dotted horizontal line. The bHLH domain is boxed in green. Brackets below mark the regions of predicted interaction with members of the ternary transcriptional complex (Myb, with Myb-type transcription factor; WDR, with WD-repeat protein; bHLH, for homo-/heterodimerization and DNA binding).

accessions that share an intermediate flower color phenotype, wherein only portions of petals, particularly vasculature, are pigmented (Fig. 4c). In the deduced b5 protein, serine replaces leucine (S497L) in a residue that is conserved in bHLH proteins throughout vascular plants. This particular leucine residue is one of 19 conserved residues that are hallmarks of the bHLH domain (Toledo-Ortiz *et al.*, 2003) and among 11 residues that form the dimerization interface, suggesting that the weak flower color phenotype of the b5 allele may derive from reduced hetero- or homodimerization capacity in the b5 variant.

To test the deduced effects of these variants on bHLH protein function, we quantified the expression of chickpea orthologs of leucoanthocyanin dioxygenase (LDOX/ANS) and anthocyanidin reductase (ANR/BANYULS), as these enzymes are known transcriptional targets of the BMW ternary complex in other plants (Appelhagen *et al.*, 2011), and their expression is indicative of anthocyanin production in flowers and pro-anthocyanidin

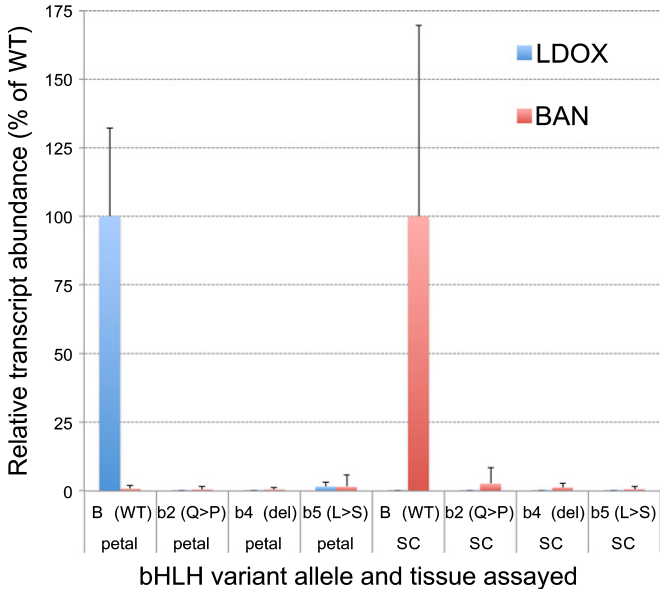

Fig. 5 Expression levels of target genes LDOX (anthocyanidin synthetase) and BAN (anthocyanidin reductase) in genotypes determined by real-time PCR in cultivated chickpea (*Cicer arietinum*) flowers and developing seed coats. Average cycle threshold (CT) values ± SD were calculated from three replicate samples of three to four biological samples (accessions) per variant allele. Transcript abundance is expressed relative to the control gene Elongation Factor-1. Deduced alterations in the basic helix–loop–helix (bHLH) protein are listed below the alleles in parentheses. WT, wild-type; SC, seed coat.

(condensed tannin) production in seed, respectively. Among all *desi* and a wild *Cicer* accession that share a pink/purple flower and dark-colored seed coat phenotypes, expression of LDOX was exclusively in petal tissues with an apparent absence in seed coats, whereas BANYULS exhibited the opposite pattern, being uniformly high in seed coats and at much lower levels in petal tissues (Fig. 5). By contrast, expression of both LDOX and BANYULS was undetected in *kabuli* accessions harboring bHLH alleles b1–4, and was low for accessions with allele b5 (Fig. 5). These results are consistent with our interpretation that deletion alleles b1, b3 and b4 and the missense allele b2 are functional null alleles, and that the missense allele 'b5' located in the dimerization interface is a weak allele of bHLH.

Association of bHLH polymorphisms with *kabuli* type in germplasm

To determine the prevalence and distribution of variant alleles within chickpea germplasm, we sequenced amplicons from an additional 178 chickpea accessions (109 *desi*, 69 *kabuli*) of the chickpea minicore collection (Upadhyaya & Ortiz, 2001), which is broadly representative of chickpea germplasm. Among these 178 accessions, only *desi* genotypes contained the wild-type B allele, 107 as homozygotes and two as heterozygotes with either the b4 or b5 allele. Conversely, all 69 *kabuli* accessions were homozygous for one of the five bHLH altered functional variants (Table 1). Two variants, the Q to P missense alteration (allele b2) and the large deletion (allele b4), occurred at high frequencies, in 36 and 26 accessions, respectively (Table 1). The L to S variant (allele b5) was observed in all five accessions that shared a weak flower color phenotype (Table 1). The 2-bp and 7-bp long reading frame deletions (alleles b1 and b3, respectively) were at low frequencies, being observed once each (Table 1). Crucially, all 69 *kabuli* accessions contained only one of the originally defined five variant haplotypes, with no evidence of interallelic recombination. This is indicative of multiple, independent origins of these functional variants and thus of the *kabuli* form(s), effectively excluding a step-wise, intragenic mutation accumulation model.

Among the low admixed set of 93 accessions, frequencies of the bHLH variants were generally similar to those observed in the unfiltered set of 222 cultivated accessions. Moreover, *kabuli* alleles were distributed among multiple groups, as would be expected for independently derived alleles. The low-frequency bHLH variants *b1*, *b3* and *b5* occurred once, each in groups otherwise composed of *desi* accessions (ISC1, Ethiopia and Iran-2, respectively; Fig. 2). High-frequency variant *b2* was predominant within the Syrian group, but also occurred within the USSR

Table 1 Basic helix–loop–helix (bHLH) transcript haplotypes in *desi* and *kabuli* accessions in cultivated chickpea (*Cicer arietinum*)

		No. of samples		Allele frequency	Allele frequency	Nucleotide position & sequence within coding region					
Type	Allele	As homozygous	As heterozygous	(overall) (%)	(within type) (%)	199–200 (2 bp)	215	238–244 (7 bp)	461–683 (223 bp)	1490	Predicted effect on protein
desi	B	107	2	60.8	99.1	CT	A	TTATCTG	Present	T	WT/functional protein
kabuli	b1	1	0	0.6	1.4	–	A	TTATCTG	Present	T	2-bp deletion; frameshift
kabuli	b2	36	0	20.1	51.4	CT	C	TTATCTG	Present	T	Mis-sense CAA → CCA; glutamine → proline
kabuli	b3	1	0	0.6	1.4	CT	A		Present	T	7-bp deletion; frameshift
kabuli	b4	26	1	14.8	37.9	CT	A	TTATCTG	Absent	T	Large internal 223-bp deletion + frameshift
kabuli	b5	5	1	3.1	7.9	CT	A	TTATCTG	Present	C	Mis-sense TTA → TCA; leucine → serine

The nature of allelic variants, their frequencies among all cultivated germplasm, and within *desi* and *kabuli* seed types, together with the predicted effects on the bHLH protein, are shown. CT, here is for the nucleotides 'C' and 'T' that occupy positions 199 and 200 within this sequence. Only variant nucleotide positions are listed. WT, wild-type.

group, where variant *b4* was also found. Variant *b4* was exceptional in being one of two frequent alleles among *kabuli* overall (Table 1), but occurring only once in the low admixed set, in the USSR clade (Fig. 2; accession IG74036 from Moldova, formerly USSR). This suggests that most *kabuli* accessions containing variant *b4*, including the reference *kabuli* accession 'CDC Frontier', have received the *b4* allele through breeding introgression. Indeed, the average admixture rate of accessions containing allele b4 was 34%, with 21 of 28 accessions with admixture rates in excess of 20%.

Discussion

The domestication of plants is shaped by human preference for traits, ranging from altered growth habit to ease of harvest, enlargement and flavor of edible parts, adjusted phenology and visual appearance of flowers or seeds. Artificial selection for preferred traits has continued since domestication. The selection and replanting of light-colored chickpea seed (i.e. the *kabuli* type of chickpea) represent the manifestation of such human preference. Indeed, condensed tannins can reduce protein digestibility and impart an astringent character to seed (Aw & Swanson, 1985; Dixon *et al.*, 2013), which may have been factors in the selection of light-colored seed and, coincidentally, white-flowered plants. Consistent with this suggestion, contemporary chickpea culinary practices mimic genetics in that dark-seeded *desi*, but not light-seeded *kabuli*, types are typically consumed after dehulling. Here, we provide genetic and transcriptional evidence that flower and seed coat color in chickpea derive from multiple, independent alleles of the same gene, selected independently to create the same light-colored phenotype throughout the existing chickpea germplasm. As is common to many other artificially selected agricultural traits (e.g. Doebley *et al.*, 2006; Olsen & Wendel, 2013), the targets of selection were loss-of-function mutations in a transcriptional regulator, the chickpea ortholog of Mendel's A gene in pea. With the exception of its heterogeneous origins, the genetic architecture underlying the loss of flower and seed coat color in cultivated chickpea germplasm is thus relatively simple and not unlike growth habit and flowering time differences between wild and cultivated chickpea, which are also controlled by large-effect loci (Millan *et al.*, 2010). Indeed, major effect mutations swept to high frequency by human selection are common in the domestication history of most crops, although increasing numbers of minor loci are also being identified (reviewed in Doebley *et al.*, 2006; Zachary *et al.*, 2007; Olsen & Wendel, 2013).

The finding that all *kabuli* genotypes contain only single bHLH variants without nesting of mutations or evidence of interallelic recombination argues against a stepwise mutation accumulation model, and instead suggests that the five alleles arose independently, a conclusion that is further supported by the largely non-overlapping distribution of bHLH mutant alleles among genetically resolved groups. The preponderance in cultivated chickpea of multiple mutant isoforms in only one component of the BMW transcriptional complex is curious, given that loss of function in other components of the BMW complex leads to comparable phenotypes in other legumes. Thus, loss-of-function mutations are the source of breeding alleles in cultivated pea at the bHLH 'A' locus and at the WDR 'A2' locus (Hellens *et al.*, 2010), and in soybean at the Myb 'W2' locus (Takahashi *et al.*, 2012). It is possible that the Myb and WDR components of the chickpea BMW complex confer additional adaptations and thus are under purifying selection for function, or that there is (partial) functional redundancy for these components. Indeed, multiple Myb paralogs in grape have the capacity to regulate structural genes in the anthocyanin and condensed tannin pathways (Appelhagen *et al.*, 2011; and references therein).

Although the significant majority of characterized natural variation in anthocyanin and/or proanthocyanidin content in seeds and flowers derives from altered transcriptional control (as in the cases of sorghum, Wu *et al.*, 2012; rice, Sweeney *et al.*, 2006; apple, Espley *et al.*, 2009; cauliflower, Chiu *et al.*, 2010), flavonoid biosynthetic genes are also potential targets of human selection in crops (e.g. soybean, Yang *et al.*, 2010) and possibly also in chickpea. Thus, in chickpea, a tandem cluster of MATE transporters, orthologous to TT12 that acts in proanthocyanidin biosynthesis in legumes (Zhao & Dixon, 2009), occurs in a narrow genomic interval that has unusually low diversity among *desi* but not *kabuli* genotypes (Varshney *et al.*, 2013), consistent with a selective sweep and purifying selection at these TT12 orthologs in dark-seeded genotypes. Likewise, variation in flavonoid biosynthetic loci could provide the range of hues of almost white through very dark brown that exist in chickpea, and indeed allelic variation in these biosynthetic genes accounts for flower color variation in some species (Broun, 2005).

Although white-flowered *kabuli* genotypes are common in cultivated chickpea, they are rare in wild *Cicer*, being known only in perennial *Cicer* species in central Asia (Uzbekistan) in the tertiary gene pool (Van Oss *et al.*, 2015), or derived from induced mutagenesis (Toker, 2009). We surveyed 23 natural populations of the cultigen's progenitor (*C. reticulatum*) from the primary gene pool, and its sister species *C. echinospermum* from the secondary gene pool, and did not encounter white-flowered individuals among > 1000 wild annual *Cicer* individuals observed throughout southeastern Anatolia. This suggests strong natural selection for colored flowers and/or seed in natural settings. Indeed, flower color can be an important determinant of pollinator activity (Stanton *et al.*, 1986, 1989), even in self-pollinated legumes, in which pollinator activity can increase seed set (Palmer *et al.*, 2009). Moreover, the products of anthocyanin and proanthocyanidin metabolism function as antioxidants and defensive chemicals and have been speculated to play a variety of ecological roles (Gould, 2004; Strauss & Whittall, 2006; Yuan *et al.*, 2012). Although it is possible that pre-existing null alleles of bHLH were introgressed into cultivated chickpea from co-occurring wild relatives in which the alleles may have been at low frequencies, we favor the hypothesis that the existing chickpea bHLH loss-of-function alleles arose from novel post-domestication mutation events during the phase of crop diversification. Indeed, the maintenance of multiple, independent loss-of-function *kabuli* bHLH alleles within cultivated germplasm is consistent with the observation that many traits associated with crop improvement, as

opposed to domestication, remain variable in modern germplasm (e.g. Pundir *et al.*, 1985).

Our results strengthen and extend the inferred phylogeny of annual *Cicer* of previous studies (Javadi *et al.*, 2007), identifying *C. reticulatum* as the likely progenitor of cultivated chickpea. Moreover, the identification of genetic groups, resolved by multiple criteria that largely correlate with geographical distribution, adds an additional dimension to our understanding of chickpea germplasm. For example, at $K = 3$ (Fig. 1), cultivated chickpea bifurcates into two groups whose origins are predominantly from the Indian subcontinent, and West/Central Asia and the Middle East, respectively. This separation may reflect the historically distinct breeding mandates of regional chickpea breeding centers (i.e. the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India and the International Center for Agricultural Research in the Dry Areas (ICARDA) in Syria), although such patterns also mirror known events in early secondary domestication. Thus, the distinctiveness of Indian and Ethiopian genotypes (Figs 1, 2) agrees with the archeological record, indicating that chickpea cultivation spread from the Middle East to South Asia at least 4000 yr ago and to Ethiopia at least 2300 yr ago (Redden & Berger, 2007); indeed, Harlan (1992) recognized India and Ethiopia as centers of chickpea diversity. In contrast with the idea of species distinctiveness, however, we identified many *C. reticulatum* accessions and several cultigen accessions that are admixtures (see Table S1), indicating gene flow between cultivated and wild forms and highlighting the potentially protracted, even ongoing, contribution of wild diversity to cultivated chickpea. Although chickpea and its relatives are predominantly selfing, some outcrossing is possible. Alternatively, these may represent contamination in collections, consistent with the lack of admixture recently found by Van Oss *et al.* (2015) in a working collection of wild material. Irrespective of their origins, from a practical point of view, hybrid accessions have the potential to serve as genetic bridges for the introgression of agronomically desired traits (e.g. biotic and abiotic stress tolerance, phenological variation) from more distantly related wild species into cultivated chickpea, which has thus far enjoyed only sporadic and unpredictable success (Hajjar & Hodgkin, 2007; Warschefsky *et al.*, 2014). Even more extensive admixture is observed with cultivated germplasm among groups with distinct geographical origins (evident in Fig. 1a–c at multiple values of *K*, and in Fig. 2); this is probably the consequence of crop breeding, and may underpin the success of crop breeding in exploiting what diversity there is within cultivated germplasm.

A recent study (Agarwal *et al.*, 2012) suggested that the origin of *kabuli* as a distinct entity (i.e. its separation from *desi* forms) dated to no less recently than 0.20–0.22 million yr ago (Mya), with evidence (mis)extrapolated from a molecular clock applied to sequence differences between single *kabuli* and *desi* genotypes. Agarwal's estimate requires the existence of distinct populations of *kabuli* genotypes > 0.2 Mya, well before the generally accepted onset of agriculture, a conjecture unsupported by evidence. Although we do not exclude the possibility that some *kabuli* alleles existed at low frequency in wild *Cicer* populations since well before the origin of agriculture, we see it as far more likely that *kabuli* genotypes arose and gained favor with farmers *c.* 6000 yr ago, potentially as a means of distinguishing crop–wild hybrids or to improve flavor by reducing the astringency of seed.

Breeders have traditionally viewed *kabuli* forms of chickpea as a genetically distinct group, indeed as a homogenous entity. Moreno & Cubero (1978) grouped large-seeded and often light-colored chickpeas as distinct from smaller seeded and dark-colored forms based exclusively on phenotypic data. Although their distinction is not one of *kabuli* vs *desi*, *per se*, there is large overlap among these two market classes and their proposed *megasperma* and *microspermum* forms. Ladizinsky & Adler's (1976) suggestion that *desi* forms represent the early domesticate, with *kabuli* being a subsequently derived type, seems more accurate. Our results provide an alternative model and one that is supported by both genetic and molecular data; namely, that *kabuli* is not a genetic entity at all, but rather a set of independently derived traits, including multiple loss-of-(or reduced)-function alleles of a transcription factor that block anthocyanin and anthocyanidin production in flowers and seeds, respectively. We suggest that the causal alleles for altered flower color and seed coat tannins arose post-domestication, in distinct subsets of the cultivated *desi* germplasm, and were then introgressed by breeders into today's preferred cultivars. This shift in viewpoint has important implications for chickpea breeding, because it dispels the notion that *kabuli* forms are a genetically distinct group and compels a shift from the common practice of focusing germplasm collections and breeding along a *desi*/*kabuli* partition. White-flowered, light-seeded chickpeas are simply mutant forms of domesticated *desi*, with the various alleles embedded within genomes that are more closely related to *desi* forms than they are to one another.

Acknowledgements

This work was supported by the following awards: National Science Foundation DBI-0605251 to D.R.C.; National Science Foundation IOS-0965531 to D.R.C. and R.K.V.; United States Department of Agriculture-Agricultural Research Service, Specific Cooperative Agreement no. 58-5348-1-220 to D.R.C. and R.V.P. We thank David L. Remington and Justen B. Whittall for critical reading of earlier drafts of the manuscript, and George Vandemark for kindly providing CRIL2 and CRIL7 populations.

Author contributions

R.V.P., E.J.B.v.W. and D.R.C. designed the research; R.V.P., N.C-G., E.M.B., L.V., B.C., M.T.K., B.K.S., S.D., A.D.F., J-M.B. and E.J.B.v.W. performed the research; R.V.P., N.C-G., E.M.B., L.V., B.C., M.T.K., B.K.S., S.D., A.D.F., J-M.B., E.J.B.v.W., C.J.C. and R.K.V. contributed to data collection, or analysis, or interpretation; R.V.P., E.J.B.v.W. and D.R.C. wrote the manuscript.

References

Agarwal G, Jhanwar S, Priya P, Singh VK, Saxena MS, Parida SK, Garg R, Tyagi AK, Jain M. 2012. Comparative analysis of *kabuli* chickpea

transcriptome with *desi* and wild chickpea provides a rich resource for development of functional markers. *PLoS One* **7**: e52443.

Appelhagen I, Jahns O, Bartelniewoehner L, Sagasser M, Weisshaar B, Stracke R. 2011. Leucoanthocyanidin dioxygenase in *Arabidopsis thaliana*: characterization of mutant alleles and regulation by MYB-BHLH-TTG1 transcription factor complexes. *Gene* **484**: 61–68.

Aw TL, Swanson BG. 1985. Influence of tannin on *Phaseolus vulgaris* protein digestibility and quality. *Journal of Food Science* **50**: 67–71.

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant Journal* **39**: 366–380.

Broun P. 2005. Transcriptional control of flavonoid biosynthesis: a complex network of conserved regulators involved in multiple aspects of differentiation in *Arabidopsis*. *Current Opinion in Plant Biology* **8**: 272–279.

Bruno WJ, Socci ND, Halpern AL. 2000. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Molecular Biology and Evolution* **17**: 189–197.

Chiu LW, Zhou X, Burke S, Wu X, Prior RL, Li L. 2010. The purple cauliflower arises from activation of a MYB transcription factor. *Plant Physiology* **154**: 1470–1480.

Dixon RA, Liu C, Jun JH. 2013. Metabolic engineering of anthocyanins and condensed tannins in plants. *Current Opinion in Biotechnology* **24**: 329–335.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309–1321.

Earl DA. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Espley RV, Brendolise C, Chagné D, Kutty-Amma S, Green S, Volz R, Putterill J, Schouten HJ, Gardiner SE, Hellens RP *et al.* 2009. Multiple repeats of a promoter segment causes transcription factor autoregulation in red apples. *Plant Cell* **21**: 168–183.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Garg R, Sahoo A, Tyagi AK, Jain M. 2010. Validation of internal control genes for quantitative gene expression studies in chickpea (*Cicer arietinum* L.). *Biochemical and Biophysical Research Communications* **396**: 283–288.

Gould KS. 2004. Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *BioMed Research International* **2004**: 314–320.

Grotewold E. 2006. The genetics and biochemistry of floral pigments. *Annual Review of Plant Biology* **57**: 761–780.

Hajjar R, Hodgkin T. 2007. The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* **156**: 1–13.

Hammer K. 1984. Das Domestikations syndrom. *Kulturpflanze* **32**: 11–34.

Harlan JR. 1992. *Crops and man*. Madison, WI, USA: American Society of Agronomy and Crop Science Society of America.

Hellens RP, Moreau C, Lin-Wang K, Schwinn KE, Thomson SJ, Friers MWE, Frew TJ, Murray SR, Hofer JMI, Jacobs JME *et al.* 2010. Identification of Mendel's white flower character. *PLoS One* **5**: e13230.

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. 2011. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany* **62**: 2465–2483.

Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD *et al.* 2012. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal* **10**: 716–732.

Hubisz MJ, Falush D, Stephens M, Pritchard JK. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**: 1322–1332.

Iruela M, Rubio J, Cubero JI, Gil J, Millán T. 2002. Phylogenetic analysis in the genus *Cicer* and cultivated chickpea using RAPD and ISSR markers. *Theoretical and Applied Genetics* **104**: 643–651.

Javadi F, Wojciechowski MF, Yamaguchi H. 2007. Geographical diversification of the genus *Cicer* (Leguminosae: Papilionoideae) inferred from molecular phylogenetic analyses of chloroplast and nuclear DNA sequences. *Botanical Journal of the Linnean Society* **154**: 175–186.

Kumar J, Vijayalakshmi VS, Rao TS. 2000. Inheritance of flower color in chickpea. *Journal of Heredity* **91**: 416–417.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* **4**: 1073–1081.

Ladizinsky G, Adler A. 1976. The origin of chickpea of *Cicer arietinum* L. *Euphytica* **25**: 211–217.

Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**: 792–794.

Millan T, Winter P, Jüngling R, Gil J, Rubio J, Cho S, Cobos MJ, Iruela M, Rajesh PN, Tekeoglu M *et al.* 2010. A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations. *Euphytica* **175**: 175–189.

Moreau C, Ambrose MJ, Turner L, Hill L, Ellis TN, Hofer JM. 2012. The b gene of pea encodes a defective flavonoid 3′,5′-hydroxylase, and confers pink flower color. *Plant Physiology* **159**: 759–768.

Moreno MT, Cubero JI. 1978. Variation in *Cicer arietinum* L. *Euphytica* **27**: 465–485.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.

Olsen KM, Wendel JF. 2013. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology* **64**: 47–70.

Palmer RG, Perez PT, Ortiz-Perez E, Maalouf F, Suso MJ. 2009. The role of crop–pollinator relationships in breeding for pollinator-friendly legumes: from a breeding perspective. *Euphytica* **170**: 35–52.

Peakall R, Smouse PE. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* **28**: 2537–2539.

Perrier X, Jacquemoud-Collet JP. 2006. *DARwin software*. [WWW document] URL http://darwin.cirad.fr/darwin [accessed 15 December 2012].

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

Pundir RPS, Rao NK, van der Maesen LJG. 1985. Distribution of qualitative traits in the world germplasm of chickpea (*Cicer arietinum* L). *Euphytica* **34**: 697–703.

Redden R, Berger JD. 2007. History and origin of chickpea. In: Yadav SS, Redden R, Chen W, Sharma B, eds. *Chickpea breeding and management*. Wallingford, UK: CABI, 1–13.

Roorkiwal M, von Wettberg EJ, Upadhyaya HD, Warschefsky EJ, Rathore A, Varshney RV. 2014. Exploring germplasm diversity to understand the domestication process in chickpea (*Cicer*) using SNP and DArT markers. *PLoS One* **9**: e102016.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.

Schefe JH, Lehmann KE, Buschmann IR, Unger T, Funke-Kaiser H. 2006. Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. *Journal of Molecular Medicine* **84**: 901–910.

Singh F, Diwakar B. 1995. Chickpea botany and production practices. *Skill Development Series* **16**: 8–9.

Stanton ML, Snow AA, Handel SN. 1986. Floral evolution: attractiveness to pollinators increases male fitness. *Science* **232**: 1625–1627.

Stanton ML, Snow AA, Handel SN, Bereczky J. 1989. The impact of a flower-color polymorphism on mating patterns in experimental populations of wild radish (*Raphanus raphanistrum* L). *Evolution* **43**: 335–346.

Strauss SY, Whittall JB. 2006. Non-pollinator agents of selection on floral traits. In: Harder LD, Barrett SCH, eds. *Ecology and evolution of flowers*. Oxford, UK: Oxford University Press, 120–138.

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. 2006. Caught redhanded: Rc encodes a basic helix–loop–helix protein conditioning red pericarp in rice. *Plant Cell* **18**: 283–294.

Takahashi R, Yamagishi N, Yoshikawa N. 2012. A MYB transcription factor controls flower color in soybean. *Journal of Heredity* **104**: 149–153.

Tanno K, Wilcox G. 2006. The origins of cultivation of *Cicer arietinum* L. and *Vicia faba* L.: early finds from Tell el-Kerkh, north-west Syria, late 10th millennium B.P. *Vegetation History and Archaeobotany* **15**: 197–204.

Tekeoglu M, Santra DK, Kaiser WJ, Muehlbauer FJ. 2000. *Ascochyta* blight resistance inheritance in three chickpea recombinant inbred line populations. *Crop Science* **40**: 1251–1256.

Toker C. 2009. A note on the evolution of kabuli chickpeas as shown by induced mutations in *Cicer reticulatum*. *Genetic Resources and Crop Evolution* **56**: 7–12.

Toledo-Ortiz G, Huq E, Quail PH. 2003. The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **8**: 1749–1770.

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 Biology* **8**: 106–117.

Upadhyaya HD, Ortiz R. 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theoretical and Applied Genetics* **102**: 1292–1298.

Van Oss R, Abbo S, Eshed R, Sherman A, Coyne CJ, Vandemark GJ, Zhang HB, Peleg Z. 2015. Genetic relationships in *Cicer* sp. expose evidence for geneflow between the cultigen and its wild progenitor. *PLoS One* **10**: e0139789.

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B *et al.* 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology* **31**: 240–246.

Warschefsky E, Penmetsa RV, Cook DR, von Wettberg EJ. 2014. Back to the wilds: tapping evolutionary adaptations for resilient crops through systematic hybridization with crop wild relatives. *American Journal of Botany* **101**: 1–10.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* **126**: 485–493.

Wu Y, Li X, Xiang W, Zhu C, Lin Z, Wu Y, Li J, Pandravada S, Ridder DD, Bai G *et al.* 2012. Presence of tannins in sorghum grains is conditioned by different natural alleles of *Tannin1*. *Proceedings of the National Academy of Sciences, USA* **109**: 10281–10286.

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. 2003. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* **299**: 396–399.

Yang K, Jeong N, Moon JK, Lee YH, Lee SH, Kim HM, Hwang CH, Back K, Palmer RG, Jeong SC. 2010. Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *Journal of Heredity* **101**: 757–768.

Yuan L, Wang L, Han Z, Jiang Y, Zhao L, Liu H, Yang L, Luo K. 2012. Molecular cloning and characterization of *PtrLAR3*, a gene encoding leucoanthocyanidin reductase from *Populus trichocarpa*, and its constitutive expression enhances fungal resistance in transgenic plants. *Journal of Experimental Botany* **63**: 2513–2524.

Zachary B, Lippman ZB, Semel Y, Zamir D. 2007. An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Current Opinion in Genetics & Development* **17**: 545–552.

Zhao J, Dixon RA. 2009. MATE transporters facilitate vacuolar uptake of epicatechin 3-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *Plant Cell* **21**: 2323–2340.

Zohary D, Hopf M, Weiss E. 2012. *Domestication of plants in the Old World: the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*, 4th edn. Oxford, UK: Oxford University Press.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Neighbor-joining dendrogram of 322 annual *Cicer* accessions from analysis of 538 SNPs (single nucleotide polymorphisms), recapitulates the taxonomic relationships among the primary, secondary and tertiary gene pools.

Table S1 Summary of annual *Cicer* germplasm used in this study, and basic helix–loop–helix (bHLH) allelic form among genotypes of cultivated chickpea

Table S2 Sequences flanking bi-allelic single nucleotide polymorphisms used for multi-locus genotyping of germplasm

Table S3 Oligonucleotides used in this study, together with the analysis for which they were used

Table S4 Summary results of hierarchical analyses of molecular variance across the wild (*Cicer reticulatum* and *C. echinospermum*) and cultivated (*C. arietinum*) species, and among cultivated groups

Table S5 Genetic diversity estimates in cultivated chickpea and its near wild relatives *Cicer reticulatum* and *C. echinospermum*

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.