Journal of Plant Biochemistry and Biotechnology (April–June 2024) 33(2):178–188
https://doi.org/10.1007/s13562-024-00884-2

ORIGINAL ARTICLE

Identification of superior haplotypes for seed protein content in pigeonpea (*Cajanus cajan* L.)

Harsha Vardhan Rayudu Jamedar[1,2] · Prasad Gandham[3] · Prasad Bajaj[1] · Srinivas Thati[4] · V. Srinivasa Rao[5] · Rajeev K. Varshney[1,6] · Rachit K. Saxena[1,7]

Received: 20 May 2023 / Accepted: 18 March 2024 / Published online: 18 April 2024

Abstract

Pigeonpea (*Cajanus cajan* L.) is an important source of quality dietary protein for over a billion people worldwide. The seeds of pigeonpea contain approximately 20–22% digestible protein, which makes it a valuable source of nutrition. Despite this, there has been little attention paid to enhancing the seed protein content (SPC) through genetic means. Recently, high-protein germplasm lines have been discovered in the secondary gene pool, which presents an opportunity to breed for high-protein cultivars. To accelerate the breeding process, genomics-assisted breeding (GAB) can be utilized. In this context, this study identified the superior haplotypes for the genes that control SPC in pigeonpea. Whole-genome re-sequencing (WGRS) data from 344 pigeonpea genotypes were analyzed to identify the superior haplotypes for 57 SPC governing genes. A total of 231 haplotypes in 43 candidate genes were identified, and haplo-pheno analysis was performed to provide superior haplotypes for 10 genes. The identification of superior haplotypes and genotypes will greatly facilitate the development of protein-rich pigeonpea seeds through the application of haplotype-based breeding (HBB).

Keywords Seed protein content (SPC) · Genes · Haplotypes · Superior haplotypes

Abbreviations

Mbp	Millions of base pairs
kg	Kilogram
ha	Hectares
SPC	Seed protein content
WGRS	Whole genome re-sequencing
GAB	Genomics-assisted breeding
CAPS	Cleaved amplified polymorphic sequences
MAS	Marker-assisted selection
MABC	Marker-assisted backcrossing
MARS	Marker-assisted recurrent selection
HBB	Haplotype-based breeding
SNP	Single nucleotide polymorphism
CcLGs	*Cajanus cajan* Linkage Groups
maf	Minor allelic frequency
ANOVA	Analysis of variance
QTL	Quantitative trait loci

Introduction

Grain legumes are among the most significant sources of dietary protein, contributing towards balanced nutrition to the human population. Among the legumes, pigeonpea (*Cajanus*

✉ Rajeev K. Varshney
rajeev.varshney@murdoch.edu.au

✉ Rachit K. Saxena
rachit.saxena@gbu.edu.in

1 International Crops Research Institute for the Semi-Arid Tropics, Patancheru, Telangana, India

2 Department of Genetics and Plant Breeding, Agricultural College, Acharya N.G. Ranga Agricultural University, Bapatla, Andhra Pradesh, India

3 School of Plant, Environmental and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA, USA

4 Regional Agricultural Research Station, Acharya N.G. Ranga Agricultural University, Maruteru, Andhra Pradesh, India

5 Department of Statistics and Computer Applications, Agricultural College, Acharya N.G. Ranga Agricultural University, Bapatla, Andhra Pradesh, India

6 Centre for Crop & Food Innovation, Food Futures Institute, WA State Agricultural Biotechnology Centre, Murdoch University, Murdoch, WA, Australia

7 Gujarat Biotechnology University, Gandhinagar, Gujarat, India

cajan L.) is a highly valued protein-rich crop having 2n = 22 chromosomes with a genome size of 833.07 Mbp (Varshney et al. 2012). It is grown on 6.36 million ha in 24 countries around the globe (FAO 2023) and not only offers nutritional food security, but its adaptability to diverse agro-ecological conditions and nitrogen-fixing ability, also makes it a crucial component of sustainable agricultural systems (Mula and Saxena 2010; Rao et al. 2010). Current pigeonpea cultivars, yielding around 700 kg/ha, are insufficient to address malnutrition effectively. To improve pigeonpea-based protein's impact on malnutrition, it is logical to increase protein harvests from existing land resources. This could be achieved by developing high-protein cultivars without sacrificing productivity (Saxena et al. 2023).

Historically, pigeonpea breeding objectives have predominantly focused on enhancing yield and crop adaptability (Mligo and Craufurd 2005; Odeny 2007; Upadhyaya et al. 2007b), with minimal emphasis on the nutritional quality. Despite this, research has indicated that sufficient genetic variability for seed protein content (SPC) exists within the cultivated gene pool for potential trait improvement (Upadhyaya et al. 2007a). The absence of selection for SPC in the past may have contributed to the loss of beneficial alleles during breeding. A further possibility is that unfavourable alleles became fixed in the population due to drift and/or hitchhiking effects during the process of selection. Utilizing the allelic diversity of genetic resources is considered essential for overcoming obstacles in the modern world due to the significance of the trait (Mayer et al. 2020). The availability of genomic resources in pigeonpea, such as a reference genome (Varshney et al. 2012; Garg et al. 2022) and whole genome re-sequencing (WGRS) data (Kumar et al. 2016; Varshney et al. 2017; Saxena et al. 2021), provide an opportunity to improve productivity and quality traits in the crop via modern breeding techniques. Nonetheless, the initial step in Genomics-assisted breeding (GAB) is the identification of molecular markers or candidate genes associated with the traits of interest (Varshney et al. 2005, 2021a). Breeding efforts for developing high seed protein pigeonpea cultivars were aided by identification of markers associated with seed protein content (Obala et al. 2019a). Genome sequencing and phenotyping data revealed sequence-based markers and candidate genes for seed protein. Screening of 16 polymorphic cleaved amplified polymorphic sequences (CAPS) markers on a F_2 population segregating for SPC identified four markers co-segregating with SPC.

GAB methods including marker-assisted selection (MAS), marker-assisted backcrossing (MABC), and marker-assisted recurrent selection (MARS), have been proposed and used to transfer or assemble superior alleles into elite genetic background(s) (Varshney et al. 2021a). Recently, haplotype-based breeding (HBB) has been proposed to develop superior cultivars in a number of crop species (Varshney et al. 2020). Attempts have been made to introduce HBB in cereal crops, including rice, for salt tolerance (Mishra et al. 2016), deep water adaptation (Kuroha et al. 2018), and improvement of 21 yield and quality traits (Abbai et al. 2019). In maize, Mayer et al. (2020) identified superior haplotypes for complex traits. In legumes, Guan et al. (2014) have identified haplotypes for salinity tolerance and Bhat et al. (2022) have identified haplotypes for plant height in soybean. Sinha et al. (2020) have identified superior haplotypes for drought tolerance in pigeonpea across 10 associated genes, and Varshney et al. (2021b) have identified superior haplotypes for 16 important traits in chickpea.

In view of the above considerations, to understand the molecular mechanism of SPC in pigeonpea, WGRS data with information on gene functions and a common variant filtering strategy were used to identify 57 candidate genes (Obala et al. 2019a). In the present study, these candidate genes have been used to examine haplotype variation across WGRS data on 344 pigeonpea genotypes. Haplotype data were combined with the SPC phenotyping data to identify superior haplotypes in 10 important genes. Furthermore, we have also identified pigeonpea genotypes carrying these superior haplotypes for SPC. Utilizing HBB, which capitalizes on both genotypes and superior haplotypes identified in this study will provide opportunities to develop pigeonpea seeds with a greater protein content.

Materials and methods

Plant material

A set of 344 pigeonpea genotypes including 179 breeding lines and 165 landraces was used to conduct haplotype analysis (ESM Table 1). These genotypes were selected based on the availability of sequence data. Further, a subset of 281 genotypes (121 breeding lines and 160 landraces) was selected for conducting haplo-pheno analysis. These 281 genotypes (ESM Table 2) and SPC phenotyping data were collected from International Crops Research Institute for the Semi-Arid Tropics genebank (ICRISAT) (http://genebank.icrisat.org/IND/Char_Pigeonpea?Crop=Pigeonpea).

Whole genome re-sequencing data and haplotype analysis

Whole genome re-sequencing data of 344 pigeonpea genotypes were collected from our previous studies including, Varshney et al. (2017) and Saxena et al. (2021). Sequencing data from the 344 pigeonpea genotypes was aligned to the improved pigeonpea reference genome (Cajca. Asha_v2.0) (Garg et al. 2022). Single nucleotide polymorphism calling from the deduplicated binary alignment map

files was performed using BCFtools version 1.9 of SAMtools (Danecek et al. 2021), thus creating the base single nucleotide polymorphism (SNP) set. Quality control of raw sequencing data was done using FastQC. The cleaned reads were aligned to the Cajca.Asha_v2.0 using Burrows-Wheeler Aligner. The aligned reads in the binary alignment map format were converted to the variant call format using SAMtools and BCFtools. Variant call format file was filtered to remove low-quality SNPs. The criteria for filtering were to have base quality of 20, mapping quality of 20, bi-allelic only, minimum depth of three, site minimum count of 80%, and 1% minor allelic frequency. Furthermore, to perform haplotype analysis, full length sequence data of 57 candidate genes governing total seed protein content in pigeonpea were used (Obala et al. 2019a). From the filtered SNPs, haplotypes present in candidate gene regions were identified using Haploview software (Barrett et al. 2005).

Haplo-pheno analysis

Using SPC phenotyping data on pigeonpea genotypes and haplotype information from the candidate genes, haplo-pheno analysis was performed. Genes with at least two haplotypes and each haplotype present in minimum three genotypes were considered for haplo-pheno analysis. The genotypes were thus categorized into haplogroups and the significance among the haplogroups was studied with the help of analysis of variance (ANOVA) and Duncan's test at 95% confidence level. The analysis was carried out using Agricolae package in R software (Team 2009).

Results

Haplotypes for seed protein content

The analysis of whole genome sequence data of 344 pigeonpea genotypes (Varshney et al. 2017; Saxena et al. 2021) provided 2,190,221 single nucleotide polymorphisms (SNPs) distributed across 11 *Cajanus cajan* Linkage Groups (CcLGs) (Fig. 1). Further SNPs were filtered for bi-allelic, minimum depth, 80% site count and 1% minor allelic frequency (maf) and provided a set of high quality 1,148,409 SNPs. Maximum number of SNPs were present on CcLG11 (231,568), whereas CcLG05 contained minimum SNPs (22,546). SNPs identified across 344 genotypes with 57 candidate genes for SPC were used to identify haplotypes by adopting the criteria of no missing bases, homozygous alleles and a minimum length of 3 bp. It resulted in the identification of 231 haplotypes across the coding regions

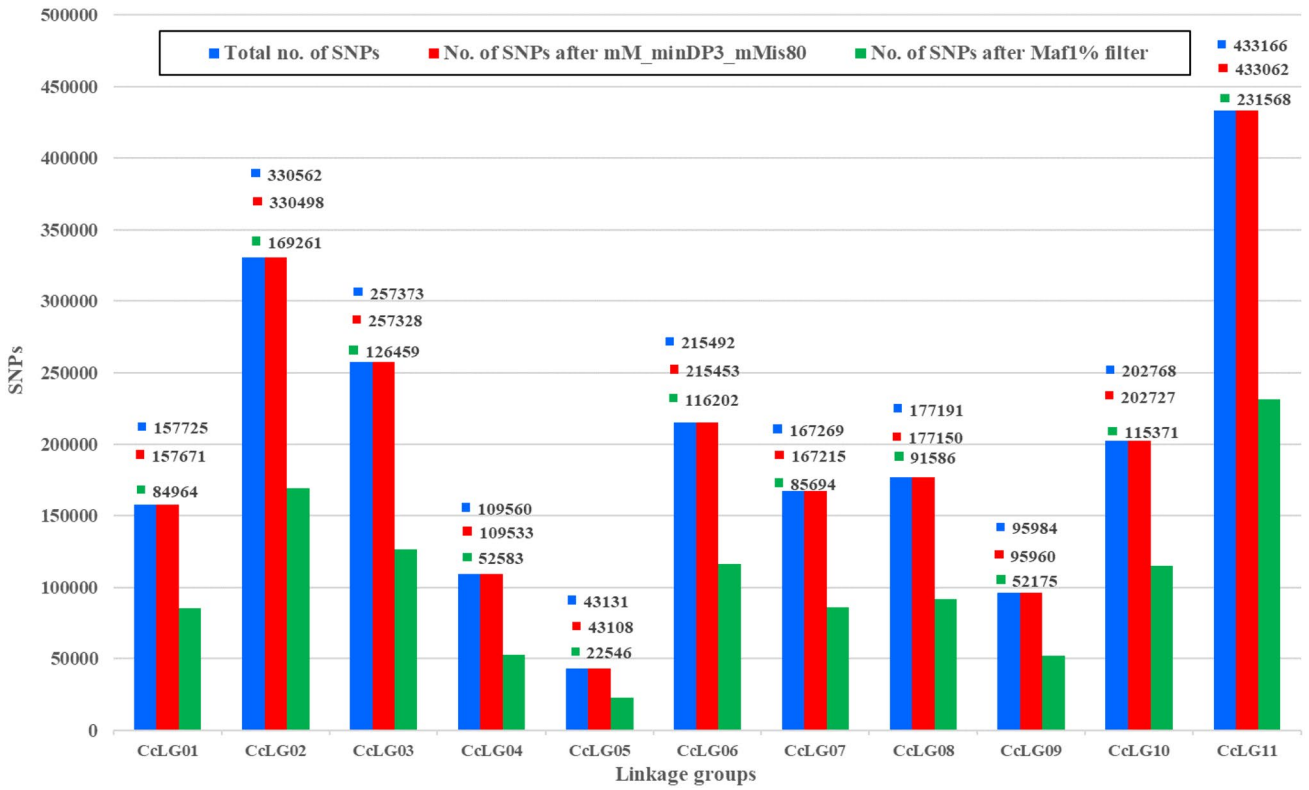

Fig. 1 Single Nucleotide Polymorphism (SNP) spectrum in re-sequencing data of 344 pigeonpea genotypes

(missense, silent and non-sense) of 43 genes distributed in all the linkage groups, except CcLG05 (Fig. 2). Maximum number of haplotypes (53) for nine candidate genes were located on CcLG01, followed by CcLG08 (40 haplotypes for eight candidate genes) and CcLG02 (40 haplotypes for seven candidate genes), while minimum number of haplotypes (2) for single candidate gene were located on CcLG10. The number of haplotypes (ESM Table 3) ranged from 2 (*C.cajan_14054*, *C.cajan_18213* and *C.cajan_18280*) to 10 (*C.cajan_07942* and *C.cajan_10048*) for different candidate genes. Varying haplotype frequencies ranging from 0.29–84.01% were identified in the present study (Fig. 3). The haplotypes with highest and lowest frequency were considered as 'major' and 'minor' haplotypes, respectively. The haplotype H1 for *C.cajan_19670* gene had recorded maximum frequency (84.01%) with 289 genotypes and hence was identified as the major haplotype, while 54 haplotypes for 23 genes had recorded minimum frequency (0.29%) and hence were identified as minor haplotypes. The minor haplotypes were observed to be mono-genotypic in constitution.

Haplotypes diversity in breeding lines and landraces

A total of 168 and 207 haplotypes were identified for the 43 candidate genes governing seed protein across 179 breeding lines and 165 landraces, respectively (ESM Table 3). From these 144 haplotypes were common in both breeding lines and landraces. While 24 and 63 unique haplotypes were found specific to breeding lines and landraces, respectively. For example, haplotype H5 for *C.cajan_19199* was present only in landraces, while H3 for *C.cajan_20717* was present only in breeding lines. The number of haplotypes for a specific gene in breeding lines ranged from one (*C.cajan_06087 and C.cajan_15499*) to eight (*C. cajan_20841*). In case of landraces, it ranged from two (*C. cajan_14054, C.cajan_18213 and C.cajan_18280*) to nine (*C.cajan_20841*). It is thus evident that for the majority of genes, maximum number of haplotypes were present in landraces in comparison to breeding lines. Few haplotypes of a gene were observed to be present in different landraces and breeding lines, while some haplotypes were present only in landraces or breeding lines and are considered as unique haplotypes. Interestingly, out of 43 genes, haplotypes of only 26 genes in breeding lines showed complete match with landraces, while 17 genes showed unique haplotypes in breeding lines which are not present in any of the landraces. The frequency of the unique haplotypes of breeding lines ranged from 11.11% (*C.cajan_04622* and *C.cajan_20842*) to 33.33% (*C.cajan_18443* and *C.cajan_20717*).

Phenotyping data for seed protein content

A subset of 281 genotypes including 121 breeding lines and 160 landraces from the total 344 genotypes was selected on the basis of available phenotyping data and representation of all 231 haplotypes for 43 genes. The SPC in 281 genotypes ranged from 16.30% to 28% (breeding lines: 16.30–27.10%; landraces: 16.30–28%) with an average value of 21.22% (breeding lines: 20.87%; landraces: 21.48%). This indicated that a significant phenotypic variation was present for SPC in the 281 genotypes studied (Fig. 4).

Superior haplotypes for seed protein content

Haplo-pheno analysis with haplotype and phenotype data on 281 genotypes provided 10 significant genes. In order

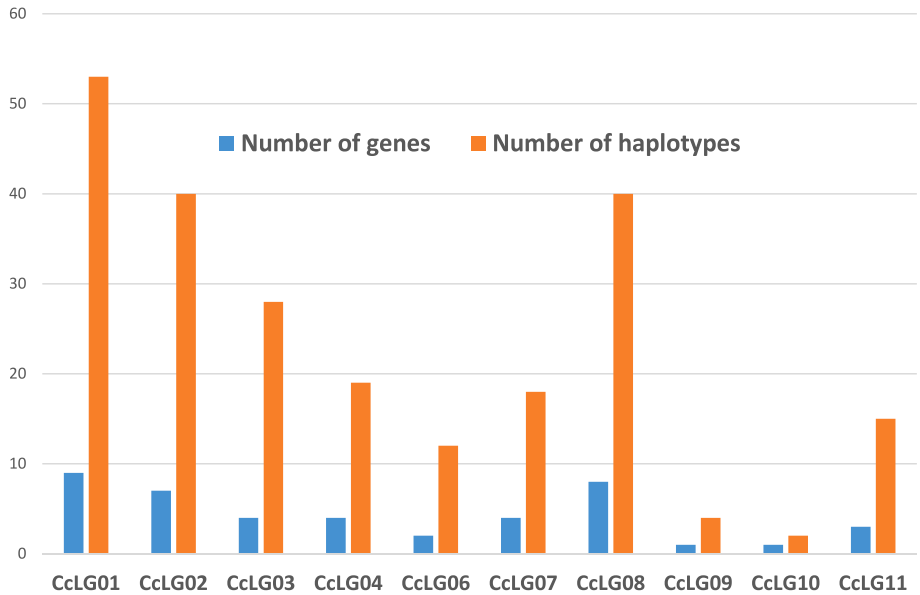

Fig. 2 Number of genes and haplotypes identified in different linkage groups of pigeonpea

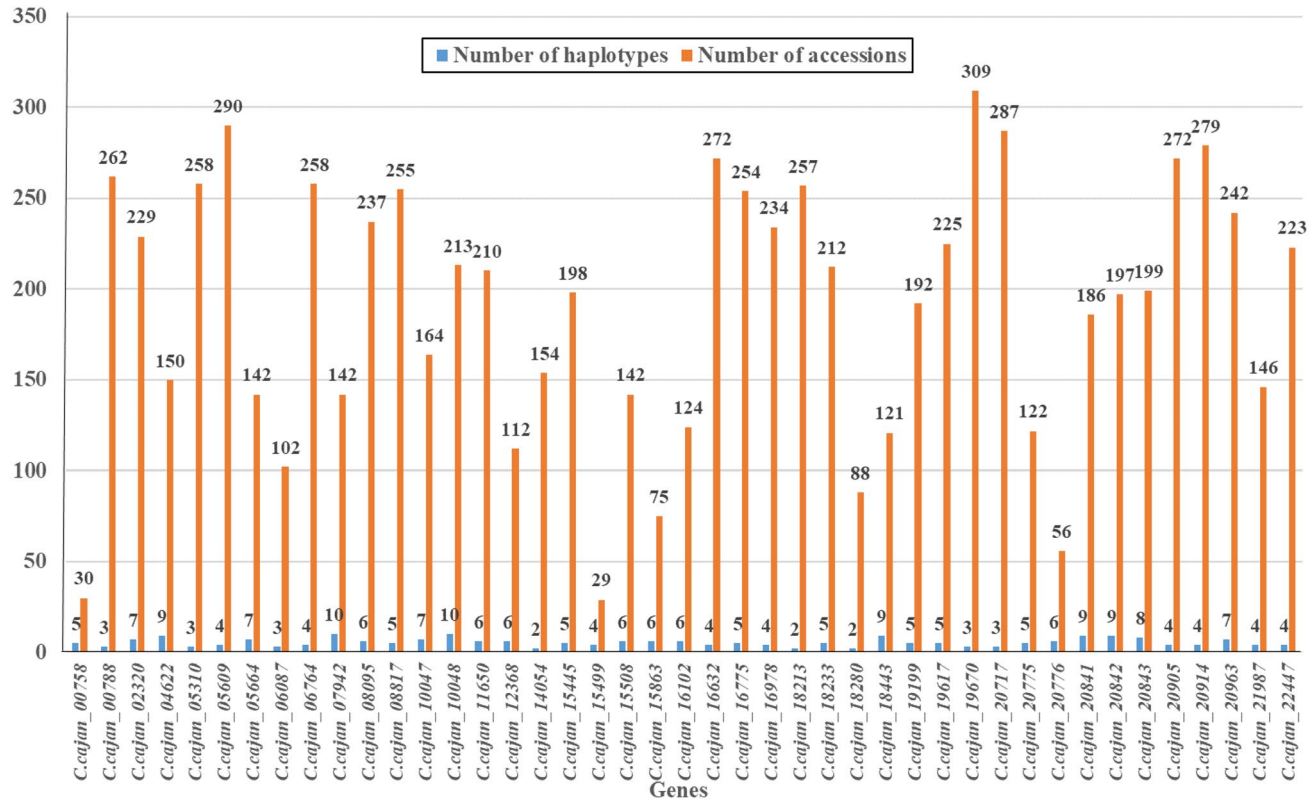

Fig. 3 Number of haplotypes and accessions / genotypes for genes governing seed protein content in pigeonpea

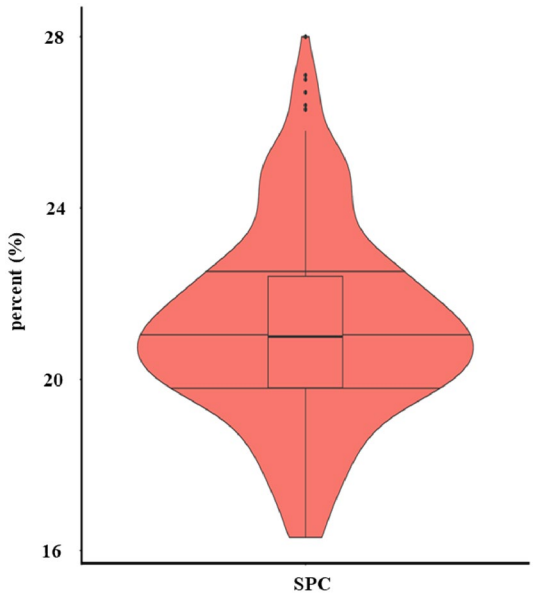

Fig. 4 Phenotypic distribution of seed protein content in 281 pigeonpea genotypes. The violin plots show the phenotypic distribution of the 281 genotypes for seed protein content. The shape of the distribution (skinny on each end and wide in the middle) indicates that the trait distribution is highly concentrated around the median

to define a superior haplotype in a given significant gene for SPC, performance of a haplogroup (group of genotypes with the same haplotype sequence) was compared to other haplogroups (Table 1). A haplotype with phenotypic performance superiority for SPC over other haplotypes has been considered as a superior haplotype in a gene. As a result, 12 superior haplotypes were identified in 10 genes regulating SPC in pigeonpea. The maximum number of superior haplotypes were present in two genes (2 haplotypes in each *C.cajan_04622* and *C.cajan_20776* and the remaining 8 genes contained one superior haplotype in each). For the 12 superior haplotypes identified, the SPC values ranged from 16.3% to 28% with an average SPC 22.33%. A total of seven genotypes, including four landraces and three breeding lines were carrying superior haplotypes for 10 SPC governing genes (*C.cajan_00758, C.cajan_02320*, *C.cajan_04622*, *C.cajan_05310*, *C.cajan_14054*, *C.cajan_15445*, *C.cajan_18280*, *C.cajan_20776*, *C.cajan_20841* and *C.cajan_20914* (Fig. 5)). In terms of individual genotypes, ICP4213 (27.1% SPC) and ICP11230 (27% SPC) were found to be superior genotypes carrying maximum number (i.e., 7 in each) of superior haplotypes. Genotypes ICP11320 and ICP7257 harboured two superior haplotypes each and the genotypes ICP6859, ICP7420 and ICP11690 carried single superior haplotype each (Table 2). Above mentioned

Table 1 Average performance of genotypes possessing superior haplotype in comparison to another group of haplotypes

Gene	Superior haplotype						Other haplotypes		
	Haplotype	Mean	Range	Promising accession	SPC (%)	Biological status	Haplotype	Mean	Range
C.cajan_00758	H1	22.52[a]	19.40–27.10	ICP7257	27.10	Breeding line	H2	17.87[b]	16.50–20.50
C.cajan_02320	H2	22.23[a]	16.7–27.1	ICP4213	27.10	Landrace	H3	20.95[b]	16.5–23.7
				ICP11230	27.00	Breeding line	H1	20.51[b]	16.3–28
C.cajan_04622	H2	22.35[a]	16.7–27.1	ICP4213	27.10	Landrace	H1	21.13[ab]	16.5–28
				ICP11230	27.00	Breeding line	H4	20.93[ab]	18.7–22.8
	H5	22.17[a]	20.3–25.1	ICP11690	25.1	Breeding line	H3	19.22[b]	16.6–22.2
C.cajan_05310	H1	21.74[a]	16.3–27.1	ICP4213	27.10	Landrace	H3	21.17[ab]	17.3–27.1
				ICP11230	27.00	Breeding line	H2	20.79[b]	16.4–26.7
C.cajan_14054	H1	21.73[a]	16.5–27.1	ICP4213	27.10	Landrace	H2	18.87[b]	17.7–19.5
				ICP11230	27.00	Breeding line			
C.cajan_15445	H2	22.17[a]	16.5–27.1	ICP4213	27.10	Landrace	H1	20.82[ab]	16.4–27.1
				ICP11230	27.00	Breeding line	H4	20.67[ab]	17.2–24.5
							H3	20.25[b]	16.6–24.4
C.cajan_18280	H1	22.73[a]	17.9–27.1	ICP4213	27.10	Landrace	H2	20.3[b]	16.5–26.3
				ICP11230	27.00	Breeding line			
C.cajan_20776	H1	22.43[a]	18.2–25.8	ICP11320	25.80	Landrace	H2	20.13[b]	16.5–25.1
	H3	21.55[a]	19.7–24.1	ICP6859	24.1	Landrace			
C.cajan_20841	H9	24.7[a]	22.9–25.8	ICP11320	25.80	Landrace	H6	22.51[b]	19.7–26.4
							H3	22.07[bc]	18.4–27.0
							H5	21.91[bc]	19.0–24.8
							H4	21.03[bc]	18.6–26.3
							H1	20.91[bc]	16.3–28.0
							H7	20.82[bc]	17.5–25.0
							H2	20.64[bc]	16.4–27.1
							H8	20.01[c]	18.0–22.9
C.cajan_20914	H1	21.32[a]	16.3–28.0	ICP7420	28.00	Landrace	H2	19.57[b]	16.6–25.0
				ICP4213	27.10	Landrace			
				ICP7257	27.10	Breeding line			
				ICP11230	27.00	Breeding line			

Duncan analysis was employed to test statistical significance ($p < 0.05$). Different alphabets (a, b, c) indicate significant differences

superior haplotypes and pigeonpea genotypes will provide new opportunities for developing high SPC pigeonpea cultivars through haplotype-based breeding (Fig. 6).

Discussion

Seed protein content (SPC) is an important nutritional trait of pigeonpea (Saxena et al. 2023). SPC is a quantitatively inherited trait and G×E interactions are further adding to this complexity (Obala et al. 2018, 2019b, 2020). Due to this trait complexity, though pigeonpea germplasm has good genetic variations available (up to 32% SPC), but could not be used efficiently to develop high yielding pigeonpea genotypes with higher SPC (Obala et al. 2019a). Therefore, few studies were undertaken recently to identify the genetic loci/QTLs for SPC in pigeonpea (Obala et al. 2019a, 2020). The present study based on haplotype approach is an advancement to our findings with respect to genetic control of SPC in pigeonpea.

Haplotype analysis is a useful tool for identifying and utilizing existing genetic diversity among target genes. In cereal crops such as rice and maize, superior haplotypes have been linked to complex traits and environmental adaptations identified (Mishra et al. 2016; Abbai et al. 2019; Mayer et al. 2020). In legumes, noteworthy progress has been made in discovering superior haplotypes for essential traits. For example, Varshney et al. (2021b) identified superior haplotypes for 16 key traits in chickpea, while in soybean, haplotypes have been detected for salinity tolerance and plant height (Guan et al. 2014; Bhat et al. 2022). In the case of pigeonpea superior haplotypes were identified for drought tolerance and related traits (Sinha et al. 2020). In the present study, we have identified superior haplotypes

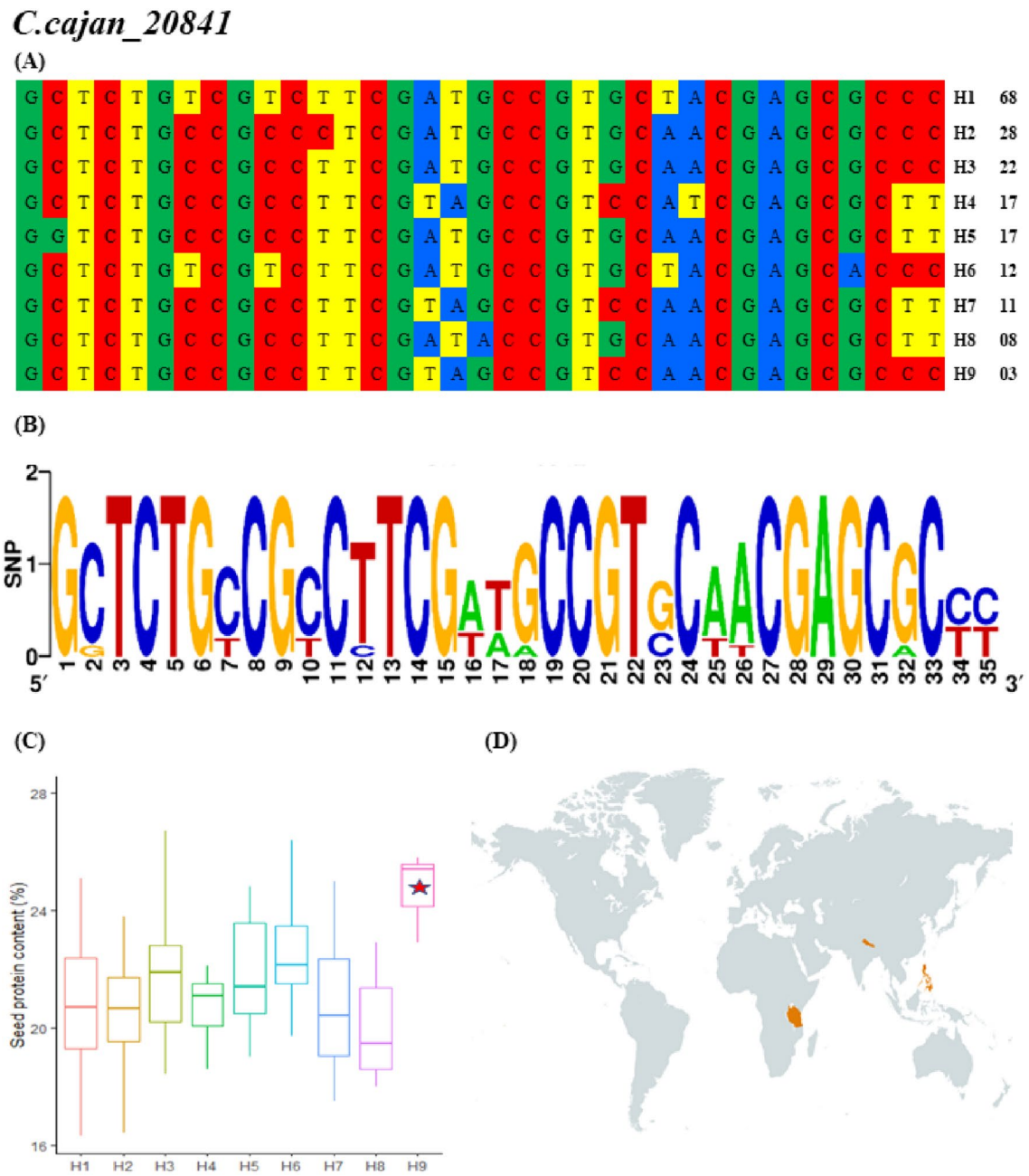

Fig. 5 Haplotype analysis of *C.cajan_20841* across the subset panel. **a** Haplotypic variation of *C.cajan_20841*, a gene associated with SPC. **b** SNP motif representation of *C.cajan_20841* haplotypes. (Height of each stack of letters represents the relative frequency of that nucleotide at a particular position). **c** Boxplot showing variation in SPC among 281 pigeonpea genotypes. Lower and upper boxes indicate the 25th and 75th percentile, respectively. The median is depicted by the horizontal line in the box. Duncan's analysis suggested H9 is the most superior haplotype of *C.cajan_20841* gene for SPC. **d** The geographical distribution of accessions harbouring H9 of *C.cajan_20841* gene

associated with seed protein content (SPC) by analyzing whole genome sequence data of 344 pigeonpea genotypes. The study identified 231 haplotypes across 43 genes, indicating considerable haplotype diversity in the pigeonpea genome for SPC.

Our study found that 100% of haplotype diversity for nine genes in landraces has already been utilized in breeding lines. While 50% to 80% of haplotype diversity harnessed from 25 genes and < 50% of haplotype diversity utilized from the remaining nine genes for SPC in breeding lines. Whereas, in an earlier study in pigeonpea for drought

Table 2 A summary of superior haplotypes for SPC candidate genes across superior pigeonpea genotypes

	SPC (%)	*C.cajan_00758*	*C.cajan_02320*	*C. cajan_04622*		*C.cajan_05310*	*C.cajan_14054*	*C.cajan_15445*	*C.cajan_18280*	*C. cajan_20776*		*C.cajan_20841*	*C.cajan_20914*
		H1	H2	H2	H5	H1	H1	H2	H1	H1	H3	H9	H1
Landraces													
ICP11320	25.8	–	–	–	–	–	–	–	–	+	–	+	–
ICP4213	27.1	–	+	+		+	+	+	+	–	–	–	+
ICP6859	24.1	–	–	–	–	–	–	–	–	–	+	–	–
ICP7420	28	–	–	–	–	–	–	–	–	–	–	–	+
Breeding lines													
ICP11230	27	–	+	+	–	+	+	+	+	–	–	–	+
ICP11690	25.1	–	–	–	+	–	–	–	–	–	–	–	–
ICP7257	27.1	+	–	–	–	–	–	–	–	–	–	–	+

+ presence of haplotype; – absence of haplotype

tolerance responsive genes have shown limited use of haplotype diversity from their ancestors in breeding lines ranging from < 20% in eight genes to > 80% in only two genes (Sinha et al. 2020). Despite extensive utilization of haplotype diversity in past and current pigeonpea breeding programs for candidate genes, there has been no notable enhancement in SPC. Most superior haplotypes were identified in landraces compared to breeding lines, suggesting their contribution to improving breeding lines with higher SPC. Our phenotyping and candidate gene-based statistical analysis identified ten strongly associated SPC genes in pigeonpea. We eventually identified twelve superior haplotypes for ten genes, and genotypes carrying these superior haplotypes were analyzed. Interestingly, eleven of the twelve superior haplotypes were transferred to breeding lines from landraces, while *C.cajan_20841* (H9) is a unique haplotype found only in landraces.

The functional annotation of identified SPC associated genes has also provided support to our findings (Detailed table presented in Obala et al. 2019a, b). For instance, *C.cajan_05310*, has been found significant in seed storage albumin protein processing, linked to increased SPC in soybean, with studies indicating its upregulation in high SPC genotypes (Gruis et al. 2002; Krishnan et al. 2007; Bolon et al. 2010). *C.cajan_04622*, essential in the first step of nitrogen assimilation, is a potential candidate for controlling grain protein content, as demonstrated in wheat (Gaur et al. 2012; Nigro et al. 2013; Guan et al. 2014). The MYB gene *C.cajan_02320* induces expression of proteinases, impacting protein content through its interaction with gibberellic acid-responsive elements (Gubler and Jacobsen 1992; Gubler et al. 1999). Genes *C.cajan_00758*, *C.cajan_14054*, and *C.cajan_18280* are involved in the processing and degradation of storage proteins in seeds, key for seed maturation and germination (Hiraiwa et al. 1997; Asakura et al. 2000; Pereira et al. 2008; Mazorra-Manzano et al. 2010). Differential expression of *C.cajan_15445* in soybean suggests its influence on SPC (Bolon et al. 2010). *C.cajan_20776*, located in a region associated with a major SPC QTL in soybean, and *C.cajan_20914*, encoding a phosphoglycerate kinase, are linked to traits like seed weight, indicating a possible dual influence on seed weight and protein content (Burstin et al. 2007; Lestari et al. 2013). *C.cajan_20841* known to regulate leaf senescence and ROS production, could be significant for understanding seed protein content in pigeonpea (Chardon et al. 2014; Li et al. 2016). Furthermore, out of 231 haplotypes identified for 43 genes, 87 were unique haplotypes found in landraces and breeding lines. The presence of unique haplotypes in breeding lines that are not found in any of the landraces suggests that these haplotypes might have originated through mutation or recombination during breeding. This scenario implies that additional

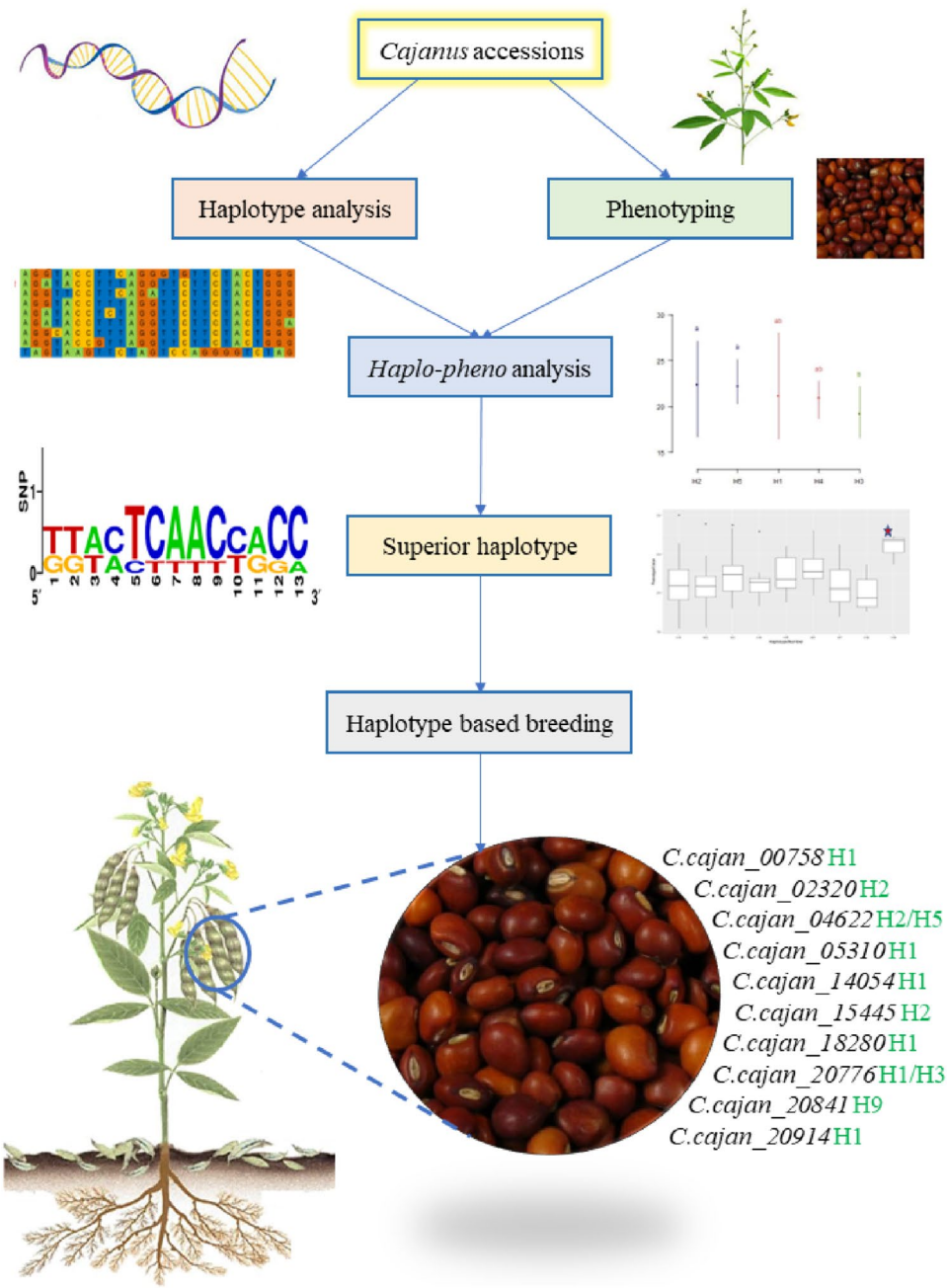

Fig. 6 Schematic representation of haplotype and haplo-pheno analysis towards developing tailored pigeonpea with superior haplotypes for seed protein content. Through haplotype-based breeding, new breeding lines can be developed with the most superior haplotype combination

efforts are required in forthcoming years to utilize available haplotype diversity to enhance SPC in pigeonpea.

To enhance the SPC of popular pigeonpea varieties or to develop better SPC varieties, a haplotype-based breeding strategy can be employed (Varshney et al. 2020). This approach can be useful in pigeonpea hybrid breeding, where parents can be selected based on superior and diverse haplotypes to develop the next generation of superior haplotypes. For instance, in a recent study by Sinha et al. (2020), an accession ICP7420 carrying a superior haplotype for *C.cajan_20914* and having a high seed protein content of 28% was also reported to carry a superior haplotype for relative water content, a drought-related trait. This accession can be used as a parent in breeding programs to further improve the SPC and drought tolerance in pigeonpea. Further investigation is required to gain a deeper understanding of the interaction between different haplotypes of SPC genes in pigeonpea. Future research should focus on validating the identified superior haplotypes in larger and more diverse pigeonpea germplasm collections to ensure their robustness across different genetic backgrounds. Additionally, studies aimed at elucidating the molecular mechanisms governing SPC in pigeonpea would further enhance our understanding of the genetic basis of this important trait and facilitate the

development of more efficient breeding strategies. Moreover, integrating the superior haplotypes identified in this study into genomic prediction models could facilitate the rapid selection of high SPC genotypes in breeding programs. Genomic prediction has emerged as a powerful approach for improving complex traits, such as SPC, which are controlled by multiple genes with small to moderate effects. By combining genomic prediction with the superior haplotypes identified in this study, breeders can accelerate the development of pigeonpea cultivars with improved SPC.

Conclusions

Haplotype analysis has been applied in various crops, resulting in the identification of superior haplotypes associated with target traits. A haplotype analysis of seed protein content in pigeonpea revealed a rich diversity of candidate genes and significant variations among different genotypes. The haplotype-specific responses to SPC found in this study highlight the inadequacy of utilizing haplotype diversity in past pigeonpea breeding programs and the need for increased efforts to utilize existing haplotype diversity to improve SPC in pigeonpea. In the future, it is expected that further functional evaluation, including the discovery of epistatic interactions between these haplotypes, and the implementation of haplotype-based breeding, will lead to the development of pigeonpea varieties with high SPC.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13562-024-00884-2.

Acknowledgements RKV performed this study as a part of his International Agriculture for Food & Nutrition Security portfolio at Murdoch University and thanks Food Future Institute for supporting this study through a start-up grant. RKS acknowledges the research support from Department of Science and Technology (DST), Government of Gujarat and Gujarat Biotechnology University (GBU). HVRJ thank the funding support from Council of Scientific and Industrial Research (CSIR), Government of India, for the award of a research fellowship.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Abbai R, Singh VK, Nachimuthu VV et al (2019) Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. Plant Biotechnol J 17:1612–1622. https://doi.org/10.1111/pbi.13087

Asakura T, Matsumoto I, Funaki J et al (2000) The plant aspartic proteinase-specific polypeptide insert is not directly related to the activity of oryzasin 1. Eur J Biochem 267:5115–5122. https://doi.org/10.1046/j.1432-1327.2000.01582.x

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. https://doi.org/10.1093/bioinformatics/bth457

Bhat JA, Karikari B, Adeboye KA et al (2022) Identification of superior haplotypes in a diverse natural population for breeding desirable plant height in soybean. Theor Appl Genet 135:2407–2422. https://doi.org/10.1007/s00122-022-04120-0

Bolon Y-T, Joseph B, Cannon SB et al (2010) Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biol 10:41. https://doi.org/10.1186/1471-2229-10-41

Burstin J, Marget P, Huart M et al (2007) Developmental genes have pleiotropic effects on plant morphology and source capacity, eventually impacting on seed protein content and productivity in pea. Plant Physiol 144:768–781. https://doi.org/10.1104/pp.107.096966

Chardon F, Jasinski S, Durandet M et al (2014) QTL meta-analysis in Arabidopsis reveals an interaction between leaf senescence and resource allocation to seeds. J Exp Bot 65:3949–3962. https://doi.org/10.1093/jxb/eru125

Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. GigaScience 10:giab008. https://doi.org/10.1093/gigascience/giab008

FAOSTAT. https://www.fao.org/faostat/en/#data/QCL. Accessed 27 Apr 2023

Garg V, Dudchenko O, Wang J et al (2022) Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement. J Adv Res 42:315–329. https://doi.org/10.1016/j.jare.2021.10.009

Gaur VS, Singh US, Gupta AK, Kumar A (2012) Influence of different nitrogen inputs on the members of ammonium transporter and glutamine synthetase genes in two rice genotypes having differential responsiveness to nitrogen. Mol Biol Rep 39:8035–8044. https://doi.org/10.1007/s11033-012-1650-8

Gruis DF, SelingerCurranJung DAJMR (2002) Redundant proteolytic mechanisms process seed storage proteins in the absence of seed-type members of the vacuolar processing enzyme family of cysteine proteases. Plant Cell 14:2863–2882. https://doi.org/10.1105/tpc.005009

Guan R, Qu Y, Guo Y et al (2014) Salinity tolerance in soybean is modulated by natural variation in GmSALT3. Plant J 80:937–950. https://doi.org/10.1111/tpj.12695

Gubler F, Jacobsen JV (1992) Gibberellin-responsive elements in the promoter of a barley high-pI alpha-amylase gene. Plant Cell 4:1435–1441. https://doi.org/10.1105/tpc.4.11.1435

Gubler F, Raventos D, Keys M et al (1999) Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant J 17:1–9. https://doi.org/10.1046/j.1365-313X.1999.00346.x

Hiraiwa N, Kondo M, Nishimura M, Hara-Nishimura I (1997) An aspartic endopeptidase is involved in the breakdown of propeptides of storage proteins in protein-storage vacuoles of plants. Eur J Biochem 246:133–141. https://doi.org/10.1111/j.1432-1033.1997.00133.x

Krishnan HB, Natarajan SS, Mahmoud AA, Nelson RL (2007) Identification of glycinin and β-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. J Agric Food Chem 55:1839–1845. https://doi.org/10.1021/jf062497n

Kumar V, Khan AW, Saxena RK, Garg V, Varshney RK (2016) First generation hapmap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnol J 14:1673–1681. https://doi.org/10.1111/pbi.12528

Kuroha T, Nagai K, Gamuyao R et al (2018) Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. Science 361:181–186. https://doi.org/10.1126/science.aat1577

Lestari P, Van K, Lee JE et al (2013) Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. Front Plant Sci 4:176

Li L, Xing Y, Chang D et al (2016) CaM/BAG5/Hsc70 signaling complex dynamically regulates leaf senescence. Sci Rep 6:31889. https://doi.org/10.1038/srep31889

Mayer M, Hölker AC, González-Segovia E et al (2020) Discovery of beneficial haplotypes for complex traits in maize landraces. Nat Commun 11:4954. https://doi.org/10.1038/s41467-020-18683-3

Mazorra-Manzano MA, Tanaka T, Dee DR, Yada RY (2010) Structure–function characterization of the recombinant aspartic proteinase A1 from Arabidopsis thaliana. Phytochemistry 71:515–523. https://doi.org/10.1016/j.phytochem.2009.12.005

Mishra S, Singh B, Panda K et al (2016) Association of SNP haplotypes of HKT family genes with salt tolerance in Indian Wild Rice Germplasm. Rice 9:15. https://doi.org/10.1186/s12284-016-0083-8

Mligo JK, Craufurd PQ (2005) Adaptation and yield of pigeonpea in different environments in Tanzania. Field Crop Res 94:43–53. https://doi.org/10.1016/j.fcr.2004.11.009

Mula MG, Saxena KB (2010) Lifting the level of awareness on pigeonpea - a global perspective. International Crops Research Institute for the Semi-Arid Tropics, Patancheru

Nigro D, Gu YQ, Huo N et al (2013) Structural analysis of the wheat genes encoding NADH-dependent glutamine-2-oxoglutarate amidotransferases and correlation with grain protein content. PLoS ONE 8:e73751. https://doi.org/10.1371/journal.pone.0073751

Obala J, Saxena RK, Singh VK et al (2018) Genetic variation and relationships of total seed protein content with some agronomic traits in pigeonpea ("Cajanus cajan" (L.) Millsp.). Aust J Crop Sci 12:1859–1865. https://doi.org/10.3316/informit.351758936228886

Obala J, Saxena RK, Singh VK et al (2019a) Development of sequence-based markers for seed protein content in pigeonpea. Mol Genet Genomics 294:57–68. https://doi.org/10.1007/s00438-018-1484-8

Obala J, Saxena RK, Singh VK et al (2019b) Genetic analysis of seed protein content and its association with seed weight and yield in pigeonpea. J Food Legumes 32:65–69

Obala J, Saxena RK, Singh VK et al (2020) Seed protein content and its relationships with agronomic traits in pigeonpea is controlled by both main and epistatic effects QTLs. Sci Rep 10:214. https://doi.org/10.1038/s41598-019-56903-z

Odeny DA (2007) The potential of pigeonpea (Cajanus cajan (L.) Millsp.) in Africa. Nat Res Forum 31:297–305. https://doi.org/10.1111/j.1477-8947.2007.00157.x

Pereira CS, da Costa DS, Pereira S et al (2008) Cardosins in postembryonic development of cardoon: towards an elucidation of the biological function of plant aspartic proteinases. Protoplasma 232:203–213. https://doi.org/10.1007/s00709-008-0288-9

Rao PP, Birthal PS, Bhagavatula S, Bantilan MCS (2010) Chickpea and Pigeonpea economies in Asia: facts, trends and outlook. http://oar.icrisat.org/191/. Accessed 20 Apr 2023

Saxena RK, Jiang Y, Khan AW et al (2021) Characterization of heterosis and genomic prediction-based establishment of heterotic patterns for developing better hybrids in pigeonpea. Plant Genome 14:e20125. https://doi.org/10.1002/tpg2.20125

Saxena KB, Reddy LJ, Saxena RK (2023) Breeding high-protein pigeonpea genotypes and their agronomic and biological assessments. Plant Breed 142(2):129–139

Sinha P, Singh VK, Saxena RK et al (2020) Superior haplotypes for haplotype-based breeding for drought tolerance in pigeonpea (Cajanus cajan L.). Plant Biotechnol J 18:2482–2490. https://doi.org/10.1111/pbi.13422

Team RDC (2009) A language and environment for statistical computing. http://www.R-project.org

Upadhyaya HD, Reddy KN, Gowda CLL, Silim SN (2007a) Patterns of diversity in pigeonpea (Cajanus cajan (L.) Millsp.) germplasm collected from different elevations in Kenya. Genet Resour Crop Evol 54:1787–1795. https://doi.org/10.1007/s10722-006-9198-x

Upadhyaya HD, Reddy KN, Sastry DVSSR, Gowda CLL (2007b) Identification of photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. J SAT Agric Res 3:1–4

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630. https://doi.org/10.1016/j.tplants.2005.10.004

Varshney RK, Chen W, Li Y et al (2012) Draft genome sequence of pigeonpea (Cajanus cajan), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83

Varshney RK, Saxena RK, Upadhyaya HD et al (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49:1082–1088. https://doi.org/10.1038/ng.3872

Varshney RK, Sinha P, Singh VK et al (2020) 5Gs for crop genetic improvement. Curr Opin Plant Biol 56:190–196. https://doi.org/10.1016/j.pbi.2019.12.004

Varshney RK, Bohra A, Yu J et al (2021a) Designing future crops: genomics-assisted breeding comes of age. Trends Plant Sci 26:631–649

Varshney RK, Roorkiwal M, Sun S et al (2021b) A chickpea genetic variation map based on the sequencing of 3,366 genomes. Nature 599:622–627. https://doi.org/10.1038/s41586-021-04066-1