

RESEARCH ARTICLE

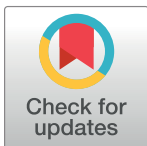
# A high-density intraspecific SNP linkage map of pigeonpea (*Cajanus cajan* L. Millsp.)

Sheetal Arora<sup>1</sup>\*, Ajay Kumar Mahato<sup>1</sup>\*, Sangeeta Singh<sup>1</sup>\*, Paritra Mandal<sup>1</sup>, Shefali Bhutani<sup>1</sup>, Sutapa Dutta<sup>1</sup>, Giriraj Kumawat<sup>1</sup>, Bikram Pratap Singh<sup>1</sup>, A. K. Chaudhary<sup>2</sup>, Rekha Yadav<sup>3</sup>, K. Gaikwad<sup>1</sup>, Amitha Mithra Sevanthi<sup>1</sup>, Subhojit Datta<sup>2</sup>, Ranjeet S. Raje<sup>3</sup>, Tilak R. Sharma<sup>1</sup>, Nagendra Kumar Singh<sup>1</sup>\*

**1** National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi, India, **2** Indian Institute of Pulses Research, Kanpur, India, **3** Division of Genetics, Indian Agricultural Research Institute, Pusa, New Delhi, India

\* These authors contributed equally to this work.

\* [nksingh4@gmail.com](mailto:nksingh4@gmail.com)



## Abstract

Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a major food legume cultivated in semi-arid tropical regions including the Indian subcontinent, Africa, and Southeast Asia. It is an important source of protein, minerals, and vitamins for nearly 20% of the world population. Due to high carbon sequestration and drought tolerance, pigeonpea is an important crop for the development of climate resilient agriculture and nutritional security. However, pigeonpea productivity has remained low for decades because of limited genetic and genomic resources, and sparse utilization of landraces and wild pigeonpea germplasm. Here, we present a dense intraspecific linkage map of pigeonpea comprising 932 markers that span a total adjusted map length of 1,411.83 cM. The consensus map is based on three different linkage maps that incorporate a large number of single nucleotide polymorphism (SNP) markers derived from next generation sequencing data, using Illumina GoldenGate bead arrays, and genotyping with restriction site associated DNA (RAD) sequencing. The genotyping-by-sequencing enhanced the marker density but was met with limited success due to lack of common markers across the genotypes of mapping population. The integrated map has 547 bead-array SNP, 319 RAD-SNP, and 65 simple sequence repeat (SSR) marker loci. We also show here correspondence between our linkage map and published genome pseudomolecules of pigeonpea. The availability of a high-density linkage map will help improve the anchoring of the pigeonpea genome to its chromosomes and the mapping of genes and quantitative trait loci associated with useful agronomic traits.

## OPEN ACCESS

**Citation:** Arora S, Mahato AK, Singh S, Mandal P, Bhutani S, Dutta S, et al. (2017) A high-density intraspecific SNP linkage map of pigeonpea (*Cajanus cajan* L. Millsp.). PLoS ONE 12(6): e0179747. <https://doi.org/10.1371/journal.pone.0179747>

**Editor:** Pawan L. Kulwal, Mahatma Phule Krishi Vidyapeeth College of Agriculture, INDIA

**Received:** October 28, 2016

**Accepted:** June 2, 2017

**Published:** June 27, 2017

**Copyright:** © 2017 Arora et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was financially supported by the NPTC Project Grant No-3052 of the Indian Council of Agricultural Research (ICAR).

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Pigeonpea (*Cajanus cajan* L. Millsp.) is a diploid species ( $2n = 2x = 22$ ) belonging to the genus *Cajanus* and is a member of the plant family Fabaceae [1]. It is a warm-season legume crop with a moderate genome size of 858 Mb. Pigeonpea is an important crop owing to its high nutritional value, drought resilience, disease resistance, and ability to improve soil carbon

content [2]. It is particularly rich in protein, vitamins, minerals, and essential secondary metabolites like isoflavonoids; therefore, pigeonpea is significant in the remediation of under-nourishment or protein-calorie malnutrition [3], which makes pigeonpea a highly promising crop for eliminating malnutrition, especially in South Asia.

Owing to various abiotic and biotic stresses and unavailability of quality seeds of improved cultivars, growth in pigeonpea yield has remained consistently low for decades. The genetics of important agronomic traits in pigeonpea have not yet been worked out and their molecular basis remains poorly understood. Furthermore, cultivated pigeonpea has a low level of DNA polymorphism in the primary gene pool and there is a lack of validated molecular markers in this species [4,5]. A Pigeonpea Genomics Initiative (PGI) was started by the Indian Council of Agricultural Research (ICAR) to address these issues under the Indo-US Agricultural Knowledge Initiative (AKI). The AKI-PGI has contributed significantly to the generation of genomic resources for pigeonpea to facilitate marker-assisted breeding (MAB) by developing numerous simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers and moderately dense linkage maps. Draft genomes of the pigeonpea variety Asha has been assembled using FLX 454 (Roche Inc., Germany) and Illumina (San Diego, USA) sequencing technologies [6,7]; however, it has not been possible to map a large number of loci in a single mapping population owing to the low level of polymorphism. Hence, to increase the number of mapped markers and construct a dense consensus map, multiple mapping populations with an acceptable number of common markers are utilized.

Recent efforts towards building a genetic map of pigeonpea have led to the development of several interspecific and intraspecific maps. The first interspecific map of pigeonpea was developed using 554 diversity arrays technology (DArT) markers covering a total map distance of 451.6 cM [8]. Recently, another interspecific linkage map was reported with 191 start codon targeted (SCoT), inter simple sequence repeats (ISSR), and random amplified polymorphic DNA (RAPD) markers mapped onto 11 linkage groups covering a total map length of 1,624.71 cM with an average marker interval of 8.51 cM [9]. An intraspecific medium density consensus map of six different mapping populations with 339 SSR loci spanning total map length of 1,059 cM has been reported [10]. Another medium density intraspecific map from a single mapping population has been reported that comprises 296 genic SNP and SSR markers covering a total map length of 1,520.22 cM with average map interval of 4.95 cM [11]. These maps have helped QTL mapping of agronomically useful traits and anchoring of the pigeonpea draft genome.

Earlier linkage maps of pigeonpea have employed mostly DArT, RAPD, ISSR, and SSR markers. Although these are useful markers, some of these have dominant type inheritance; therefore, in the present study, we have exploited more informative co-dominantly inherited SSR and SNP markers. In the last few years, SNPs have become the markers of choice for all species owing to their abundance in the genome, high-throughput assays, and low cost per data point. In recent years, SNP genotyping by restriction site associated DNA (RAD) sequencing has been favored because it makes available a genome-wide representation of all the sites of a particular restriction enzyme at an affordable cost. Therefore, we chose SNP markers to prepare a high-density linkage map of pigeonpea comprising 932 markers. The component linkage maps were developed mainly with SNP markers using the Illumina's GoldenGate assay and reduced representation RAD genotyping-by-sequencing (GBS) approaches [12].

## Materials and methods

### Plant material and DNA extraction

Three different intraspecific  $F_2$  mapping populations, Asha/UPAS 120 (population A) consisting of 92 individuals, Pusa Dwarf/H2001-4 (population R) consisting of 94 individuals, and

Pusa Dwarf/HDM04-1 (population G) consisting of 186 individuals [11], were used for the construction of the pigeonpea linkage map. The seeds of original parents were obtained from ICAR-IARI Division of Genetics, New Delhi and ICAR-IIPR, Kanpur, India. Genomic DNA was isolated from the fresh young leaves of F<sub>2</sub> individuals grown in net house along with their parents using a modified CTAB method [13].

### Genotyping using SSR markers

A total of 222 SSR markers including 71 genic arhar SSR (ASSR) [14] and 151 genomic highly variable arhar SSR (HASSR) [15] were used for genotyping 92 F<sub>2</sub> individuals in population A. The PCR reactions were carried out in a PTC225 Gradient Cycler (MJ Research, Waltham, USA). Each PCR reaction in a final reaction volume of 15 µl consisted of 1.5 µl 10x reaction buffer, 0.20 µl 10 mM dNTPs (133 µM), 1.5 µl each forward and reverse primers (10 pmol each), and 2.0 µl (70 ng) template genomic DNA, 0.15 µl (0.75 U) Taq DNA polymerase (Vivantis Technologies, CA, USA). The PCR thermal profile included an initial denaturation at 94°C for 5 min. followed by 35 cycles of 94°C for 1 min., 55°C for 1 min., and 72°C for 1 min., and finally, 10 min at 72°C for a final extension. The PCR products were resolved by electrophoresis in 4% metaphor agarose gels or 8% polyacrylamide gels. After electrophoresis, PCR products were visualized and photographed in a Fluorchem™ 5500 (Alfa Innotech Crop, USA) gel documentation system. In population G, of the 595 SSR markers screened for parental polymorphism between Pusa Dwarf/HDM04-1, only 29 (4.8%) were found to be polymorphic [11].

### SNP genotyping using GoldenGate assays

Two different multiplexed SNP assays using 768 and 1,536 genic SNP markers were used, employing Illumina GoldenGate bead-array technology (<http://www.illumina.com>), as described earlier [11]. The 1,536-plex assay was used for genotyping of all three pigeonpea populations (A, G and R), whereas the 768-plex assay was used for the populations A and G only.

### SNP genotyping by RAD sequencing

RAD sequencing was performed on two sets of 96-plex libraries using restriction enzymes *ApeKI* and *PstI* on 92 F<sub>2</sub> individuals and their parents using the raw sequence reads that were generated for the optimized RAD library with Illumina True-Seq V3 paired-end chemistry and average read lengths of 100 bp on an Illumina HiSeq platform. Bioinformatics analysis included selecting high quality reads from the raw data at a threshold quality score of Q20 and the removal of adapter sequences, mapping high quality reads, and SNP detection from GBS data with TASSEL version 3.0 [16]. The reference sequence used in this analysis covered approximately 60% of the pigeonpea genome. In the TASSEL software, the reads were sorted after alignment and identical reads were collapsed into a single tag. Default parameters were used for the previous steps, parameters used during SNPs calling included: inbreeding coefficient (mnf) = -1, minimum minor allele frequency (mnMAF) = 0.1, minimum locus coverage (mnLCov) = 0.05, minimum site coverage (mnSCov) = 0.70, minimum taxon coverage (mnTCov) = 0.50, and the “hLD” option was invoked once for each of the two enzymes. TASSEL version 4.0 was used for the integration of data from both runs, once for each enzyme to obtain a single HapMap output file.

### Construction of individual and consensus linkage maps

The F<sub>2</sub> mapping population A of Asha/UPAS 120 was used to generate the genetic reference map for pigeonpea. Genotyping data was generated from 92 F<sub>2</sub> individuals for 10,786 polymorphic

marker loci using different marker systems that included Illumina GoldenGate genic SNP, RAD-GBS SNP, ASSR, and HASSR markers. JoinMap 4.0 [17] was used for linkage map construction. A chi-square test was performed on the genotyping data for this population to test the goodness of fit with the expected segregation ratio of 1:2:1. To include more markers in the map, the cut-off chi-square value was relaxed to 13.5. The grouping of markers was done on the basis of cut off Rf value of 0.25 and a logarithm of the odds (LOD) score of 3.0. In the F<sub>2</sub> mapping population R (Pusa Dwarf/H2001-4), the chi-square cut-off was set at 9.2 and all markers followed the expected segregation ratio of 1:2:1. The linkage map of Pusa Dwarf/HDM04-1 (population G) was previously drawn using MapDisto software [18] and has been used for mapping QTLs for plant height, stem branching, and maturity time [11]. This map was redrawn in this study using JoinMap software for uniformity with the purpose of integrating all three maps into a single consensus map. The chi-square cutoff used for this population was also 9.2 and all the markers showed the expected segregation ratio. The grouping of chromosomes was done on the basis of a recombination frequency of 0.35 and markers that did not conform to the stable order were discarded.

The three individual genetic maps were then merged to generate the consensus linkage map using MergeMap online [19] by converting the individual maps into directed acyclic graphs (DAGs) that were then merged into a consensus graph on the basis of their share vertices. While constructing the consensus linkage map, “weights” were assigned to the three populations based on the level of confidence we had on their marker orders. Hence, population G was given a weight of 5.0, followed by population A with a weight of 3.0, and population R with a weight of 1.0. The final map was drawn using software MapChart [20].

## Comparison of consensus linkage groups with published genome pseudomolecules of pigeonpea

A comparative analysis was carried out using sequence information from the markers included in the present consensus linkage map with published pigeonpea genome pseudomolecules of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) [7]. Mapped SNP flanking sequences from the 1,536-plex and 768-plex Illumina GoldenGate assays, primer sequences of SSRs, and flanking sequences of 319 RAD-SNP markers were used for comparative analysis. The sequences of all 932 markers in the consensus map were searched for in the ICRISAT chromosome pseudomolecules using BLASTn with earlier published [21] pre-optimized parameters (-v 1, -b 1, -E 1, -q 1, -r 1, -F, F-e, and 1e-10). A chromosome wise list of our (NRCPB) markers matched in the ICRISAT pseudomolecules was also prepared.

## Results

The main objective of this study was to construct a high-density SNP linkage map of pigeonpea for which we first constructed two new component linkage maps using two different mapping populations, (i) Asha/UPAS 120 (Population A) and (ii) Pusa Dwarf/H2001-4 (Population R). The linkage map of population A comprised 725 markers and that of population R comprised 136 markers. These two new maps were then integrated with a third population, a redrawn map of Pusa Dwarf/HDM04-1 population (Population G) [11], to develop a high-density consensus linkage map of the 11 chromosomes of pigeonpea.

### Linkage map of Asha/UPAS 120

Genotyping data were generated from 92 F<sub>2</sub> individuals of population A (Asha/UPAS 120) to identify 10,786 marker loci using different marker systems including GoldenGate genic SNP, RAD-GBS SNP, ASSR and HASSR markers (Table 1). From the 768-plex GoldenGate assay,

**Table 1. Features of the component genetic maps along with their contribution of markers to the consensus pigeonpea genetic map.**

Number of markers that contributed to the component genetic map (Number of markers used in the consensus map)				Consensus Map	
Linkage Group	Asha/ UPAS 120 (population A)	Pusa Dwarf/ H2001-4 (population R)	Pusa Dwarf/ HDM04-1 (population G)	Number of Markers	Map Distance (cM)
1	94(60)	22(-)	56(29)	127	136.9
2	124(94)	22(-)	35(16)	144	191.6
3	24(17)	7(-)	32(24)	52	115.03
4	65(53)	13(-)	27(21)	89	124.76
5	58(50)	10(-)	21(17)	78	102.23
6	58(45)	8(-)	20(12)	71	117.361
7	81(68)	6(-)	22(11)	94	119.02
8	81(35)	8(1)	23(11)	63	104.99
9	43(32)	11(1)	18(12)	58	135.41
10	53(38)	15(-)	16(12)	67	156.48
11	77(61)	14(1)	21(9)	89	140.82
<b>Total</b>	<b>725(553)</b>	<b>136(3)</b>	<b>291(174)</b>	<b>932</b>	<b>1411.83</b>

<https://doi.org/10.1371/journal.pone.0179747.t001>

159 polymorphic markers were analyzed, of which 106 were included in the linkage map, whereas from the 1,536-plex GoldenGate assay, 382 markers were found to be polymorphic, of which 197 were included in the linkage map. Overall, out of the 2,304 SNP loci from the two GoldenGate assays 541 loci (23.48%) were polymorphic between parental varieties Asha and UPAS 120. We were expecting higher polymorphism because SNPs in the GoldenGate assay were identified by an in-silico comparison of Asha and UPAS 120 transcript sequences; however, the lower than expected success may be attributed to heterogeneous sample of variety Asha that was used in the transcript sequencing. For GBS-RAD based genotyping 26.5 Gb and 21.3 Gb high quality reads were generated with the *ApeKI* and *PstI* enzymes, respectively. In total 7,500 SNPs were identified, of which only 1,207 SNPs could be used for map making, the remaining markers were discarded owing to segregation distortion and a high number of missing data points due to lack of commonly genotyped markers across the 92 individuals in the mapping population A. Finally, only 319 GBS-RAD SNPs were included in the map of population A (S1 Table). Of the 71 ASSR and 151 HASSR markers screened, only 65 (29%) were found to be polymorphic and 41 were included in the linkage map. The number of mapped markers in the 11 different linkage groups (LGs) ranged from 24 in LG3 to 124 in LG1. The minimum map length was observed in LG3 (62.0 cM) whereas the maximum map length was observed in LG11 (127.2 cM). The largest interval was observed in LG9 (13.2 cM) and the shortest interval was observed in LG08 (4.3 cM). In total, 725 markers were included in the map with population A, which covered a map length of 1,102.2 cM and an adjusted map length of 1,105.5 cM with an average map interval of 1.39 cM (Table 1; S1 Fig). This provides a new dense genetic map of pigeonpea based on a single intraspecific mapping population.

### Linkage map of Pusa Dwarf/H2001-4

The population R (Pusa Dwarf/H2001-4) included 94 F<sub>2</sub> individuals that were genotyped using only the 1,536-plex GoldenGate SNP assay, and 186 polymorphic loci (12.1%) were found, owing to the narrow genetic base of pigeonpea varieties. Only 136 of these loci were included in the linkage map, with the remaining 50 markers discarded owing to higher missing values in the genotyping data and segregation distortion. The number of mapped markers in the 11 different linkage groups ranged from 6 in LG7 to 22 in LG1 and LG2. The minimum

map length was observed in LG7 (39.66 cM), whereas the maximum map length was observed in LG2 (174.1 cM). The largest interval was observed in LG3 (34.80 cM) and the shortest interval was observed in LG01 (10.7 cM). The total map length for population R was 929.82 cM with an adjusted map length of 943.58 cM and average marker interval of 8.05 cM (S2 Fig). This is also a new intraspecific molecular linkage map of pigeonpea with 136 SNP markers.

### Linkage map of Pusa Dwarf/HDM04-1

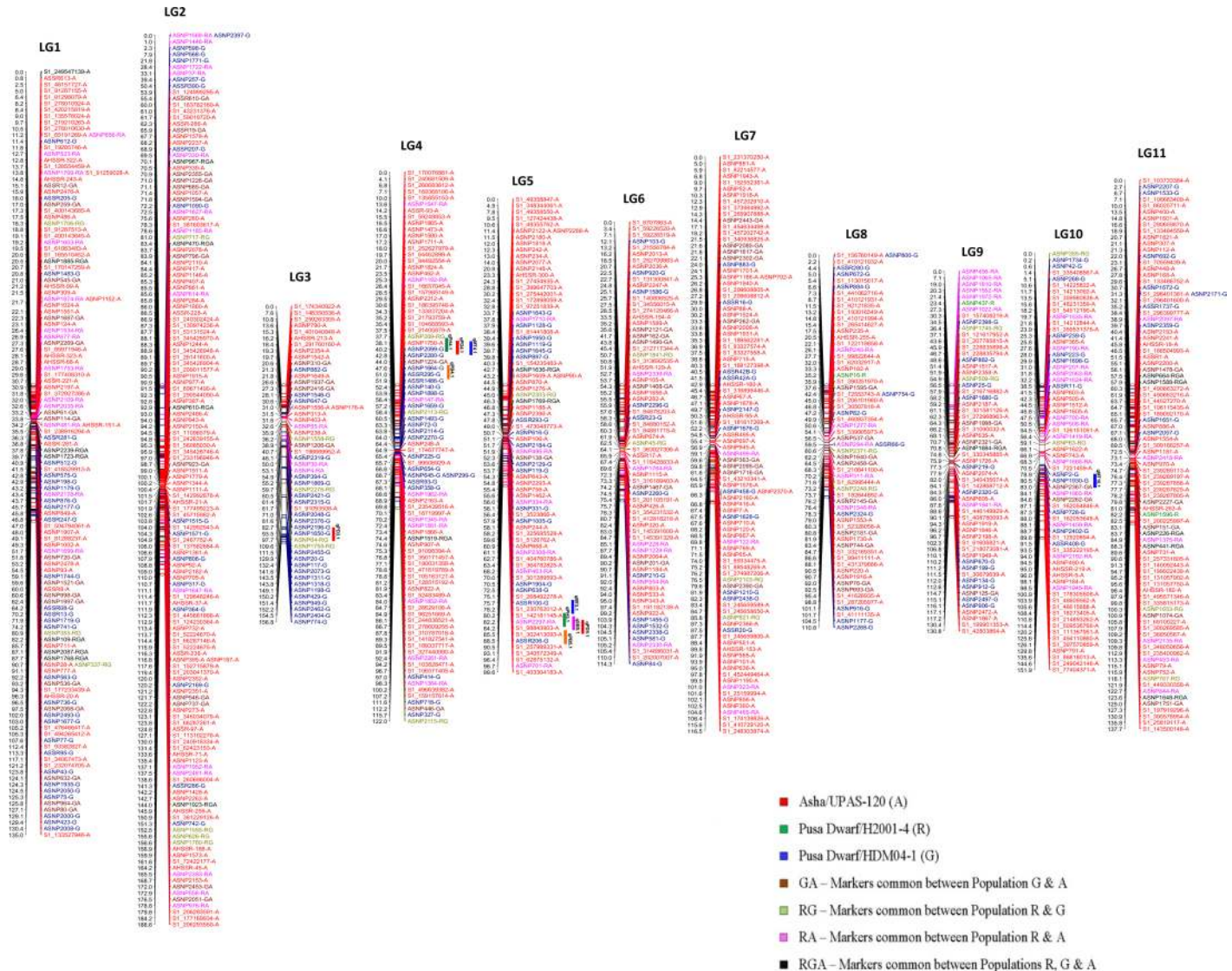
The linkage map of population G (Pusa Dwarf/HDM04-1) with 296 markers was previously drawn using MapDisto software and has been used for the mapping of QTLs for plant height, stem branching, and maturity time in pigeonpea [11]. The map was redrawn in this study using JoinMap software to preserve uniformity when integrating the three maps into a single consensus map. Seven of the 296 markers were discarded from the earlier published map and two new markers were added, providing a total number of 291 markers. The seven markers that were removed in the current map were ASNP1310 from LG4; ASNP1818, ASNP234, ASNP242, and ASSR109 from LG5; ASNP280 from LG8; and ASSR81 from LG1. These markers were removed because they did not conform to the most stable marker order. Excluding these markers and including the two additional ones provided the most stable marker order for the individual map of population G using JoinMap.

The order of the markers was similar to the previously published map but the total map distance was reduced from 1,406.7 cM to 923.9 cM (S3 Fig). The map length was also reduced because JoinMap uses a different algorithm than MapDisto and the current linkage distances were slightly smaller than those in the previously published map. The number of mapped markers in the 11 different linkage groups ranged from 16 in LG11 to 56 in LG1 (Table 1). The minimum map length was observed in LG11 (52.2 cM) and maximum map length was observed in LG1 (196.6 cM). The largest gap size was observed in LG10 (25.4 cM) and the smallest one was observed in LG6 (15.1 cM). The reconstructed map has a total map length of 923.9 cM and adjusted map length of 930.3 cM, with average marker interval of 2.93 cM (S3 Fig).

### Consensus linkage map

Three individual genetic maps of populations A, R, and G were merged to generate an intraspecific consensus map using MergeMap software by converting the individual population maps into DAGs that were then merged into a consensus graph on the basis of their shared vertices. In total, 1,152 markers (1,083 SNP and 69 SSR) were included in the three linkage maps, of which 158 conflicting markers were discarded during the process of merging the individual maps using MergeMap. The consensus linkage map has 932 markers that included 863 SNP and 69 SSR loci (Table 1; Fig 1). In the consensus map a total of 202 (21.6%) markers were common between the populations, out of which 69 markers were common between population G and A, 87 were common between populations R and A, 27 were common between populations R and G, and 19 markers were common to all three populations; leaving the remaining 730 (78.4%) that were unique to the individual mapping populations themselves.

To avoid confusion with multiple nomenclatures, we have retained the original numbering of linkage groups LG1 to LG11 from our previously published map with population G [11]. The total number of markers in the individual linkage groups in the consensus map ranged from 52 in LG3 to 144 in LG2. LG2 exhibited the maximum adjusted map length of 191.60 cM, whereas LG5 with 78 markers showed the minimum map length of 102.23 cM. The average marker interval ranged from 1.07 cM for LG1 to 2.3 cM for LG9 and LG10. Non-uniform distribution of markers was evident in all of the linkage groups. The largest map interval was



**Fig 1. The consensus linkage map of 932 markers that are both common and unique to the three pigeonpea mapping populations.** QTLs for plant height (PH), maturity time (MT), number of pods (PD), number of primary branches (PB), and number of secondary branches (SB) identified by Giriraj et al. [11] are indicated by different colored bars. Green, blue, red, orange, and magenta colored bars represent the QTLs for PH, MT, PD, PB, and SB, respectively.

<https://doi.org/10.1371/journal.pone.0179747.g001>

27.42 cM in LG1 followed by 13.25 cM in LG9; however, two are more markers were collocated at 17 different positions in the consensus map (S2 Table). The detailed statistics for each linkage group in the consensus linkage maps along with number of markers contributed by each component genetic map is shown in Table 2.

The consensus linkage map covered a total genome map length of 1,408.85 cM, which converted to an adjusted map length of 1,411.83 cM using method 4 from Chakravarti et al. [22]. The average marker interval was 1.51 cM, which was very useful for anchoring the genome and genetic studies in pigeonpea. From earlier published QTLs in pigeonpea by Giriraj et al (thirteen) [11], Bohra et al (four) [10] and Gnanesh et al (six) [23], we could easily place 11 out of 13 QTLs reported by Giriraj et al. [11] for plant height (PH), maturity time (MT), number of pods (PD), number of primary branches (PB), and number of secondary branches (SB)

**Table 2. Chromosome-wide details of a predominantly SNP based intraspecific consensus linkage map of pigeonpea constructed using three different F<sub>2</sub> populations: Asha/UPAS 120, Pusa Dwarf/H2001-4, and Pusa Dwarf/HDM04-1.**

Linkage group	No. of markers	Map length (cM)	Adjusted map length (cM)	Average map interval (cM)	Largest map interval (cM)
LG1	127	135.01	136.90	1.07	27.42
LG2	144	188.57	191.60	1.33	6.71
LG3	52	110.69	115.03	2.21	8.09
LG4	89	121.99	124.76	1.40	6.97
LG5	78	99.64	102.23	1.31	6.93
LG6	71	114.34	117.61	1.65	6.14
LG7	94	116.51	119.02	1.26	9.17
LG8	63	101.71	104.99	1.66	4.24
LG9	58	130.82	135.41	2.33	13.25
LG10	67	151.88	156.48	2.33	9.05
LG11	89	137.69	140.82	1.58	7.06
<b>Total</b>	<b>932</b>	<b>1,408.8</b>	<b>1,411.83</b>	<b>1.51</b>	<b>27.42</b>

<https://doi.org/10.1371/journal.pone.0179747.t002>

in our consensus map. We could not place other QTLs reported by Bohra et al. [10] and Gnanesht et al. [23], due to the lack of common markers/marker sequence information (Fig 1).

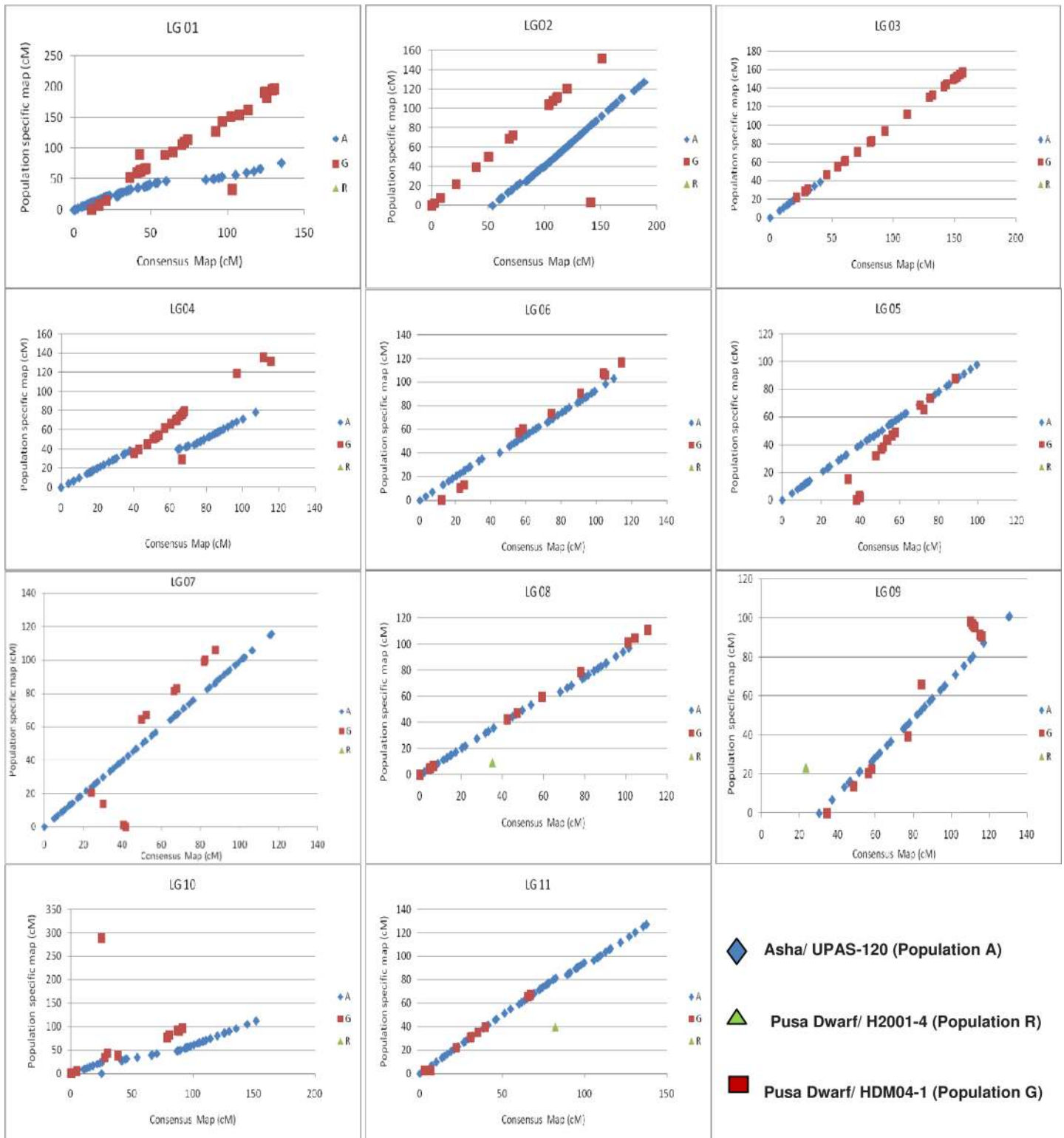
A detailed comparison of all linkage groups across all the maps is presented in Fig 2. A high degree of correlation was observed for most of the LGs between consensus and component LGs. A high degree of correlation was observed for all the LGs and showed a good agreement for both marker orders and marker positions between consensus and component LGs.

The consistency of marker order and possible rearrangement between the component maps and consensus genetic map was compared using the Genetic Map Comparator software [24]. Between the three maps, a total of 19 markers were common to all the maps and 183 markers were common in pairs of maps. For example, the correspondence among consensus and component maps of LG01 is shown in Fig 3. The component maps for the remaining linkage groups is shown in S4 Fig.

### Correspondence between NRCPB and ICRISAT linkage groups in pigeonpea

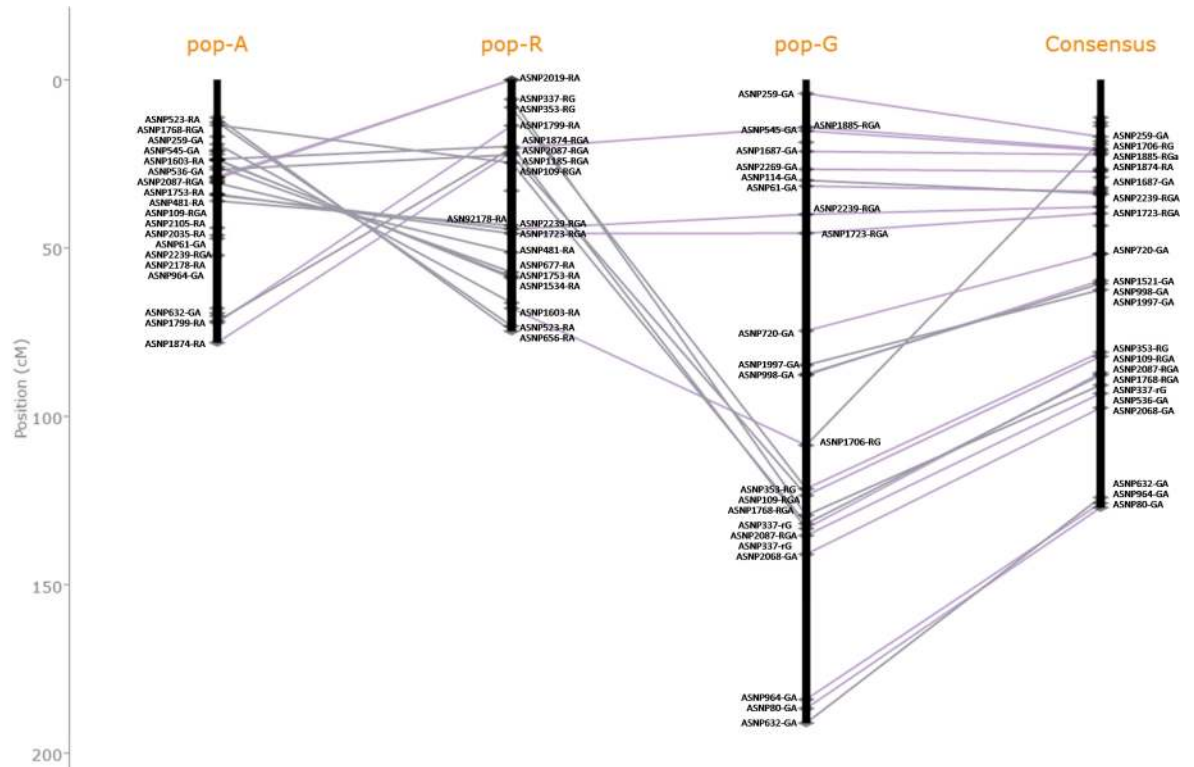
Prior to the Indo-US AKI Pigeonpea Genomics Initiative (PGI) there was no published molecular linkage map of pigeonpea [25]. However, after the AKI-PGI project commenced, two different groups at the National Research Centre on Plant Biotechnology (NRCPB) in New Delhi, India, and the International Crop Research Institute for Semi-Arid Tropics (ICRISAT) in Hyderabad, India, published independent pigeonpea molecular linkage maps; hence, it was necessary to examine the consistency between the NRCPB and ICRISAT linkage maps. The numbering used for linkage groups in the consensus linkage map described here is consistent with the published NRCPB linkage map [11]. The nucleotide sequences of 932 markers in the present map were compared using a BLAST search of the ICRISAT pigeonpea genome pseudomolecules CcLG1-CcLG11. Clear identities were obtained for only 391 of the markers, indicating that the ICRISAT pseudomolecules cover only approximately 41% of the pigeonpea genome when considering random genomic distribution of the NRCPB markers. Furthermore, each of the NRCPB linkage groups matched with an ICRISAT linkage group, but with different numbering except for NRCPB LG2, which had identical numbering as that in the two maps (Table 3). Interestingly, ICRISAT CcLG2, in addition to having similarity to the NRCPB LG2, also matched with NRCPB LG6. This suggests that either NRCPB LG2 and LG6 were falsely separated into two different chromosomes, or more likely, because of the very tightly





**Fig 2. Graphical representation of correspondence between component maps and consensus map.** Scatter plot showing the extent of the correlations between the consensus genetic map and component genetic maps. The markers from component genetic maps of the populations A, G, and R are shown as blue diamonds, red squares, and green triangles, respectively.

<https://doi.org/10.1371/journal.pone.0179747.g002>



**Fig 3. Correspondence among consensus and component maps of LG1.** Depiction of common markers between the consensus and individual genetic maps for the LG01 pigeonpea linkage group aligned together using the Genetic Map Comparator software.

<https://doi.org/10.1371/journal.pone.0179747.g003>

**Table 3. Correspondence between NRCPB pigeonpea linkage groups and ICRISAT chromosome pseudomolecules (CcLG1-11).**

Linkage Group	NRCPB linkage map		ICRISAT Pseudomolecule	
	No. of markers	No. (%) of markers matching in ICRISAT pseudomolecules	Linkage group (no. of matched markers)	Size of the best matched LG (Mb)
LG1	127	49 (38)	<b>LG11(44)</b> , LG02(02), LG07(02), LG06(01)	48
LG2	144	77(53)	<b>LG02(51)</b> , LG03(11), LG07(04), LG09(03) LG11 (03), LG06(02), LG10(02), LG01(01)	36
LG3	52	23(44)	<b>LG08(18)</b> , LG02(03), LG09(01), LG06(01)	20
LG4	89	40(44)	<b>LG03(26)</b> , LG11(09), LG08(02), LG10(02), LG02 (01)	29
LG5	78	29(37)	<b>LG07(13)</b> , LG11(06), LG03(03), LG02(03), LG01 (02), LG08(01)	19
LG6	71	16(22)	<b>LG02(11)</b> , LG09(02), LG01(02), LG11(01)	36
LG7	94	25(26)	LG11(05), LG08(05), LG01(04), LG09(04), LG05 (03), LG10(02), LG02(01), LG06(01)	-
LG8	63	32(50)	<b>LG06(27)</b> , LG11(20), LG10(02), LG07(01)	23
LG9	58	26(44)	<b>LG10(10)</b> , LG11(03), LG01(02), LG02(01), LG03 (01), LG08(01)	22
LG10	67	36(53)	<b>LG04(24)</b> , LG06(04), LG03(03), LG11(02), LG10 (02)	12
LG11	89	38(42)	<b>LG01(28)</b> , LG10(05), LG11(02), LG08(02), LG02 (01)	17
<b>Total</b>	<b>932</b>	<b>391 (41%)</b>		

<https://doi.org/10.1371/journal.pone.0179747.t003>

linked markers in the NRCPB genetic map, the ICRISAT pseudomolecules for CcLG2 may in fact be a fusion of two distinct chromosomes. To resolve this potential inaccuracy, we took five markers from each of the terminal ends of NRCPB LG2 and NRCPB LG6, reanalyzed using JoinMap that have shown that these markers belong to two different groups and that the ICRISAT CcLG2 was probably a false fusion of the two linkage groups. There was a large number of cross-matching of markers from a single NRCPB linkage group with multiple pseudomolecules of ICRISAT, suggesting poor assembly of the pseudomolecules. This needs further assessment to generate a high-quality reference genome for pigeonpea.

## Discussion

Availability of genomic resources including high-resolution genetic and physical maps is essential for accelerating the breeding of improved crop varieties. In this study, we report an intraspecific consensus linkage map based on three different  $F_2$  mapping populations (A, R, and G) of pigeonpea that surpasses the previously published genetic maps with a significantly higher number (932) of markers, including a large number of highly reproducible SNP markers covering a total map length of 1,411.83 cM. We expected a high proportion of polymorphic loci in population A because the SNPs in the assays were identified after *in silico* comparison of the parental varieties Asha and UPAS 120 transcriptome sequences. Although the observed 20.87% polymorphism rate in the GoldenGate assay for population A was less than expected due to heterogeneous samples used in the generation of transcriptome data, it was still significantly higher than the populations R and G, where *in-silico* polymorphism searched was not feasible.

Unfortunately, a majority of RAD-GBS markers were discarded owing to a high number of missing data in the progeny. A low level of useful polymorphism was observed with the RAD-GBS sequencing method as compared to the GoldenGate genotyping method, mainly because it was difficult to find a large number of common markers in the population genotyped by RAD sequencing.

The linkage map of population G has already been published using MapDisto software. Hence, we re-constructed this map using JoinMap for uniformity between component maps and found that the loci order and their map positions were quite similar to that of the published map, with the exception of the exclusion of seven markers and inclusion of two additional markers. These differences resulted in a total of 291 markers in our new component map of population G, instead of 296 markers in the previous map.

The success of building a dense consensus linkage map by merging multiple individual genetic maps depends on the availability of common markers between the populations for each of the linkage groups. For this study the nomenclature of markers should be the same and their position should be similar across all populations [26]. We followed the convention of at least three common markers in each linkage group and then proceeded to make the final map using MergeMap in a single run. It was observed that a consensus order was not possible unless the markers with conflicting positions between maps were removed. Hence, the marker order was filtered and rearranged until the most agreeable order was established, incorporating the maximum possible number of retained markers. In this way, a consensus linkage order among markers was achieved for 932 markers and the remaining 158 markers were discarded owing to their conflicting locations between maps. The total adjusted length of the consensus genetic map was 1,411.83 cM covered by the 11 linkage groups. This new consensus map is now has the highest marker density among the published linkage maps of pigeonpea.

The 932 marker sequences included in the present map were compared to the ICRISAT pigeonpea pseudomolecules CcLG1-CcLG11; however, we observed significant matches with

only 391 of the markers, indicating that the ICRISAT pseudomolecules cover only approximately 41% of the pigeonpea genome, considering that there is a random genomic distribution of markers in the NRCPB map. There was one-to-one correspondence in nine of the eleven linkage groups between NRCPB and ICRISAT maps, although the numbering of seven linkage groups was different from that of the ICRISAT linkage groups. There were large numbers of cross matches of markers from a single NRCPB linkage group with multiple ICRISAT pseudomolecules, suggesting that the poor assembly of the pseudomolecules needs further attention to generate a high-quality reference genome for pigeonpea that can be used for genetic and breeding applications.

Our results suggest that the LG2 pseudomolecule is the fusion of two different linkage groups that correspond to the NRCPB linkage groups two and six, which may have been joined together due to a misassembled scaffold. When we analyzed the markers at the junction of NRCPB LG2 and LG6 homology, on the ICRISAT LG2 we found that these markers belong to different linkage groups, as they do not show any linkage to each other. The NRCPB LG7 did not clearly match with any ICRISAT linkage group; but instead it showed matches with several ICRISAT chromosomes.

In summary, a largely SNP-based, high-density, intraspecific consensus linkage map of the pigeonpea genome was constructed, which included 932 loci with an average marker interval of 1.51 cM. This is the first map of pigeonpea using RAD-SNP markers and it will be helpful in genetic mapping and breeding applications for this species and for anchoring the reference genome of pigeonpea. Furthermore, the GBS-RAD SNP approach was found to be tedious and the least productive because a large proportion of SNPs was excluded from the map due to a lack of genotyping data owing to low fold sequence coverage. The present map is the first high-density, intraspecific linkage map of pigeonpea that is based on a large number of genic-SSR and SNP markers that cover a high genome length of 1,411 cM. Our results also emphasize the need for utilizing SNP markers for the creation of ultra high-density linkage maps of pigeonpea because of the poor polymorphism of SSR markers. With the availability of a high-density consensus linkage map of pigeonpea, molecular breeding applications for pigeonpea will be improved by the use of DNA markers linked to the traits of agronomic importance.

## Supporting information

**S1 Table. Sequence information of the TASSEL data used in population A.**

(XLS)

**S2 Table. Markers collocated at 17 positions in the consensus map.**

(XLS)

**S1 Fig. Details of the component genetic map of Pop A.**

(TIF)

**S2 Fig. Details of the component genetic map of Pop R.**

(TIF)

**S3 Fig. Details of the component genetic map of Pop G.**

(TIF)

**S4 Fig. Correspondence among consensus and component maps of Linkage Groups LG02 to LG11.**

(PDF)

## Acknowledgments

We gratefully acknowledge the financial and infrastructural support of Indian Council of Agricultural Research (ICAR).

## Author Contributions

**Conceptualization:** Nagendra Kumar Singh.

**Data curation:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal.

**Formal analysis:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal.

**Funding acquisition:** Nagendra Kumar Singh.

**Investigation:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal.

**Methodology:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal, Shefali Bhutani, Sutapa Dutta, Giriraj Kumawat, Bikram Pratap Singh.

**Project administration:** Nagendra Kumar Singh.

**Resources:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal, Bikram Pratap Singh, A. K. Chaudhary, Rekha Yadav, Subhojit Datta, Ranjeet S. Raje.

**Software:** Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal.

**Supervision:** Tilak R. Sharma, Nagendra Kumar Singh.

**Validation:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh.

**Visualization:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh.

**Writing – original draft:** Sheetal Arora, Ajay Kumar Mahato, Sangeeta Singh, Paritra Mandal, Amitha Mithra Sevanthi.

**Writing – review & editing:** K. Gaikwad, Amitha Mithra Sevanthi, Subhojit Datta, Tilak R. Sharma, Nagendra Kumar Singh.

## References

1. Greilhuber J, Obermayer R. Genome size variation in *Cajanus cajan* (Fabaceae): a reconsideration. *Plant Syst Evol.* 1998; (212) 135–141. <https://doi.org/10.1007/BF00985225>
2. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al. The Challenge of Food Security. *Science.* 2012; 327: 812. <https://doi.org/10.4337/9780857939388>
3. Cannon SB, May GD, Jackson SA. Three Sequenced Legume Genomes and Many Crop Species: Rich Opportunities for Translational Genomics. *Plant Physiol.* 2009; 151: 970–977. <https://doi.org/10.1104/pp.109.144659> PMID: 19759344
4. Yang S, Pang W, Ash G, Harper J, Carling J, Wenzl P, et al. Low level of genetic diversity in cultivated Pigeonpea compared to its wild relatives is revealed by diversity arrays technology. *Theor Appl Genet.* 2006; 113: 585–595. <https://doi.org/10.1007/s00122-006-0317-z> PMID: 16845522
5. Odeny DA, Jayashree B, Gebhardt C, Crouch J. New microsatellite markers for pigeonpea (*cajanus cajan* (L.) millsp.). *BMC Res Notes.* 2009; 2: 35. <https://doi.org/10.1186/1756-0500-2-35> PMID: 19284532
6. Singh NK, Gupta DK, Jayaswal PK, Mahato AK, Dutta S, Singh S, et al. The first draft of the pigeonpea genome sequence. *J Plant Biochem Biotechnol.* 2012; (21) 98–112. <https://doi.org/10.1007/s13562-011-0088-8> PMID: 24431589
7. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter J, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotechnol.* 2012; 30: 83–89. <https://doi.org/10.1038/nbt.2022> PMID: 22057054
8. Yang SY, Saxena RK, Kulwal PL, Ash GJ, Dubey A, Harper JDI, et al. The first genetic map of pigeon pea based on diversity arrays technology (DArT) markers. *J Genet.* 2011; 90: 103–109. <https://doi.org/10.1007/s12041-011-0050-5> PMID: 21677394

9. Sahu AR, Mishra RR, Rath SC, Panigrahi J. Construction of interspecific genetic linkage map of pigeonpea using SCoT, RAPD, ISSR markers and simple inherited trait loci. *Nucl.* 2015; 58: 23–31. <https://doi.org/10.1007/s13237-015-0129-0>
10. Bohra A, Saxena RK, Gnanesh BN, Saxena K, Byregowda M, Rathore A, et al. An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor Appl Genet.* 2012; 125: 1325–1338. <https://doi.org/10.1007/s00122-012-1916-5> PMID: [22772726](https://pubmed.ncbi.nlm.nih.gov/22772726/)
11. Kumawat G, Raje RS, Bhutani S, Pal JK, Mithra ASVCR, Gaikwad K, et al. Molecular mapping of QTLs for plant type and earliness traits in pigeonpea (*Cajanus cajan* L. Millsp.). *BMC Genet.* 2012; 13: 84. <https://doi.org/10.1186/1471-2156-13-84> PMID: [23043321](https://pubmed.ncbi.nlm.nih.gov/23043321/)
12. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 2011; 6: 1–10. <https://doi.org/10.1371/journal.pone.0019379> PMID: [21573248](https://pubmed.ncbi.nlm.nih.gov/21573248/)
13. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980; 8: 4321–4326. <https://doi.org/10.1093/nar/8.19.4321> PMID: [7433111](https://pubmed.ncbi.nlm.nih.gov/7433111/)
14. Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, et al. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 2011; 11: 17. <https://doi.org/10.1186/1471-2229-11-17> PMID: [21251263](https://pubmed.ncbi.nlm.nih.gov/21251263/)
15. Dutta S, Mahato AK, Sharma P, Raje RS, Sharma TR, Singh NK: Highly variable “Arhar” simple sequence repeat markers for molecular diversity and phylogenetic studies in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Plant Breed.* 2013; 132: 191–196. <https://doi.org/10.1111/pbr.12014>
16. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007; 23: 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> PMID: [17586829](https://pubmed.ncbi.nlm.nih.gov/17586829/)
17. Ooijen Van. JoinMap® 4.0, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands.
18. Lorieux M. MapDisto: Fast and efficient computation of genetic linkage maps. *Mol Breed.* 2012; 30: 1231–1235. <https://doi.org/10.1007/s11032-012-9706>
19. MergeMap Online. [<http://mergemap.org>]
20. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002; 93: 77–78. <https://doi.org/10.1093/jhered/93.1.77> PMID: [12011185](https://pubmed.ncbi.nlm.nih.gov/12011185/)
21. Singh NK, Raghuvanshi S, Srivastava SK, Gaur A, Pal AK, Dalal V, et al. Sequence analysis of the long arm of rice chromosome 11 for rice-wheat synteny. *Funct Integr Genomics.* 2014; 4: 102–117. <https://doi.org/10.1007/s10142-004-0109-y> PMID: [15085449](https://pubmed.ncbi.nlm.nih.gov/15085449/)
22. Chakravarti A, Lasher LK, Reefer JE. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics.* 1991; 128: 175–182. PMID: [2060775](https://pubmed.ncbi.nlm.nih.gov/2060775/)
23. Gnanesh BN, Bohra A, Sharma M, Byregowda M, Pandey S, Wesley V, et al. Genetic mapping and quantitative trait locus analysis of resistance to sterility mosaic disease in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *J. Field Crops Research.* 2011; 123. pp. 53–61.
24. Holtz Y, David JL, Ranwez V. The genetic map comparator: a user-friendly application to display and compare genetic maps. *Bioinformatics.* 2017; 33 (9): 1387–1388. <https://doi.org/10.1093/bioinformatics/btw816>. <https://doi.org/10.1093/bioinformatics/btw816> PMID: [28453680](https://pubmed.ncbi.nlm.nih.gov/28453680/)
25. Varshney RK, Penmetsa R V., Dutta S, Kulwal PL, Saxena RK, Datta S, et al. Pigeonpea genomics initiative (PGI): An international effort to improve crop productivity of pigeonpea (*Cajanus cajan* L.). *Mol Breed.* 2010; 26: 393–408. <https://doi.org/10.1007/s11032-009-9327-2> PMID: [20976284](https://pubmed.ncbi.nlm.nih.gov/20976284/)
26. Ratnaparkhe MB, Gupta VS, Ven Murthy MR, Ranjekar PK. Genetic fingerprinting of pigeonpea [*Cajanus cajan* (L.) Millsp.] and its wild relatives using RAPD markers. *Theor Appl Genet.* 1995; 91: 893–898. <https://doi.org/10.1007/BF00223897> PMID: [24169974](https://pubmed.ncbi.nlm.nih.gov/24169974/)