

## InDels in an intronic region of gene *Ccsmd04* coding for dormancy/auxin-associated protein control sterility mosaic disease resistance in pigeonpea<sup>☆</sup>

Sagar Krushnaji Rangari<sup>a,b</sup>, Namita Dube<sup>a</sup>, Vinay Sharma<sup>a</sup>, Sunil S. Gangurde<sup>a</sup>, Mamta Sharma<sup>a</sup>, Prakash I. Gangashetty<sup>a</sup>, Rachit K. Saxena<sup>a,c</sup>, Abhinav Moghiya<sup>a</sup>, Vinay K. Sharma<sup>b</sup>, Karma L. Bhutia<sup>b</sup>, Ravi Kant<sup>d</sup>, Mahendar Thudi<sup>e</sup>, S.J. Satheesh Naik<sup>f</sup>, Aditya Pratap<sup>f</sup>, Girish P. Dixit<sup>f</sup>, Sean Mayes<sup>a</sup>, Manish K. Pandey<sup>a,\*</sup> 

<sup>a</sup> Center of Excellence in Genomics & Systems Biology (CEGSB), and Center for Pre-Breeding Research (CPBR), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, Hyderabad, India

<sup>b</sup> Department of Agricultural Biotechnology and Molecular Biology, Dr Rajendra Prasad Central Agricultural University, Pusa 848125, Bihar, India

<sup>c</sup> ICAR-Indian Agricultural Research Institute, Dirpai Chapari 787034, Assam, India

<sup>d</sup> Department of Genetics and Plant Breeding, Tirhut College of Agriculture (TCA), Dholi, Dr Rajendra Prasad Central Agricultural University, Pusa 848125, Bihar, India

<sup>e</sup> College of Agriculture, Family Sciences and Technology, 1005 State University Dr Fort Valley State University, Fort Valley, GA, USA

<sup>f</sup> ICAR-Indian Institute of Pulses Research (IIPR), Kanpur, Kalyanpur 208024, Kanpur, Uttar Pradesh, India

### ARTICLE INFO

**Keywords:**  
Pigeonpea  
QTL-Seq  
ΔSNP-index  
InDels  
KASP genotyping

### ABSTRACT

Sterility mosaic disease (SMD) presents a significant challenge to pigeonpea cultivation in the Indian subcontinent, potentially leading to total crop failure. The development of diagnostic molecular markers for SMD resistance can aid in developing SMD-resistant varieties. In this context, a QTL-seq approach identified genomic regions associated with SMD resistance using a recombinant inbred line population generated from ICP8863 × ICPL87119. In total, 6105 high-confidence variants were identified in the genomic region, *smdCc04* based on the ΔSNP index. A genomic region *smdCc04* on chromosome Cc04 spans 3.2 Mb (9.3–12.5 Mb) comprises 5 missense variants and 9 InDels. A total of 211 candidate genes were identified from this region. 1 bp insertion, 21 bp insertion, 9 bp deletion, and 3 bp insertion at different intronic positions in 22 susceptible genotypes leads to downregulation of *Dormancy/auxin associated protein (Ccsmd04 gene)* resulting in loss of signaling in disease resistance pathways. The identified sites recognize four important disease and plant growth related transcription factors. A total of 4 InDels and 8 SNPs were validated from *smdCc04* genomic region using whole genome re-sequencing data and kompetitive allele specific polymerase chain reaction (KASP) genotyping in resistant and susceptible pigeonpea genotypes. Gene-based markers *snpCC00149* and *snpCC00150* present in the genomic region *smdCc04* can be used in pigeonpea breeding programs for developing SMD resistant varieties.

### 1. Introduction

Pigeonpea or red gram (*Cajanus cajan* L.) emerging as a climate-resilient legume due to its ability to withstand a wide temperature range from 8.2 °C to 43 °C and its ability to grow with a limited annual rainfall of 65 cm [1–3]. It serves as an important source of protein and considered a ‘life savior crop’ in Asia and Africa due to its significance in food and feed production [2]. Globally pigeonpea is cultivated on 6.3

million hectares area with production of 5.32 million tons [4]. India is the leading producer, contributing 4.34 million, despite this, India imports 0.88 million tons of pigeonpea. The productivity of pigeonpea has significantly been impacted by major diseases such as fusarium wilt (FW), sterility mosaic disease (SMD) and phytophthora blight. SMD, caused by pigeonpea sterility mosaic virus (PSMV) corresponding to the emaraviruses, transmitted by the eriophyid mite (*Aceria cajani*) [5]. >90 % of yield losses have been reported when SMD infects crops at an

<sup>☆</sup> This article is part of a Special issue entitled: ‘Nucleic Acid Research’ published in International Journal of Biological Macromolecules.

\* Corresponding author.

E-mail address: [manish.pandey@icrisat.org](mailto:manish.pandey@icrisat.org) (M.K. Pandey).

<https://doi.org/10.1016/j.ijbiomac.2025.145777>

Received 17 May 2025; Received in revised form 21 June 2025; Accepted 3 July 2025

Available online 9 July 2025

0141-8130/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

early stage [6]. In the major pigeonpea growing areas of Indian sub-continent, FW-resistant varieties are available, although many remain susceptible to SMD. For instance, elite cultivars such as Maruti (South India) and JKM189 (Central India) are susceptible to SMD, further exacerbating the spread of mites and viruses.

Molecular markers have been successfully used in various genomics-assisted breeding programs in legumes for improving traits such as late leaf spot (LLS) and rust in groundnut [7–9] and fusarium wilt and ascochyta blight in chickpea [7]. Therefore, the development of diagnostic markers is a key to accelerating breeding efforts through marker-assisted breeding. To develop reliable and promising markers, QTL-Seq is one of the most rapid and cost-effective approaches and can be implemented in early ( $F_2$ ) as well as in advanced generations ( $F_6$ ). This approach was initially developed in rice and used to identify the genomic regions associated with blast resistance and seedling vigor [10]. Subsequently, it is deployed in a variety of crops for multiple traits [11] including groundnut for rust and LLS resistance [12].

Earlier, in pigeonpea, QTL-seq was used to identify genomic regions for FW and SMD resistance in recombinant inbred line (RIL) population (ICPL20096 × ICPL332) but no success [13]. However, the use of common resistant and susceptible bulks for both FW and SMD might have hindered with QTL detection. The same dataset was then used to identify 26 genome-wide InDels affecting 26 candidate genes [14]. Again, three populations [ICPL20096 × ICPL332, ICPL20097 × ICP8863, and ICP8863 × ICPL87119 ( $F_2$ ) population] were used to map genomic regions for SMD resistance using genotyping-by-sequencing (GBS) [15]. Among these, two major QTL regions were identified in ICP8863 × ICPL87119 on chromosome Cc03 ( $qSMD3.1$ ) with 34.3 % PVE and LOD 2.8, and on chromosome Cc11 ( $qSMD11.3$ ) with 24.2 % PVE and LOD 5.8. Therefore, to further investigate these QTL regions, we advanced ICP8863 × ICPL87119  $F_2$  population to  $F_8$  generation to develop a RIL population comprised of 153 lines. All the above mentioned studies were performed using the older version of ‘Asha’ pigeonpea reference genome v1.0, assembled using short reads sequencing [16], therefore the identified genomic regions for SMD resistance have shown less applicability. Interestingly, in this study, we have used the improved telomere-to-telomere ‘Asha’ reference genome v2.0 assembled using Hi-Fi sequencing in combination with Hi-C (High throughput chromosome conformation capture) sequencing approach [17].

In this study, QTL-Seq analysis using improved version of pigeonpea genome assembly and advanced RIL population (ICP8863 × ICPL87119) identified highly significant ( $\geq 99$  % confidence interval) genomic region on chromosome Cc04, harboring non-synonymous variants (SNPs and InDels) and potential disease resistance genes. From this study, KASP markers were designed from identified genomic regions and validated in the entire RIL population (ICP8863 × ICPL87119) as well as on the 21 resistant and 22 susceptible genotypes of WGRS panel to deploy in genomics-assisted breeding programs to improve SMD resistance.

## 2. Material and methods

### 2.1. Background of the parents and RIL population

In this study, a RIL population comprised of 153 RILs was developed from a cross between susceptible ICP8863 (Maruti) (SMD severity score: 100 %) and resistant ICPL87119 (Asha) (SMD severity score: 2 %) genotypes. RIL population and both parents were phenotyped for SMD severity at ICRISAT, Hyderabad, India. The male parent ICPL87119 (Asha), is a medium-duration, high-yielding, and resistant to FW and SMD. While the ICP8863 (Maruti), a popular FW resistant cultivar but susceptible to SMD was used as a female parent.

### 2.2. Phenotyping for the sterility mosaic disease and construction of bulks

RIL population along with parents were planted in sick plot specially designed for SMD screening at ICRISAT, Hyderabad (17.5111° N, 78.32752° S) during rainy-2023. 15 days old plants were inoculated at the two-leaf stage using the leaf stapling technique [18]. SMD incidence was recorded at three growth stages in one particular season. The first observation was recorded for 30 days post-inoculation (DPI), followed by the second observation at 60 DPI and third was recorded at the flowering stage. Disease incidence scale ranging from 0 % (highly resistant) to 100 % (highly susceptible) was used to categorize disease incidence into four groups: 0–10 % (resistant), 11–20 % (moderately susceptible), 21–40 % (susceptible), >40 % (highly susceptible). The overall disease incidence was calculated using the below-mentioned formula [18]:

$$\text{Disease incidence(\%)} = \frac{\text{Number of disease seedlings}}{\text{Total number of seedlings}} \times 100$$

Based on the phenotyping data for SMD disease incidence (%) generated on the RIL population, a total of 18 resistant and 18 susceptible RILs were used to construct resistant (R-bulk) and susceptible (S-bulk), respectively (Supplementary Table 1, Fig. 1a).

### 2.3. DNA extraction, library construction and sequencing of bulks and parents

High quality DNA was extracted from each line in the RIL population including the parents using the NucleoSpin® 96 Plant II (Macherey-Nagel) Kit, and DNA quality was checked on a 0.8 % agarose gel and quantity was estimated using Nanodrop spectrophotometer (Thermo-fisher Scientific, USA) as explained in Pandey et al. [19]. Based on the SMD disease incidence (%) on 153 RILs, the DNA of 18 resistant and 18 susceptible RILs was pooled to construct R-bulk and S-bulk respectively (Fig. 1a) as detailed in Pandey et al. [12]. Quality of pooled DNA for both bulks was checked on a 0.8 % agarose gel before library preparation. Further, a single library was prepared for each sample using TruSeq DNA Sample Prep kit, and the quality of libraries [Two bulks and susceptible parent (ICP8863)] was checked using Qubit fluorometer and sequenced using the Illumina HiSeq 2500 platform. The sequencing data for resistant parent (ICPL87119) was already available (17).

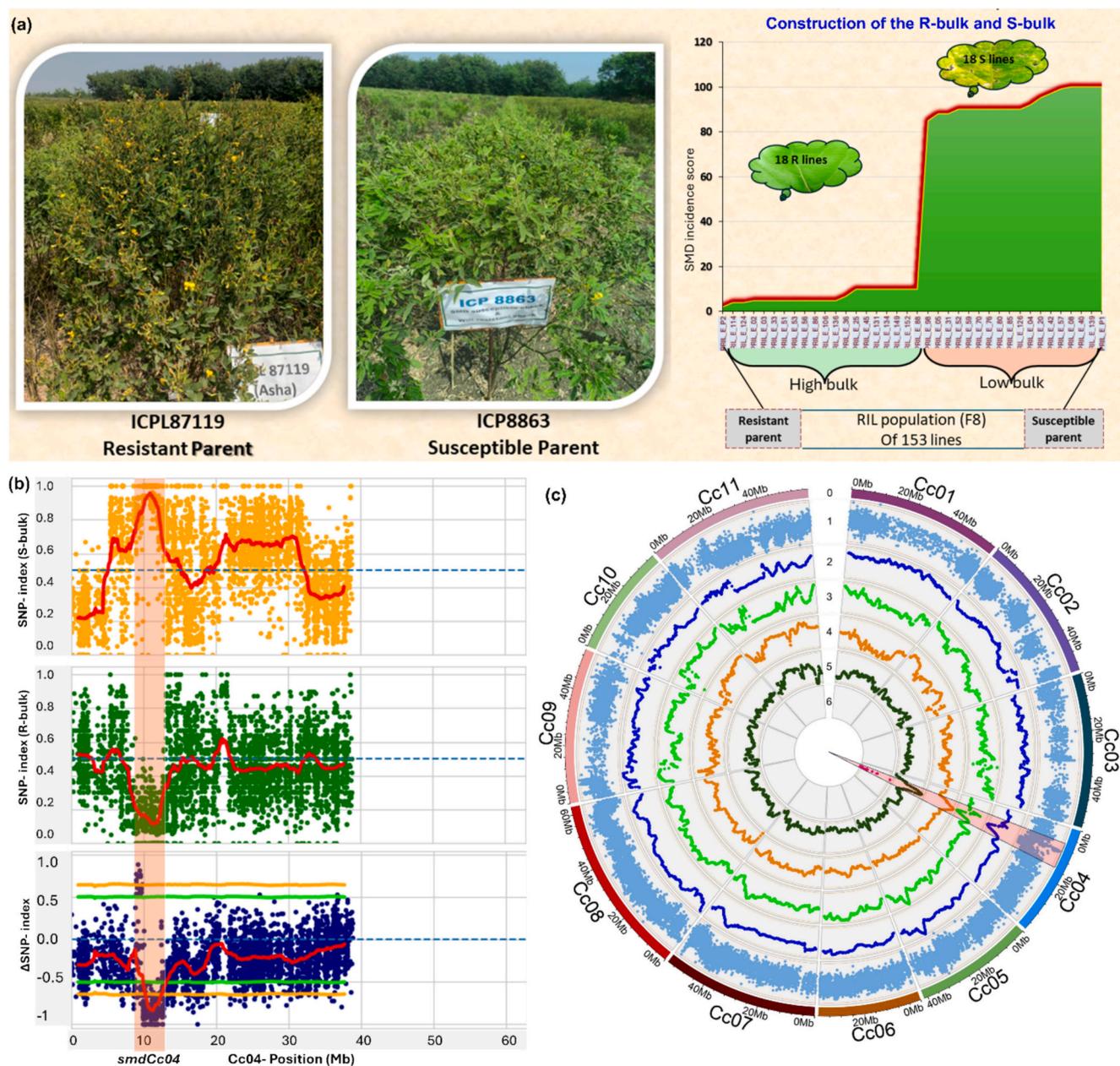
### 2.4. Development of reference guided assembly and bi-directional QTL-Seq analysis

The resistant parent ICPL87119 ‘Asha’ used in this study is already being used for developing reference genome assembly in pigeonpea [16]. Therefore, the susceptible parent, ICP8863 was aligned with the improved reference genome assembly [17] to identify the variants between the parents. The reads of R-bulk and S-bulk were aligned with both parents, and variants were called for both bulks to calculate SNP-index (Table 1, Supplementary Fig. 1, Supplementary Fig. 2). The statistics of generated sequencing reads were estimated using the raspberry tool of NGS-QCbox [20].

QTL-seq analysis was performed by using QTL-seq pipeline developed by Takagi et al. [10]. All the SNPs and InDels above the 99 % confidence interval (CI) and 95 % CI from the significant region (peak region) were extracted as guided by the sliding windows (0.2 Mb). A SNP variation index for each SNP position was calculated for resistant and susceptible bulks by using the following formula [10]:

$$\text{SNP index (at position)} = \frac{\text{Count of alternate base}}{\text{Count of reads aligned}}$$

SNPs with homozygous alleles in susceptible and resistant bulks, were used for the calculation of the  $\Delta$ SNP index. For the calculation of  $\Delta$ SNP index, the SNP index of S-bulk was subtracted from the SNP index of the



**Fig. 1.** QTL-seq analysis and identification of associated genomic region on chromosome Cc04 for SMD resistance **a.** Phenotype of resistant and susceptible parents and construction of resistant or R-bulk and susceptible or S-bulk by selecting extremely resistant and susceptible RILs for sterility mosaic disease from RIL population ICP8863  $\times$  ICPL87119.

**b.** SNP-index and  $\Delta$ SNP index plot of Cc04. SNP-index plot for low bulk (Susceptible) in orange color, high bulk (Resistant) in green color, and  $\Delta$ SNP index plot in blue illustrates the genomic region smdCc04 (highlighted in light orange color). **c.** Circos plot illustrates genome-wide variants identified for SMD resistance in pigeonpea. The tracks go from outside to inside. 0. 11 chromosomes of pigeonpea in different colors in a scale of Mb. 1. Genome-wide variants identified on each chromosome. 2.  $\Delta$ SNP index ( $P > 0.01$ ) using resistant parent ICPL87119. 3.  $\Delta$ SNP index ( $P > 0.01$ ) of susceptible parent ICP8863. 4. SNP index for S-bulk. 5. SNP index for R-bulk. 6. Non-synonymous SNPs identified on chromosome Cc04.

R-bulk (Supplementary Fig. 3). The SNPs having  $\Delta$ SNP index =  $-1$  and near to  $-1$ , were considered as favorable alleles, contributed by resistant parent. While the alleles with  $\Delta$ SNP index =  $1$  and near to  $1$ , were considered as unfavorable alleles contributed by susceptible parent. Among these SNPs and InDels, the synonymous and non-synonymous variants were investigated using snpeff version 5.2c [21] and snpsift version 5.2c [22]. All the genes were extracted from the peak region and their functions were identified using the information available in the annotation file and by using the protein sequences of the respected genes. Preferentially, the function of candidate genes having non-synonymous variants were identified and their relationship with

disease resistance was also identified (Table 2). Circos plot illustrating the SNP index and  $\Delta$ SNP index calculated from both the parents and bulks was developed using ShinyCircoseV2.0 ([https://venyao.xyz/shiny\\_circos/](https://venyao.xyz/shiny_circos/)) [23] (Fig. 1c).

### 2.5. Allele mining for validation of candidate genes in the identified genomic region

The markers from the genomic region (smdCc04) on chromosome Cc04 were mined for favorable and un-favourable alleles in SMD resistant and susceptible genotypes using whole genome re-sequencing (WGRS)

**Table 1**  
Details of sequencing data generated for both bulks and parents.

Sample	Pedigree	Phenotyping of SMD (%)	Illumina sequencing				Data processing		
			Data generated (Gb)	Reads (Million reads)	Read depth (X)	Genome coverage (%)	Trim bases (Gb)	Trim reads (Million reads)	Alignment (%)
<b>Parents</b>									
ICP8863	Selection from landrace	100	12.42	82.27	22.35	92.5	11.67	78.06	98.35
ICPL87119	C11 9 ICP1-6-W3-W	2	25.14	167.58	45.33	91.35	23.77	160.93	98.21
<b>Mapping population</b>									
R-bulk	ICP8863 × ICPL87119	0–10	13.61	90.11	24.42	93.12	86.14	86.14	98.11
S-bulk	ICP8863 × ICPL87119	85–100	15.00	99.33	26.9	93.2	94.35	94.35	98.08
<b>Total</b>			<b>91.31</b>	<b>606.87</b>	<b>164.33</b>		<b>239.7</b>	<b>580.41</b>	

**ICP8863:** Susceptible parent, **ICPL87119:** Resistant parent, **R-bulk:** Resistant bulk, **S-bulk:** Susceptible bulk.

data (Supplementary Table 2). A total of 4 insertions (8bp, 21bp, 3bp, and 3 bp), 2 deletions (1bp and 9 bp), and 8 SNPs from *smdCc04* region were used for allele mining in resistant and susceptible lines. The functional validation of the identified InDels was done by identifying regulatory elements using Plantpan4.0 webtool (<https://plantpan.itps.ncku.edu.tw/>).

### 3. Results

#### 3.1. Phenotypic variability for sterility mosaic disease in RIL population

SMD disease incidence (%) of the RIL population ranged between 5 % to 100 % with a mean of 47.3 %. Based on SMD disease incidence (%), 18 susceptible and 18 resistant RILs were selected from the 153 RILs for the construction of susceptible and resistant bulks. The phenotypic score for the resistant bulk ranged from 5 % to 10 % with a mean of 6.6 %, and for susceptible bulks 85 % to 100 % with a mean of 88 % (Fig. 1; Supplementary Table 1).

#### 3.2. Whole genome sequencing and variant calling

A total of 13.61 Gb data was generated (92.5 % coverage) for the R-bulk with 90.11 million reads at a depth of 24.42X. For S-bulk, 15 Gb data was generated (91.75 % coverage) with 99.33 million reads at a depth of 26.9X. For susceptible parent ICP8863, 12.42 Gb data was generated (92.5 % coverage) with 82.27 million reads at a read depth of 22.35X. The resistant parent ICPL87119 (Asha) was sequenced earlier with coverage (91.35 %) [17]. The trimmed and filtered high-quality reads of the R-bulk (86.14 million reads) and S-bulk (94.35 million reads) mapped against the newly developed reference guided assembly of ICPL87119. The alignment percentage of R-bulk was 98.11 % followed by S-bulk 98.08 %. At a read depth  $\geq 7$  in both the bulks and  $\geq 0.3$  SNP index in either of the bulks were considered as a selection criterion for resistant loci (Table 1).

A total of 603,135 genome-wide variants were identified, with 137,019 heterozygous and 465,916 homozygous variants between R-bulk and S-bulk. A total of 29,780 polymorphic SNPs and 9256 polymorphic InDels identified after aligning the resistant and susceptible bulks with the reference-guided assembly. Minimum 1581 SNPs were obtained on chromosome Cc01 and maximum 4975 SNPs on chromosome Cc08. Similarly, a minimum of 520 InDels were obtained on chromosome Cc03 and maximum 1346 InDels were obtained on chromosome Cc08 (Supplementary Table 3).

#### 3.3. Identification of genomic regions and candidate genes for sterility mosaic disease resistance

QTL-seq analysis identified one potential genomic region on chromosomes Cc04 (*smdCc04*). On chromosome Cc04, the genomic region *smdCc04* of 3.2 Mb (9.3–12.5 Mb) contained 3540 variants including 2324 SNPs, and 1103 InDels. Among these variants, 266 (189 SNPs and 77 InDels) were identified at the 95 % confidence interval (CI) ( $p < 0.05$ ), and 229 (173 SNPs and 56 InDels) were identified at 99 % CI ( $p < 0.01$ ) (Fig. 1c, Supplementary Table 4).

The identified genomic region was mined for candidate gene discovery using ICPL87119 (Asha) genome v2.0. Genes containing genome variants such as non-synonymous mutations, intron splicing region variants, and premature start codon mutations were referred as candidate genes. In particular, non-synonymous variants including missense mutations, frameshift mutations, stop-gained, stop-lost, start-gain, and start-lost variants were considered as potential causative mutations due to their impact on the function of candidate genes. (Supplementary Table 5).

In the genomic region *smdCc04*, on chromosome Cc04, a total of 211 genes were identified, of which 18 are involved in the pathway of SMD resistance mechanism (Table 3, Supplementary Table 6, Supplementary Table 8). In this region, missense variants were corresponding to important candidate genes including *Dormancy/auxin associated protein* (*Cc\_09569*), *exoribonuclease 3-like* (*Cc\_09596*), *E3 ubiquitin-protein ligase* (*Cc\_09614*), *YABBY transcription factor* (*Cc\_09623*), *formin-like protein 3* (*Cc\_09685*), *QWRF motif-containing protein 3-like* (*Cc\_09721*). Splice-region-variants were found in probable *CCR4-associated factor* (*Cc\_09591*), *mitogen-activated protein kinase A-like* (*Cc\_09636*), and *aconitate hydratase* (*Cc\_09636*) genes. Interestingly, six frameshift variants and one missense were identified in the region of *Dormancy/auxin associated protein* (*Cc\_09569*) gene, now named as "**Ccsmd04**" which is one of the signal receptor genes (Table 2, Fig. 2). This gene has seven significant variants, six are InDels and one is missense SNP variation.

#### 3.4. Identification of SNPs and mining favorable alleles in whole genome re-sequencing data of resistant and susceptible genotypes

Allele-mining of putative markers from the *smdCc04* region was performed in the whole genome re-sequencing data of resistant and susceptible genotypes. Alleles from the resistant and susceptible parents were mined in the 19 resistant and 22 susceptible genotypes for Cc04\_9918236 (1bp deletion) and Cc04\_9918564 (21bp insertion) markers. For marker Cc04\_9919016 (9bp insertion) and Cc04\_9923664 (3bp insertion) alleles from the resistant and susceptible parents were mined in the 21 resistant and 22 susceptible genotypes. Moreover, for

**Table 2**

List of non-synonymous/Indel variants and corresponding candidate genes identified for SMD resistance.

SN	Chromosome	Gene ID	Position	ICPL87119	R-bulk	S-bulk	Gene name	Effect	Function of gene
1.	Cc04	<i>Cc_09569</i>	9,926,160	C	C	A	<i>Dormancy/auxin associated protein</i>	Missense variant	Response protein for hormones and play role in disease resistance
2.	Cc04	<i>Cc_09569</i>	9,918,236	CA	CA	C	<i>Dormancy/auxin associated protein</i>	Deletion	Response protein for hormones and play role in disease resistance
3.	Cc04	<i>Cc_09569</i>	9,918,253	A	A	AACG	<i>Dormancy/auxin associated protein</i>	Insertion	Response protein for hormones and play role in disease resistance
4.	Cc04	<i>Cc_09569</i>	9,918,438	TATA	TATA	T	<i>Dormancy/auxin associated protein</i>	Deletion	Response protein for hormones and play role in disease resistance
5.	Cc04	<i>Cc_09569</i>	9,918,564	C	C	CAACATGTACTCGAAGATGTTG	<i>Dormancy/auxin associated protein</i>	Insertion	Response protein for hormones and play role in disease resistance
6.	Cc04	<i>Cc_09569</i>	9,919,016	CATATCCACT	CATATCCACT	C	<i>Dormancy/auxin associated protein</i>	Deletion	Response protein for hormones and play role in disease resistance
7.	Cc04	<i>Cc_09569</i>	9,923,664	T	T	TTTA	<i>Dormancy/auxin associated protein</i>	Insertion	Response protein for hormones and play role in disease resistance
8.	Cc04	<i>Cc_09569</i>	10,885,462	C	C	CAA	<i>Dormancy/auxin associated protein</i>	Insertion	Response protein for hormones and play role in disease resistance
9.	Cc04	<i>Cc_09569</i>	11,562,282	GT	GT	G	<i>Dormancy/auxin associated protein</i>	Deletion	Response protein for hormones and play role in disease resistance
10.	Cc04	<i>Cc_09596</i>	10,230,356	T	T	C	<i>Exoribonuclease 3-like</i>	Missense variant	5'-3' exonuclease type-II is a post transcription silencer
11.	Cc04	<i>Cc_09614</i>	10,563,669	G	G	A	<i>E3 ubiquitin-protein ligase</i>	Missense variant	E3-ubiquitin regulates the expression of the pathogenesis-related (PR) genes
12.	Cc04	<i>Cc_09623</i>	10,697,972	A	A	G	<i>YABBY transcription factor</i>	Missense variant	Activates other disease resistance genes
13.	Cc04	<i>Cc_09685</i>	11,558,588	T	T	C	<i>Formin-like protein 3</i>	Missense variant	Formins are a group of proteins that work with the ARP2/3 complex to help in plants resistant
14.	Cc04	<i>Cc_09721</i>	12,151,350	ACGATC	ACGATC	ACGATCGATC	<i>QWRF motif-containing protein 3-like</i>	Frameshift variant	Play role against pathogen attack

one missense SNP marker Cc04\_9,926,160 allele mining was done in 5 resistant and 5 susceptible genotypes. All these markers were corresponding to *dormancy/auxin associated protein* (Cc\_09569) gene. Auxin significantly influences plant-microbe interactions, particularly between plant hosts and disease-causing pathogenic microorganisms (Fig. 4).

### 3.5. Validation of the InDel and SNP markers and their functional validation

We have identified a total of six InDel markers in the gene *Ccsmd04* coding for *dormancy/auxin associated protein* (Cc\_09569), namely,

**Table 3**List of disease resistance related genes from *smdCc\_04* SMD genomic region.

Gene ID	Start position	End position	Gene annotation
<i>Cc_09532</i>	9,469,600	9,471,933	Pentatricopeptide repeat-containing protein
<i>Cc_09534</i>	9,483,639	9,488,448	Histidine-containing phosphotransferase protein 2
<i>Cc_09544</i>	9,669,963	9,674,076	Serine/threonine-protein phosphatase 2A activator
<i>Cc_09554</i>	9,793,075	9,795,082	F-box/LRR-repeat protein
<i>Cc_09569</i>	9,916,754	9,926,688	Dormancy/auxin associated protein
<i>Cc_09570</i>	9,948,089	9,949,664	Senescence regulator
<i>Cc_09571</i>	9,960,261	9,961,782	Pollen ole-e1
<i>Cc_09572</i>	9,975,376	9,976,586	NAC domain-containing protein 90
<i>Cc_09603</i>	10,284,457	10,290,444	Serine/threonine-protein kinase Nek5
<i>Cc_09604</i>	10,307,910	10,310,363	Apoptosis-inducing factors like A
<i>Cc_09607</i>	10,332,542	10,336,694	Receptor protein kinase
<i>Cc_09613</i>	10,529,538	10,530,964	E3 ubiquitin-protein ligase PUB24-like
<i>Cc_09631</i>	10,816,217	10,821,035	Condensin complex subunit 2-like isoform X1
<i>Cc_09642</i>	10,943,300	10,944,491	Aquaporin SIP1-1-like
<i>Cc_09657</i>	11,127,795	11,133,797	Glutamate receptor
<i>Cc_09671</i>	11,318,596	11,320,919	Peroxidase 3-like
<i>Cc_09699</i>	11,697,844	11,700,609	G-type lectin S-receptor-like serine/threonine-protein kinase
<i>Cc_09704</i>	11,832,924	11,834,516	Myb-related protein Myb4

*Cc04\_9918236*, *Cc04\_9918564*, *Cc04\_9919016*, *Cc04\_9923664*, *Cc04\_9918429*, and *Cc04\_9918252* (Fig. 2). Among these, 4 are InDel markers validated in WGRS sequencing panel, namely, *Cc04\_9918236*, *Cc04\_9918564*, *Cc04\_9919016*, *Cc04\_9923664*. These four functional markers are housed in the intronic region of the *dormancy/auxin associated protein* (*Cc\_09569*). *Cc04\_9918236* is identified in Intron1, *Cc04\_9918564* and *Cc04\_9919016* in Intron2, and *Cc04\_9923664* is in Intron4. InDel *Cc04\_9918236* disrupted the transcription factor recognition site (caaaaAAAAAaatgc) for RTV1 transcription factor, InDel *Cc04\_9918564* disrupted the transcription factor site (CAACAG) for MYB2 transcription factor, InDel *Cc04\_9919016* disrupted the transcription factor site (ACACATAT) for bHLH transcription factor, and InDel *Cc04\_9923664* disrupted the transcription factor site (TTTTTttttc) for the AT-hook (Fig. 4a).

We have identified, developed and validated 15 SNPs (KASP assay). Among 15 SNPs, a total of 8 markers (snpCC00149, snpCC00150, snpCC00151, snpCC00152, snpCC00153, snpCC00154, snpCC00155, and snpCC00156) were validated on the validation panel. Three SNP markers (*Cc04\_9920859*, *Cc04\_9926160*, and *Cc04\_9926339*) are housed in the *dormancy/auxin associated protein* (*Cc\_09569*). While rest of the five SNP markers (*Cc04\_9933395* (snpCC00154), *Cc04\_9934586*, *Cc04\_9940694*, *Cc04\_9945283*, and *Cc04\_9968188* (snpCC00156)) are housed in the intergenic region between *dormancy/auxin associated protein* (*Cc\_09569*), senescence regulator (*Cc\_09570*), and protein ole e-1 (*Cc\_09571*). A total of 151 lines (out of 153 lines) of the ICP8863 × ICPL87119 RIL population were sent for validation. A total of 69 RIL resistant lines (out of 76 resistant lines) have been confirmed with the presence of resistant allele, and 58 susceptible lines (out of 75 susceptible lines) have been confirmed with the presence of susceptible allele. 11 lines have shown heterozygous alleles (Fig. 5 and Supplementary Table 9). Using this data, PVE of 52 % was calculated for *smdCc04* region. Resistant and susceptible bulks used in this study contain 18 resistant and 18 susceptible lines, respectively. Out of 18 resistant lines, 16 have shown resistant alleles, while out of 18 susceptible lines, 16 have shown susceptible alleles (Fig. 3.b). Additionally, 87 putative variants were also identified from *smdCc04* region (Supplementary Table 7).

#### 4. Discussions

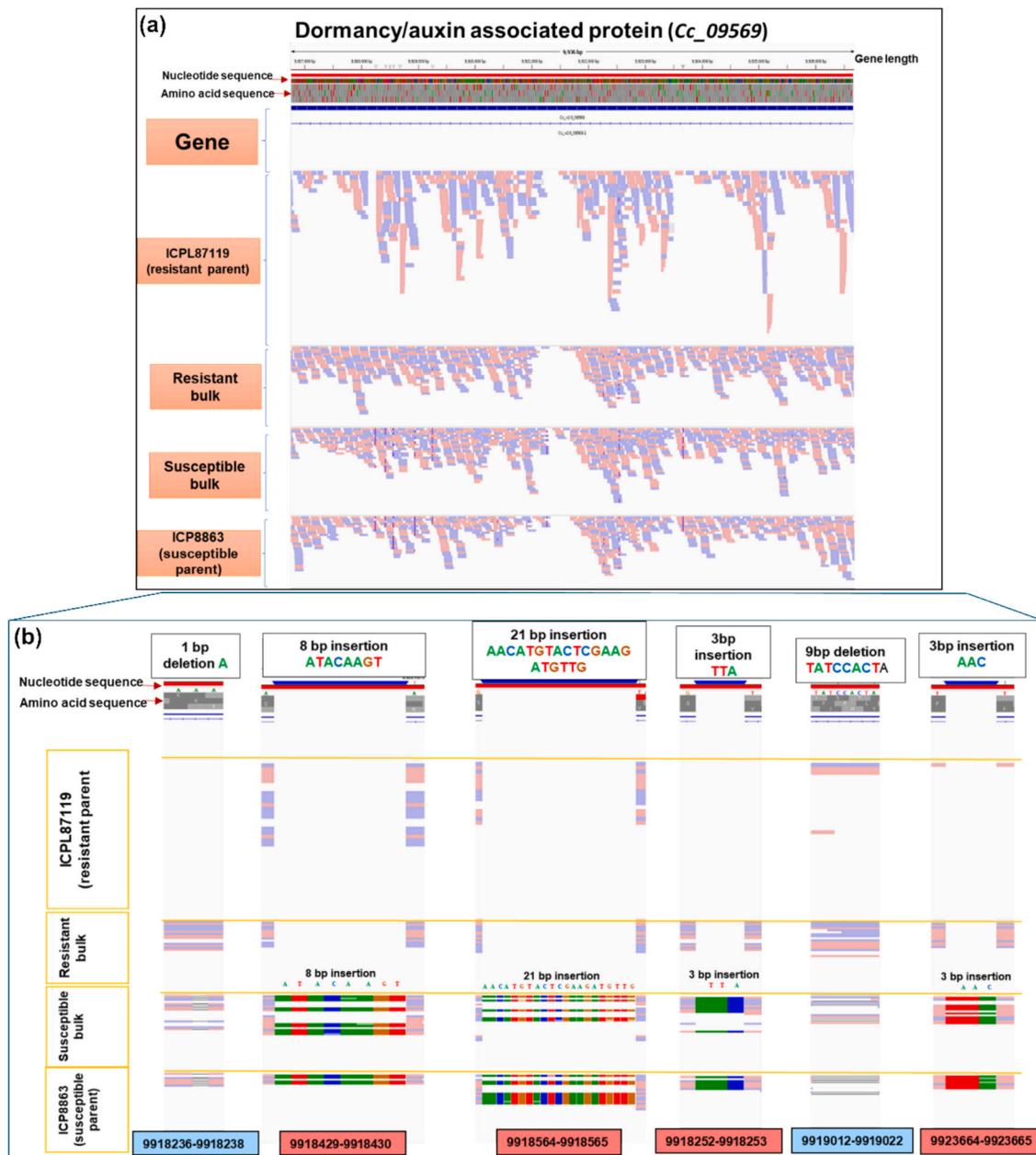
Pigeonpea is no longer an orphan crop and has now ample genomic resources for gene discovery and genomic-assisted breeding (GAB). SMD

is a devastating disease that results >90 % yield penalty in the Indian sub-continent [6]. It is essential to identify associated genome regions resistant genes from various resistance sources and introgress them into elite cultivars to mitigate resulting yield losses due to SMD. The whole genome sequencing and re-sequencing approaches have proven to be reliable and effective for high-resolution trait mapping [24,25]. In the past decade user-friendly bioinformatics pipelines such as QTL-Seq, Mut-Map, and Indel-Seq have increased the accuracy of trait mapping and marker discovery for complex traits in various crops such as rice [10], pigeonpea [13,26] and groundnut [12,27,28]. The QTL-seq approach is resource efficient because it requires sequencing of fewer samples, a time-efficient method that yields high-quality results for the targeted oligogenic trait.

Earlier, QTL-Seq on ICPL20096 × ICPL332 RIL population was used in common resistant and susceptible bulks for FW and SMD to identify genomic regions for FW and SMD using an old version of genome ICPL87119 (Asha) developed using short reads [13]. No QTL could be detected maybe because common bulks were made for both the traits, and the phenotypic effect of QTL was minor/less. QTL-Seq study on ICP8863 × ICPL87119 RIL population identified one genomic region associated with SMD resistance on chromosomes Cc04. We selected this RIL population because major QTLs detected in F<sub>2</sub> mapping population of our previous study [15]. Further, we used the improved genome assembly of 'ICPL87119' (Asha) as a reference genome [17] using RIL population (ICP8863 × ICPL87119, F<sub>8</sub> generation). This RIL population was advanced from the F<sub>2</sub> mapping population previously utilized in the GBS-based QTL identification study for SMD resistance [15]. The initial genome assembly, covering 40.86 % of the genome [16], was recently upgraded to achieve 91.35 % genome recovery [17], enhancing the reliability of sequencing-based approaches.

A region on Cc04 (*smdCc04*), spanning 9.3–12.5 Mb, contains variants above the 99 % confidence interval (CI) and 14 non-synonymous variants (5 SNPs and 9 InDels) impacting the function of six disease resistance-related genes. Among 14 variants, one missense SNP and six InDel variants were identified in the *Dormancy/auxin associated protein coding* gene. This protein is generally responsive to hormones (Such as auxin) involve in biotic defense response [29]. A missense variant (*Cc04\_10563669*) was identified in *E3-ubiquitin protein ligase* (*Cc\_09614*), which plays a critical role in the expression of pathogenesis-related (PR) genes during plant disease resistance from pathogen perception to execution [30,31]. A missense variant at (*Cc04\_10697972*) was identified in *YABBY-protein* (*Cc\_09623*). The *YABBY transcription* family interacts with other disease-resistance genes like *CRR3*, playing a crucial role in plant disease resistance [32]. In pineapple plants, the *YABBY transcription factor* triggers the accumulation of ABA and ethylene in response to pathogen attack [33]. A splice region variant (*Cc04\_10885462*) was identified in the region of mitogen-activated protein kinase A-like (MAPK) (*Cc\_09636*). MAPKs play a crucial role in signaling plant defense against pathogens and microbes by developing MAPK cascade. For instance, in cotton *MPK9*, *MPK13*, and *MPK25* are upregulated when inoculated with verticillium wilt causing *Verticillium dahliae*. Silencing these genes makes the plant more susceptible to *V. dahlia* [34]. The MAK3/MPK6 cascade and calcium signaling pathway, activated during plant immunity, converge on *WRKY33* to regulate the biosynthesis of camalexin, a crucial phytoalexin for disease resistance in *Arabidopsis* [35].

Interestingly, four InDels (*Cc04\_9918236*, *Cc04\_9918564*, *Cc04\_9919016*, and *Cc04\_9923664*) were identified in the *Dormancy/auxin associated protein* (*Cc\_ARP*) (*Cc\_09569*). These mutations are responsible for the disruption of *Cc\_ARP* leading to the hinderance of auxin signaling pathways. The downregulation of *Dormancy/auxin associated protein* enhances the plant susceptibility to infection by *Tobacco mosaic virus* (TMV), *Pectobacterium carotovorum* subsp. *carotovora*, and *Phytophthora parasitica* var. *nicotianae*. Similarly, the knockdown of *TaARP* gene resulted in higher susceptibility to fungus *Blumeria graminis* f. sp. *tritici* in *Triticum aestivum* [36,37]. *Dormancy/auxin associated protein* gene is responsive to the signaling hormones involved in plant disease



**Fig. 2.** Insertion and Deletions identified in associated genomic region of gene *Ccsmd04* for SMD resistance **a.** Sequencing data of resistant and susceptible parents with their respective bulks aligned to genomic region of gene ID *Cc\_09569* in pigeonpea genome version 2 (ICPL87119). Sequence represents nucleotide sequence of genes with their translated amino acid sequences. Red asterisks are for stop codon; light green color coded letter “M” is for methionine. Forward strand reads are blue in color, while reverse strand reads are red color. **b.** Represents Insertion and Deletion for the gene ID *Cc\_09569* in resistant and susceptible parent with their respective bulks. Sequence represents nucleotide sequence of gene and amino acid sequence of respected protein. The range of Deletions is indicated in blue color, while the range of Insertions is indicated in crimson red color. The variant depth and read length are indicated in brick boxes for the identified potential variants in *smdCc04* region.

resistance, growth, and development, such as salicylic acid (SA), jasmonic acid (JA), ethylene, and auxins [29]. It is reported that the activation of auxin signaling enhances the resistance against plant viruses [38] (Fig. 4b). The overexpression of auxin signaling repressor in lines of rice showed susceptible response to the Rice black streaked dwarf virus (RBSDV) [39]. This study also revealed that the jasmonic acid pathway may also be influenced by the auxin-signaling mediated response to the RBSDV in rice. The ARF factor was targeted by several viral proteins and in each case, it is found that these interactions benefited for the viral infections. This suggests that manipulation in auxin signaling is the common pathogenicity strategy in plant viruses [40].

Hormones play a most crucial role in the development of

reproductive tissues by initiating and regulating the growth and maturation of reproductive organs and controlling the timing of important reproductive events [41]. The proper signaling of these hormones is mediated by response proteins such as *Dormancy/auxin associated protein* [42]. Therefore, the sterility in the reproductive tissues of pigeonpea might be due to the loss in the function of *Dormancy/auxin associated protein* gene. *Dormancy/auxin associated protein* (*Cc\_09569*) has 10 variants, Out of these, 4 InDels and 3 SNPs were successfully validated on WGRS sequencing panel and through KASP genotyping respectively. The gene family of this gene shows the important link between response to pathogen and plant growth and development. Therefore, this gene (*Ccsmd04*) could be considered as the most important candidate gene in

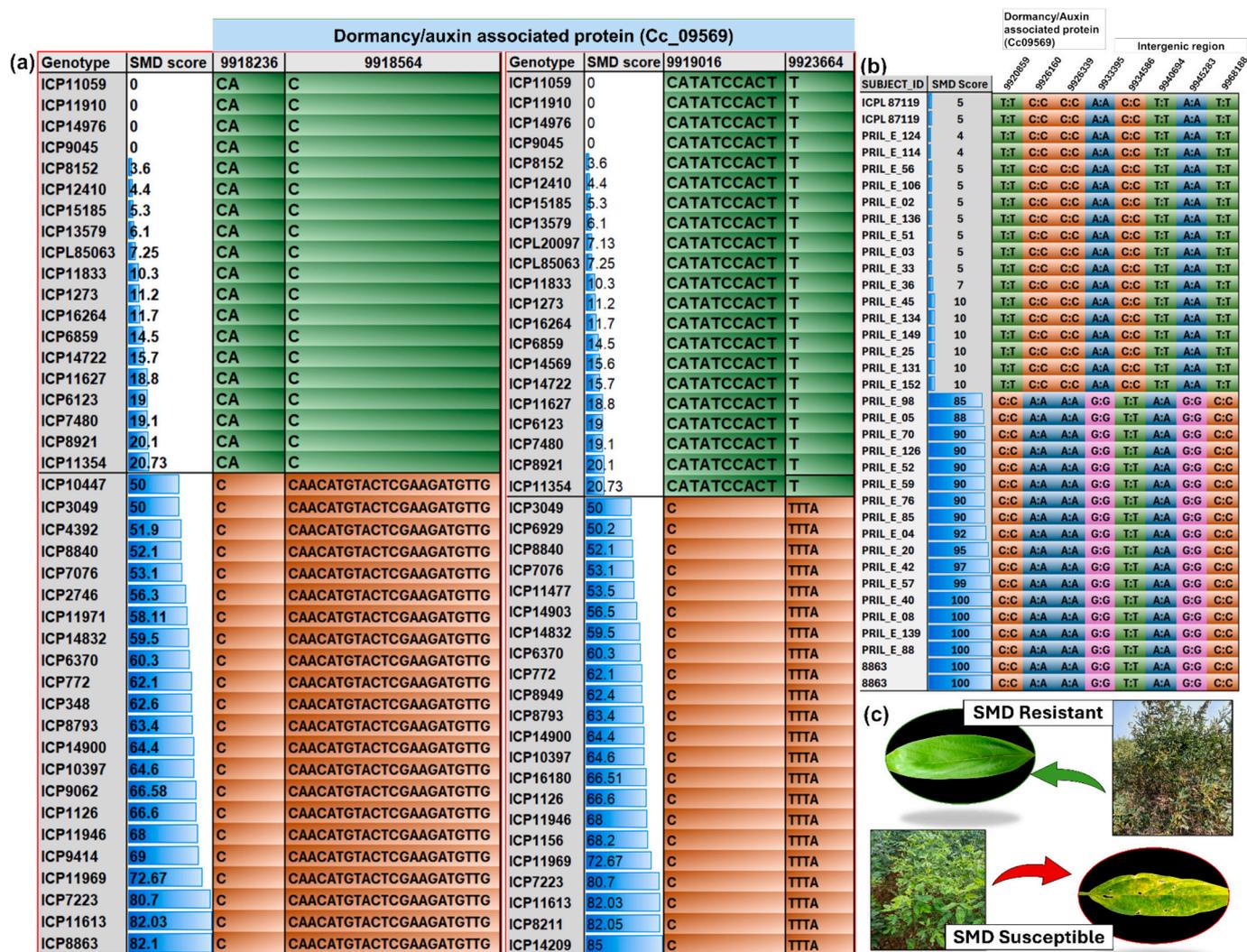
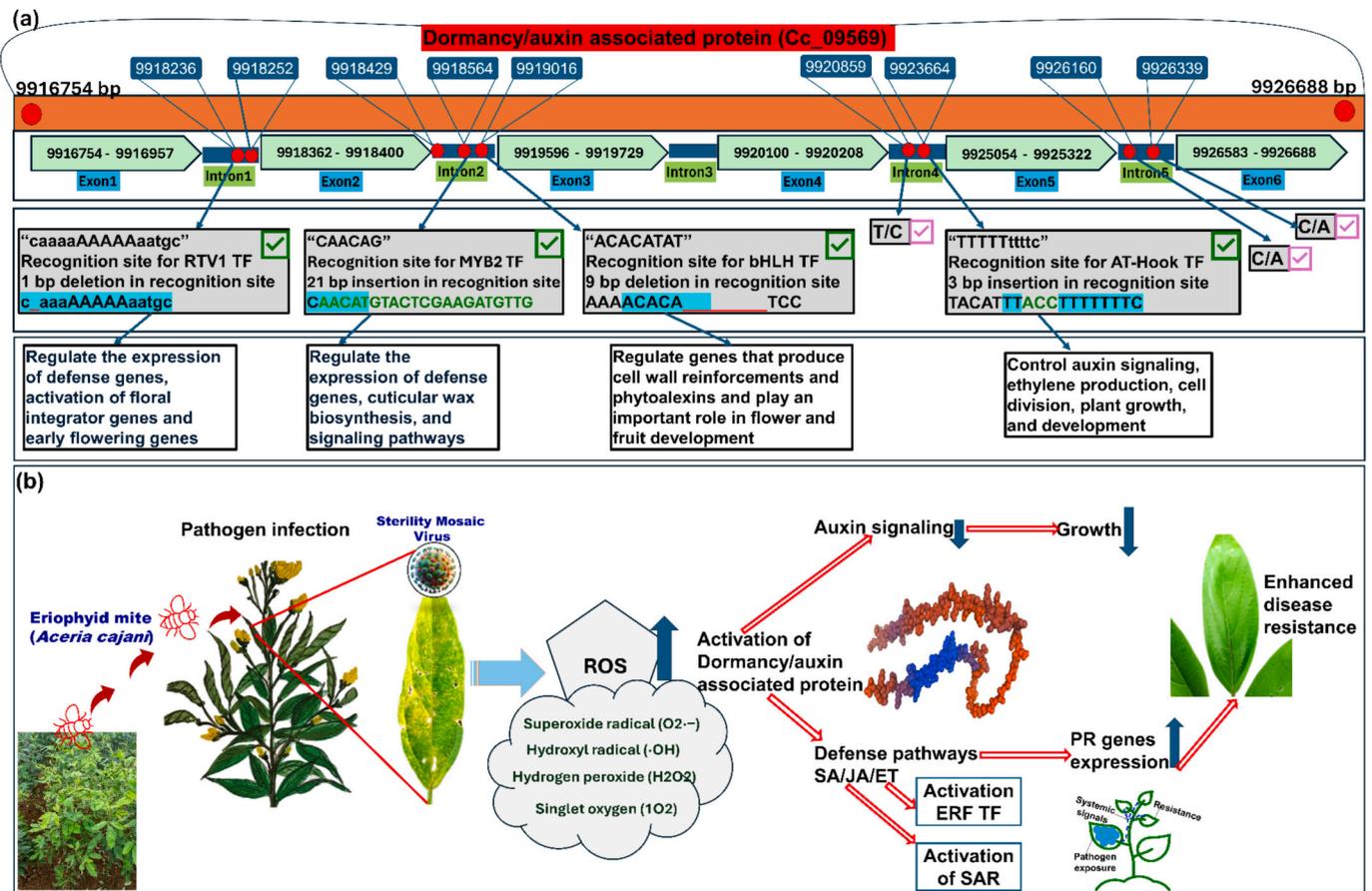


Fig. 3. Validation of InDels and missense SNP present in Dormancy/auxin associated protein (*Cc\_09569*) from *smdCc04* (9.3–12.5 Mb) region a. Allele-mining and validation of the *smdCc04* markers on a set of genotypes contrasting phenotype for SMD. From *Dormancy/auxin associated protein (Cc\_09569)* one Deletion and one Insertion marker validated on a one set of 19 resistant and 22 susceptible genotypes, and one Insertion and one Deletion marker validated on another set of 21 resistant and 22 susceptible genotypes. SMD scores are mentioned in horizontal maya blue color bar graphs. The favorable alleles are mentioned in natural green color and un-favorable calls are indicated in vivid color. At position 9,918,236, 1 bp deletion, at position 9,918,564, 21 bp insertion, at position 9,919,016, 9 bp deletion, while at position 9,923,664, 3 bp insertion was observed in the susceptible genotypes. b. Validation of KASP markers. From *Dormancy/auxin associated protein (Cc\_09569)* three intronic SNP validated on RIL population. Here we have shown the favorable and un-favorable calls in 16 resistant and 16 susceptible lines of ICP8863 × ICPL87119 population respectively. The other five validated markers are housed in the intergenic region. SMD scores are mentioned in horizontal maya blue color bar graphs. In this figure, SMD incidence score between 5 and 10 have shown the favorable calls, while the un-favorable calls are between 85 and 100. c. Phenotype of SMD resistant and SMD susceptible plants.

this region and further transcriptomics, proteomics, and metabolomics studies may be performed to explore the actual mechanism in the pigeonpea resistant and susceptible genotypes. In summary, genomic region *smdCc04* identified for SMD resistance uncovered important disease resistance genes and markers. This study indicated that SMD susceptibility might be due to the disruption of *Dormancy/auxin associated protein* gene (*Ccsmd04*) under infection sterility mosaic disease (SMD).

Identification and validation of the functional markers is the major outcome of any study done by using genomic resources. InDels have potential to disturb the function of associated genes. In our study, we have identified 4 functional InDel markers, namely, Cc04\_9918236, Cc04\_9918564, Cc04\_9919016, Cc04\_9923664. These markers are housed in the intronic region of the *dormancy/auxin associated protein (Cc\_09569)*. Cc04\_9918236 is identified in Intron1, Cc04\_9918564 and Cc04\_9919016 in Intron2, and Cc04\_9923664 is in Intron4. InDel

Cc04\_9918236 have disrupted the transcription factor recognition site (caaaaAAAAAaatgc) for RTV1 transcription factor, InDel Cc04\_9918564 have disrupted the transcription factor site (CAACAG) for MYB2 transcription factor, InDel Cc04\_9919016 have disrupted the transcription factor site (ACACATAT) for Basic helix-loop-helix (bHLH) transcription factor, and InDel Cc04\_9923664 have disrupted the transcription factor site (TTTTTttttc) for the AT-hook. RTV1 transcription factor is a DNA binding domain (B3 domain family) generally controls the flowering time [43]. MYB2 transcription factor regulates the expression of genes that are involved in pathogens and pests' resistance pathways. This TF also facilitates the biosynthesis of lignin, flavonoids, and cuticular wax. Moreover, this TF regulates the signaling of plant hormones involved in plant disease resistance [44]. The bHLH transcription factors fight plant pathogen by sensing H<sub>2</sub>O<sub>2</sub>, lignin and phytoalexin biosynthesis [45]. AT-hook motif nuclear localized (AHL) family plays an important role in plant disease resistance by regulating the expression of defense-related



**Fig. 4.** Identification of significant variants in smdCc04 region and its significance in SMD resistance. a. Four functionally validated InDel markers (mentioned by green check mark) housed in the intronic region of the dormancy/auxin associated protein (*Cc\_09569*). Cc04\_9918236 is identified in Intron1, Cc04\_9918564 and Cc04\_9919016 in Intron2, and Cc04\_9923664 is in Intron4. InDel Cc04\_9918236 disrupted the transcription factor recognition site (caaaaAAAAaatgc) for RTV1 transcription factor, InDel Cc04\_9918564 disrupted the transcription factor site (CAACAG) for MYB2 transcription factor, InDel Cc04\_9919016 disrupted the transcription factor site (ACACATAT) for bHLH transcription factor, and InDel Cc04\_9923664 disrupted the transcription factor site (TTTTTtttc) for the AT-hook. Moreover, three validated SNP markers (mentioned by purple check mark) housed in the intronic region of dormancy/auxin associated protein (*Cc\_09569*). Cc04\_9920859 is identified in Intron4, Cc04\_9926160 and Cc04\_9926339 are identified in Intron5. b. Schematic representation to illustrate the role of smdCc04 region in SMD resistance. Following are the abbreviations used in this schematic representation: ROS- Reactive oxygen species; SA- Salicylic acid; JA- Jasmonic acid; ET- Ethelene; ERF- Ethelene response factor; TF- Transcription factor; SAR- Systemic acquired resistance; PR- Pathogenesis related.

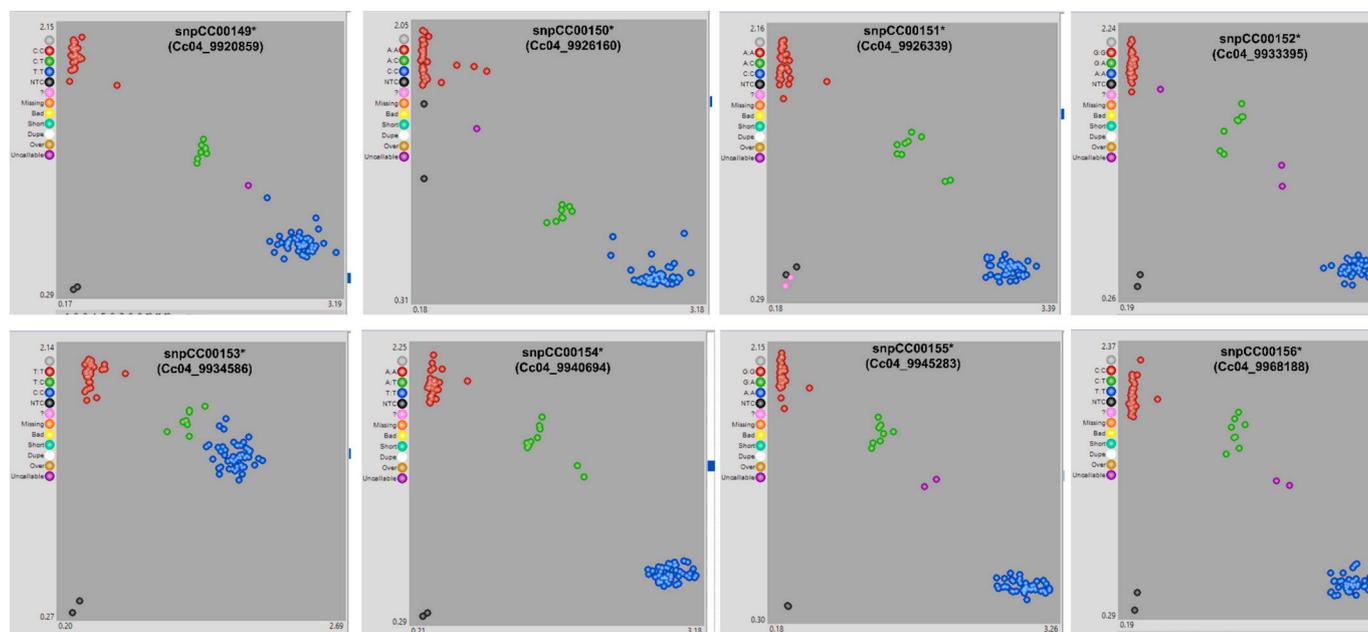
genes [46].

The expression of the gene is controlled through the binding of transcription factors (TFs) to regulatory genomic regions [47]. Some of the studies report that the presence of the transcription factor recognition site in the intronic region of the gene could play an important role in the gene expression regulation by influencing splicing mechanisms and potentially modifying the rate of transcription elongation within the body of gene. These sites can act as regulatory elements within the intron to control the transcription of the gene and their processing into mature mRNA [48]. Few earlier studies have confirmed that the conserved transcription factor binding sites are required for the appropriate expression of the genes in *Caenorhabditis elegans* [49–52]. In *Arabidopsis thaliana*, the second intron of AGAMOUS (AG) gene contains cis-acting DNA sequences that bind the transcription factors LEAFY and WUSCHEL which control floral fate and floral-specific expression of AG, respectively [53]. Such types of introns that act as a co and post-transcriptional regulation are called as intron-mediated enhancement (IME). Earlier, this phenomenon was reported in maize genes *Alcohol dehydrogenase1* and *Bronze1* [54], and *Shrunken1* [55]. In *Arabidopsis thaliana* it was reported in gene *Phosphoribosylanthranilate transferase1* [56].

## 5. Conclusion

In earlier study, QTL-seq on ICPL20096 × ICPL332 RIL population was used in common resistant and susceptible bulks for FW and SMD to identify genomic regions for FW and SMD using an old version of ICPL87119 (Asha) developed using short reads [13]. No potential genomic region could be detected maybe because common bulks were made for both the traits, and the use of older genome assembly. In our QTL-Seq study on ICP8863 × ICPL87119 RIL population identified one genomic region associated with SMD resistance on chromosomes Cc04 because we used an improved genome assembly of ‘Asha’ as a reference genome [17] and only extreme bulks for SMD were used. The initial genome assembly, covering 40.86 % of the genome [16], was recently upgraded to achieve 91.35 % genome recovery [17], enhancing the reliability of sequencing-based approaches.

This study identified candidate genes and markers (SNPs and InDels) for SMD resistance in pigeonpea. Most of the genes identified in this study play an important role in Plant-pathogen interaction, growth, development, and signaling. The identified four frameshift mutations (InDel) in *Dormancy/auxin associated protein* gene showed 1 bp insertion, 21 bp insertion, 9 bp deletion, and 3 bp insertion at different positions in 22 susceptible lines, producing disturbance in the recognition site for important transcription sites in the intronic region of *Dormancy/auxin*



**Fig. 5.** The pictorial representation of the KASP assays used for genotyping 153 lines of RIL population (ICP8863 × ICPL87119)153 including parents for SNP verification and validation. Red color indicates the favorable allele; blue color indicates the un-favorable and green color indicates the heterozygotes; dark pink color indicates uncallable; faint pink color is for missing; black color indicated non-targeted control. Star indicates Intertek IDs. Total 8 markers are validated using KASP genotyping platform.

associated protein resulting into loss of signaling in disease resistance pathways. *Dormancy/auxin associated protein* also works as a response protein for signaling molecules important in growth and development of plants. Therefore, disruption in the function of gene *Ccsm04* encoding *Dormancy/auxin associated protein* in SMD susceptible genotypes might be resulting in pollen sterility, pollen abortion or seed shriveling/sterility. The InDels and SNP variants are already validated on the resistant and susceptible genotypes of WGRS panel and RIL population (Entire used in this study) respectively. These validated Indel markers are located in the intronic region of the *Dormancy/auxin associated protein* and hinder the recognition sites for RTV1 transcription factor, MYB2 transcription factor, Basic helix-loop-helix (bHLH) transcription factor, and AT-hook transcription factor. These transcription factors regulate the expression of *Dormancy/auxin associated protein* gene. These validated markers (snpCC00149, snpCC00150, snpCC00151, snpCC00152, snpCC00153, snpCC00154, snpCC00155, and snpCC00156) could be converted into diagnostic markers and could be used to identify the SMD resistant and susceptible lines/genotypes for their use in improving SMD resistant pigeonpea cultivars. The identified important genes from this study could also be exploited for gene cloning, functional validation of genes for the viral disease resistance in pigeonpea or other legume species. Moreover, we have proved that the use of updated version of the reference genome assembly of pigeonpea will be helpful to identify the robust and reliable results for future research.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2025.145777>.

#### CRediT authorship contribution statement

**Sagar Krushnaji Rangari:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation. **Namita Dube:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Visualization, Data curation. **Vinay Sharma:** Writing – review & editing, Writing – original draft, Formal analysis. **Sunil S. Gangurde:** Writing – review & editing, Writing – original draft. **Mamta Sharma:** Writing – review & editing, Data curation. **Prakash I. Gangashetty:** Writing – review & editing,

Methodology. **Rachit K. Saxena:** Writing – review & editing, Resources. **Abhinav Moghiya:** Writing – review & editing, Formal analysis. **Vinay K. Sharma:** Writing – review & editing, Supervision. **Karma L. Bhutia:** Writing – review & editing, Supervision. **Ravi Kant:** Writing – review & editing, Supervision. **Mahendar Thudi:** Writing – review & editing, Supervision. **S.J. Satheesh Naik:** Writing – review & editing, Data curation. **Aditya Pratap:** Writing – review & editing, Data curation. **Girish P. Dixit:** Writing – review & editing. **Sean Mayes:** Writing – review & editing. **Manish K. Pandey:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Manish K. Pandey reports financial support was provided by International Crops Research Institute for the Semi-Arid Tropics. Manish K. Pandey reports a relationship with International Crops Research Institute for the Semi-Arid Tropics that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

This work was partially supported by ICAR-ICRISAT collaborative project and Bill and Melinda Gates Foundation (BMGF), USA.

#### Data availability

All the data generated in the present study is provided in the Supplementary Information and sequencing data deposited as Bioproject ID PRJNA1184965, PRJNA575817, PRJNA383013, PRJNA57990 in NCBI.

## References

- [1] M. Musokwa, P. Mafongoya, Pigeonpea yield and water use efficiency: a savior under climate change-induced water stress, *Agronomy* 11 (2020) 5.
- [2] CropTrust. <https://www.croptrust.org/news-events/news/pigeonpea-food-for-drought>, 2025 (accessed 24 February 2025).
- [3] ICRISAT. [www.icrisat.org/crops/pigeonpea/overview](http://www.icrisat.org/crops/pigeonpea/overview), 2024 (accessed 24 February 2025).
- [4] FAOSTAT, Food and Agriculture Organization of the United Nations. FAOSTAT Statistical Database, FAO, Rome, 2022 <https://www.fao.org/faostat/en/#home>. (accessed 24 February 2025).
- [5] B.R. Sayiprathap, A.K. Patibanda, M. Mantesh, S. Hiremath, N. Sagar, C.N. L. Reddy, C.R.J. Basha, S.E.D. Reddy, M.K. Rao, R.M. Nair, Sterility mosaic disease of pigeonpea (*Cajanus cajan* (L.) Huth): current status, disease management strategies, and future prospects, *Plants* 13 (2024) 2146.
- [6] Chowdary L. Rajesh, P. Hariitha, Rani B. Pramila, Performance of pigeonpea varieties against sterility mosaic disease (SMD) for rainfed regions of Prakasam district, Andhra Pradesh, India, *Int J plant, Soil Sci.* 35 (2023) 1215–1220, <https://doi.org/10.9734/ijps/2023/v35i203919>.
- [7] R.K. Varshney, S.M. Mohan, P.M. Gaur, S.K. Chamarthi, V.K. Singh, S. Srinivasan, N. Swapna, M. Sharma, S. Pande, S. Singh, Marker-assisted backcrossing to introgress resistance to fusarium wilt race 1 and ascochyta blight in C214, an elite cultivar of chickpea, *Plant Genome* 7 (2014), 2013 10.
- [8] D.B. Deshmukh, B. Marathi, H.K. Sudini, M.T. Variath, S. Chaudhari, S.S. Manohar, C.V.D. Rani, M.K. Pandey, J. Pasupuleti, Combining high oleic acid trait and resistance to late leaf spot and rust diseases in groundnut (*Arachis hypogaea* L.), *Front. Genet.* 11 (2020) 514.
- [9] Y. Shasidhar, M.T. Variath, M.K. Vishwakarma, S.S. Manohar, S.S. Gangurde, M. Sriswathi, H.K. Sudini, K.L. Dobariya, S.K. Bera, T. Radhakrishnan, Improvement of three popular Indian groundnut varieties for foliar disease resistance and high oleic acid using SSR markers and SNP array in marker-assisted backcrossing, *Crop J.* 8 (2020) 1–15.
- [10] H. Takagi, A. Abe, K. Yoshida, S. Kosugi, S. Natsume, C. Mitsuoka, A. Uemura, H. Utsushi, M. Tamiru, S. Takuno, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, *Plant J.* 74 (2013) 174–183.
- [11] A. Majeed, P. Johar, A. Raina, R.K. Salgotra, X. Feng, J.A. Bhat, Harnessing the potential of bulk segregant analysis sequencing and its related approaches in crop breeding, *Front. Genet.* 13 (2022) 944501.
- [12] M.K. Pandey, A.W. Khan, V.K. Singh, M.K. Vishwakarma, Y. Shasidhar, V. Kumar, V. Garg, R.S. Bhat, A. Chitkineeni, P. Janila, QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.), *Plant Biotechnol. J.* 15 (2017) 927–941.
- [13] V.K. Singh, A.W. Khan, R.K. Saxena, V. Kumar, S.M. Kale, P. Sinha, A. Chitkineeni, L.T. Pazhamala, V. Garg, M. Sharma, Next-generation sequencing for identification of candidate genes for fusarium wilt and sterility mosaic disease in pigeonpea (*Cajanus cajan*), *Plant Biotechnol. J.* 14 (2016) 1183–1194.
- [14] V.K. Singh, A.W. Khan, R.K. Saxena, P. Sinha, S.M. Kale, S. Parupalli, V. Kumar, A. Chitkineeni, S. Vechalapu, C.V. Sameer Kumar, Indel-seq: a fast-forward genetics approach for identification of trait-associated putative candidate genomic regions and its application in pigeonpea (*Cajanus cajan*), *Plant Biotechnol. J.* 15 (2017) 906–914.
- [15] R.K. Saxena, V.K. Singh, S.M. Kale, R. Rathineni, S. Parupalli, V. Kumar, V. Garg, R. Das, M. Sharma, K.N. Yamini, Construction of genotyping-by-sequencing based high-density genetic maps and QTL mapping for fusarium wilt resistance in pigeonpea, *Sci. Rep.* 7 (2017) 1911.
- [16] R.K. Varshney, W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T. A. Donoghue, S. Azam, G. Fan, A.M. Whaley, Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers, *Nat. Biotechnol.* 30 (2012) 83.
- [17] V. Garg, O. Dudchenko, J. Wang, A.W. Khan, S. Gupta, P. Kaur, K. Han, R. K. Saxena, S.M. Kale, M. Pham, Chromosome-length genome assemblies of six legume species provide insights into genome organization, evolution, and agronomic traits for crop improvement, *J. Adv. Res.* 42 (2022) 315–329.
- [18] M. Sharma, R. Ghosh, R. Telangre, A. Rathore, M. Saifulla, D.M. Mahalinga, D. R. Saxena, Y.K. Jain, Environmental influences on pigeonpea-fusarium wilt interactions and stability of genotypes to fusarium wilt, *Front. Plant Sci.* 7 (2016) 253.
- [19] M.K. Pandey, S.S. Gangurde, Y. Shasidhar, V. Sharma, S.M. Kale, A.W. Khan, P. Shah, P. Joshi, R.S. Bhat, P. Janila, High-throughput diagnostic markers for foliar fungal disease resistance and high oleic acid content in groundnut, *BMC Plant Biol.* 24 (2024) 262.
- [20] M.A. Katta, A.W. Khan, D. Doddamani, M. Thudi, R.K. Varshney, NGS-QCbox and raspberry for parallel, automated and rapid quality control analysis of large-scale next generation sequencing (Illumina) data, *PLoS One* 10 (2015) e0139868.
- [21] P. Cingolani, V.M. Patel, M. Coon, T. Nguyen, S.J. Land, D.M. Ruden, X. Lu, Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift, *Front. Genet.* 3 (2012) 35.
- [22] P. Cingolani, A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3, *Fly (Austin)* 6 (2012) 80–92.
- [23] Y. Wang, L. Jia, G. Tian, Y. Dong, X. Zhang, Z. Zhou, X. Luo, Y. Li, W. Yao, Shinycircos-V2.0: leveraging the creation of circos plot with enhanced usability and advanced features, *Imeta* 2 (2023) e109.
- [24] R.K. Varshney, M.K. Pandey, A. Bohra, V.K. Singh, M. Thudi, R.K. Saxena, Toward the sequence-based breeding in legumes in the post-genome sequencing era, *Theor. Appl. Genet.* 132 (2019) 797–816.
- [25] M.K. Pandey, S. Chaudhari, D. Jarquin, P. Janila, J. Crossa, S.C. Patil, S. Sundravandana, D. Khare, R.S. Bhat, T. Radhakrishnan, Genome-based trait prediction in multi-environment breeding trials in groundnut, *Theor. Appl. Genet.* 133 (2020) 3101–3117.
- [26] V. Singh, P. Sinha, J. Obala, A.W. Khan, A. Chitkineeni, R.K. Saxena, R.K. Varshney, QTL-seq for the identification of candidate genes for days to flowering and leaf shape in pigeonpea, *Heredity* (Edinb) 128 (2022) 411–419.
- [27] R. Kumar, P. Janila, M.K. Vishwakarma, A.W. Khan, S.S. Manohar, S.S. Gangurde, M.T. Variath, Y. Shasidhar, M.K. Pandey, R.K. Varshney, Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut, *Plant Biotechnol. J.* 18 (2020) 992–1003.
- [28] S.S. Gangurde, A.W. Khan, P. Janila, M.T. Variath, S.S. Manohar, P. Singam, A. Chitkineeni, R.K. Varshney, M.K. Pandey, Whole-genome sequencing based discovery of candidate genes and diagnostic markers for seed weight in groundnut, *Plant Genome* 16 (2023) e20265.
- [29] G.B. de Souza, T.A. de Oliveira Mendes, P.P. Fontes, V. de Almeida Barros, A. B. Goncalves, T. de Freitas Ferreira, M. D.-B.L. Costa, M.S. Alves, L.G. Fietto, Genome-wide identification and expression analysis of dormancy-associated gene 1/auxin repressed protein (DRM1/ARP) gene family in *Glycine max*, *Prog. Biophys. Mol. Biol.* 146 (2019) 134–141.
- [30] E.J. Andersen, S. Ali, E. Byamukama, Y. Yen, M.P. Nepal, Disease resistance mechanisms in plants, *Genes* (Basel) 9 (2018) 339.
- [31] S. Gazzani, T. Lawrenson, C. Woodward, D. Headon, R. Sablowski, A link between mRNA turnover and RNA interference in Arabidopsis, *Science* (New York, N.Y.) 306 (2004) 1046–1048.
- [32] S. Zhao, Y. Zhang, M. Tan, J. Jiao, C. Zhang, P. Wu, K. Feng, L. Li, Identification of YABBY transcription factors and their function in ABA and salinity response in *Nelumbo nucifera*, *Plants* 12 (2023) 380.
- [33] Z. Li, G. Li, M. Cai, S.V.G.N. Priyadarshani, M. Aslam, Q. Zhou, X. Huang, X. Wang, Y. Liu, Y. Qin, Genome-wide analysis of the YABBY transcription factor family in pineapple and functional identification of AcYABBY4 involvement in salt stress, *Int. J. Mol. Sci.* 20 (2019) 5863.
- [34] Y. Zhu, M. Zhao, T. Li, L. Wang, C. Liao, D. Liu, H. Zhang, Y. Zhao, L. Liu, X. Ge, Interactions between *Verticillium dahliae* and cotton: pathogenic mechanism and cotton resistance mechanism to *Verticillium wilt*, *Front. Plant Sci.* 14 (2023) 1174281.
- [35] M. Zhang, S. Zhang, Mitogen-activated protein kinase cascades in plant signaling, *J. Integr. Plant Biol.* 64 (2022) 301–341.
- [36] J. Song, Z. Niu, Q. Li, Y. Bao, X. Ma, H. Wang, L. Kong, D. Feng, Isolation and identification of differentially expressed genes from wheat in response to *Blumeria graminis* f. sp. tritici (Bgt), *Plant Mol. Biol. Report.* 33 (2015) 1371–1380.
- [37] Y. Zhao, C. Li, J. Ge, M. Xu, Q. Zhu, T. Wu, A. Guo, J. Xie, H. Dong, Recessive mutation identifies auxin-repressed protein ARP1, which regulates growth and disease resistance in tobacco, *Mol. Plant-Microbe Interact.* 27 (2014) 638–654.
- [38] B.N. Kunkel, J.M.B. Johnson, Auxin plays multiple roles during plant-pathogen interactions, *Cold Spring Harb. Perspect. Biol.* 13 (2021) a040022.
- [39] H. Zhang, X. Tan, L. Li, Y. He, G. Hong, J. Li, L. Lin, Y. Cheng, F. Yan, J. Chen, Suppression of auxin signalling promotes rice susceptibility to Rice black streaked dwarf virus infection, *Mol. Plant Pathol.* 20 (2019) 1093–1104.
- [40] H. Zhang, L. Li, Y. He, Q. Qin, C. Chen, Z. Wei, X. Tan, K. Xie, R. Zhang, G. Hong, Distinct modes of manipulation of rice auxin response factor OsARF17 by different plant RNA viruses for infection, *Proc. Natl. Acad. Sci. USA* 117 (2020) 9112–9121.
- [41] Z.H. Lee, T. Hirakawa, N. Yamaguchi, T. Ito, The roles of plant hormones and their interactions with regulatory genes in determining meristem activity, *Int. J. Mol. Sci.* 20 (2019) 4065.
- [42] X. Tan, N. Zheng, Hormone signaling through protein destruction: a lesson from plants, *Am. J. Physiol. Endocrinol. Metab.* 296 (2009) E223–E227.
- [43] J.B. Heo, S. Sung, S.M. Assmann, Ca<sup>2+</sup>-dependent GTPase, extra-large G-protein 2 (XLG2), promotes activation of DNA-binding protein related to vernalization 1 (RTV1), leading to activation of floral integrator genes and early flowering in *Arabidopsis*, *J. Biol. Chem.* 287 (2012) 8242–8253.
- [44] Y. Yu, S. Zhang, Y. Yu, N. Cui, G. Yu, H. Zhao, X. Meng, H. Fan, The pivotal role of MYB transcription factors in plant disease resistance, *Planta* 258 (2023) 16.
- [45] H. Liao, Y. Fang, J. Yin, M. He, Y. Wei, J. Zhang, S. Yong, J. Cha, L. Song, X. Zhu, Rice transcription factor bHLH25 confers resistance to multiple diseases by sensing H<sub>2</sub>O<sub>2</sub>, *Cell Res.* (2025) 1.
- [46] S. Kim, Y. Kim, E.S. Seong, Y. Lee, J.M. Park, D. Choi, The chili pepper CaATL1: an AT-hook motif-containing transcription factor implicated in defence responses against pathogens, *Mol. Plant Pathol.* 8 (2007) 761–771.
- [47] J.I. Fuxman Bass, A.M. Tamburino, A. Mori, N. Beittel, M.T. Weirauch, J.S. Reece-Hoyes, A.J.M. Walhout, Transcription factor binding to *Caenorhabditis elegans* first introns reveals lack of redundancy with gene promoters, *Nucleic Acids Res.* 42 (2014) 153–162.
- [48] T.C. Lin, C.H. Tsai, C.K. Shiau, J.H. Huang, H.K. Tsai, Predicting splicing patterns from the transcription factor binding sites in the promoter with deep learning, *BMC Genomics* 25 (2024) 830.
- [49] S.B. Hwang, J. Lee, Neuron cell type-specific SNAP-25 expression driven by multiple regulatory elements in the nematode *Caenorhabditis elegans*, *J. Mol. Biol.* 333 (2003) 237–247.
- [50] R. Voutev, R. Keating, E.J.A. Hubbard, L.G. Vallier, Characterization of the *Caenorhabditis elegans* islet LIM-homeodomain ortholog, *lim-7*, *FEBS Lett.* 583 (2009) 456–464.

- [51] S.G. Meyers, A.K. Corsi, C. *elegans* twist gene expression in differentiated cell types is controlled by autoregulation through intron elements, *Dev. Biol.* 346 (2010) 224–236.
- [52] C. Brabin, P.J. Appleford, A. Woollard, The *Caenorhabditis elegans* GATA factor ELT-1 works through the cell proliferation regulator BRO-1 and the fusogen EFF-1 to maintain the seam stem-like fate, *PLoS Genet.* 7 (2011) e1002200.
- [53] J.A. Casas-Mollano, N.T. Lao, T.A. Kavanagh, Intron-regulated expression of SUVH3, an *Arabidopsis* Su (var) 3-9 homologue, *J. Exp. Bot.* 57 (2006) 3301–3311.
- [54] J. Callis, M. Fromm, V. Walbot, Introns increase gene expression in cultured maize cells, *Genes Dev.* 1 (1987) 1183–1200.
- [55] M. Clancy, L.C. Hannah, Splicing of the maize Sh1 first intron is essential for enhancement of gene expression, and a T-rich motif increases expression without affecting splicing, *Plant Physiol.* 130 (2002) 918–929.
- [56] A.B. Rose, R.L. Last, Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1, *Plant J.* 11 (1997) 455–464.