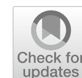
Theoretical and Applied Genetics (2020) 133:737–749
https://doi.org/10.1007/s00122-019-03504-z

ORIGINAL ARTICLE



Genotyping-by-sequencing and multilocation evaluation of two interspecific backcross populations identify QTLs for yield-related traits in pigeonpea

Rachit K. Saxena[1] · Sandip Kale[2] · Reyazul Rouf Mir[3] · Nalini Mallikarjuna[1] · Pooja Yadav[1] · Roma Rani Das[1] · Johiruddin Molla[1] · Muniswamy Sonnappa[4] · Anuradha Ghanta[5] · Yamini Narasimhan[5] · Abhishek Rathore[1] · C. V. Sameer Kumar[5] · Rajeev K. Varshney[1]

Received: 26 April 2019 / Accepted: 6 December 2019 / Published online: 16 December 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This study has identified single-nucleotide polymorphism (SNP) markers associated with nine yield-related traits in pigeonpea by using two backcross populations (BP) developed through interspecific crosses and evaluating them at two locations and 3 years. In both the populations, markers have shown strong segregation distortion; therefore, a quantitative trait locus (QTL) mapping mixed model was used. A total of 86 QTLs explaining 12–21% phenotypic variation were detected in BP-1. On the other hand, 107 QTLs explaining 11–29% phenotypic variation were detected in BP-2. Although most QTLs were environment and trait specific, few stable and consistent QTLs were also detected. Interestingly, 11 QTLs in BP-2 were associated with more than one trait. Among these QTLs, eight QTLs associated with days to 50% flowering and days to 75% maturity were located on CcLG07. One SNP "S7_14185076" marker in BP-2 population has been found associated with four traits, namely days to 50% flowering, days to 75% maturity, primary branches per plant and secondary branches per plant with positive additive effect. Hence, the present study has not only identified QTLs for yield-related traits, but also discovered novel alleles from wild species, which can be used for improvement of traits through genomics-assisted breeding.

Introduction

Pigeonpea (*Cajanus cajan* (L.) Millspaugh) is the only domesticated species of genus *Cajanus*. It is a highly valued protein-rich legume crop grown on about 7 million hectares as a field and/or backyard crop in about 82 countries around the globe. The genus *Cajanus* contains a total of 32 species (van der Maesen 1986, 1990) grouped into three gene pools (GP) with cultivated pigeonpea in primary gene pool (GP1), cross-compatible wild species in the secondary gene pool (GP2) and the cross-incompatible wild species in the tertiary gene pool (GP3) (Bohra et al. 2010). Pigeonpea productivity has remained stagnant at around 700–750 kg/ha (Saxena et al. 2015) although further gain in productivity should be possible by enhancing the genetic base of cultivated material either through complex crossing schemes such as multi-parent advanced generation intercrossing (MAGIC) or through bringing novel alleles from wild species (Saxena et al. 2018c).

Wild *Cajanus* species are the reservoir of many useful genes including resistance to diseases and insect pests that were eliminated from the cultivated gene pools during

Communicated by Jiankang Wang.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-019-03504-z) contains supplementary material, which is available to authorized users.

✉ Rajeev K. Varshney
r.k.varshney@cgiar.org

[1] International Crops Research Institute for the Semi-Arid Tropics (ICRSAT), Patancheru, Telangana 502324, India

[2] The Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstr. 3, 06466 Seeland, OT Gatersleben, Germany

[3] Sher-E-Kashmir University of Agricultural Sciences and Technology of Kashmir (SKUAST-K), Wadura Campus, Sopore, Kashmir 193201, India

[4] Agricultural Research Station (UAS-Raichur), Gulbarga, Karnataka 585101, India

[5] Professor Jayashankar Telangana State Agricultural University, Rajendranagar, Hyderabad, Telangana 500030, India

the domestication and breeding process (Varshney et al. 2017). Therefore, wild species carrying favorable alleles possess great potential for pigeonpea improvement (Aruna et al. 2005). Development of interspecific populations such as backcross populations (BP), introgression lines (ILs), chromosome segment substitution lines (CSSLs) and near-isogenic lines (NILs) via backcrossing is a well-accepted approach being used for introgression of favorable alleles into cultivated genotypes. Backcross populations are advanced through a series of backcrossing the donor and recipient parents and multiple rounds of selfing to reach the stage of ILs and CSSLs. While early-generation backcross populations such as BC_1F_1, BC_1F_2s (Khera et al. 2019) and BC_2F_2s (Swamy et al. 2012) have been useful in trait mapping studies in different crops, advanced backcross populations such as ILs, CSSLs and NILs have been proven invaluable resource for mapping, cloning and gene interactions for the beneficial genomic segments or quantitative trait loci (QTLs) governing important agronomic traits in a number of crop species. However, precise genotyping and phenotyping information on above-mentioned populations is required to identify the short chromosome segments and associate them with trait value for their applications in breeding. For instance, IL populations were used for identification and isolation of several candidate genes responsible for agronomic traits in tomato (Rousseaux et al. 2005; Schauer et al. 2006; Bermudez et al. 2008; Perez-Fons et al. 2014). In the case of rice, CSSLs and NILs were used in fine mapping of QTL regions for grain length (Wan et al. 2006), grain width (Wan et al. 2008), flowering time (Thomson et al. 2006) and other economically important traits (Wang et al. 2007; Xu et al. 2010; Chen et al. 2014; Subudhi et al. 2015; Qiu et al. 2017). Barley is another crop where CSSLs were extensively used for identification of QTLs for morphological traits (Gyenis et al. 2007), agronomic traits (von Korff et al. 2006; Schmalenbach et al. 2009; Schmalenbach and Pillen 2009), disease resistance (Yun et al. 2006), etc. Similar examples are also available in many other crop species. Because of limitations arising due to incompatibility barriers, limited recombination and linkage drag, only few wild species can be used for pigeonpea improvement (Saxena et al. 2018a). Further, effective use of wild germplasm in crop improvement will also require surveys of accessions at the genome scale, identification of adaptive alleles to environmental extremes and incorporation of the novel alleles from wild species into cultivated backgrounds.

Application of genomics has accelerated the development/characterization of backcross/advanced backcross populations and facilitated introgression of favorable alleles into elite breeding backgrounds. This approach is proved useful in discovery of novel alleles from wild species for agronomical important traits (Wang et al. 1992; Song et al. 1995; Zamir 2001; Periyannan et al. 2013). However, no such efforts have been undertaken in pigeonpea using interspecific backcross populations. On the other hand, availability of draft genome (Varshney et al. 2012), re-sequencing data (Kumar et al. 2016; Varshney et al. 2017), molecular markers (Bohra et al. 2011, 2012; Saxena et al. 2012, 2018c), coupled with advances in next-generation sequencing including genotyping-by-sequencing (GBS) (Saxena et al. 2017a, c, 2018b) and Axiom *Cajanus* SNP array (Saxena et al. 2018c; Yadav et al. 2019), has further reduced the cost of high-throughput genotyping and increased the resolution of trait mapping in a number of intraspecific populations in pigeonpea. For instance, economically important traits including disease resistance (Saxena et al. 2017a, c), restoration of fertility (Saxena et al. 2018b), growth habit (Saxena et al. 2017b), seed protein content (Obala et al. 2019), high-selfing flower and seed quality traits (Yadav et al. 2019) have been mapped in pigeonpea.

In view of the above, this study used two backcross populations (BP) developed using *C. acutifolius* and *C. cajanifolius* and cultivated pigeonpea (Sharma and Upadhyaya 2016; Sharma 2017). *Cajanus acutifolius* (ICPW 15613) and the interspecific derivatives *C. acutifolius* × *C. cajan* have shown resistance to *Helicoverpa armigera* (Jadhav et al. 2012), whereas *C. cajanifolius* (ICPW 29) is the immediate ancestor of *C. cajan* (Saxena et al. 2014; Varshney et al. 2017) and has been used in the development of A4-based cytoplasmic male sterility system in pigeonpea (Saxena et al. 2005). In this study, we report molecular and phenotypic characterization of two backcross populations and their use in identification of QTLs/novel alleles for several yield-related traits.

Material and methods

Plant material and field evaluation

Two backcross populations derived from "ICPL 87119 × ICPW 15613" were designated as BP-1 with 149 lines, and "ICPL 87119 × ICPW 29" was designated as BP-2 population with 181 lines (Sharma and Upadhyaya 2016). The common recurrent parent ICPL 87119, popularly known as "Asha," is a high-yielding, large-seeded, widely adapted, resistant to fusarium wilt and sterility mosaic diseases pigeonpea variety, released in the Central and South zones of India in 1992 (ICRISAT 1993). It is important to note that a number of genetic and genomic resources including draft genome have been developed for this variety (Varshney et al. 2012). Two *Cajanus* species, namely *C. acutifolius* (ICPW 15613) and *C. cajanifolius* (ICPW 29), were used as donor parents in developing two pigeonpea backcross populations (Sharma and Upadhyaya 2016). Both the populations were developed by backcrossing with the recurrent parent

to generate BC_2F_2 population (Fig. 1). The two backcross populations were evaluated for 3 years and two locations (continuously 2 years at one location).

BP-1 and BP-2 were grown at ICRISAT, Patancheru, for 2012–2013 and 2013–2014 cropping seasons. Then, BP-1 was grown at PJTSAU, ARS, Tandur, for 2013–2014 and 2014–2015 cropping seasons and BP-2 was grown at UAS-R, ARS, Gulbarga, for 2013–2014 and 2014–2015 cropping seasons. These populations were evaluated for nine yield and yield-related traits, viz. days to 50% flowering, days to 75% maturity, plant height, primary branches per plant, secondary branches per plant, number of pods per plant, number of seeds per pod, seed weight and yield. During these 3 years, the BP-1 and BP-2 populations were evaluated in BC_2F_2, BC_2F_3 and BC_2F_4 generations. In order to correlate the BC_2F_3 and BC_2F_4 generations phenotyping data with the BC_2F_2 plants, we have followed a line-to-progeny approach, which is a widely accepted approach. In summary, we have taken phenotyping data on 5–20 plants in three replications following alpha design in advanced generations representing line (where genotyping performed) from the previous generation.

Phenotyping data analysis

Analysis of variance was performed at individual locations for both the seasons considering random effect for replication and lines. Combined analysis of variance across locations and years was carried out using REML procedure of GenStat for Windows v19 (VSN International 2017) considering all effects as random and modeling individual environment (combination of year and location) error variances. Best linear unbiased predictors (BLUPs) were calculated for lines, environment and their interaction ($G \times E$) effects. Broad-sense heritability, genetic advance and other genetic parameters were estimated for both individual and combined analyses.

Genotyping-by-sequencing (GBS)

Two to three young leaves from individual plants in both the populations and parental lines were used to isolate the genomic DNA using NucleoSpin Plant II kit (Macherey–Nagel, Düren, Germany). The quality and quantity of DNA were checked on 0.8% agarose gel followed by Qubit 2.0 fluorometer (Thermo Fisher Scientific Inc., USA). Ten nanograms of genomic DNA from each sample was restriction digested using *Ape*KI (recognition site: G/CWCG) endonuclease. The digested product was ligated with uniquely bar-coded adaptors using T4 DNA ligase enzyme and was further incubated at 22 °C for 1 h and heated at 65 °C for 30 min to inactivate the T4 ligase. Such digested ligated products from each sample were mixed in equal proportion to construct the GBS libraries, which were then amplified and purified to remove excess adapters. The DNA libraries were sequenced on HiSeq 2500 platform (Illumina Inc., San Diego, CA, USA) to generate genome-wide sequence reads. The generated GBS data were analyzed using a set of analytical and decision support tools for the identification and genotyping of single-nucleotide polymorphism (SNP) as mentioned in Saxena et al. (2017a, c).

Single-nucleotide polymorphism (SNP) identification

A very straightforward strategy was utilized to analyze GBS data generated on both the populations. The SNPs were identified using TASSEL-GBS pipeline (Bradbury et al. 2007). The GBS and whole genome re-sequencing (WGRS) (Kumar et al. 2016) data on parental lines were compared, and wherever possible, the missing SNPs from parental lines were called based on allele calls in WGRS data. Further, the SNPs with > 10% missing data and samples with > 20% missing data were excluded from downstream analysis.

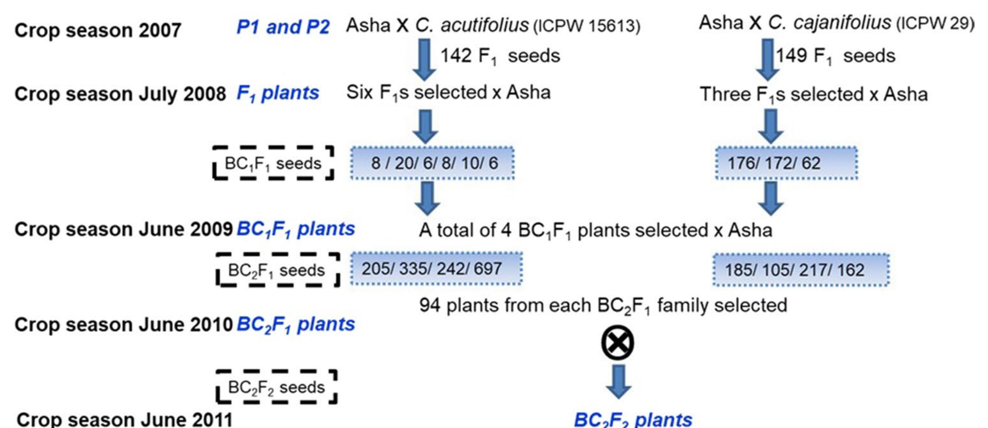


Fig. 1 Scheme for the development of the two pigeonpea backcross (BP) populations. The BP-1 population was derived from ICPL 87119 × ICPW 15613 and BP-2 population derived from ICPL 87119 × ICPW 29

Quantitative trait loci (QTL) mapping

Two approaches, viz. chromosome segment substitution line (CSSL) QTL mapping implemented in CSSL finder (Lorieux 2005) and mixed model approach implemented in TASSEL, were used for QTL analysis. In CSSL-based QTL studies, F test was used for each marker and for a given trait to evidence positions of QTLs. In the present study, a total of 500 permutations were carried out and F test threshold value 10 was used to identify QTL. In mixed model approach, first kinship relationships among samples were determined based on genotyping data. This kinship matrix along with genotyping and phenotyping data was used to identify significant marker trait associations (MTAs). The false-positive MTAs were removed using Bonferroni correction ($P<0.05$).

Results

Development of backcross populations

Two backcross populations were developed successfully, namely BP-1 (ICPL 87119 × ICPW 15613) and BP-2 (ICPL 87119 × ICPW 29) (Sharma and Upadhyaya 2016). While developing the BP-1 and BP-2 populations, six and three F_1s were selected, respectively, based on polymorphic markers between the parents. Subsequently selected F_1s were backcrossed with ICPL 87119 to produce BC_1F_1 seeds. The BC_1F_1 plants in both the crosses were raised and confirmed based on morphological features. Further, confirmed BC_1F_1 plants were used for making second backcross with the cultivated parent to produce BC_2F_1s. These BC_2F_1 plants were selfed to produce BC_2F_2s (Fig. 1). A representative set of 149 (in BP-1 population) and 181 (in BP-2 population) BC_2F_2s were then phenotyped for 3 years and two locations.

Phenotypic analysis

Mean square and coefficient of variation (CV) for each of the studied traits were analyzed in BP-1 population (Table 1) and BP-2 population (Table 2). Among the agronomic traits, plant height and seed yield per plant in both populations have shown the maximum variation in the pooled data analysis across the locations and years. However, the number of seeds per pod and primary branches per plant have shown least variation in both populations. In BP-1 population, phenotypic variance (σ^2P) had more pronounced effect on target traits as compared to genotypic variance (σ^2G) and environment* genotype variance (σ^2E*G) (Table 1). In BP-2 population, other than the plant height and yield per plant, all other traits have shown nonsignificant differences among σ^2P, σ^2G and σ^2E*G values (Table 2). The heritability (H^2) in BP-1 population ranged from 5.11 (secondary branches per plant) to 47.45 (seed weight) (Table 1) and in BP-2 population 17.35 (number of seeds per pod) to 82.53 (seed weight) (Table 2). The phenotyping data analysis suggested that agronomic traits measured in BP-2 population had lesser phenotype and environment effects on genotype and high heritability as compared to BP-1 population.

GBS-based SNP discovery

A total of 18.6 Gb (203,115,257 million reads) and 53.7 Gb (17,228,019,343 million reads) GBS reads were generated using HiSeq2500 platform for BP-1 and BP-2 populations, respectively. The reads from individual progenies ranged from 1009 to 2,634,414 reads in BP-1 population (Supplementary Table S1) and 11 to 419,575,017 reads in BP-2 population (Supplementary Table S2). Also, a total of 2,051,107 (ICPL 87119) and 2,501,773 (ICPW 15613) reads of BP-1 population parents and 1,498,201 (ICPL 87119) and 1,868,189 (ICPW 29) reads of BP-2 population parents were generated. To identify the genome-wide SNPs, TASSEL-GBS pipeline was used

Table 1 Estimates of broad-sense heritability, genotypic and phenotypic coefficients of variation for nine traits in BP-1 population

Trait	σ^2G	σ^2E*G	σ^2E	σ^2P	H^2	Grand mean	Minimum	Maximum	Range
100 seed weight (gm)	0.12	0.30	0.69	0.25	47.45	9.94	9.02	10.82	1.80
Days to 50% flowering	1.40	1.40	56.68	6.47	21.63	125.06	123.10	127.57	4.46
Days to 75% maturity	1.50	1.80	60.38	6.98	21.49	168.14	165.81	170.90	5.09
Number of pods per plant	0.52	3.66	7.68	2.59	20.05	13.79	10.96	17.02	6.05
Number of seeds per pod	0.00	0.00	0.01	0.00	13.11	1.84	1.80	1.87	0.07
Plant height (cm)	31.70	62.40	396.37	80.33	39.46	149.72	133.32	163.28	29.96
Primary branches per plant	0.00	0.03	0.13	0.03	15.66	2.85	2.54	3.05	0.50
Secondary branches per plant	0.01	0.26	0.57	0.16	5.11	4.43	3.90	5.07	1.17
Yield per plant (gm)	36.93	46.35	886.32	207.82	17.77	49.12	37.69	68.53	30.84

σ^2G, Genetic variance; σ^2E*G, environment * genotype variance; σ^2E, variance error; σ^2P, phenotypic variance; H^2, broad-sense heritability

Table 2 Estimates of broad-sense heritability, genotypic and phenotypic coefficients of variation for nine traits in BP-2 population

Trait	σ^2G	σ^2E*G	σ^2E	σ^2P	H^2	Grand mean	Minimum	Maximum	Range
100 seed weight (gm)	0.39	0.23	0.31	0.47	82.53	10.10	7.75	11.70	3.95
Days to 50% flowering	11.86	11.70	8.04	15.45	76.74	129.80	114.38	140.16	25.78
Days to 75% maturity	11.76	15.16	9.34	16.33	72.02	176.01	161.18	187.69	26.52
Number of pods per plant	0.71	0.62	0.80	1.01	70.67	13.04	9.80	16.35	6.55
Number of seeds per pod	0.00	0.00	0.00	0.00	17.35	1.89	1.88	1.90	0.02
Plant height (cm)	70.11	89.85	112.47	101.94	68.77	170.01	142.71	190.65	47.94
Primary branches per plant	0.01	0.02	0.06	0.02	28.11	3.59	3.20	3.84	0.64
Secondary branches per plant	0.04	0.11	0.12	0.09	43.55	3.55	2.83	4.32	1.48
Yield per plant (gm)	36.62	48.73	144.82	68.95	53.11	53.11	36.12	76.61	40.48

σ^2G, Genetic variance; σ^2E*G, environment * genotype variance; σ^2E, variance error; σ^2P, phenotypic variance; H^2, broad-sense heritability

with sequence data for all parents along with respective lines. As a result, 148,487 SNPs in BP-1 population and 168,056 SNPs in BP-2 population were identified (Table 3). Detected SNPs were nonuniformly distributed across different *Cajanus cajan* linkage groups (CcLGs). For instance, 4051 (CcLG05) to 32,416 (CcLG11) SNPs in BP-1 population and 4547 (CcLG05) to 36,663 (CcLG11) SNPs in BP-2 population were distributed across the linkage groups (Table 3). A total of 101,512 SNPs from 148,487 SNPs and 64,801 SNPs from 168,056 SNPs were polymorphic in parental lines of BP-1 population and BP-2 population, respectively (Table 3). Polymorphic SNPs in parental lines and individual lines were subjected to filtering criteria as mentioned in Material and methods section. Finally, in BP-1 population 26,006 SNPs and 134 lines and in BP-2 population 16,052 SNPs and 165 lines could be retained for downstream analysis.

Genomic composition of backcross populations

BP-1 population (ICPL 87119 × ICPW 15613)

A total of 26,006 SNPs from 134 lines were used for inquiry of genome composition in BP-1 population. An average of 2364.18 SNPs were placed in each CcLG. A wide range of chromosomal segments from ICPW 15613 (donor parent) in different CcLGs across 134 lines were observed (Table 4). For instance, in CcLG05, a minimum range of zero to five and in CcLG11 maximum range of four to 25 donor chromosomal segments were found across lines (Table 4). The average coverage per CcLG of the ICPL 87119 genome by the lines was 49.34% with a minimum coverage of 27.97% (CcLG05) to maximum coverage of 60.82% (CcLG04). These lines carried 4.15% (CcLG04) to 17.34% (CcLG01) donor genomes (ICPW 15613) with an average of 10.22% per CcLG. In terms of heterozygosity, lines have shown 33.44% (CcLG08) to 54.99% (CcLG05) with an average of 40.44% heterozygous genome (Table 4).

Table 3 SNPs detected in BP-1 population (ICPL 87119 × ICPW 15613) and BP-2 population (ICPL 87119 × ICPW 29)

CcLGs	BP-1 (ICPL 87119 × ICPW 15613)			BP-2 (ICPL 87119 × ICPW 29)		
	Total	Polymorphic between parents	After filtering	Total	Polymorphic between parents	After filtering
CcLG01	10,809	6403	1684	12,223	4693	1196
CcLG02	21,212	13,831	3513	24,228	9794	2269
CcLG03	16,233	12,278	3070	18,324	6980	1683
CcLG04	7650	5658	1528	8502	3144	828
CcLG05	4051	2984	963	4547	1675	479
CcLG06	12,467	7374	1848	14,141	5431	1413
CcLG07	11,221	8114	2035	12,580	4878	1281
CcLG08	10,215	7715	1990	11,789	4409	1020
CcLG09	6444	4156	1064	7103	2721	690
CcLG10	15,769	9887	2411	17,956	7151	1620
CcLG11	32,416	23,112	5900	36,663	13,925	3573
Total	1,48,487	1,01,512	26,006	1,68,056	64,801	16,052

Table 4 Genome composition of BP-1 population based on SNP markers

CcLGs	# of donor segments			% recurrent genome			% donor genome			% heterozygous genome		
	Min	Max	Average	Min	Max	Average	Min	Max	Average	Min	Max	Average
CcLG01	3	13	7.19	14.41	59.56	38.43	6.05	31.09	17.34	24.71	65.66	44.23
CcLG02	5	23	11.16	26.03	65.56	47.61	4.14	21.23	10.27	20.92	61.92	42.12
CcLG03	2	16	7.31	15.98	78.60	55.62	2.34	16.20	7.24	12.57	78.15	37.14
CcLG04	0	10	1.93	15.90	87.46	60.82	0.00	55.05	4.15	7.67	80.92	35.03
CcLG05	0	5	1.99	9.47	44.96	27.97	0.00	38.34	17.03	28.17	74.08	54.99
CcLG06	4	14	8.51	24.20	62.48	44.69	5.79	24.10	12.47	21.47	64.38	42.83
CcLG07	2	12	5.80	13.16	75.95	54.82	2.42	18.62	8.10	15.82	78.15	37.08
CcLG08	2	11	4.78	16.19	81.57	60.38	2.26	12.78	6.18	8.67	79.21	33.44
CcLG09	1	8	4.39	18.16	67.27	45.00	2.18	26.00	14.00	11.43	61.50	41.00
CcLG10	2	11	5.77	30.82	68.80	53.50	2.24	16.70	8.22	20.03	64.09	38.28
CcLG11	4	25	12.31	28.40	75.33	53.87	2.41	33.30	7.40	17.60	66.81	38.72

BP-2 population (ICPL 87119 × ICPW 29)

A total of 16,052 SNPs from 165 lines were used for inquiry of genome composition in BP-2 population. An average of 1459.27 SNPs were placed in each CcLG. Similar to BP-1 population, a wide range of chromosomal segments in BP-2 population from ICPW 29 (donor parent) in different CcLGs across 165 lines were observed (Table 5). For instance, in CcLG05 a minimum range of zero to four and in CcLG11 a maximum range of 24–42 donor chromosomal segments were found across lines (Table 5). The average coverage per CcLG of the recurrent parent genome by the lines was 40.30% with a minimum coverage of 30.27% (CcLG08) to a maximum coverage of 55.43% (CcLG04). These lines carried 4.55% (CcLG05) to 36.69% (CcLG09) donor genomes (ICPW 29) with an average of 23.63% per CcLG. In terms of heterozygosity, lines have shown 29.79% (CcLG09) to 42.39% (CcLG10) with an average of 36.07% heterozygous genome (Table 5).

SNP-based QTL mapping of complex traits

Phenotyping data collected for 3 years (consecutively 2 years at one location) together with SNP genotyping data were used for QTL analysis in BP-1 and BP-2 populations implemented in CSSL finder and mixed model approach in TASSEL. The significance of QTLs was determined by using a threshold with FDR adjusted P values < 0.05. In BP-1 population, a total of 86 significant QTLs have been detected for all the traits under study with phenotypic variance explained (PVE) ranging from 12 to 21% (Table 6, Supplementary Table S3, Supplementary figure 1). Out of a total of 86 QTLs, four were detected for year 1, i.e., cropping season 2012–2013 [@ location 1 (Patancheru)], 60 for year 2, i.e., cropping season 2013–2014 [@ location 1(Patancheru) and 2 (ARS, Tandur)], and the remaining 22 for year 3 data, i.e., cropping season 2014–2015 [@ location 2 (ARS, Tandur)]. In terms of locations, 47 QTLs were detected at location 1 during years 1 and 2 of phenotyping and 39 QTLs at location 2 during years 2 and 3 of phenotyping. Most of the QTLs were co-localized on CcLG11,

Table 5 Genome composition of BP-2 population based on SNP markers

CcLGs	# of donor segments			% recurrent genome			% donor genome			% heterozygous genome		
	Min	Max	Average	Min	Max	Average	Min	Max	Average	Min	Max	Average
CcLG01	7	17	12.18	23.56	52.24	35.39	13.32	39.14	24.28	16.90	54.28	40.32
CcLG02	15	29	22.12	30.65	56.79	46.12	10.59	33.86	21.83	14.25	50.35	32.05
CcLG03	9	25	17.99	29.53	54.26	40.62	11.33	33.34	24.14	17.78	45.61	35.24
CcLG04	3	10	6.05	23.14	68.48	55.43	5.37	23.45	13.58	10.18	59.96	30.99
CcLG05	0	4	1.46	27.18	73.33	54.28	0.00	23.74	4.55	21.45	60.39	41.17
CcLG06	9	21	15.82	24.44	45.85	33.83	16.00	42.93	30.10	19.86	52.38	36.07
CcLG07	6	17	12.64	23.86	57.08	41.88	10.37	41.56	25.44	16.74	53.71	32.69
CcLG08	8	21	15.15	21.19	45.54	30.27	12.17	52.02	30.71	21.06	57.04	39.01
CcLG09	4	11	7.92	16.50	50.12	33.52	18.39	52.54	36.69	9.75	57.73	29.79
CcLG10	9	19	13.92	19.60	50.04	34.76	12.03	39.63	22.85	25.58	59.48	42.39
CcLG11	24	42	32.85	29.81	51.42	37.22	16.52	34.71	25.70	20.10	48.92	37.08

Table 6 Summary of QTLs detected for target traits in BP-1 and BP-2 populations evaluated under two locations

Trait	Population	Year	Location	CcLG01	CcLG02	CcLG03	CcLG04	CcLG05	CcLG06	CcLG07	CcLG08	CcLG09	CcLG10	CcLG11	Total	PVE_min	PVE_max
Days to 50% flowering	BP-1	2012–2013	Patancheru	0	0	0	0	0	0	0	0	0	0	2	2	0.16	0.16
		2013–2014	ARS, Tandur	1	0	0	0	0	0	0	2	0	0	0	3	0.12	0.14
	BP-2	2013–2014	ARS, Gulbarga	0	1	0	2	0	0	9	0	0	0	2	14	0.12	0.28
		2013–2014	Patancheru	6	4	0	2	1	3	11	0	1	1	0	29	0.13	0.29
		2014–2015	ARS, Gulbarga	0	1	0	0	0	0	3	0	0	0	2	6	0.12	0.19
Days to 75% maturity	BP-2	2013–2014	ARS, Gulbarga	0	0	0	0	0	0	8	0	5	0	2	15	0.11	0.26
		2013–2014	Patancheru	0	0	0	0	0	0	3	0	0	0	2	5	0.12	0.17
		2014–2015	ARS, Gulbarga	0	0	0	0	0	0	2	0	0	0	1	3	0.12	0.13
Grain weight per plot	BP-2	2012–2013	Patancheru	0	0	0	0	0	0	5	0	0	1	0	6	0.12	0.13
		2013–2014	Patancheru	0	0	2	0	0	0	0	0	0	0	0	2	0.13	0.13
Number of pods per plant	BP-1	2013–2014	Patancheru	0	0	0	0	0	0	0	0	0	4	0	4	0.18	0.18
		2014–2015	ARS, Tandur	0	1	0	0	0	0	0	0	0	0	0	1	0.16	0.16
Number of seeds per pod	BP-1	2013–2014	Patancheru	0	0	4	0	0	0	2	0	1	1	7	15	0.14	0.20
Plant height	BP-1	2014–2015	ARS, Tandur	0	0	0	0	0	0	0	0	0	0	4	4	0.17	0.17
	BP-2	2012–2013	Patancheru	0	0	0	1	0	0	0	0	0	3	0	4	0.12	0.14
Primary branches per plant	BP-1	2013–2014	ARS, Tandur	0	0	0	0	0	1	0	0	0	0	0	1	0.16	0.16
		2013–2014	Patancheru	0	0	2	2	2	0	0	1	0	0	17	24	0.15	0.21

Table 6 (continued)

Trait	Population	Year	Location	CcLG01	CcLG02	CcLG03	CcLG04	CcLG05	CcLG06	CcLG07	CcLG08	CcLG09	CcLG10	CcLG11	Total	PVE_min	PVE_max
	BP-2	2014–2015	ARS, Gulbarga	0	0	0	0	0	0	1	0	0	0	0	1	0.13	0.13
Secondary branches per plant	BP-1	2013–2014	ARS, Tandur	0	2	0	0	0	0	0	0	0	0	11	13	0.14	0.15
	BP-2	2014–2015	ARS, Gulbarga	0	0	0	0	0	0	1	0	0	0	0	1	0.11	0.11
Seed weight per plant	BP-1	2014–2015	ARS, Tandur	0	0	0	0	0	0	0	0	0	1	0	1	0.13	0.13
	BP-2	2012–2013	Patancheru	0	0	0	0	0	6	0	3	0	4	0	13	0.13	0.28
		2013–2014	Patancheru	0	0	0	0	0	4	0	4	0	0	0	8	0.12	0.19
Yield per plant	BP-1	2014–2015	ARS, Tandur	0	3	0	0	2	6	0	0	0	2	3	16	0.15	0.17

i.e., 51.2% (44/86). The lowest number of QTLs were found on CcLG01 and CcLG09, with one QTL each. It has been observed that many identified QTLs in 1 year/location were environment specific and showed relatively weak or no associations in the other year/location. In the case of BP-2 population, a total of 107 significant QTLs have been detected for all the traits under study with PVE ranging from 11 to 29% (Table 6, Supplementary Table S4, Supplementary figure 2). Out of a total of 107 QTLs, 23 were detected for year 1, i.e., cropping season 2012–2013 [@ location 1 (Patancheru)], 73 for year 2, i.e., cropping season 2013–2014 [@ location 1(Patancheru) and 2 (ARS, Gulbarga)], and the remaining 11 for year 3 data, i.e., cropping season 2014–2015 [@ location 2 (ARS, Gulbarga)]. In terms of locations, 67 QTLs were detected at location 1 during year 1 and 2 of phenotyping and 40 QTLs at location 2 during year 2 and 3 of phenotyping. Most of the QTLs were co-localized on CcLG07, i.e., 40.18% (43/107). However, the minimum number of QTLs (one) was found on CcLG05. The identified QTLs were also classified as stable (present in more than one locations) and consistent QTLs (present in more than 1 year). For both populations, details on QTLs identified have been explained below.

QTLs for days to 50% flowering

From the total identified QTLs, ~28% QTLs were detected for days to 50% flowering. In BP-1 population, only five QTLs (two in year 1 and three in year 2) with PVE ranging from 12 to 16% were identified. In BP-2 population, 49

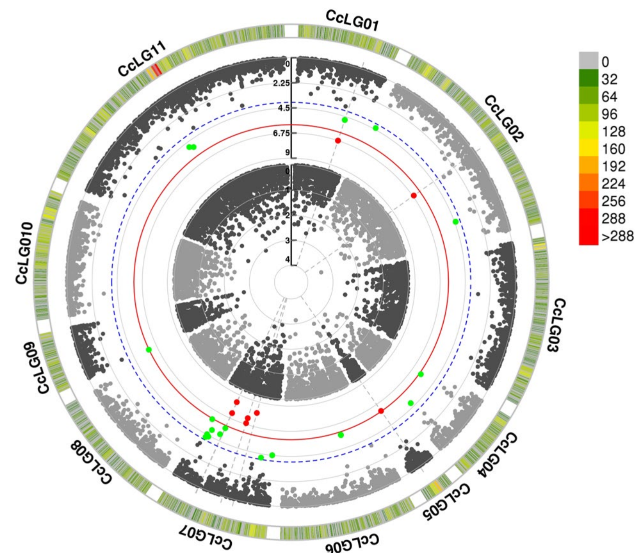


Fig. 2 Manhattan plot of the BP-2 population (ICPL 87119 × ICPW 29) showing genomic regions with QTLs for days to 50% flowering. Gray represents all SNPs. Similarly, green circles represent significant MTAs at 0.05, while red circles represent significant MTAs at 0.01 *P* value

QTLs were detected for days to 50% flowering with PVE ranging from 12 to 29%. From these, 43 QTLs (14 QTLs at ARS, Gulbarga, and 29 QTLs at Patancheru) were present in years 2013–2014 (Fig. 2). Remaining six QTLs were present at ARS, Gulbarga, in year 2014–2015. Across locations 1 and 2, four stable QTLs were identified. From all the 54 QTLs detected, only six QTLs in BP-2 were consistent (Table 7).

QTLs for days to 75% maturity

In BP-2 population, 23 QTLs with PVE ranging from 11 to 26% were identified. From these QTLs, 20 were identified in year 2013–2014 (15 at ARS, Gulbarga, and five at Patancheru). Only three QTLs could be identified in year 2014–2015 at ARS, Gulbarga. Across locations 1 and 2, three stable and consistent QTLs were identified.

QTLs for number of primary branches per plant

A total of 26 QTLs (25 in BP-1 population and one in BP-2 population) with PVE ranging from 13 to 21% were identified. In BP-1 population, all the 25 QTLs were present in year 2013–2014 (24 at Patancheru and one at ARS, Tandur), whereas in BP-2 population, only one QTL was detected in year 2014–2015 at ARS, Gulbarga.

QTLs for number of secondary branches per plant

In BP-1 population 13 QTLs with PVE ranging from 14 to 15% were detected. The maximum number of QTLs, i.e., 11 QTLs, was found on CcLG11 and the remaining two on CcLG02. These QTLs were present in year 2013–2014 at ARS, Tandur. In BP-2 population, only one QTL was present on CcLG07 in year 2014–2015 at ARS, Gulbarga.

QTLs for seed weight per plant

Only one QTL in BP-1 population and 21 QTLs in BP-2 population with PVE ranging from 12 to 28% were detected. From the total QTLs identified in BP-2 population, seven QTLs were found to be consistent across years 2012–2013 and 2013–2014 at Patancheru location.

QTLs associated with more than one trait

Only one SNP (S2_7683317) on CcLG02 was found to be associated with more than one traits in BP-1 population, i.e., with seed yield per plant (PVE ~ 15%) and pods per plant (PVE ~ 16%) (Supplementary Table S3). However, in BP-2 population, 11 QTLs were associated with more than one trait. Importantly, the SNP S7_14185076 (on CcLG07) was concomitantly associated with days to 50% flowering, days

Table 7 Stable and consistent QTLs identified in BP-2 population

Trait	MTAs	Location			
		Patancheru		ARS, Gulbarga	
		2012–2013	2013–2014	2013–2014	2014–2015
Days to 50% flowering	S2_20166271	–	–	Yes	Yes
	S7_14185076	–	Yes	Yes	Yes
	S7_14276210	–	Yes	Yes	-
	S7_14775218	–	Yes	Yes	-
	S7_17156984	–	Yes	Yes	-
	S7_17744567	–	Yes	Yes	-
	S7_6897487	–	Yes	Yes	Yes
	S7_9191635	–	Yes	Yes	Yes
	S11_22788684	–	Yes	Yes	Yes
	S11_24158029	–	-	Yes	Yes
Days to 75% maturity	S7_6897487	–	Yes	Yes	Yes
	S7_9191635	–	Yes	Yes	Yes
	S11_22788684	–	Yes	Yes	Yes
Seed weight per plant	S6_19362124	Yes	Yes	–	–
	S6_19414531	Yes	Yes	–	–
	S6_19532741	Yes	Yes	–	–
	S6_19532818	Yes	Yes	–	–
	S8_62044	Yes	Yes	–	–
	S8_648217	Yes	Yes	–	–
	S8_724256	Yes	Yes	–	–

to 75% maturity, primary branches per plant and secondary branches per plant, which accounted for 11.9–28% PVE. The remaining ten QTLs were found to be associated with only two traits, i.e., days to 50% flowering and days to 75% maturity. Four (S7_14185076, S7_6897487, S7_9191635 and S11_22788684) out of 11 QTLs have shown consistent and stable behavior, whereas other four (S7_14276210, S7_14775218, S7_17156984 and S7_17744567) QTLs have just shown stable behavior. One QTL (S11_24158029) was consistent at location 2. The remaining two QTLs (S7_2665107 and S9_46575) were either location/year specific or associated with different traits in different years (Supplementary Table S5).

Discussion

Yield is the primary requirement for pigeonpea growers. However, it is the most complex trait for the genetic improvement of the crops. In the case of pigeonpea, a number of traits have been found to be involved in realizing the final yield of a plant, which include secondary branches per plant, number of pods per plant, number of seeds per pod, etc. The wild species of pigeonpea have shown many favorable unutilized variations for diseases and insect resistance as well as for the yield-related traits (Saxena et al. 2018a). However, the major concern to utilize the wild species is to selectively transfer useful variations while avoiding linkage drag. In the present scenario, this can be achieved with molecular breeding strategies (Saxena et al. 2015).

Interspecific crosses are important in introducing novel variations into the cultivated genotypes, specifically if the genetic variation for the desired trait is not available in the cultivated germplasm. In pigeonpea, the wild species have been used to transfer alleles for high protein content, insect resistance, new plant type to the cultivated types and also in developing cytoplasmic male sterility system (Saxena et al. 2018a). The yield potential of elite pigeonpea varieties can further be improved when such superior alleles from wild species are transferred into elite cultivars. Thus, backcross populations provide an opportunity for the efficient use of the genetic potential of the wild species (Tanksley and Nelson 1996). The use of ILs derived from interspecific backcross populations with better yield per se or yield-related traits can be one of the key strategies for improving pigeonpea yield. In the present study, two backcross populations developed (Sharma and Upadhyaya, 2016) with two rounds of backcrossing and two cycles of selfing were used for further evaluation. Advanced generations (BC_2F_2) selected for the present study were fully fertile. The presence of a uniform background in backcross lines with different donor segments allowed us to estimate the introgressed and native alleles. Both the populations have shown a large range of variability for the yield-related traits. These populations were used to dissect genetic architecture of the yield-related traits in this study and will serve as a vital genomic resource for pigeonpea breeding and QTL fine mapping.

In the present study, GBS was used for the generation of high-density marker data in both the populations. Imputation of missing alleles in GBS data was carried out with the help of WGRS data available (Kumar et al. 2016) on parental lines following Saxena et al. (2017a, c, 2018a). The number of polymorphic markers identified in specific combinations of parents was highest as compared to previous markers-based studies on biparental mapping populations in pigeonpea (Yang et al. 2011; Gnanesh et al. 2011; Saxena et al. 2012, 2017a, c, 2018a; Bohra et al. 2011, 2012). For instance, a minimum 52 to a maximum of 3941 markers were polymorphic in different parents' combinations of mapping populations in pigeonpea (Bohra et al. 2011; Saxena et al. 2017a). However, between parents of BP-1 and between parents of BP-2, 101,512 and 64,801 markers were polymorphic, respectively. Even after applying stringent filtering criteria, 26,006 and 16,052 markers in BP-1 and BP-2 were retained for further analysis, respectively. The filtered sets of markers were also many folds higher as compared to previously identified total polymorphic markers (filtered or unfiltered) (Yang et al. 2011; Gnanesh et al. 2011; Bohra et al. 2011, 2012; Saxena et al. 2012, 2017a, c, 2018a). This shows the usefulness of GBS approach for developing high-density genotyping data. The genetic divergence between crossing parents used in the present study might have also played an important role in identification of higher number of polymorphic markers. Nonetheless, we would expect more variation, greater disruption of recombination, etc., between parents of BP-1 than between parents of BP-2. This was due to the genetic closeness of *C. cajanifolius* with cultivated pigeonpea as compared to any other wild species (Saxena et al. 2014; Varshney et al. 2017). Consistent with this hypothesis, the higher number of polymorphic markers was identified in parents of BP-1 (101,512 SNPs) as compared to BP-2 (64,801 SNPs). Thus, the high-density genotyping data generated make it possible to define the size and positions of donor and background introgressions and to identify small donor introgressions that might have missed using lower density marker datasets. In both BP-1 (49.34%) and BP-2 (40.30%) populations, the average recurrent genome recovered was far below the expected theoretical percentage in BC_2 generation line (87.5%). The average donor genome in BP-1 (10.22%) was quite similar to the expected value (12.5%). In contrast to BP-1, segments of donor genomes were higher in BP-2 (23.63%). The deviation from the expected values of recurrent and donor genomes in BP-1 and BP-2 may be partly due to genome coverage by GBS which takes random sites of genome (Elshire et al. 2011) and possibly because the populations used for selection during

the development cycles were not large enough to enable us to identify all the combination of target introgression(s) from the donor parent in the genetic background of recurrent parent.

Backcross populations generated using wild species of pigeonpea were not found appropriate for generating the genetic maps due to segregation distortion, which might be the result of the selection of the target traits during the generation advancement. Many alleles have shown segregation distortion across the 11 CcLGs in both BP-1 and BP-2. The severe segregation distortion in backcross populations developed using wild species is also observed in barley (Mora et al. 2016), cotton (Li et al. 2018), rice (Xu et al. 1997), potato (Manrique-Carpintero et al. 2016), etc. Therefore, as proposed by Malosetti et al. (2011) for such populations with segregation distortion, kinship matrix was used for QTL identification. Additionally, the GBS method is prone to large amount of missing data and heterozygous calls (Swarts et al. 2014), which further exaggerate segregation distortion issue. Most of the identified QTLs in BP-1 and BP-2 were environment specific. These findings are consistent with the previous pigeonpea studies (Saxena et al. 2017a, c), in which most of the QTLs for different traits varied between environments. It might be due to the inherent nature of pigeonpea crop which is photoperiod as well as temperature-sensitive crop and whose phenology changes with the environmental conditions. Therefore, it is recommended to breed varieties with high agronomic performance for specific regions. However, stable and/or consistent QTLs were also detected for three traits (days to 50% flowering, days to 75% maturity and seed weight) in BP-2. Moreover, 11 QTLs were associated with more than one trait in BP-2. Interestingly, eight QTLs on CcLG07 were associated with days to 50% flowering and days to 75% maturity. Out of these eight QTLs, three QTLs had negative and five QTLs had positive additive genetic effects. The negative additive effect revealed maternal parent, and positive additive effect showed paternal parent as the source of alleles. In BP-1 and BP-2, 27 and 51 QTLs have shown positive additive genetic effects, respectively, for targeted traits. These QTLs can be considered as novel and useful QTLs as the favorable alleles have been contributed from the wild species. The SNP "S7_14185076" has been found to be associated with four traits, namely days to 50% flowering, days to 75% maturity, primary branches per plant and secondary branches per plant with positive additive effect. These results have shown the importance of interspecific backcross populations in bringing/combining favorable alleles for the target traits, especially for days to 50% flowering and days to 75% maturity. Moreover, few lines from these backcross populations have also been selected based on their performance for yield and disease resistance at different locations (data have not provided) and were incubated in varietal release program of All India Coordinated Research Project on Pigeonpea (AICRP-Pigeonpea), Indian Council of Agricultural Research, India. For instance, ICPL 1028 (AGL-124), ICPL 15031(AGL-141), ICPL 15048 (AGL-870) from BP-1 population were recommended by AICRP- Pigeonpea and AGL-870 from BP-1 population and ICPL 15061 (AGL-1896), ICPL 15067 (AGL-2217), AGL-1640 from BP-2 were recommended in state-level technical program of Professor Jayashankar Telangana State Agricultural University, Telangana, India, for commercial cultivation. Apart from these, AGL-2240 and AGL-1891 from BP-2 population also yielded higher than checks in station trials at Professor Jayashankar Telangana State Agricultural University (data have not provided). In summary, this study has provided several lines for varietal release and useful information for marker-assisted selection (MAS) for yield-related traits. Identified significant QTLs after validation can be used for developing high-yielding pigeonpea lines with synchronized flowering and maturity.

Acknowledgement Authors are thankful for the Department of Agriculture Cooperation & Farmers' Welfare, Ministry of Agriculture & Farmers Welfare, Government of India and United States Agency for International Development (USAID). This work is also partially funded as part of the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives," which is supported by the Government of Norway. The project is managed by the Global Crop Diversity Trust. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes and Dryland Cereals (GLDC). ICRISAT is a member of CGIAR Consortium.

Author contribution statement RKV, RRM and RKS designed the experiments; NM led and RKV, RRM and RKS contributed to develop backcross populations; CVSK, NM, MS, AG, YN and RKS generated phenotyping data; RKV, RKS and SK generated genotyping data; RRD and AR analysed the phenotyping data; SK, PY, JM analysed the genotyping data and performed QTL analysis with support of RKS and RKV. All authors reviewed and approved the submission.

Compliance with ethical standards

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

References

Aruna R, Manohar RD, Reddy LJ, Upadhyaya HD, Sharma HC (2005) Inheritance of trichomes and resistance to pod borer (*Helicoverpa armigera*) and their association in interspecific crosses between cultivated pigeonpea (*Cajanus cajan*) and its wild relative *C. scarabaeoides*. Euphytica 145:247–257

Bermudez L, Urias U, Milstein D, Kamenetzky L, Asis R, Fernie AR, Van Sluys MA, Carrari F, Rossi M (2008) A candidate gene survey of quantitative trait loci affecting chemical composition in tomato fruit. J Exp Bot 59:2875–2890

Bohra A, Mallikarjuna N, Saxena KB, Upadhyaya HD, Vales I, Varshney RK (2010) Harnessing the potential of crop wild relatives through genomics tools for pigeonpea improvement. J Plant Biol 37:1–16

Bohra A, Dubey A, Saxena RK, Penmetsa RV, Poornima KN, Kumar N, Farmer AD, Srivani G, Upadhyaya HD, Gothalwal R, Ramesh R, Singh D, Saxena KB, Kavikishor PB, Town CD, May GD, Cook DR, Varshney RK (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

Bohra A, Saxena RK, Gnanesh BN, Saxena KB, Byregowda M, Rathore A, Kavi Kishor PB, Cook DR, Varshney RK (2012) An intra-specific consensus genetic map of pigeonpea (*Cajanus cajan* (L.) Millsp.) derived from six mapping populations. Theor Appl Genet 125:1325–1338

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

Chen J, Li X, Cheng C, Wang Y, Qin M, Zhu H, Zeng R, Fu X, Liu Z, Zhang G (2014) Characterization of epistatic interaction of QTLs *LH8* and *EH3* controlling heading date in rice. Sci Rep 4:4263

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Gnanesh BN, Bohra A, Sharma M, Byregowda M, Pande S, Wesley V, Saxena RK, Saxena KB, Kavi Kishor PB, Varshney RK (2011) Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. Field Crop Res 123:53–61

Gyenis L, Yun SJ, Smith KP, Steffenson BJ, Bossolini E, Muehlbauer GJ (2007) Genetic architecture of quantitative trait loci associated with morphological and agronomic trait differences in a wild by cultivated barley cross. Genome 50:714–723

ICRISAT (1993) Pigeonpea variety ICPL 87119. Plant material description no.43. Documentation. International Crops Research Institute for the Semi-Arid Tropics. https://oar.icrisat.org/566/1/PMD_43.pdf. Accessed 5 Apr 2019

Jadhav DR, Mallikarjuna N, Sharma HC, Saxena KB (2012) Introgression of *Helicoverpa armigera* resistance from *Cajanus acutifolius*—a wild relative from secondary gene pool of pigeonpea (*Cajanus cajan*). Asian J Agric Sci 4:242–248

Khera P, Pandey MK, Mallikarjuna N, Sriswathi M, Roorkiwal M, Janila P, Sharma S, Shilpa K, Sudini H, Guo B, Varshney RK (2019) Genetic imprints of domestication for disease resistance, oil quality, and yield component traits in groundnut (*Arachis hypogaea* L.). Mol Genet Genom 294:365–378

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

Li P, Kirungu JN, Lu H, Magwanga RO, Lu P, Cai X, Zhou Z, Wang X, Hou Y, Wang Y, Xu Y (2018) SSR-Linkage map of interspecific populations derived from *Gossypium trilobum* and *Gossypium thurberi* and determination of genes harbored within the segregating distortion regions. PLoS ONE 13:e0207271

Lorieux M (2005) CSSL finder: a free program for managing introgression lines. https://mapdisto.free.fr/CSSLFinder/. Accessed 20 Aug 2018

Malosetti M, van Eeuwijk FA, Boer MP, Casas AM, Elía M, Moralejo M, Bhat PR, Ramsay L, Molina-Cano JL (2011) Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. Theor Appl Genet 122:1605–1616

Manrique-Carpintero NC, Coombs JJ, Veilleux RE, Buell CR, Douches DS (2016) Comparative analysis of regions with distorted segregation in three diploid populations of potato. G3 Gen Genom Genet 6:2617–2628

Mora F, Quitral YA, Matus I, Russell J, Waugh R, del Pozo A (2016) SNP-based QTL mapping of 15 complex traits in barley under rain-fed and well-watered conditions by a mixed modeling approach. Front Plant Sci 7:909

Obala J, Saxena RK, Singh V, Sameer Kumar CV, Saxena KB, Tongoona P, Sibiya J, Varshney RK (2019) Development of sequence-based markers for seed protein content in pigeonpea. Mol Gen Genom 294:57–68

Perez-Fons L, Wells T, Corol DI, Ward JL, Gerrish C, Beale MH et al (2014) A genome-wide metabolomic resource for tomato fruit from *Solanum pennellii*. Sci Rep 4:3859

Periyannan S, Moore J, Ayliffe M, Bansal U, Wang X, Huang L, Deal K, Luo M, Kong X, Bariana H, Mago R, McIntosh R, Dodds P, Dvorak J, Lagudah E (2013) The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science 341:786–788

Qiu X, Chen K, Lv W, Ou X, Zhu Y, Xing D, Yang L, Fan F, Yang J, Xu J, Zheng T, Li Z (2017) Examining two sets of introgression lines reveals background-independent and stably expressed QTL that improve grain appearance quality in rice (*Oryza sativa* L.). Theor Appl Genet 130:951–967

Rousseaux MC, Jones CM, Adams D, Chetelat R, Bennett A, Powell A (2005) QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. Theor Appl Genet 111:1396–1408

Saxena KB, Kumar RV, Srivastava N, Bao S (2005) A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. Euphytica 145:289–294

Saxena RK, Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD, Cook DR, Varshney RK (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

Saxena RK, von Wettberg E, Upadhyaya HD, Sanchez V, Songok S, Saxena KB, Kimurto P, Varshney RK (2014) Genetic diversity and demographic history of *Cajanus* spp. Illustrated from genome-wide SNPs. PLoS ONE 9:e88568

Saxena RK, Saxena KB, Pazhamala LT, Patel K, Parupalli S, Sameerkumar CV, Varshney RK (2015) Genomics for greater efficiency in pigeonpea hybrid breeding. Front Plant Sci 6:793

Saxena RK, Kale SM, Kumar V, Parupalli S, Joshi S, Singh VK, Garg V, Das RR, Sharma M, Yamini KN, Ghanta A, Rathore A, Sameer Kumar CV, Saxena KB, Varshney RK (2017a) 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

Saxena RK, Obala J, Sinjushin A, Sameer Kumar CV, Saxena KB, Varshney RK (2017b) Characterization and mapping of *Dt1* locus which co-segregates with *CcTFL1* for growth habit in pigeonpea. Theor Appl Genet 130:1773–1784ss

Saxena RK, Singh VK, Kale SM, Tathineni R, Parupalli S, Kumar V, Garg V, Das RR, Sharma M, Yamini KN, Muniswamy S, Ghanta A, Rathore A, Sameer Kumar CV, Saxena KB, Kavi Kishor PB, Varshney RK (2017c) Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for *fusarium* wilt resistance in pigeonpea. Sci Rep 7:1911

Saxena KB, Saxena RK, Sharma S, Sameer Kumar CV, Sultana R, von Wettberg EB, Varshney RK (2018a) The drivers and methodologies for exploiting wild *Cajanus* genome in pigeonpea breeding. Euphytica 214:222

Saxena RK, Patel K, Sameer Kumar CV, Tyagi K, Saxena KB, Varshney RK (2018b) Molecular mapping and inheritance of restoration of fertility (Rf) in A4 hybrid system in pigeonpea (*Cajanus cajan* (L.) Millsp.). Theor Appl Genet 131:1605–1614

Saxena RK, Rathore A, Bohra A, Yadav P, Das RR, Khan AW, Singh VK, Chitikineni A, Singh IP, Sameer Kumar CV, Saxena KB, Varshney RK (2018c) Development and application of high density Axiom®*CajanusSNP* Array with 56K SNPs to understand the

genome architecture of released cultivars and founder genotypes for redefining future pigeonpea breeding programs. Plant Genome 11:180005. https://doi.org/10.3835/plantgenome2018.01.0005

Schauer N, Semel Y, Roessner U et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol 24:447–454

Schmalenbach I, Pillen K (2009) Detection and verification of malting quality QTLs using wild barley introgression lines. Theor Appl Genet 118:1411–1427

Schmalenbach I, Leon J, Pillen K (2009) Identification and verification of QTLs for agronomic traits using wild barley introgression lines. Theor Appl Genet 118:483–497

Sharma S (2017) Prebreeding using wild species for genetic enhancement of grain legumes at ICRISAT. Crop Sci 57:1132–1144

Sharma S, Upadhyaya (2016) Interspecific hybridization to introduce useful genetic variability for pigeonpea improvement. Ind J Genet 76:496–503

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270:1804–1806

Subudhi PK, De Leon T, Singh PK, Parco A, Cohn MA, Sasaki T (2015) A chromosome segment substitution library of weedy rice for genetic dissection of complex agronomic and domestication traits. PLoS ONE 10:e0130650

Swamy BP, Kaladhar K, Shobha Rani N, Prasad GS, Viraktamath BC, Reddy GA, Sarla N (2012) QTL analysis for grain quality traits in 2 BC_2F_2 populations derived from crosses between *Oryza sativa* cv Swarna and 2 accessions of *O. nivara*. J Hered 103:442–452

Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome. https://doi.org/10.3835/plantgenome2014.05.0023

Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191–203

Thomson MJ, Edwards JD, Septiningsih EM, Harrington SE, McCouch SR (2006) Substitution mapping of *dth1.1*, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. Genetics 172:2501–2514

van der Maesen LJG (1986) *Cajanus* DC and *Alylosia* W. & A (Leguminosae). Agricultural University Wageningen papers 85-4 (1985). Agricultural University Wageningen, the Netherlands, p 222

van der Maesen LJG (1990) Pigeonpea: origin, history, evolution and taxonomy. In: Nene YL, Hall SD, Sheila VK (eds) The pigeonpea. CAB International, Wallingford, pp 44–86

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotechnol 30:83–89

Varshney RK, Saxena RK, Upadhyaya HD, Khan AW, Yu Y, Kim C, Rathore A, Kim D, Kim J, An S, Kumar V, Anuradha G, Yamini KN, Zhang W, Muniswamy S, Kim JS, Penmetsa RV, von Wettberg EB, Datta SK (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49:1082–1088

von Korff M, Wang H, Leon J, Pillen K (2006) AB-QTL analysis in spring barley: II Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. *spontaneum*). Theor Appl Genet 112:1221–1231

VSN International (2017) Genstat for Windows, 19th edn. VSN International, Hemel Hempstead, UK. Web page: Genstat.co.uk

Wan XY, Wan JM, Jiang L, Wang JK, Zhai HQ, Weng JF, Wang HL, Lei CL, Wang JL, Zhang X, Cheng ZJ, Guo XP (2006) QTL analysis for rice grain length and fine mapping of an identified QTL with stable and major effects. Theor Appl Genet 112:1258–1270

Wan X, Weng J, Zhai H, Wang J, Lei C, Liu X, Guo T, Jiang L, Su N, Wan J (2008) Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5* in a recombination hotspot region on chromosome 5. Genet 179:2239–2252

Wang ZY, Second G, Tanksley SD (1992) Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor Appl Genet 83:565–581

Wang J, Wan X, Li H, Pfeiffer W, Crouch J, Wan J (2007) Application of identified QTL-marker associations in rice quality improvement through a design-breeding approach. Theor Appl Genet 115:87–100

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F_2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet 253:535–545

Xu J, Zhao Q, Du P, Xu C, Wang B, Feng Q, Liu Q, Tang S, Gu M, Han B, Liang G (2010) Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.). BMC Genom 11:656

Yadav P, Saxena KB, Hingane A, Kumar CV, Kandalkar VS, Varshney RK, Saxena RK (2019) An "Axiom *Cajanus* SNP Array" based high density genetic map and QTL mapping for high-selfing flower and seed quality traits in pigeonpea. BMC Genom 20:235

Yang SY, Saxena RK, Kulwal PL, Ash GJ, Dubey A, Harpe JDI, Upadhyaya HD, Gothalwal R, Kilian A, Varshney RK (2011) The first genetic map of pigeonpea based on Diversity Arrays Technology (DArT) markers. J Genet 90:103–109

Yun SJ, Gyenis L, Bossolini E, Hayes PM, Matus I, Smith KP, Steffenson BJ, Tuberosa R, Muehlbauer GJ (2006) validation of quantitative trait loci for multiple disease resistance in barley using advanced backcross lines developed with a wild barley. Crop Sci 46:1179–1186

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.