Published online October 18, 2018

Development and Application of High-Density Axiom *Cajanus* SNP Array with 56K SNPs to Understand the Genome Architecture of Released Cultivars and Founder Genotypes

Rachit K. Saxena, Abhishek Rathore, Abhishek Bohra, Pooja Yadav, Roma Rani Das, Aamir W. Khan, Vikas K. Singh, Annapurna Chitikineni, Indra P. Singh, C. V. Sameer Kumar, K. B. Saxena, and Rajeev K. Varshney*

R.K. Saxena, A. Rathore, P. Yadav, R.R. Das, A.W. Khan, V.K. Singh, A. Chitikineni, C.V.S. Kumar, K.B. Saxena, R.K. Varshney, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Telangana, India; A. Bohra, I.P. Singh, ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, Uttar Pradesh, India.

ABSTRACT As one of the major outputs of next-generation sequencing (NGS), a large number of genome-wide single-nucleotide polymorphisms (SNPs) have been developed in pigeonpea [*Cajanus cajan* (L.) Huth.]. However, SNPs require a genotyping platform or assay to be used in different evolutionary studies or in crop improvement programs. Therefore, we developed an Axiom *Cajanus* SNP array with 56K SNPs uniformly distributed across the genome and assessed its utility in a genetic diversity study. From the whole-genome resequencing (WGRS) data on 104 pigeonpea lines, ~2 million sequence variations (SNPs and insertion–deletions [InDels]) were identified, from which a subset of 56,512 unique and informative sequence variations were selected to develop the array. The Axiom *Cajanus* SNP array developed was used for genotyping 103 pigeonpea lines encompassing 63 cultivars released between 1960 and 2014 and 40 breeding, germplasm, and founder lines. Genotyping data thus generated on 103 pigeonpea lines provided 51,201 polymorphic SNPs and InDels. Genetic diversity analysis provided in-depth insights into the genetic architecture and trends in temporal diversity in pigeonpea cultivars. Therefore, the continuous use of the high-density Axiom *Cajanus* SNP array developed will accelerate high-resolution trait mapping, marker-assisted breeding, and genomic selection efforts in pigeonpea.

Abbreviations: CcLG, *Cajanus cajan* linkage group or pseudomolecule; GBS, genotyping-by-sequencing; InDels, insertion–deletions; NGS, next-generation sequencing; QTL, quantitative trait loci; RP, release period; SNP, single-nucleotide polymorphism; SSR, simple-sequence repeat; WGRS, whole-genome resequencing.

CORE IDEAS

- Axiom *Cajanus* SNP array revealed genetic architecture and temporal diversity in pigeonpea varieties.

Pigeonpea is one of the most important pulse crops in the tropics and subtropics of Asia and Africa. Because of its vital economic and nutritional value, tremendous research efforts have led to the development of a large number of improved pigeonpea cultivars during the past decades (Saxena, 2008). The development and application of genomics information, particularly DNA markers and draft genome sequence, represent major achievements (Saxena et al., 2016; Varshney et al., 2012). A number of marker systems, including restriction fragment-length polymorphisms, amplified fragment-length polymorphisms, random amplified polymorphic DNA, single-feature polymorphism, and simple-sequence repeats (SSRs) have been developed and used to assess

Saxena, R.K., A. Rathore, A. Bohra, P. Yadav, R.R. Das, A.W. Khan, V.K. Singh, A. Chitikineni, I.P. Singh, C.V.S. Kumar, K.B. Saxena, and R.K. Varshney. 2018. Development and Application of High-Density Axiom *Cajanus* SNP Array with 56K SNPs to Understand the Genome Architecture of Released Cultivars and Founder Genotypes. Plant Genome 11:180005. doi: 10.3835/plantgenome2018.01.0005

Received 28 Jan. 2018. Accepted 24 July 2018.
*Corresponding author (r.k.varshney@cgiar.org).

genetic diversity (Saxena et al., 2010), construct genetic maps (Bohra et al., 2011), and analyze quantitative trait loci (QTL) (Gnanesh et al., 2011; Bohra et al., 2012) in pigeonpea. Of late, SNPs have been identified in a large number of pigeonpea lines, especially as a result of the NGS-based resequencing of diverse germplasm (Kumar et al., 2016; Varshney et al., 2017). The SNPs have advantages over other existing marker systems such as uniformity and high density across the genome and their amenability to automatic, high-throughput and cost-effective genotyping for genetic polymorphism and QTL analyses in pigeonpea (Saxena et al., 2014, 2017a,b).

Routine and extensive application of SNPs in practical pigeonpea breeding largely depends on ease of use and cost-effectiveness. In this context, a number of SNP genotyping assays such as KASPar (Saxena et al., 2012) and GoldenGate (Roorkiwal et al., 2013) were developed in pigeonpea. However, they proved to be cost-effective only with a limited number of SNPs and genotypes. Another NGS-based approach to identify and assay SNPs is genotyping-by-sequencing (GBS) that has been used recently in trait mapping studies in pigeonpea (Saxena et al., 2017a,b). While GBS can generate SNP data on large sets of genotypes in less time and in a cost-effective manner, it has a limitation of missing values in a large number of lines on a specific data point across the tested population. Given this scenario, SNP arrays with candidate and informative SNPs are a better option for generating high-throughput SNP genotyping data across a population. Data generated through SNP arrays also require less computational knowledge and resources for analysis, which is why high-density SNP arrays have been developed and used for a variety of genetic and breeding applications in many crop species (for a review, please see Rasheed et al., 2017). The availability of the draft genome sequence (Varshney et al., 2012) and resequencing of several hundred lines (Kumar et al., 2016, Varshney et al., 2017) have made it possible to identify millions of SNPs in pigeonpea.

The current study identified a set of ~56K most informative and high-quality SNPs for pigeonpea and developed an Axiom *Cajanus* SNP array. We demonstrated the utility of the SNP array in assessing temporal genetic diversity and pedigree analyses of released cultivars in pigeonpea. We anticipate the Axiom *Cajanus* SNP array to have extensive utility in in-depth germplasm analysis, advancing genetic and genomics research, and breeding applications in pigeonpea.

MATERIALS AND METHODS

Plant Material and DNA Isolation

A set of 103 pigeonpea lines was used. The set included 63 released cultivars and 40 donor, germplasm lines, landraces, or founder parents (Supplemental Table S1). Genomic DNA was isolated from young leaves of individual plants of each pigeonpea line using a NucleoSpin Plant II kit (Macherey-Nagel). The quality of DNA was checked on 0.8% agarose gel and DNA quantity assessed on Qubit 2.0 Fluorometer (Life Technologies, Thermo Fisher Scientific Corp.).

Single-Nucleotide Polymorphism Selection and Array Design

Available data on resequencing of 104 pigeonpea lines (unpublished data, 2017) were used to select sequence variations (SNPs and InDels). The 35-bp sequences flanking both sides of selected sequence variations were extracted using custom script. Flanking sequences of selected sequence variations were filtered according to the criterion mentioned in Pandey et al. (2017). Selected sequence variations representing 11 pseudomolecules (CcLGs) were subjected to in silico validation and their *p*-convert values generated using Affymetrix power tool AxiomGTv1 algorithm (http://www.affymetrix.com/estore/partners_programs/programs/developer/tools/powertools).

For each sequence variation, forward and reverse probes were designed and *p*-convert values assigned. Probes were selected if *p*-convert was >0.6, with no wobbles, and poly count equaled 0. Finally, the selected probes were used to design the Axiom *Cajanus* SNP array.

Genotyping

The selected lines mentioned above were used to generate genotyping data using the newly developed Axiom *Cajanus* SNP array. The Affymetrix GeneTitan platform was used for genotyping. Target probes were prepared using high-quality DNA (20 µL of 10 ng μL^{-1}) from each line following Affymetrix Axiom 2.0 procedure. Further, DNA samples were amplified, fragmented, and hybridized on the chip followed by single-base extension through DNA ligation and signal amplification. Affymetrix GeneTitan was used for staining and scanning the samples.

Allele Calling and Data Analysis

Alleles for specific SNPs were detected through Axiom Analysis Suite version 1.0 (http://media.affymetrix.com/support/downloads/manuals/Axiom_analysis_suite_user_guide.pdf). The best-practices workflow built in Axiom Analysis Suite version 1.0 was used for quality control analysis of samples. The genotyping workflow was used to perform genotyping on the imported .CEL files. Finally, the summary-only workflow was used to produce a summary of the details of the intensities for the probe sets for use in copy number analysis tools. It also allowed the export of SNP data after the analysis.

Diversity Analysis

Polymorphism information content value and minor allele frequency for each SNP were calculated to measure the usefulness of a marker for genetic studies. Further, AMOVA was estimated to examine population variability and nucleotide diversity ($\theta\pi$). Additionally, F_{ST} values between populations grouped in different release periods (RPs) were calculated using Weir and Cockerham (1984) method. At first, we calculated genome-wide diversity for each line; population-specific mean diversities were

Table 1. Details on Axiom *Cajanus* SNP array.

Pseudomolecule	Total sequence variations identified	Sequence variations passed Axiom GTv1	Sequence variations placed on array	No. of polymorphic SNPs	No. of polymorphic InDels	Total polymorphic sequence variations
CcLG01	6360	4644	4638	4353	11	4364
CcLG02	11,744	8516	8506	7771	20	7791
CcLG03	7867	5882	5869	5419	16	5435
CcLG04	5482	4138	4134	3489	10	3499
CcLG05	1255	926	922	857	8	865
CcLG06	6244	4803	4784	4413	36	4449
CcLG07	5948	4330	4321	3909	12	3921
CcLG08	7117	5204	5178	4706	33	4739
CcLG09	5093	3615	3615	3061	0	3061
CcLG10	8432	5949	5928	5235	33	5268
CcLG11	17,297	12,160	8617	7753	56	7809
Total	82,839	60,167	56,512	50,966	235	51,201
Average	7531	5470	5137	4633	21	4655

calculated as the arithmetic mean across the lines. All values were calculated using R language.

Phylogenetic and Structure Analyses

Neighbor joining tree construction and principal component analysis were performed based on a distance matrix using R language. Population structure was estimated using STRUCTURE V2.3.452 software. The tested *K* was set from 1 to 9 and analyses were repeated five times with 50000 Markov Chain Monte Carlo replicates and 10,000 burn-ins. The SNPs with ≥5% minor allele frequency were used for the analyses.

RESULTS

Single-Nucleotide Polymorphism Selection and Axiom *Cajanus* SNP Array Design

The resequencing data generated on 104 lines (unpublished data, 2017) were aligned with the pigeonpea reference genome (Varshney et al., 2012) to identify sequence variations (SNPs and InDels). The alignment resulted in the identification of ~2 million sequence variations. Further, a set of 1554 informative SNPs and 385 InDels showing their possible association with grain protein content, fertility restoration, resistance to sterility mosaic, and fusarium wilt diseases in different pigeonpea studies were also added in the initial set of sequence variations. After applying stringent filtering criteria (see Material and Methods section), a total of 82,839 high-quality sequence variations were selected. These were passed through in silico validation using the Axiom GTv1 algorithm. As a result, 60,167 sequence variations were selected for further processing. Finally, the Axiom *Cajanus* SNP array with 56,512 sequence variations (56,127 SNPs and 385 InDels) was developed (Table 1).

Genome-Wide Distribution of Selected Sequence Variations

The 56,512 sequence variations were distributed on 11 CcLGs of the pigeonpea genome with an average of 5137 sequence variations per CcLG (Table 1). A maximum number of sequence variations came from CcLG11 (8617), while a minimum number came from CcLG05 (922). With respect to genomic positions of the SNPs fixed on the Axiom *Cajanus* SNP array, a total of 39,862 sequence variations were present in the intergenic regions and 7340 in the intronic regions. The other major groups of SNPs include synonymous coding (6299) followed by nonsynonymous coding (2867) regions. In terms of sequence variations effects, they were classified into 2903 missense, 1722 silent, and 73 nonsense mutations, whereas the remaining 51,814 sequence variations were grouped into the 'other' category (Supplemental Table S2).

Grouping Released Cultivars

To assess temporal trends in diversity, all the released cultivars were grouped into four temporal groups: (i) cultivars released before 1980 (hereafter referred to as RP1), (ii) cultivars released between 1981 and 1990 (RP2), (iii) cultivars released between 1991 and 2000 (RP3), and (iv) cultivars released 2001 onward (RP4). These cultivars also represent different pigeonpea growing zones of India as established by the Indian Council of Agricultural Research (ICAR) (http://www.envfor.nic.in/divisions/csurv/geac/Biology_of_Cajanus_cajan_Pigeon_pea.pdf), which include the North Western Plain Zone, North Eastern Plain Zone, Central Zone, and Southern Zone (Supplemental Table S1).

Based on the pedigree of the cultivars, it was observed that 42 were derived from conventional hybridization and pedigree selection programs, while 21 were direct selections from landraces or mutations (Supplemental Table S1). All seven cultivars in RP1 were selections from landraces, as the development of pigeonpea cultivars through hybridization and selection began later in RP2, and such cultivars were released only after 1981. A number of medium- and late-maturing cultivars, in addition to showing good adaptation and yield, were resistant to the two most devastating diseases of wilt (caused by a fungus *Fusarium udum* Butler) and sterility mosaic virus. In RP3 and RP4, a total of 18 and 24 cultivars, respectively, were grouped together and a majority of them were developed through hybridization.

Genetic Diversity and Phylogenetic Relationships

A total of 56,512 sequence variations present on the Axiom *Cajanus* SNP array were used for polymorphism screening on 103 pigeonpea lines representing 63 released cultivars and 40 other lines including elite germplasm and landraces. From the tested sequence variations, a total of 51,201 were found polymorphic across 103 lines, and this included 50,966 SNPs and 235 InDels (Table 1). On average, 4654.64 sequence variations (4633.27 SNPs and 21.36 InDels) per CcLG were identified, and the number of polymorphic sequence variations ranged from 865 (CcLG05) to 7809 (CcLG11). From the identified polymorphic sequence variations, SNPs ranged from 857 (CcLG05) to 7771 (CcLG02), wherein the number of InDels ranged from 0 (CcLG09) to 56 (CcLG11). The polymorphic information content value for the markers ranged from 0.01 to 0.38, with an average of 0.25 for all the examined lines.

Genotyping data obtained for polymorphic loci with ≥5% minor allele frequency were used to assess the genetic relatedness among the lines by calculating pairwise genetic distances (Fig. 1). Based on these polymorphic data, all 103 lines were classified into two main clusters: Cluster I contained 32 lines and Cluster II contained the remaining 71 lines. Cluster I predominantly comprised of early-maturing cultivars (14) and elite germplasm and landraces (18). Cluster II was further divided into subclusters Cluster IIa and Cluster IIb. While Cluster IIa comprised of all 10 late-maturing cultivars, six medium-maturing cultivars, and 14 elite germplasm and landraces, Cluster IIb included 22 medium-maturing cultivars, 11 early-maturing cultivars, and eight elite germplasm and landraces. The STRUCTURE program was used to assess the clustering of released cultivars, elite germplasm, and landraces. Admixture observed in individuals of subpopulation explained the involvement of different parental lines in developing pigeonpea cultivars (Fig. 2). Principal coordinate analysis was also used to distinguish between the released cultivars, elite germplasm, and landraces. A substantial overlap among them was observed (Fig. 3).

Temporal Diversity Trends

The SNPs among the subgroups (temporal groups or released periods) RP1 (34,520), RP2 (39,532), RP3 (42,320), and RP4 (43,445) accounted for 67.4 (percentage in total SNPs detected across RPs), 77.2, 82.6, and 84.8%, respectively, of the total polymorphic SNPs (Table 2). The highest number of unique SNPs was identified within RP4 cultivars (672), followed by RP3 (631), RP2 (298), and RP1 (93) (Fig. 4). A total of 24,154 SNPs were common to all the RPs, suggesting that the breeding materials had relatively low levels of genetic variability (Fig. 4). In pairwise comparisons, 32,461 SNPs were common in RP3 and RP4; while between RP2 and RP4 the number of common SNPs was 31,039. Similarly, in RP2 and RP3 and in RP1 and RP4, 30,994 and 26,356 SNPs were common, respectively (Fig. 4).

To assess the genetic differences within the RPs, pairwise F_{ST} values were estimated (Supplemental Table S3). Results generally showed negative F_{ST} values with the exceptions of RP2 and RP4 (F_{ST}: 0.019) and RP3 and RP4 (F_{ST}: 0.009), where relatively low but positive F_{ST} values were detected. These negative F_{ST} values should be effectively considered as zero, which implies that there is no genetic subdivision between the populations considered (Weir 1996). The negative F_{ST} values indicated the presence of higher heterozygosity levels in cultivars released during the periods of different RPs. However, the level of average heterozygosity decreased in the initial phase of pigeonpea breeding in India, that is, in RP2 (0.048) compared with RP1 (0.062) (Supplemental Table S4). This may have been because the cultivars released in RP1 were direct selections from landraces and possessed higher heterozygosity than RP2 cultivars, which were based on a hybridization and selection program. In RP3, the average heterozygosity increased slightly (0.054) compared with RP2, but it was less than that in RP1. In the case of RP4, the average heterozygosity increased significantly (0.093) compared with all other RPs. The highest number of heterozygous SNPs (13,957) was identified in GT101, an early-maturing cultivar released in 2002 for the Central Zone (Supplemental Table S4). Five other cultivars, namely PKV Tara (10,457 SNPs), TJT501 (10,105 SNPs), GC1139 (10,057 SNPs), CO6 (8971 SNPs), and BSMR853 (8512 SNPs) from RP4 also exhibited high levels of heterozygosity compared with other released cultivars in different RPs. Ten pigeonpea cultivars showed 6 to 16% heterozygous SNPs, while the remaining cultivars (47) showed ≤5% (Supplemental Table S4). Further analysis of molecular variance (AMOVA) revealed 75% variation between the samples of different RPs, while the molecular variation within RP samples was 24.69% (Table 3).

The genotyping data on all the 63 released pigeonpea cultivars and 40 elite germplasm and landraces revealed a nonsignificant increase in diversity in the released cultivars ($\theta\pi = 0.54$) compared with the elite germplasm and landraces ($\theta\pi = 0.53$) (Table 2). This may be an artifact of grouping of varying sizes, as many elite lines could not be used in the present study. In addition, different released varieties grouped in RPs were used to calculate nucleotide diversity. Almost similar nucleotide diversity was detected in cultivars released in RP1 ($\theta\pi = 0.43$), RP2 ($\theta\pi = 0.48$), RP3 ($\theta\pi = 0.49$), and RP4 ($\theta\pi = 0.48$) (Table 2), which indicated that similar diversity levels had been maintained during the course of pigeonpea breeding in the past decades.

DISCUSSION

Genomic approaches based on molecular markers are being used routinely to study genetic diversity, evolution, association mapping, diagnostics, and fingerprinting in genetic enhancement programs in different crops. In the past decade, SNPs have emerged as one of the most potent markers because of their abundance in the genome and feasibility for high-throughput genotyping for genetic studies. The SNP arrays allow the profiling of thousands of markers in a given set of genotypes in a short time to identify associated markers for target traits. Many SNP-based arrays have been developed and used in a number of food crops such as rice (*Oryza*

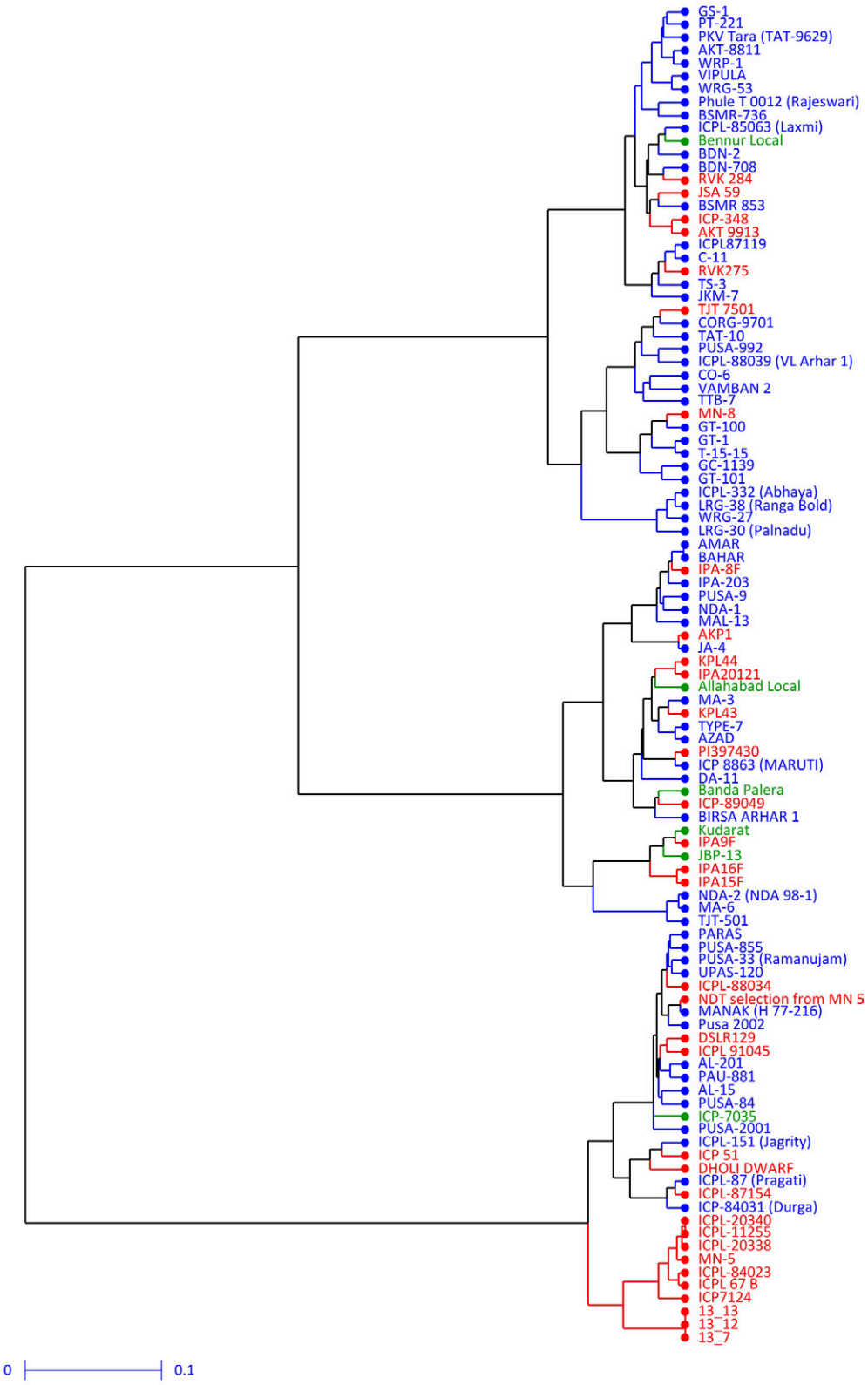

Fig. 1. Neighbor-joining tree analysis of 103 pigeonpea genotypes [63 released cultivars in blue and 40 donor, germplasm lines, and founder parents in red and green (specific to landraces)] using Axiom *Cajanus* SNP array genotyping data.

sativa L.) (McCouch et al., 2010; Chen et al., 2014; Singh et al., 2015), sunflower (*Helianthus annuus* L.) (Bachlava et al., 2012), soybean [*Glycine max* (L.) Merr.] (Song et al., 2013), oil palm (*Elaeis guineensis* Jacq.) (Kwong et al., 2016), maize (*Zea mays* L.) (Ganal et al., 2011), wheat (*Triticum aestivum* L.) (Wang et al., 2014), chickpea (*Cicer arietinum* L.) (Roorkiwal et al., 2017), and groundnut (*Arachis hypogaea* L.) (Pandey et al., 2017).

Given the limited genetic diversity available in cultivated pigeonpea germplasm (Saxena et al., 2014),

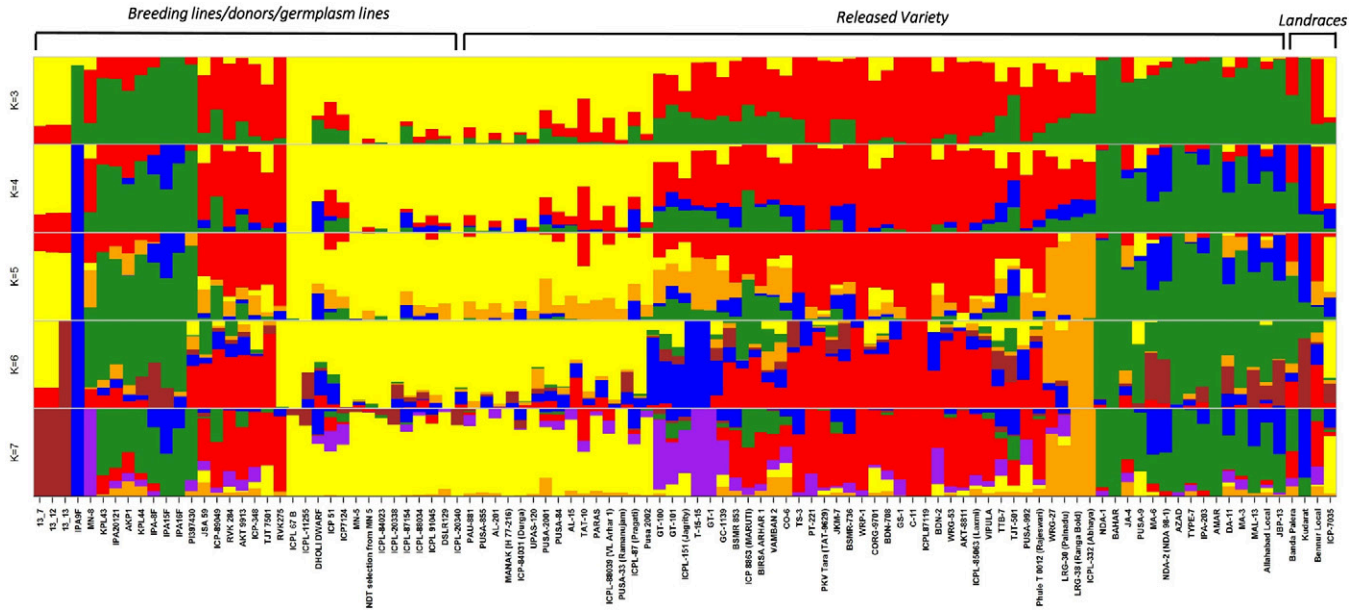

Fig. 2. Clustering of released cultivars, elite germplasm, and landraces in structure analysis using *K*-values from 3 to 7.

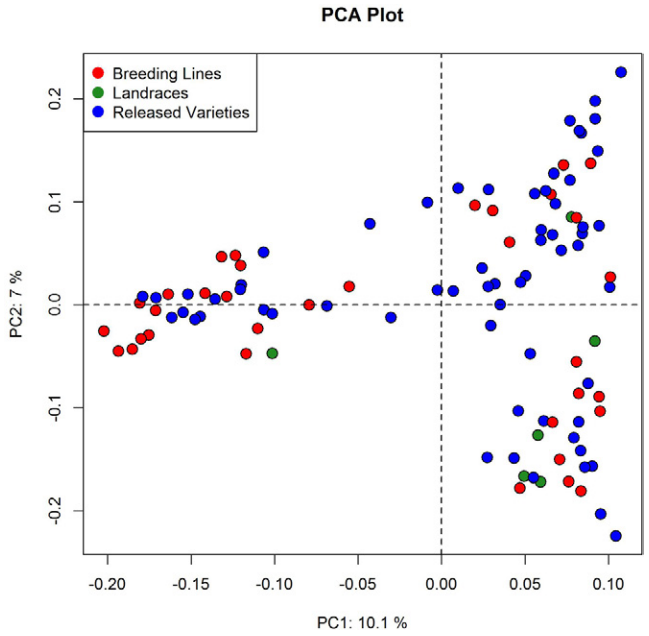

Fig. 3. Principal component analysis (PCA) of 103 pigeonpea genotypes where PC1 explains 10.1% and PC2 explains 7% of total variation.

developing SNP arrays using the most informative markers was required to accelerate genetic and breeding applications. Initially, SNPs were filtered from WGRS data obtained from cultivated pigeonpea lines (unpublished data, 2017). The advantage of having WGRS data from a cultivated pool was that a comprehensive subset of polymorphic SNPs that can work across the cultivated lines could be selected for the array design, which could be highly useful. A potential set of 1554 SNPs and 385 InDels associated with key agronomic traits was also included to design the array. This will provide the opportunity to identify novel haplotypes for key traits to improve pigeonpea. Therefore, we developed Axiom *Cajanus* SNP array

Table 2. Diversity features across pigeonpea cultivars and subgroups.

	No. of genotypes	Polymorphic loci	Nucleotide diversity ($\theta\pi$)
Across cultivars	103	51,201	0.544
Across released cultivars	63	45,266	0.541
RP1	7	34,520	0.483
RP2	14	39,532	0.525
RP3	18	42,320	0.536
RP4	24	43,445	0.533
Across donor, germplasm lines, and founder parents	40	43,931	0.532

with 56,512 sequence variations (SNPs and InDels) that showed a high call rate of 99.9% and very little (<0.10) missing data.

In the last 50 yr, 80 pigeonpea cultivars have been released in India (Singh et al., 2016). However, they did not influence crop productivity to the expected levels, and productivity remained stagnant with a yield of ~700 kg ha^{-1}. This undesirable yield plateau is critical in view of the country's growing population at ~2% per annum and various nutritional security concerns (Prasad 2013). In this context, many pigeonpea scientists believe that breeders' failure in enhancing yield by breeding new cultivars is related to the use of genetic materials with a narrow genetic base. These observations were supported by Kumar et al. (2003) when they showed that the released pigeonpea cultivars largely had common parentage. Hence, it became important to reveal facts about the genetic diversity of the cultivars and redefine new breeding approaches using various molecular tools to overcome this constraint. A number of morphological and molecular methods have been developed in different crop species for this purpose (Liu et al., 1992), and DNA markers with high levels of polymorphism in closely related genotypes have been found to be the best approach (Liu et al., 1992; Lyudmyll et al., 2007;

Table 3. Analysis of molecular variance (AMOVA) in pigeonpea released cultivars using 10,000 permutations.

Source of variation†	df	Sum of squares	Mean square	Variance components	Variation	*p*-value
					%	
Between group	3	37105.83	12368.61	20.91	0.31	0.28327
Between samples within group	59	692881.19	11743.75	5041.95	75.00	0.00009999*
Within samples	63	104570.50	1659.85	1659.85	24.69	0.00009999*

* Significant at the 0.05 probability level.

† Phi: Phi samples total j_{it}, 0.753; Phi samples group j_{is}, 0.752; Phi group total j_{st}, 0.003.

Singh et al., 2016). Therefore, the Axiom *Cajanus* SNP array was deployed to study genetic diversity among 103 pigeonpea lines including 63 released cultivars and 40 elite germplasm and landraces. The identified marker polymorphism (90.6%) reported in this study is higher than previous marker-based diversity studies in pigeonpea (Saxena et al., 2012, 2014). The efficiency of the Axiom *Cajanus* SNP array in assessing polymorphism in pigeonpea is comparable with such arrays developed in rice (96.2%) (Chen et al., 2014) and soybean (90.9%) (Song et al., 2013) and relatively higher than the arrays in sunflower (71.8%) (Bachlava et al., 2012) and groundnut (77.6%) (Pandey et al., 2017). This could be attributed to the selection of the SNPs from the initial discovery panel, which was comprised mainly of pigeonpea cultivars.

It has been suggested that genetic diversity in a number of crop species gradually decreases during the processes of domestication and pure-line breeding (Tanksley and McCouch 1997; Varshney et al., 2017). Investigations related to the genetic diversity in pigeonpea lines using Axiom *Cajanus* SNP array technology revealed that there were no significant differences in the overall genetic diversity among the groups of released cultivars, elite germplasm, and landraces. Similar observations have been made in maize, pea (*Pisum sativum* L.) (Le Clerc et al., 2006), and barley (*Hordeum vulgare* L.) (Koebner et al., 2003; Malysheva-Otto et al., 2007). These observations suggest that either the parental materials used in breeding the cultivars had limited genetic diversity of the favorable alleles or that the breeders failed to bring together the favorable alleles spread across the genome in a single genotype. No significant differences in allelic richness have been reported in cultivated barley collected over a period of 40 to 50 yr (Khlestkina et al., 2006). On the other hand, Russell et al. (2000), in northern European spring barley, and Ordon et al. (2005), in German winter barley six-rowed cultivars, reported loss of genetic diversity in modern cultivars compared with landrace or foundation genotypes. Similarly, Roussel et al. (2004, 2005) revealed a significant decrease in allelic diversity within the French bread wheat germplasm accessions and in European wheat cultivars. In contrast, increased allelic diversity was reported in 75 Nordic spring wheat cultivars from 1900 to 1940 and again from the 1960s (Christiansen et al., 2002) and in 198 Nordic bread wheat landraces and cultivars from the 19th to the 21st centuries (Hysing et al., 2008). In 135 Italian rice accessions (Mantegazza et al., 2008), representing genotypes from 1880 to 2001, and in chickpea (Thudi et al., 2016), 100

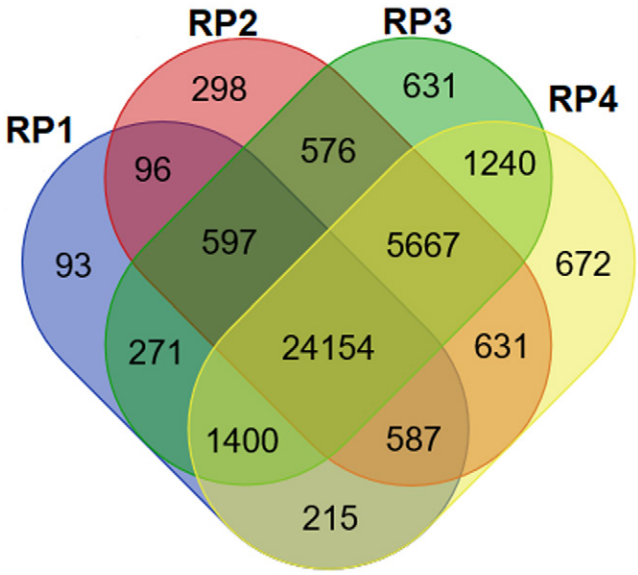

Fig. 4. Shared and unique sequence variations in 63 released cultivars grouped in RP1, RP2, RP3, and RP4.

cultivars released between 1948 and 2012 showed significant enhancement in genetic diversity.

In the present study, the nucleotide diversity in the specified temporal groups slightly varied over the time period (Table 2). This is similar to the observations recorded in our previous simple sequence repeat–based study (Bohra et al., 2017), where 59 Indian pigeonpea cultivars were analyzed and no significant changes in genetic diversity were noticed. The number of cultivars studied in the first two RPs was low compared with RP3 and RP4. The nonlinear increase in unique sequence variations or unique alleles in the population with an increase of the population size confirmed the steep increase in unique alleles in the recent breeding (RP3 and RP4). The majority of sequence variations were common in all RPs (~47.2%). Though a majority of alleles from the first temporal group (RP1) were preserved, unique sequence variations were constantly acquired in the latter RPs, with 17.6, 37.2, and 39.6% of total novel alleles in RP2, RP3, and RP4, respectively. The induction of unique sequence variations in RP2, RP3, and RP4 may also be a reflection of use of more genotypes as well as hybridization breeding, while RP1 has all varieties developed through selection. Since the total genetic diversity within RPs was found to be more or less similar, it can be postulated that there is not enough evidence that breeding had any (negative or positive) impact on genetic diversity in pigeonpea.

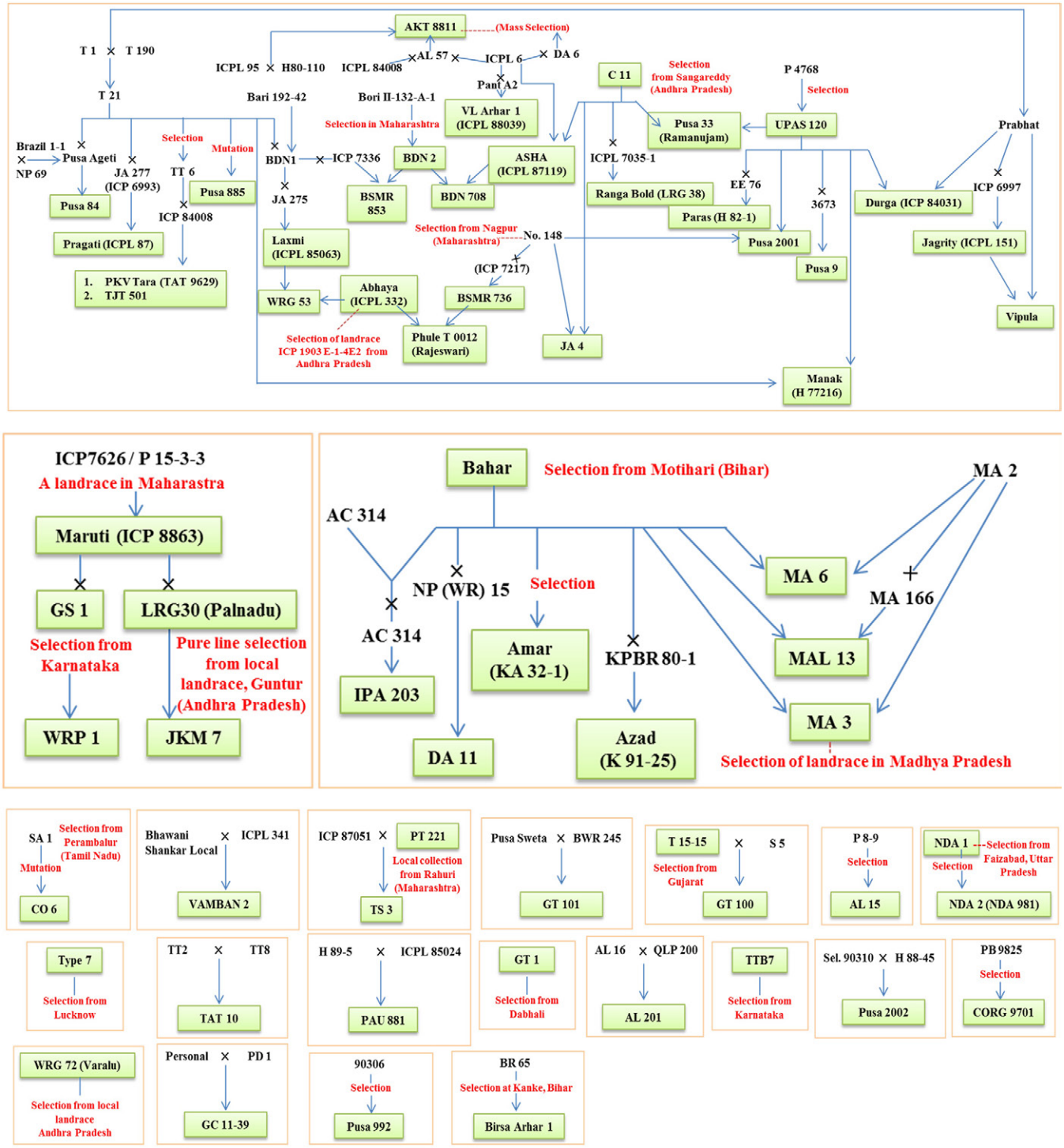

Fig. 5. A status of the genetic base in 63 released pigeonpea cultivars.

Implications of Genome Architecture of Cultivars on Future Breeding

Results of the present study have implications on pigeonpea improvement programs. Although there has not been a reduction in genetic diversity in released cultivars versus founder parents, the genetic base as revealed by the pedigree of released cultivars appears to be narrow because of the frequent use of the same or related parents and their derivatives in breeding programs. A number of improved cultivars have been developed through mainly pure-line selection (21) and hybridization (42) using cultivated species. The 63 cultivars analyzed were developed from 58 founder lines (Fig. 5). The top six founders—T1, T190, P 4768, Bahar (selections from Motihari, Bihar), C11 (selection from Sangareddy, Telangana), and No. 148 (selection from Nagpur, Maharashtra)—contributed 50.8% to the genetic base of the released cultivars. Founders T1, T190, and Bahar appeared in 28.6% of the cultivars. Thus the

amount of variability available to pigeonpea breeders has been limited because of the founder effect and domestication bottleneck as well as limited exploitation of wild species' diversity and other causes. The number of founder lines that have been used in pigeonpea improvement programs appears to constitute an insignificant part of the germplasm in genebanks (Kumar et al., 2003).

Therefore, there is a need to use more germplasm accessions from different genepools through strategic crossing schemes. To this end, the Axiom *Cajanus* SNP array will be helpful in identifying the best parental combinations and in eradicating or minimizing linkage drag in segregating populations, especially when crosses involve unadapted germplasm lines. Crossing programs should involve more genotypes besides those that have been used extensively in the past. Breeding lines derived from multiparent crossing such as MAGIC (multi-parent advanced generation inter-crossing) populations may be used in new crossing programs. In addition, emphasis should be on generating larger populations than were generated earlier or are currently being generated so that selection intensity that is a direct function of genetic gain can be enhanced. Breeding programs should also explore the possibility of using genomic selection that requires multilocation phenotyping data and high-density genotyping data. The Axiom *Cajanus* SNP array will be more effective and useful than other genotyping platforms for this purpose. Furthermore, the Axiom *Cajanus* SNP array will also be useful for undertaking genome-wide association study to identify markers associated with trait for pigeonpea improvement.

In summary, we have reported the development of the Axiom *Cajanus* SNP array with 56K informative SNPs and its successful use in assessing the genetic diversity of released pigeonpea cultivars and founder parents. The array is an important genomic resource for the pigeonpea community that will be useful in implementing new approaches and new ways to accelerate genetic gains in pigeonpea improvement programs. It may be useful in understanding evolutionary biology in the *Cajanus* species as well as the relationship between germplasm collections stored in genebanks.

Supplemental Information Available

Supplemental information is available with the online version of this manuscript.

Conflict of Interest

The authors declare that there is no conflict of interest.

ACKNOWLEDGMENTS

The authors wish to thank the Science and Engineering Research Board (SERB), Government of India; United States Agency for International Development (USAID) and Bill & Melinda Gates Foundation for funding this study. The work was undertaken as part of the CGIAR Research Program on Grain Legumes and Dryland Cereals.

REFERENCES

Bachlava, E., C.A. Taylor, S. Tang, J.E. Bowers, J.R. Mandel, J.M. Burke, and S.J. Knapp. 2012. SNP discovery and development of a high-density genotyping array for sunflower. PLoS One 7:e29814. doi:10.1371/journal.pone.0029814

Bohra, A., A. Dubey, R.K. Saxena, R.V. Penmetsa, K.N. Poornima, N. Kumar, A.D. Farmer, G. Srivani, H.D. Upadhyaya, S.R. Ramesh, D. Singh, K.B. Saxena, P.B. KaviKishor, N.K. Singh, C.D. Town, G.D. May, D.R. Cook, and R.K. Varshney. 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). BMC Plant Biol. 11:56–70. doi:10.1186/1471-2229-11-56

Bohra, A., R. Jha, G. Pandey, P.G. Patil, R.K. Saxena, I.P. Singh, D. Singh, R.K. Mishra, A. Mishra, F. Singh, R.K. Varshney, and N.P. Singh. 2017. New hypervariable SSR markers for diversity analysis, hybrid purity testing and trait mapping in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Front. Plant Sci. 8:377. doi:10.3389/fpls.2017.00377

Bohra, A., R.K. Saxena, B.N. Gnanesh, K.B. Saxena, M. Byregowda, A. Rathore, P.B. Kavi Kishor, D.R. Cook, and R.K. Varshney. 2012. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. Theor. Appl. Genet. 125:1325–1338. doi:10.1007/s00122-012-1916-5

Chen, H., W. Xie, H. He, H. Yu, W. Chen, J. Li, R. Yu, Y. Yao, W. Zhang, Y. He, X. Tang, F. Zhou, X.W. Deng, and Q. Zhang. 2014. A high density SNP genotyping array for rice biology and molecular breeding. Mol. Plant 7:541–553. doi:10.1093/mp/sst135

Christiansen, M.J., S.B. Andersen, and R. Ortiz. 2002. Diversity changes in an intensively bred wheat germplasm during the 20th century. Mol. Breed. 9:1–11. doi:10.1023/A:1019234323372

Ganal, M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, A. Charcosset, J.D. Clarke, E.M. Graner, M. Hansen, J. Joets, M.L. Paslier, M.D. McMullen, P. Montalent, M. Rose, C.C. Schön, Q. Sun, H. Walter, O.C. Martin, and M. Falque. 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6:e28334. doi:10.1371/journal.pone.0028334

Gnanesh, B.N., A. Bohra, M. Sharma, M. Byregowda, S. Pande, V. Wesley, R.K. Saxena, K.B. Saxena, P.B. Kavi Kishor, and R.K. Varshney. 2011. Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crops Res. 123:53–61. doi:10.1016/j.fcr.2011.04.011

Hysing, S.C., T. Säla, H. Nybom, E. Liljeroth, A. Merker, S. Orford, and R.M.D. Koebner. 2008. Temporal diversity changes among 198 Nordic bread wheat landraces and cultivars detected by retrotransposon-based S-SAP analysis. Plant Genet. Resour. 6:113–125. doi:10.1017/S1479262108983544

Khlestkina, E.K., R.K. Varshney, M.S. Röder, A. Graner, and A. Börner. 2006. A comparative assessment of genetic diversity in cultivated barley collected in different decades of the last century in Austria, Albania and India by using genomic and genic SSR markers. Plant Genet. Resour. 4:125–133. doi:10.1079/PGR2006109

Koebner, R.M., P. Donini, J.C. Reeves, R.J. Cooke, and J.R. Law. 2003. Temporal flux in the morphological and molecular diversity of UK barley. Theor. Appl. Genet. 106:550–558. doi:10.1007/s00122-002-1065-3

Kumar, S., S. Gupta, S. Chandra, and B.B. Singh. 2003. How wide is the genetic base of pulse crops? M. Ali, B.B. Singh, S. Kumar, and V. Dhar, editors, Pulses in new perspectives. Proceedings of National symposium on crop diversification and natural resource management. Indian Institute of Pulses Research, Kanpur, India. p. 211–221.

Kumar, V., A.W. Khan, R.K. Saxena, V. Garg, and R.K. Varshney. 2016. First-generation HapMap in *Cajanus* spp. reveals untapped variations in parental lines of mapping populations. Plant Biotechnol. J. 14:1673–1681. doi:10.1111/pbi.12528

Kwong, Q.B., C.K. Teh, A.L. Ong, H.Y. Heng, H.L. Lee, M. Mohamed, J.Z. Low, S. Apparow, F.T. Chew, S. Mayes, H. Kulaveerasingam, M. Tammi, and D.R. Appleton. 2016. Development and validation of a high density SNP genotyping array for African oil palm. Mol. Plant 9:1132–1141. doi:10.1016/j.molp.2016.04.010

Le Clerc, V., V. Cadot, M. Canadas, J. Lallemand, D. Guerin, and F. Boulineau. 2006. Indicators to assess temporal genetic diversity in the French catalogue: No losses for maize and peas. Theor. Appl. Genet. 113:1197–1209. doi:10.1007/s00122-006-0368-1

Liu, Y.G., T.M. Ikeda, and K. Tsunewaki. 1992. Moderately repeated, dispersed, and highly variable (MRDHV) genomic sequences of common wheat usable for cultivar identification. Theor. Appl. Genet. 84:535–543. doi:10.1007/BF00224149

Lyudmyll, M.-O., M.W. Ganal, J.R. Law, J.C. Reeves, and M.S. Röder. 2007. Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.). Mol. Breed. 20:309–322. doi:10.1007/s11032-007-9093-y

Malysheva-Otto, L., M.W. Ganal, J.R. Law, J.C. Reeves, and M.S. Röder. 2007. Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.). Mol. Breed. 20:309–322. doi:10.1007/s11032-007-9093-y

Mantegazza, R., M. Biloni, F. Grassi, B. Basso, B.R. Lu, X.X. Cai, F. Sala, and A. Spada. 2008. Temporal trends of variation in Italian rice germplasm over the past two centuries revealed by AFLP and SSR markers. Crop Sci. 48:1832–1840. doi:10.2135/cropsci2007.09.0532

McCouch, S.R., K. Zhao, M. Wright, C-W. Tung, K. Ebana, M. Thomson, A. Reynolds, D. Wang, G. DeClerck, Md.A. Liakat, A. McClung, G. Eizenga, and C. Bustamante. 2010. Development of genome-wide SNP assays for rice. Breed. Sci. 60:524–535. doi:10.1270/jsbbs.60.524

Ordon, F., J. Ahlemeyer, K. Werner, W. Köhler, and W. Friedt. 2005. Molecular assessment of genetic diversity in winter barley and its use in breeding. Euphytica 146:21–28. doi:10.1007/s10681-005-5192-1

Pandey, M.K., G. Agarwal, S.M. Kale, J. Clevenger, S.N. Nayak, M. Sriswathi, A. Chitikineni, C. Chavarro, X. Chen, H.D. Upadhyaya, and M.K. Vishwakarma. 2017. Development and evaluation of a high density genotyping 'Axiom_*Arachis*' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci. Rep. 7:40577. doi:10.1038/srep40577

Prasad, R. 2013. Population growth, food shortages and ways to alleviate hunger. Curr. Sci. 105:32–46.

Rasheed, A., Y. Hao, X. Xia, A. Khan, Y. Xu, R.K. Varshney, and Z. He. 2017. Crop breeding chips and genotyping platforms: Progress, challenges, and perspectives. Mol. Plant 10:1047–1064. doi:10.1016/j.molp.2017.06.008

Roorkiwal, M., A. Jain, S.M. Kale, D. Doddamani, A. Chitikineni, M. Thudi, and R.K. Varshney. 2017. Development and evaluation of high-density Axiom®*CicerSNP* Array for high-resolution genetic mapping and breeding applications in chickpea. Plant Biotechnol. J. 16:890–901 doi:10.1111/pbi.12836

Roorkiwal, M., S.L. Sawargaonkar, A. Chitikineni, M. Thudi, R.K. Saxena, H.D. Upadhyaya, M.I. Vales, O. Riera-Lizarazu, and R.K. Varshney. 2013. Single nucleotide polymorphism genotyping for breeding and genetics applications in chickpea and pigeonpea using the BeadXpress platform. Plant Genome 6:1–10. doi:10.3835/plantgenome2013.05.0017

Roussel, V., J. Koenig, M. Beckert, and F. Balfourier. 2004. Molecular diversity in French bread wheat accessions related to temporal trends and breeder origin. Theor. Appl. Genet. 108:920–930. doi:10.1007/s00122-003-1502-y

Roussel, V., L. Leisova, F. Exbrayat, Z. Stehno, and F. Balfourier. 2005. SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor. Appl. Genet. 111:162–170. doi:10.1007/s00122-005-2014-8

Russell, J.R., R.P. Ellis, W.T.B. Thomas, R. Waugh, J. Provan, A. Booth, J. Fuller, P. Lawrence, G. Young, and W. Powell. 2000. A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol. Breed. 6:553–568. doi:10.1023/A:1011372312962

Saxena, K.B. 2008. Genetic improvement of pigeonpea: A review. Trop. Plant Biol. 1:159–178. doi:10.1007/s12042-008-9014-1

Saxena, R.K., S.M. Kale, V. Kumar, S. Parupalli, S. Joshi, V.K. Singh, V. Garg, R.R. Das, M. Sharma, K.N. Yamini, A. Ghanta, A. Rathore, C.V. Sameer Kumar, K.B. Saxena, and R.K. Varshney. 2017b. Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea. Sci. Rep. 7:1813. doi:10.1038/s41598-017-01535-4

Saxena, R.K., R.V. Penmetsa, H.D. Upadhyaya, A. Kumar, N. Carrasquilla-Garcia, J.A. Schlueter, A. Farmer, A.M. Whaley, B.K. Sarma, G.D. May, D.R. Cook, and R.K. Varshney. 2012. Large-scale development of cost-effective SNP marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. DNA Res. 19:449–461. doi:10.1093/dnares/dss025

Saxena, R.K., K.B. Saxena, and R.K. Varshney. 2010. Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. Mol. Breed. 26:371–380. doi:10.1007/s11032-010-9459-4

Saxena, R.K., V.K. Singh, S.M. Kale, R. Tathineni, S. Parupalli, V. Kumar, V. Garg, R.R. Das, M. Sharma, K.N. Yamini, S. Muniswamy, A. Ghanta, A. Rathore, C.V. Sameer Kumar, K.B. Saxena, P.B. Kavi Kishor, and R.K. Varshney. 2017a. Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for *fusarium* wilt resistance in pigeonpea. Sci. Rep. 7:1911. doi:10.1038/s41598-017-01537-2

Saxena, R.K., M. Thudi, and R.K. Varshney. 2016. Genomics, trait mapping and molecular breeding in pigeonpea and chickpea. Indian J. Genet. Plant Breed. 76:504–511. doi:10.5958/0975-6906.2016.00068.7

Saxena, R.K., E. von Wettberg, H.D. Upadhyaya, V. Sanchez, S. Songok, K.B. Saxena, P. Kimurto, and R.K. Varshney. 2014. Genetic diversity and demographic history of *Cajanus* spp. illustrated from genome-wide SNPs. PLoS One 9:e88568. doi:10.1371/journal.pone.0088568

Singh, N., D.R. Choudhury, G. Tiwari, A.K. Singh, S. Kumar, K. Srinivasan, R.K. Tyagi, A.D. Sharma, N.K. Singh, and R. Singh. 2016. Genetic diversity trend in Indian rice varieties: An analysis using SSR markers. BMC Genet. 17:127. doi:10.1186/s12863-016-0437-7

Singh, N., P.K. Jayaswal, K. Panda, P. Mandal, V. Kumar, B. Singh, S. Mishra, Y. Singh, R. Singh, V. Rai, A. Gupta, T.R. Sharma, and N.K. Singh. 2015. Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. Sci. Rep. 5:11600. doi:10.1038/srep11600

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One 8:e54985. doi:10.1371/journal.pone.0054985

Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. Science 277:1063–1066. doi:10.1126/science.277.5329.1063

Thudi, M., A. Chitikineni, X. Liu, W. He, M. Roorkiwal, W. Yang, J. Jian, D. Doddamani, P.M. Gaur, A. Rathore, S. Samineni, R.K. Saxena, D. Xu, N.P. Singh, S.K. Chaturvedi, G. Zhang, J. Wang, S.K. Datta, X. Xu, and R.K. Varshney. 2016. Recent breeding programs enhanced genetic diversity in both desi and kabuli varieties of chickpea (*Cicer arietinum* L.). Sci. Rep. 6:38636. doi:10.1038/srep38636

Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T. Donoghue, S. Azam, G. Fan, A.M. Whaley, A.D. Farmer, J. Sheridan, A. Iwata, R. Tuteja, R.V. Penmetsa, W. Wu, H.D. Upadhyaya, S.P. Yang, T. Shah, K.B. Saxena, T. Michael, W.R. McCombie, B. Yang, G. Zhang, H. Yang, J. Wang, C. Spillane, D.R. Cook, G.D. May, X. Xu, and S.A. Jackson. 2012. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat. Biotechnol. 30:83–89. doi:10.1038/nbt.2022

Varshney, R.K., R.K. Saxena, H.D. Upadhyaya, A.W. Khan, Y. Yu, C. Kim, A. Rathore, D. Kim, J. Kim, S. An, V. Kumar, G. Anuradha, K.N. Yamini, W. Zhang, S. Muniswamy, J.S. Kim, R.V. Penmetsa, E. von Wettberg, and S.K. Datta. 2017. Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat. Genet. 49:1082–1088. doi:10.1038/ng.3872

Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao, B.E. Huang, M. Maccaferri, S. Salvi, S.G. Milner, L. Cattivelli, A.M. Mastrangelo, A. Whan, S. Stephen, G. Barker, R. Wieseke, J. Plieske, International Wheat Genome Sequencing Consortium, M. Lillemo, D. Mather, R. Appels, R. Dolferus, G. Brown-Guedira, A. Korol, A.R. Akhunova, C. Feuillet, J. Salse, M. Morgante, C. Pozniak, M.C. Luo, J. Dvorak, M. Morell, J. Dubcovsky, M. Ganal, R. Tuberosa, C. Lawley, I. Mikoulitch, C. Cavanagh, K.J. Edwards, M. Hayden, and E. Akhunov. 2014. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol. J. 12:787–796. doi:10.1111/pbi.12183

Weir, B.S. 1996. Genetic data analysis II: Methods for Discrete Popuation Genetic Data. Sinauer Assoc., Inc., Sunderland, MA.

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370. doi:10.2307/2408641