

# SCIENTIFIC REPORTS



OPEN

## Genotyping-by-sequencing of three mapping populations for identification of candidate genomic regions for resistance to sterility mosaic disease in pigeonpea

Rachit K. Saxena<sup>1</sup>, Sandip M. Kale<sup>1</sup>, Vinay Kumar<sup>1</sup>, Swathi Parupali<sup>1</sup>, Shourabh Joshi<sup>2</sup>, Vikas Singh<sup>1</sup>, Vanika Garg<sup>1</sup>, Roma R. Das<sup>1</sup>, Mamta Sharma<sup>1</sup>, K. N. Yamini<sup>2</sup>, Anuradha Ghanta<sup>2</sup>, Abhishek Rathore<sup>1</sup>, C. V. Sameerkumar<sup>1</sup>, K. B. Saxena<sup>1</sup> & Rajeev K. Varshney<sup>1,3</sup>

Sterility mosaic disease (SMD) is one of the serious production constraints that may lead to complete yield loss in pigeonpea. Three mapping populations including two recombinant inbred lines and one  $F_2$ , were used for phenotyping for SMD resistance at two locations in three different years. Genotyping-by-sequencing approach was used for simultaneous identification and genotyping of SNPs on above mentioned populations. In total, 212,464, 89,699 and 64,798 SNPs were identified in ICPL 20096  $\times$  ICPL 332 (PRIL\_B), ICPL 20097  $\times$  ICP 8863 (PRIL\_C) and ICP 8863  $\times$  ICPL 87119 ( $F_2$ ) respectively. By using high-quality SNPs, genetic maps were developed for PRIL\_B (1,101 SNPs; 921.21 cM), PRIL\_C (484 SNPs; 798.25 cM) and  $F_2$  (996 SNPs; 1,597.30 cM) populations. The average inter marker distance on these maps varied from 0.84 cM to 1.65 cM, which was lowest in all genetic mapping studies in pigeonpea. Composite interval mapping based QTL analysis identified a total of 10 QTLs including three major QTLs across the three populations. The phenotypic variance of the identified QTLs ranged from 3.6 to 34.3%. One candidate genomic region identified on CcLG11 seems to be promising QTL for molecular breeding in developing superior lines with enhanced resistance to SMD.

Pigeonpea [*Cajanus cajan* (L.) Millspaugh] is the sixth most important legume crop grown predominantly in the tropical and sub-tropical regions of the world. It is a protein rich (20–23%), versatile crop and thus an important source of income for smallholder farmers. Along with that, it helps to increase soil fertility by fixing atmospheric nitrogen. India ranks first in both area under cultivation (5.06 Mha) and in production (3.29 Mt) (<http://faostat3.fao.org/home/>, as of August 2016), however, over the last six decades, there is not much increase in the crop yield and its productivity has remained less than one ton/ha. This is predominantly because of susceptibility of majority of cultivated varieties to various biotic stresses such as sterility mosaic disease (SMD) and *Fusarium* wilt (FW). SMD is caused by pigeonpea sterility mosaic virus (PPSMV) that is transmitted by a mite (*Aceria cajani*) and FW is caused by a fungus *Fusarium udum*. Annual losses due to SMD and FW diseases have been reported to be US\$ 113 million<sup>1</sup>. SMD is characterized by mosaic symptoms on leaves, excessive vegetative growth and cessation of reproductive organs resulting in the sterility of the plant<sup>2</sup>. Since its emergence, SMD has huge negative effect on pigeonpea productivity<sup>3</sup>. Although, application of sprays in order to control mite populations can limit spread of the disease, identification and introgression of genomic segments attributing disease resistance through genomics-assisted breeding (GAB) programmes would be an important strategy for development of disease resistant pigeonpea varieties.

<sup>1</sup>International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, 502 324, India. <sup>2</sup>Institute of Biotechnology, Professor Jayshankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad, 500 030, India. <sup>3</sup>School of Plant Biology and Institute of Agriculture, The University of Western Australia, Crawley, WA, 6009, Australia. Correspondence and requests for materials should be addressed to R.K.V. (email: [r.k.varshney@cgiar.org](mailto:r.k.varshney@cgiar.org))

Prior to availability of the pigeonpea draft genome sequence, pigeonpea crop was considered as an orphan crop as very few genetic and genomic resources were available<sup>4</sup>. Moreover, low level of genetic diversity in the primary gene pool further made it challenging to develop genomic resources in pigeonpea<sup>4,5</sup>. As a result, limited numbers of genetic maps were available in pigeonpea<sup>6</sup>. Although efforts were made to identify molecular markers associated with SMD resistance<sup>5,7,8</sup>, so far only one study has reported quantitative trait loci (QTLs) for SMD resistance in pigeonpea<sup>9</sup>. However, the availability of the genome sequence coupled with advances in the next generation sequencing (NGS) technology has enhanced the pace of pigeonpea genomics and genetics research<sup>10</sup>. As a result, large-scale genomic resources have been developed in pigeonpea<sup>11</sup>. Also ample information on single nucleotide polymorphism (SNP) has been generated in pigeonpea lines from the cultivated gene pool. SNPs are more advantageous than many other markers due to their high abundance in the genome, ease of automation in genotyping and ubiquitous distribution throughout the genome. Very recently, sequencing-based trait mapping approach has been utilized for the identification of SNPs associated with resistance to SMD and FW resistance through re-sequencing of extreme pools of selected RILs along with the resistant parent<sup>12</sup>. In addition, a number of new approaches such as restriction site associated DNA sequencing (RADseq)<sup>13</sup>, genotyping-by-sequencing (GBS)<sup>14</sup>, etc. have also been developed for simultaneous SNP discovery and genotyping of thousands of samples in cost effective manner in several crops. Among different approaches, GBS has been considered as the simplest and cost effective approach because of its simple library preparation procedure and high level multiplexing capacity<sup>14</sup>. GBS approach has been used widely for diversity studies, trait mapping and genome-wide association studies (GWAS)<sup>15</sup> in a number of crops like chickpea<sup>16</sup>, common bean<sup>17</sup>, wheat<sup>18</sup>, cabbage<sup>19</sup>, etc.

In the present study, GBS approach was used for SNP identification and genotyping of three different mapping populations segregating for SMD. The identified SNPs were used in construction of high density genetic maps for three populations. The SNP genotyping data together with multi-location and multi-year phenotyping data on these populations were used to identify QTLs governing SMD resistance in pigeonpea.

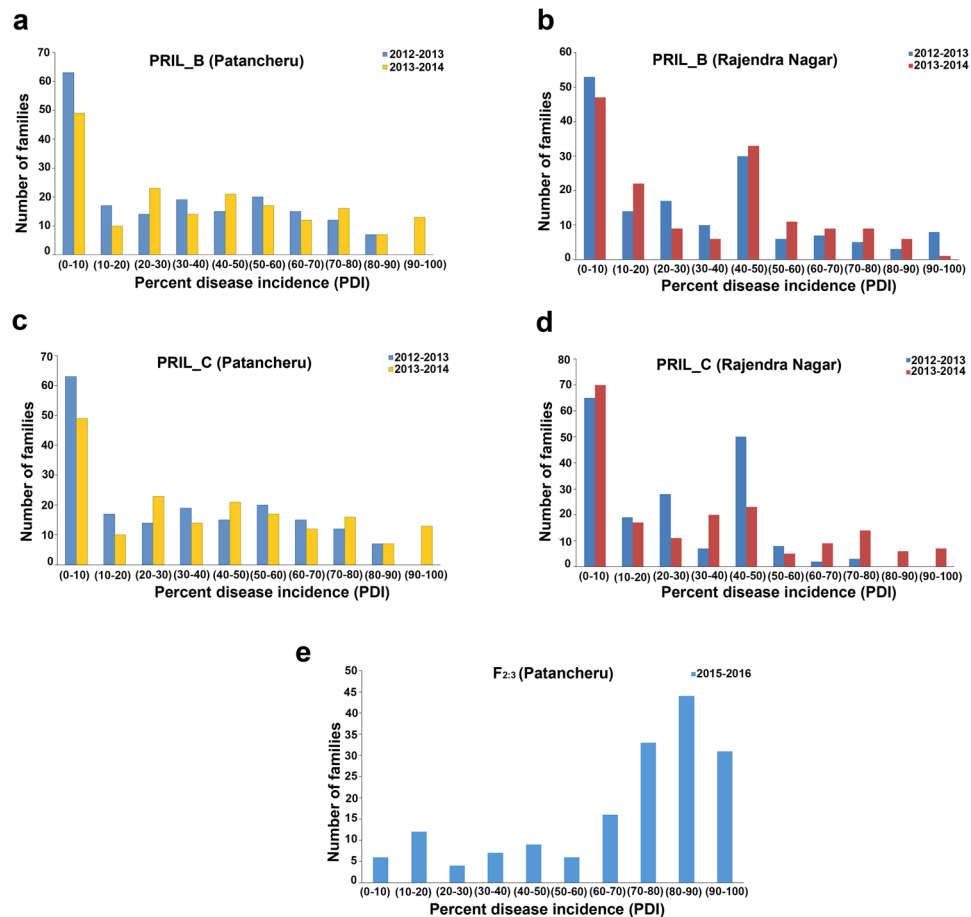
## Results

**Phenotypic evaluation.** The cumulative disease occurrence at 90 days of sowing was considered as the final percent diseases incidence (PDI). The PDI on the PRIL\_B population ranged from 0 to 100 at both Patancheru and Rajendranagar, Hyderabad location (Supplementary Table 1). The parents of the mapping population showed contrasting phenotypic reaction for SMD resistance. ICPL 20096 exhibited mean PDI score of 2.08 and 5.00 at Patancheru and Rajendranagar, Hyderabad locations respectively, whereas susceptible parent ICPL 332 showed PDI score of 100 at both Patancheru and Rajendranagar, Hyderabad locations. Similarly, in PRIL\_C, the PDI score ranged from 0 to 100 at both Patancheru and Rajendranagar, Hyderabad locations (Supplementary Table 1) wherein the mean PDI score of resistant parent ICPL 20097 was 0 and of susceptible parent ICP 8863 was 100 and 66.67 at Patancheru and Rajendranagar, Hyderabad locations respectively. In the case of F<sub>2,3</sub> population, the mean PDI score ranged from 5.1 to 98.2, whereas the parents exhibited 100 (ICP 8863) and 0 (ICPL 87119) PDI score at Patancheru location. The frequency distributions corresponding to each of the populations are presented in the histograms (Fig. 1), which indicate the presence of several genes governing SMD resistance. Higher numbers of lines were found to be resistant in case of both the PRILs whereas, susceptibility was found to be predominant in F<sub>2,3</sub> population, indicating the role of different genetic factors contributing to SMD resistance in pigeonpea. The arcsine transformed means of SMD were considered for QTL analysis in order to avoid distortion abnormalities.

**SNP discovery and genotyping.** High throughput sequencing of three mapping populations and parental lines on HiSeq 2500 platform provided large amount of sequencing data. For instance, 34.37 Gb (340.30 million reads; Supplementary Table 2), 42.32 Gb (419.04 million reads; Supplementary Table 3) and 28.76 Gb (284.77 million reads; Supplementary Table 4) data were generated for PRIL\_B, PRIL\_C and F<sub>2</sub> respectively. For all three populations, samples having less than 80 Mb data were discarded from further analysis in order to reduce missing data error. The number of reads obtained within a given population varied from 0.80 million to 5.46 million, 0.80 million to 5.43 million and 0.84 million to 8.19 million in PRIL\_B, PRIL\_C and F<sub>2</sub> respectively (Supplementary Fig. 1). Further, filtered sequencing reads in each population were analysed separately for SNP identification using TASSEL-GBS pipeline. As a result 212,464; 89,699 and 64,798 SNPs were identified in PRIL\_B, PRIL\_C and F<sub>2</sub> populations, respectively. The SNP density on each pseudomolecule or chromosome has been represented using circos diagram (Fig. 2).

Considering the different levels of heterozygosity in PRILs and F<sub>2</sub> populations, different criteria were used to filter SNPs identified in these populations. For PRIL\_B and PRIL\_C, the SNPs with  $\geq 50\%$  missing data and minor allele frequency (MAF) of  $\leq 0.3$  were filtered out. Additionally, lines in PRILs having more than 50% missing data were removed. Further, imputation of missing data was carried out using FSFHap algorithm<sup>20</sup> implemented in TASSEL v4.0. This resulted in the identification of 1,789 SNPs in 153 lines of PRIL\_B and 507 SNPs in 182 lines of PRIL\_C (Table 1). The number of SNPs identified ranged from 26 (CcLG05) to 427 (CcLG11) in PRIL\_B while those in PRIL\_C ranged from 2 (CcLG05) to 100 (CcLG02) (Table 1). In the case of the F<sub>2</sub> population, 3,941 SNPs with contrasting alleles in parental genotypes and having  $< 30\%$  missing data were retained for further study (Table 1). Identified SNPs were also scattered in all CcLGs with maximum 908 SNPs on CcLG11 and minimum 87 SNPs on CcLG05 in F<sub>2</sub> population.

**Genetic maps for three populations.** The imputed SNPs from PRIL\_B, PRIL\_C and F<sub>2</sub> were used for constructing high-density genetic maps. In the case of PRIL\_B population, 1,101 (61.54%) SNPs out of 1,789 SNPs were mapped on 11 linkage groups, while 484 (95.46%) SNPs out of 507 SNPs were mapped in the case of PRIL\_C population (Table 1). This resulted in the construction of genetic maps of lengths 921.20 cM (Fig. 3; Table 1) and 798.25 cM (Fig. 4; Table 1) sizes in PRIL\_B and PRIL\_C respectively. The length of individual linkage groups varied from 30.26 cM (CcLG05) to 118.40 cM (CcLG11) in the PRIL\_B population while 15.06 cM

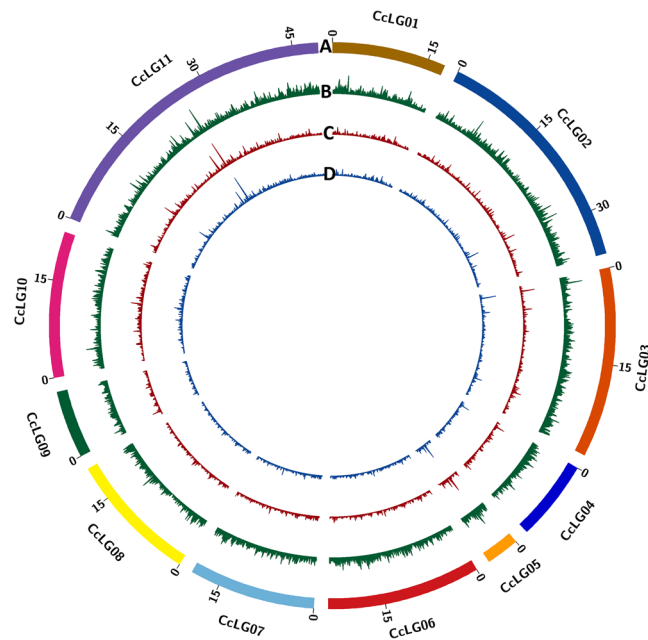


**Figure 1.** Frequency distribution of percent disease incidence (PDI) for Patancheru SMD isolate in various populations at different locations and years. The disease scoring was done on the basis of percentage of affected plants wherein 0% means complete resistance while 100% means complete susceptibility to SMD. The PDI was monitored for two consecutive years (2012–2013, 2013–2014) in ICPL 20096 × ICPL 332 (PRIL\_B) and ICPL 20097 × ICP 8863 (PRIL\_C) population while for one year (2015–2016) in ICP 8863 × ICPL 87119 (F<sub>2</sub>) population. The PDI was divided into 10 categories and number of families falling in each category were plotted as bar plot. The PDI in ICPL 20096 × ICPL 332 (PRIL\_B) population at Patancheru location and Rajendranagar, Hyderabad location is shown in **a** and **b**. Figure **c** and **d** represent PDI in ICPL 20097 × ICP 8863 (PRIL\_C) population at Patancheru and Rajendranagar, Hyderabad location respectively, while, the Figure **e** represents the PDI in ICP 8863 × ICPL 87119 (F<sub>2</sub>) population at Patancheru location.

(CcLG09) to 148.51 cM (CcLG02) in the case of PRIL\_C population (Table 1). A maximum of 382 (CcLG11) and 99 (CcLG02) SNPs, while minimum of 10 (CcLG05) and 2 (CcLG05) SNPs were mapped in PRIL\_B and PRIL\_C populations, respectively. In the case of PRIL\_B population, an average inter marker distance per CcLG ranged from 0.31 cM (CcLG11) to 3.09 cM (CcLG01) with overall average of 0.84 cM (Table 1). Similarly, in PRIL\_C, an average inter marker distance per CcLG ranged from 0.89 cM (CcLG07) to 8 cM (CcLG05) with overall average of 1.65 cM (Table 1).

As segregation distortion of SNPs was the major issue in F<sub>2</sub> population, the strategy described by Chen *et al.*<sup>21</sup> was used to filter out highly distorted markers. In this way, 996 SNPs showing expected segregation at *P* value of  $<10^{-9}$  were retained and used for construction of a genetic map of 1597.3 cM (Fig. 4; Table 1). Out of 996 SNPs, maximum 321 SNPs were mapped on CcLG11 while minimum five SNPs were mapped on CcLG09. An average inter marker distance per CcLG ranged from 0.63 cM on CcLG11 to 8.78 cM on CcLG09 with an overall average of 1.60 cM (Table 1).

**QTLs for SMD resistance.** Phenotypic data collected for two consecutive years at two different locations for PRIL\_B and PRIL\_C while those collected for F<sub>2:3</sub> population (for one year and one location) along with respective genotypic data were used for identification of QTLs for SMD resistance. A single consistent (appeared in more than 1 year/season) and stable (appeared in more than one location) QTL, *qSMD11.1* on CcLG11, explaining about 13% phenotypic variation (%PV) was identified in the case of PRIL\_B population (Table 2; Fig. 3). In the case of PRIL\_C, four QTLs were identified, out of which three were present on CcLG02 and one was present on CcLG10 (Table 3; Fig. 4). The QTL on CcLG10 (*qSMD10.1*) was found stable as well as consistent. The



**Figure 2.** Genome-wide distribution of SNPs identified in ICPL 20096 × ICPL 332 (PRIL\_B), ICPL 20097 × ICP 8863 (PRIL\_C) and ICP 8863 × ICPL 87119 (F<sub>2</sub>) populations in pigeonpea. The number of SNPs identified within 100 Kb interval were calculated and plotted as a smooth line curve. The height of the curve is proportional to the number of SNPs within that 100 Kb interval. (A) Pigeonpea pseudomolecules, labelled as CcLG01 to CcLG11 and each pseudomolecule is shown in different colour. The numbers on arches represent the scale for the size of pseudomolecule in Mb. (B) Genome-wide distribution of SNPs identified in ICPL 20096 × ICPL 332 (PRIL\_B) population (C) Genome-wide distribution of SNPs identified in ICPL 20097 × ICP 8863 (PRIL\_C) population and (D) Genome-wide distribution of SNPs identified in ICP 8863 × ICPL 87119 (F<sub>2</sub>) population in pigeonpea.

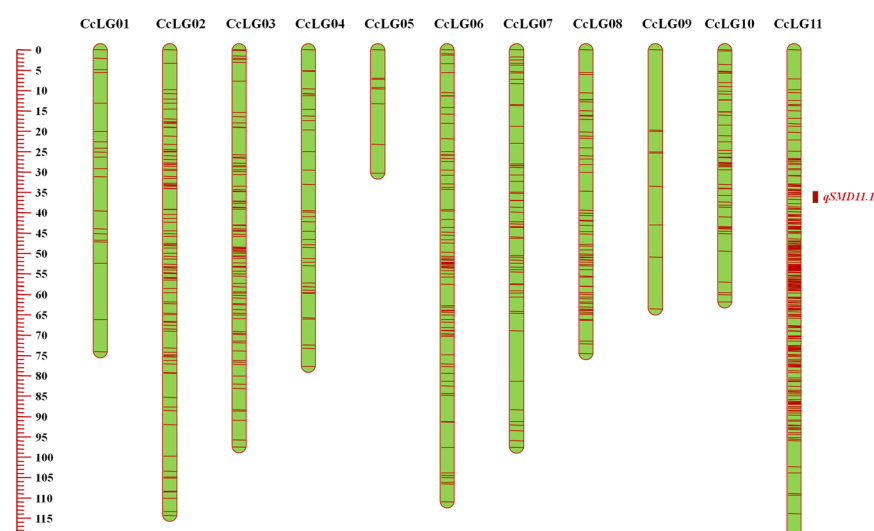
%PV explained by each QTL varied from 6.64 to 7.61 in which the maximum %PV was explained by the QTL from CcLG10 (*qSMD10.1*). In the case of F<sub>2</sub> population, a total of five QTLs one each on CcLG03 (*qSMD3.1*) and CcLG07 (*qSMD7.1*) while three on CcLG11 (*qSMD11.2*, *qSMD11.3*, *qSMD11.4*) were identified (Table 4; Fig. 5). The %PV explained by the QTLs in F<sub>2</sub> population varied from 3.6 to 34.3, wherein the QTL on CcLG03 (*qSMD3.1*) showed maximum 34.3% PV. However, no common QTL was identified among three populations for SMD resistance. Physical positions of SNPs in genome flanking the QTL region were used to determine the size of each QTL in Mb. QTLs sizes range from 1.27 to 16.66 Mb wherein the smallest QTL (*qSMD10.1*) was identified on CcLG10 in PRIL\_C population while largest QTL (*qSMD2.1*) was identified on CcLG02 in PRIL\_C population (Supplementary Table 5). Moreover, physical positions were also used to determine the number of genes present within each QTL region and minimum 125 genes were identified within QTL, *qSMD10.1*, on CcLG10 while maximum 1,556 genes were identified within QTL, *qSMD2.1*, on CcLG02 in PRIL\_C population (Supplementary Table 5). Refinement of these QTLs regions with high density genotyping would help to identify candidate genes for SMD resistance in pigeonpea. In parallel, the positions of QTLs identified in earlier study<sup>9</sup> were also determined and compared with the QTLs identified in present study to identify common QTLs. For this, the BAC-end sequences (BES)<sup>22</sup> of BES-SSR markers were extracted from NCBI database and searched against the pigeonpea reference genome using blastn program. The location of the best hit obtained for each sequence was considered as location of respective BES-SSR marker. Two QTLs identified on LG 9 in ICP 8863 × ICPL 20097 were found to be located on CcLG03 pseudomolecule (Supplementary Table 6) while, the QTLs identified in TTB 7 × ICP 7035 population were found to be distributed on unassembled scaffolds of pigeonpea reference genome (Supplementary Table 6). However, no common QTL has been identified between earlier and present study.

## Discussion

Sterility mosaic disease (SMD) also known as “green plague of pigeonpea”, is one of the severe threats in pigeonpea production causing up to 95% yield loss<sup>3</sup>. Development of SMD resistant pigeonpea varieties is an important strategy for sustainable agriculture development. Marker assisted selection (MAS) has proved its importance in development of improved cultivars<sup>23–25</sup>, however, availability of sufficient genetic and genomic resources is the basic pre-requisite. Although, some efforts were taken in recent past to develop genetic and genomic resources in pigeonpea, limited or no success has been achieved to develop disease resistant pigeonpea cultivars through MAS. Till date, one QTL study<sup>9</sup> employing SSR based linkage map and one NGS based Seq-BSA<sup>12</sup> study has been carried out to identify QTLs/candidate genes for SMD resistance in pigeonpea. In this regard, genetic maps with reasonable marker-density and genomic regions associated with SMD resistance identified in the present study are important milestones in pigeonpea breeding.

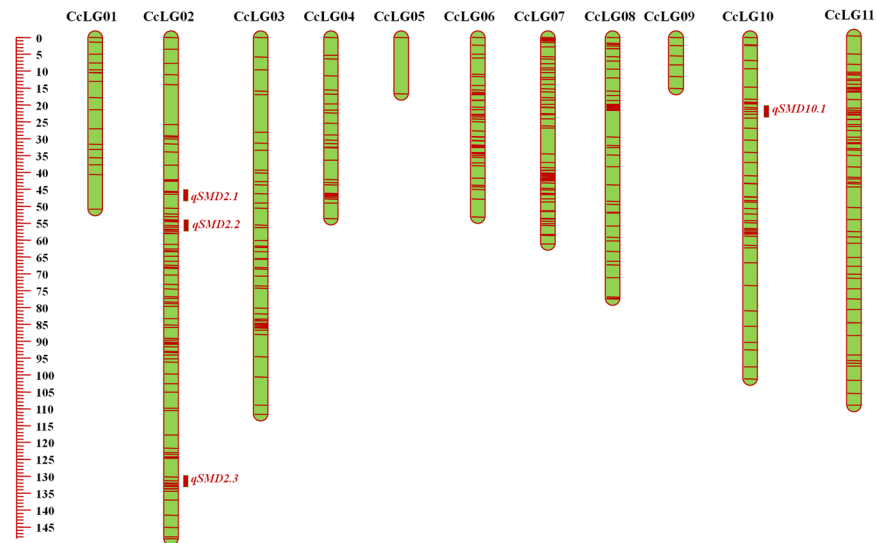
Chromosome	ICPL 20096 × ICPL 332 (PRIL_B)				ICPL 20097 × ICP 8863 (PRIL_C)				ICP 8863 × ICPL 87119 (F <sub>2</sub> )			
	SNPs	SNPs mapped	Distance (cM)	Avg. inter marker distance (cM)	SNPs	SNPs mapped	Distance (cM)	Avg. inter marker distance (cM)	SNPs	SNPs mapped	Distance (cM)	Avg. inter marker distance (cM)
CcLG01	75	24	74.07	3.09	27	21	50.88	2.42	117	40	120.00	3.00
CcLG02	219	139	114.43	0.82	100	99	148.51	1.50	641	178	178.10	1.00
CcLG03	172	123	97.53	0.79	48	48	111.63	2.33	386	81	167.00	2.06
CcLG04	87	52	77.68	1.49	41	27	53.70	1.99	251	19	114.40	6.02
CcLG05	26	10	30.26	3.03	2	2	16.00	8.00	87	18	88.00	4.89
CcLG06	205	119	111.05	0.93	51	51	53.21	1.04	336	75	187.00	2.49
CcLG07	148	69	97.64	1.42	69	69	61.18	0.89	323	74	173.00	2.34
CcLG08	174	94	74.65	0.79	46	45	77.49	1.72	282	49	144.00	2.94
CcLG09	74	22	63.60	2.89	8	7	15.06	2.15	200	5	43.90	8.78
CcLG10	182	67	61.91	0.92	52	52	101.23	1.95	410	136	179.90	1.32
CcLG11	427	382	118.39	0.31	63	63	109.36	1.74	908	321	202.00	0.63
<b>Total</b>	<b>1789</b>	<b>1101</b>	<b>921.21</b>	<b>0.84</b>	<b>507</b>	<b>484</b>	<b>798.25</b>	<b>1.65</b>	<b>3,941</b>	<b>996</b>	<b>1597.30</b>	<b>1.60</b>

**Table 1.** Features of genetic maps in ICPL 20096 × ICPL 332 (PRIL\_B), ICPL 20097 × ICP 8863 (PRIL\_C) and ICP 8863 × ICPL 87119 (F<sub>2</sub>) populations in pigeonpea.



**Figure 3.** Genetic and QTL map comprising 1,101 SNPs and spanning 921.21 cM in ICPL 20096 × ICPL 332 (PRIL\_B) population in pigeonpea. The scale on left side represents map distance in cM. The eleven linkage groups are shown as vertical bars and each horizontal line on the bar represent single SNP marker. Aggregation on horizontal lines indicate higher marker density on that particular linkage group. The single, consistent QTL (*qSMD11.1*) identified for SMD resistance on CcLG11 is shown by coloured rectangle.

**GBS for faster mapping in pigeonpea.** Earlier markers based studies reported exceptionally low level of polymorphism in cultivated pigeonpea genotypes<sup>9,22,26</sup>. Therefore, limited numbers of markers were available for genetic mapping which in turn resulted in producing very low density genetic maps in cultivated pigeonpea<sup>6,9</sup>. For instance, the number of SSR markers in individual component genetic maps ranged from 59 to 140 while the consensus map constructed using six component maps had only 339 markers<sup>6</sup>. These studies warranted the use of high-throughput genotyping for genetic mapping studies in pigeonpea. In fact, the importance of SNPs in genetic mapping in pigeonpea was already demonstrated by Saxena *et al.*<sup>27</sup> wherein a genetic map with 875 PKAMs (pigeonpea KASPar assay markers) was developed. However, GBS approach can generate thousands of SNPs in faster and cost effective manner. So far, the GBS approach has not been used in the case of pigeonpea. Therefore, this is the first report of use of GBS for simultaneous identification and genotyping of SNPs in pigeonpea. Using this approach, although several thousand SNPs were identified between parental genotypes of each of three populations, stringent criteria for considering high-quality SNPs for genetic mapping restricted construction of genetic maps with few hundred loci. However, presence of high amount of missing data is an inherent problem of GBS approach<sup>14</sup> and similar results have been reported in most of the GBS based linkage mapping studies<sup>16,19</sup>. Nevertheless, sequencing at higher depth and optimization of sequencing and analysis processes, may provide larger number of SNPs useful for constructing genetic maps in near future.



**Figure 4.** Genetic and QTL map constructed using 484 SNPs and of 798.25 cM length in ICPL 20097 × ICP 8863 (PRIL\_C) population in pigeonpea. The scale on left side represents map distance in cM. The eleven linkage groups are shown as vertical bars and each horizontal line on the bar represent single SNP marker. Aggregation on horizontal lines indicate higher marker density on that particular linkage group. Three and one rectangles on right side of CcLG02 and CcLG10 represent the four QTLs identified for SMD resistance in PRIL\_C population.

**Better genetic maps for pigeonpea.** Three genetic maps containing 484, 996 and 1,101 SNP markers with a map distance of 798.27, 1,597.30 and 921.20 cM were constructed using PRIL\_C, F<sub>2</sub> and PRIL\_B populations, respectively. Number of markers mapped in individual genetic maps was higher than the earlier SSR based genetic maps in pigeonpea<sup>6,9</sup>. The genetic map for PRIL\_C population contained less markers (484) compared to other maps constructed in the present study. This may be partly attributed to the low level of diversity present between the parental lines (ICPL 20097 and ICP 8863). Earlier study using SSR markers also reported low polymorphism between ICPL 20097 and ICP 8863 genotypes<sup>9</sup>. The average inter-marker distance varied from 0.84 cM (PRIL\_B) to 1.65 cM (PRIL\_C) and were 5 to 6 fold lesser than those reported in SSR based genetic mapping studies<sup>6,9</sup>. Furthermore, the average inter marker distance observed in PRIL\_B (0.84 cM) was the lowest in all the genetic mapping studies conducted in pigeonpea so far<sup>6,9,27</sup> which indicates better saturation of available genetic maps. These better maps developed in the present study will be an important resource for not only QTL identification but also in QTL cloning and identification of candidate genes. As a very small number of common markers (7) were found between genetic maps developed for the PRIL\_B and PRIL\_C populations, a consensus map couldn't be developed. This might be due to the steps followed in the GBS protocol such as restriction digestion, amplicon enrichment and sequencing of genomic DNA were completely random and therefore probability of enrichment and sequencing of similar fragment from the entire sample was very less. Combined analysis of all the GBS data and then imputation of missing data will help to reduce above issues; however, the imputation methods require extensive and accurate haplotype data and could only be possible in very few plant species such as Arabidopsis, maize, rice etc. where large scale sequencing data on different set of lines are available. Nevertheless, efforts are being taken to generate such data in pigeonpea and imputation of large amount of missing data could be possible in near future with refined genome assembly.

**Distinct sources of SMD resistance in pigeonpea.** Genetic maps and multi-location phenotypic data were utilized for the identification of QTLs for SMD resistance in pigeonpea. A total of 10 QTLs, one in PRIL\_B, four in PRIL\_C and five in F<sub>2</sub> populations, were identified. Earlier, bulk segregant analysis using AFLP<sup>5</sup>, RAPD and SCAR<sup>7</sup> markers to identify markers linked to SMD resistance in pigeonpea. Further, microsatellite markers were also used for single marker analysis<sup>8</sup> to identify markers linked to SMD resistance in pigeonpea. However, none of the study reported major QTLs (with % PV ≥ 10) for SMD. Further, QTL analysis was carried out using early generation (F<sub>2</sub> population) of PRIL\_C and identified two QTLs for resistance to Patancheru SMD isolate on linkage group 9 (LG 9)<sup>9</sup>. On the other hand, four QTLs on two different linkage groups were identified in PRIL\_C population in present study. However, no common QTL was identified between earlier and present study. This could be because of several reasons such as population type, number of markers used, robustness of phenotyping, etc. For instance, the QTLs reported in earlier study were based on the early generation (F<sub>2</sub>) population, using very less number of markers (120) and using only one season phenotypic data. In contrast, four fold more markers (484) and phenotypic data collected at two different locations (Patancheru and Rajendranagar, Hyderabad) for two consecutive years have been utilized for QTL identification in genetically stable RIL population (PRIL\_C). Further, whole genome re-sequencing (WGRS) based bulk segregant analysis using PRIL\_B population has identified a total of four SNPs, two on CcLG02 and one each on CcLG08 and CcLG11, associated with SMD resistance

QTL	Location	Year	Chromosome	Position (cM)	Marker interval	QTL size (cM)	% PV* explained	Additive effect	LOD
<i>qSMD11.1</i>	Patancheru	2012–2013	CcLG11	36.81	S11_30004779–S11_36027138	1.71	9.89	–14.46	3.07
	Patancheru	2013–2014	CcLG11	36.81	S11_30004779–S11_36027138	1.71	9.46	–13.80	3.36
	Rajendranagar, Hyderabad	2012–2013	CcLG11	36.81	S11_30004779–S11_36027138	1.71	12.99	–12.98	4.62
	Rajendranagar, Hyderabad	2013–2014	CcLG11	36.81	S11_30004779–S11_36027138	1.71	9.54	–10.06	3.39

**Table 2.** Results of QTL analysis in ICPL 20096 × ICPL 332 (PRIL\_B) population in pigeonpea. \*%Phenotypic variation.

QTL	Location	Year	Chromosome	Position (cM)	Marker interval	QTL size (cM)	% PV* explained	Additive effect	LOD
<i>qSMD2.1</i>	Patancheru	2012–2013	CcLG02	45.91	S2_16997696–S2_336378	4.4	6.64	–0.04	2.68
<i>qSMD2.2</i>	Patancheru	2012–2013	CcLG02	55.61	S2_959921–S2_17370903	2.9	7.36	–0.05	2.62
<i>qSMD2.3</i>	Patancheru	2012–2013	CcLG02	130.31	S2_33097589–S2_36345936	10.9	6.74	–0.03	2.81
<i>qSMD10.1</i>	Rajendranagar, Hyderabad	2013–2014	CcLG10	22.81	S10_13987842–S10_15260172	5.061	7.41	0.10	3.34
	Rajendranagar, Hyderabad	2013–2014	CcLG10	22.81	S10_13987842–S10_15260172	5.061	7.61	8.10	3.47
	Patancheru	2013–2014	CcLG10	23.91	S10_13987842–S10_15260172	5.061	6.80	8.24	2.57

**Table 3.** Summary of QTL analysis in ICPL 20097 × ICP 8863 (PRIL\_C) population in pigeonpea. \*%Phenotypic variation.

QTL	Location	Year	Chromosome	Position (cM)	Marker interval	QTL size (cM)	% PV* explained	Additive effect	Dominant	LOD
<i>qSMD3.1</i>	Patancheru	2015–2016	CcLG03	98.41	S3_18837756–S3_5324938	1.4	34.3	–18.4	25.7	2.8
<i>qSMD7.1</i>	Patancheru	2015–2016	CcLG07	162.21	S7_14725598–S7_7547477	9.7	14	17.7	–1.5	2.9
<i>qSMD11.2</i>	Patancheru	2015–2016	CcLG11	6.11	S11_22689650–S11_24071417	1.8	3.6	17.2	10.8	2.9
<i>qSMD11.3</i>	Patancheru	2015–2016	CcLG11	46.61	S11_16365686–S11_5757417	0.9	24.2	–14.8	16.8	5.8
<i>qSMD11.4</i>	Patancheru	2015–2016	CcLG11	104.71	S11_11799692–S11_5757598	0.4	5.2	1.8	–24.2	3

**Table 4.** Summary of QTLs for SMD resistance, position of the QTL on the map, percentage of phenotypic variance explained (% PV), additive and dominant effects, and LOD scores in ICP 8863 × ICPL 87119 (F<sub>2</sub>) population in pigeonpea. \*% phenotypic variation.

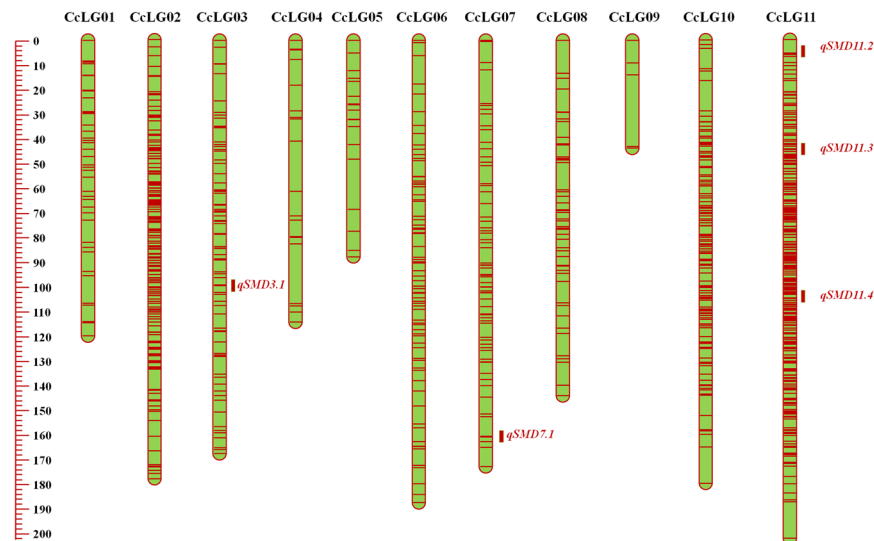
in pigeonpea<sup>12</sup>. Out of the four SNPs, one SNP at 19,958,148 bp position on CcLG11 was found to be present within the QTL (*qSMD11.1*) region identified on CcLG11 in present study in PRIL\_B population.

Interestingly, no common QTL was identified among three different populations, indicating distinct genomic regions are associated with SMD resistance in three populations. Similar observation has been made in the earlier study<sup>9</sup>. This could also be because of variations in marker density in three genetic maps which may lead to less number of common markers between three genetic maps. Therefore, integration of more markers is necessary to confirm the above results. Two major QTLs (*qSMD11.1* and *qSMD11.3*) identified in the present study were present on CcLG11 suggesting importance of these regions on CcLG11 in governing SMD resistance in pigeonpea. Additionally, a whole genome re-sequencing (WGRS) based Seq-BSA approach has also identified one SNP within the QTL region identified on CcLG11 in PRIL\_B population supporting the present results<sup>12</sup>. Detailed analysis of QTLs on CcLG11 could help to understand the genetic basis of SMD resistance in pigeonpea. Moreover, these regions can be integrated into elite lines through marker assisted breeding to improve the SMD resistance. From the previous and present studies, it is clear that SMD is very complex trait and governed by various small effect QTLs and breeding strategies involving multiple parents such as multi-parent advanced generation inter cross (MAGIC) need to be employed in order to increase accumulation of all the favorable small effect QTLs into superior genotypes.

In summary, GBS was found a very effective approach to generate large scale SNP genotyping data. Based on these data, three genetic maps with better marker density and genome coverage have been developed for PRIL\_B (1,101 markers with 921.20 cM), PRIL\_C (484 markers with 798.25 cM) and F<sub>2</sub> (996 SNP markers with 1597.30 cM) populations. This study has generated the highest saturated genetic maps so far in pigeonpea and also identified one candidate genomic region on CcLG11 associated with SMD resistance for deployment in genomics-assisted breeding.

## Materials and Methods

**Plant material.** Five parental lines viz., ICPL 20096, ICPL 332, ICPL 20097, ICP 8863 and ICPL 87119 with contrasting SMD resistance were selected for development of mapping populations. Out of these lines, ICPL 20096, ICPL 20097 and ICPL 87119 were resistant to SMD while ICPL 332 and ICP 8863 were susceptible to SMD.



**Figure 5.** Genetic and QTL map constructed using 996 SNPs and of 1,596.30 cM length in ICP 8863 × ICPL 87119 ( $F_2$ ) population in pigeonpea. The scale on left side represents map distance in cM. The eleven linkage groups are shown as vertical bars and each horizontal line on the bar represent single SNP marker. Aggregation on horizontal lines indicate higher marker density on that particular linkage group. The QTLs identified for SMD resistance on various linkage groups have been shown as brown colored rectangle.

Two recombinant inbred lines (RILs) populations, each of 188 lines were generated by crossing ICPL 20096 × ICPL 332 (PRIL\_B) and ICPL 20097 × ICP 8863 (PRIL\_C)<sup>1</sup> while one early generation ( $F_2$ ) mapping population of 168 lines was generated by crossing ICP 8863 × ICPL 87119 genotypes.

**Phenotypic evaluation and statistical analysis.** All the three populations along with parental lines were evaluated for SMD resistance using leaf stapling technique<sup>28</sup>. Disease reactions for all three populations were carried out against the Patancheru isolate. The plants were artificially inoculated by stapling the infected leaves to allow the transmission of mite vector. Resistant and susceptible controls viz. ICPL 2376 and ICP 8863 were sown at regular intervals to monitor the disease infection. Disease scoring for SMD was done periodically at 30 days interval; however final cumulative data was taken at 90 days as disease can be easily distinguished as patches of bushy, pale green plants without flowers or pods by this time. The disease severity was determined on the basis of percentage of affected plants wherein 0% means complete resistance while 100% means complete susceptibility to SMD. The PRIL\_B and PRIL\_C populations were evaluated for SMD at two different locations viz. International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, India and Professor Jayashankar Telangana State Agricultural University (PJTSAU), Rajendranagar, Hyderabad, India for two consecutive years (2012–2013 and 2013–2014).

SMD screening is destructive and may lead to complete loss of seeds from susceptible plants. Therefore to avoid this issue in  $F_2$  population, the plants were selfed to generate  $F_{2,3}$  plants which were subsequently phenotyped for SMD resistance at ICRISAT during year 2015–2016. 10  $F_{2,3}$  plants representing each  $F_2$  plant were sown in two replications following randomized complete block design (RCBD). As a result, a total of 3360  $F_{2,3}$  plants were considered for disease scoring. BLUPs were estimated from multi-locations data and the arcsine transformed values have been utilized for performing QTL analysis.

**DNA extraction and genotyping-by-sequencing (GBS).** DNA from the parental genotypes and from individual progenies were extracted using NucleoSpin Plant II kit (Macherey-Nagel, Dren, Germany). The quality and quantity of DNA was checked on 0.8% agarose gel and then using Qubit 2.0 fluorometer (Thermo Fisher Scientific Inc., USA).

For GBS approach, 10 ng genomic DNA from each sample was restriction digested using *ApeKI* (recognition site: G/CWCG) endonuclease. The digested product was ligated with uniquely barcoded adaptors using T4 DNA ligase enzyme. Such digested ligated products from each sample were mixed in equal proportion to construct the GBS libraries, which were then amplified, purified to remove excess adaptors and used for sequencing on HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA).

**SNP identification and genotyping.** The sequence reads from raw FASTQ files were used for SNP identification and genotyping using reference based GBS analysis pipeline implemented in TASSEL v4.0<sup>29</sup>. Briefly, the sequencing reads were searched for perfectly matched barcode with the expected four base remnant of the enzyme cut site. The barcode containing reads were sorted, de-multiplexed according to barcode sequence and trimmed to first 64 bases starting from enzyme cut site. Further, those reads containing ‘N’ within first 64 bases were rejected. The remaining good quality reads (called as tags) were aligned against the draft genome sequence of pigeonpea<sup>10</sup> using Bowtie 2 software<sup>30</sup>. The alignment file was then processed through GBS analysis pipeline for SNP calling and genotyping. An allele was considered only if it was supported with a minimum tag count value of



10. The SNPs identified were further filtered to remove missing data and such filtered SNPs were used for genetic mapping and QTL analysis.

**Genetic mapping and QTL analysis.** The filtered SNPs from each population were used for the construction of high density genetic maps using JoinMap V4.0<sup>31</sup>. The Chi-square ( $\chi^2$ ) values calculated for each SNP marker using Joinmap V4.0 were used to determine segregation distortion and highly distorted and unlinked markers were excluded from further analysis. The grouping and ordering of markers was carried out using regression mapping algorithm with maximum recombination frequency of 0.4 at minimum logarithm of odds (LOD) value of 3. The marker orders after incorporation of new marker was confirmed using ripple command. Finally, Kosambi mapping function was used to convert recombination fraction into map units<sup>32</sup>. Mapchart 2.30<sup>33</sup> was used for the visualization of the linkage groups.

The genotyping data along with phenotyping data collected for each population were used for QTL analysis using QTL Cartographer V.2.5 software<sup>34</sup>. The composite interval mapping (CIM) method with model 6 and other default values were used for QTL identification. The empirical LOD thresholds for each trait were determined by 1,000 permutations at the  $P \leq 0.05$  level<sup>35</sup>. The LOD score values were used to determine the significance of QTL intervals.

## References

- Saxena, R. K., Saxena, K. B., Kumar, R. V., Hoisington, D. A. & Varshney, R. K. Simple sequence repeat-based diversity in elite pigeonpea genotypes for developing mapping populations to map resistance to Fusarium wilt and sterility mosaic disease. *Plant Breed* **129**, 135–141, doi:10.1111/j.1439-0523.2009.01698.x (2010).
- Kumar, P. L., Jones, A. T. & Reddy, D. V. R. A novel mite-transmitted virus with a divided RNA genome closely associated with pigeonpea sterility mosaic disease. *Phytopathol* **93**, 71–81, doi:10.1094/Phyto.2003.93.1.71 (2003).
- Kannaiyan, J., Nene, Y. L., Reddy, M. V., Ryan, J. G. & Raju, T. N. Prevalence of pigeonpea disease and associated crop losses in Asia, Africa and America. *Trop Pest Manag* **30**, 62–71, doi:10.1080/09670878409370853 (1984).
- Varshney, R. K. *et al.* Pigeonpea genomics initiative (PGI): an international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol Breed* **26**, 393–408, doi:10.1007/s11032-009-9327-2 (2010).
- Ganapathy, K. N. *et al.* Identification of AFLP markers linked sterility mosaic disease in pigeonpea *Cajanus cajan* (L.) Millsp. *Int J Integr Biol* **7**, 145–149 (2009).
- Bohra, A. *et al.* 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 (2012).
- Prasanthi, L. R. *et al.* RAPD and SCAR marker linked to the sterility mosaic disease resistance gene in pigeonpea (*Cajanus cajan* L. Millsp.). *Asian Australian J Plant Sci Biotechnol* **3**, 16–20 (2009).
- Patil, P. G. *et al.* Microsatellite markers linked to sterility mosaic disease resistance in pigeonpea (*Cajanus cajan* L. Millsp.). *Legume Genomics Genet* **7**, doi:10.5376/lgg.2016.07.0006 (2016).
- Gnanesh, B. N. *et al.* 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, doi:10.1016/j.fcr.2011.04.011 (2011).
- Varshney, R. K. *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol* **30**, 83–U128, doi:10.1038/nbt.2022 (2012).
- Pazhamala, L. *et al.* Genomics-assisted breeding for boosting crop improvement in pigeonpea (*Cajanus cajan*). *Front Plant Sci* **6**, doi:10.3389/fpls.2015.00050 (2015).
- Singh, V. K. *et al.* Next-generation sequencing for identification of candidate genes for Fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*). *Plant Biotechnol J* **14**, 1183–1194, doi:10.1111/pbi.12470 (2016).
- Baird, N. A. *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *Plos One* **3**, e3376, doi:10.1371/journal.pone.0003376 (2008).
- Elshire, R. J. *et al.* A robust, simple genotyping-by-Sequencing (GBS) approach for high diversity species. *Plos One* **6**, e19379, doi:10.1371/journal.pone.0019379 (2011).
- Poland, J. A. & Rife, T. W. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome-Us* **5**, 92–102, doi:10.3835/plantgenome2012.05.0005 (2012).
- Jaganathan, D. *et al.* Genotyping-by-sequencing based intra-specific genetic map refines a “QTL-hotspot” region for drought tolerance in chickpea. *Mol Genet Genomics* **290**, 559–571, doi:10.1007/s00438-014-0932-3 (2015).
- Hart, J. P. & Griffiths, P. D. Genotyping-by-sequencing enabled mapping and marker development for the by-2 potyvirus resistance allele in common bean. *Plant Genome-Us* **8**, doi:10.3835/plantgenome2014.09.0058 (2015).
- Gao, L. L. *et al.* Development of genotyping by sequencing (GBS)- and array-derived SNP markers for stem rust resistance gene Sr42. *Mol Breed* **35**, doi:10.1007/S11032-015-0404-4 (2015).
- Lee, J. *et al.* Genotyping-by-sequencing map permits identification of clubroot resistance QTLs and revision of the reference genome assembly in cabbage (*Brassica oleracea* L.). *DNA Res* **23**, 29–41, doi:10.1093/dnares/dsv034 (2016).
- Swarts, K. *et al.* Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *Plant Genome-Us* **7**, doi:10.3835/plantgenome2014.05.0023 (2014).
- Chen, Z. L. *et al.* An ultra-high density bin-map for rapid QTL mapping for tassel and ear architecture in a large F-2 maize population. *Bmc Genomics* **15**, doi:10.1186/1471-2164-15-433 (2014).
- Bohra, A. *et al.* 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**, doi:10.1186/1471-2229-11-56 (2011).
- Varshney, R. K. *et al.* Fast-track introgression of “QTL-hotspot” for root traits and other drought tolerance traits in JG 11, an elite and leading variety of chickpea. *Plant Genome-Us* **6**, doi:10.3835/plantgenome2013.07.0022 (2013).
- Varshney, R. K. *et al.* Marker-assisted backcrossing to introgress resistance to fusarium wilt race 1 and ascochyta blight in C 214, an elite cultivar of chickpea. *Plant Genome-Us* **7**, doi:10.3835/plantgenome2013.10.0035 (2014).
- Varshney, R. K. *et al.* Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). *Theor Appl Genet* **127**, 1771–1781, doi:10.1007/s00122-014-2338-3 (2014).
- Saxena, R. K. *et al.* Novel SSR markers for polymorphism detection in pigeonpea (*Cajanus* spp.). *Plant Breed* **129**, 142–148, doi:10.1111/j.1439-0523.2009.01680.x (2010).
- Saxena, R. K. *et al.* Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. *DNA Res* **19**, 449–461, doi:10.1093/dnares/dss025 (2012).
- Nene, Y. L. & Reddy, M. V. Leaf stapling technique to screen pigeonpea for resistance to sterility mosaic. *Indian Phytopathol* **30** (1977).
- Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635, doi:10.1093/bioinformatics/btm308 (2007).

30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–U354, doi:[10.1038/Nmeth.1923](https://doi.org/10.1038/Nmeth.1923) (2012).
31. Van Ooijen, J. W. & Voorrips, R. E. JoinMap version4.0: software for the calculation of genetic linkage maps. *Wageningen: Plant Res Int* (2006).
32. Kosambi, D. D. The estimation of map distances from recombination values. *Annals of Human Genet* **12**, 172–175, doi:[10.1111/j.1469-1809.1943.tb02321.x](https://doi.org/10.1111/j.1469-1809.1943.tb02321.x) (1943).
33. Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**, 77–78, doi:[10.1093/jhered/93.1.77](https://doi.org/10.1093/jhered/93.1.77) (2002).
34. Wang, S., Basten, C. J. & Zeng, Z. B. Windows QTL cartographer version 2.5. *Statistical genetics. North Carolina State University, Raleigh* (2005).
35. Churchill, G. A. & Doerge, R. W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).

## Acknowledgements

Authors are thankful to the United States Agency for International Development (USAID)-India Mission and Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India for the financial support. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of CGIAR Consortium.

## Author Contributions

R.K.S. and S.M.K. performed most of the experiments; V.K. and S.P. generated the sequence data; V.G. and S.M.K. analysed the GBS data; M.S., K.N.Y., S.J. and A.G. contributed in the phenotyping of mapping populations; R.R.D. and A.R. analysed the phenotypic data; R.K.S., S.M.K., V.K.S. and S.P. performed QTL-mapping analysis; R.K.S., S.M.K. and R.K.V. wrote the manuscript; C.V.S.K. and K.B.S. contributed the genetic materials; R.K.V. conceived, designed and supervised the study and finalized the manuscript; All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-01535-4](https://doi.org/10.1038/s41598-017-01535-4)

**Competing Interests:** The authors declare that they have no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017