

Association of *nad7a* Gene with Cytoplasmic Male Sterility in Pigeonpea

Pallavi Sinha, K. B. Saxena, Rachit K. Saxena, Vikas K. Singh, V. Suryanarayana, C.V. Sameer Kumar, Mohan A.V.S. Katta, Amir W. Khan, and Rajeev K. Varshney*

Abstract

Cytoplasmic male sterility (CMS) has been exploited in the commercial pigeonpea [*Cajanus cajan* (L.) Millsp.] hybrid breeding system; however, the molecular mechanism behind this system is unknown. To understand the underlying molecular mechanism involved in A_4 CMS system derived from *C. cajanifolius* (Haines) Maesen, 34 mitochondrial genes were analyzed for expression profiling and structural variation analysis between CMS line (ICRISAT Pigeonpea A line, ICPA 2039) and its cognate maintainer (ICPB 2039). Expression profiling of 34 mitochondrial genes revealed nine genes with significant fold differential gene expression at $P \leq 0.01$, including one gene, *nad4L*, with 1366-fold higher expression in CMS line as compared with the maintainer. Structural variation analysis of these mitochondrial genes identified length variation between ICPA 2039 and ICPB 2039 for *nad7a* (subunit of *nad7* gene). Sanger sequencing of *nad4L* and *nad7a* genes in the CMS and the maintainer lines identified two single nucleotide polymorphisms (SNPs) in upstream region of *nad4L* and a deletion of 10 bp in *nad7a* in the CMS line. Protein structure evaluation showed conformational changes in predicted protein structures for *nad7a* between ICPA 2039 and ICPB 2039 lines. All above analyses indicate association of *nad7a* gene with the CMS for A_4 cytoplasm in pigeonpea. Additionally, one polymerase chain reaction (PCR) based Indel marker (*nad7a_del*) has been developed and validated for testing genetic purity of A_4 derived CMS lines to strengthen the commercial hybrid breeding program in pigeonpea.

CYTOPLASMIC MALE STERILITY is a unique, maternally inherited trait that has been identified in more than 150 flowering plant species. Formation of nonfunctional pollen occurs with CMS, though female fertility usually remains unaffected (see Schnable and Wise, 1998). Cytoplasmic male sterility is determined by recombination events in the mitochondrial genome that leads to formation of chimeric open reading frames (ORFs; Chase, 2007). These rearranged chimeric ORFs encode deleterious proteins resulting in reduced respiration and other associated mitochondrial defects that ultimately lead to pollen sterility (Igarashi et al., 2013; Luo et al., 2013). Effect of pollen sterility, however, can be suppressed by crossing CMS line with the line carrying nuclear fertility restorer (*Rf*) genes (Wang et al., 2006; Eckardt, 2006; Tang et al., 2014). The *Rf* genes encode pentatricopeptide repeat proteins that counteract the mitochondrial defects and reduce or remove the deleterious properties of the CMS-associated genes (Hu et al., 2012; Igarashi et al., 2013).

Cytoplasmic male sterile lines, together with nuclear fertility restoration, are exploited for commercial production of hybrid seeds in various economically important crops such as rice (*Oryza sativa* L.; Cheng et al., 2004; Ellur et al., 2013), wheat (*Triticum aestivum* L.; Longin

P. Sinha, K.B. Saxena, R.K. Saxena, V.K. Singh, V. Suryanarayana, C.V. Sameer Kumar, M.A.V.S. Katta, A.W. Khan, and R.K. Varshney, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502324, India; R.K. Varshney, School of Plant Biology and Institute of Agriculture, The Univ. of Western Australia, 35 Stirling Hwy., Crawley, WA, 6009, Australia. Received 17 Nov. 2014. Accepted 26 Feb. 2015. *Corresponding author (r.k.varshney@cgiar.org).

Published in The Plant Genome 8
doi: 10.3835/plantgenome2014.11.0084
© Crop Science Society of America
5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Abbreviations: CMS, cytoplasmic male sterile/sterility; ETS, electron transport system; ICPA, ICRISAT Pigeonpea A line; ICPB, ICRISAT Pigeonpea B line; ORF, open reading frame; PCR, polymerase chain reaction; qRT-PCR, quantitative real-time polymerase chain reaction; *Rf*, nuclear fertility restorer genes; SNP, single nucleotide polymorphism.

et al., 2013), barley (*Hordeum vulgare* L.; Mühleisen et al., 2013), maize (*Zea mays* L.; Crow, 1998; Duvick et al., 2004), sorghum [*Sorghum bicolor* (L.) Moench; Reddy et al., 2013], and pearl millet [*Pennisetum glaucum* (L.) R. Br.; Rai et al., 1999]. In the case of pigeonpea, significant efforts have been made to identify CMS lines using wide hybridization technology. For example, eight different cytoplasm systems (A₁ to A₈) have been identified (Saxena et al., 2010a; Saxena et al., 2013). Among these, the CMS line derived from *Cajanus cajanifolius*, designated as A₄ cytoplasm has shown stability in variable environments and excellent performance in hybrid seed production (Saxena et al., 2005; Sawargaonkar, 2011). Using this system, the first CMS based pigeonpea hybrid (ICPH 2671) showing up to 47% higher yield over control varieties in multilocation trials has been released in India for commercial cultivation (Saxena et al., 2013). However, to further accelerate pigeonpea hybrid breeding programs in terms of yield and quality traits, understanding the molecular basis of CMS is important.

To understand the molecular basis of CMS in A₄ cytoplasm, we sequenced mitochondrial genomes of four *Cajanus* genotypes: the CMS line ICPA 2039, its cognate maintainer line ICPB 2039, the hybrid line ICPH 2433, and the wild relative ICPW 29 (accession from *Cajanus cajanifolius*), source of A₄ cytoplasm. Sequence annotation of mitochondrial genome of ICPA 2039 line predicted a total of 51 genes, including 34 protein coding, 14 tRNA, and 3 rRNA genes. Comparative analysis of four different *Cajanus* mitochondrial genomes revealed 31 chimeric ORFs, among which 13 were identified between CMS and maintainer lines (Tuteja et al., 2013). Chimeric ORFs found in the proximity of known mitochondrial protein coding genes were reported to be associated with the male sterility in a number of plant species (Schnable and Wise, 1998; Hanson and Bentolila, 2004; Igarashi et al., 2013). Considering the fact that chimeric ORFs are associated with CMS phenotype, the markers developed based on the structural differences between CMS and maintainer lines have been frequently utilized for genetic purity testing of CMS lines (Narayanan et al., 1996; Sane et al., 1997; Yashitola et al., 2004; Rajendrakumar et al., 2007; Saxena et al., 2010b). Mitochondrial DNA based markers can be used to differentiate CMS and maintainer lines at the early stage of plant growth. This can replace time consuming and labor-intensive traditional grow-out test that involves field trials, with accurate detection of contaminants in CMS.

In view of above, this study analyzes expression and structural variation of 34 protein coding mitochondrial genes between CMS (ICPA 2039) and its maintainer (ICPB 2039) lines to identify the CMS associated gene(s) for A₄ cytoplasm derived from *C. cajanifolius*. Most importantly one mitochondrial gene based Indel marker that can differentiate the CMS line from its fertile counterpart has been developed in this study.

Table 1. Plant materials used for validation gene based markers in cytoplasmic male sterile (CMS) and its maintainer lines.

Serial No.	Source cytoplasm	CMS line	Maintainer line	Backcross generation
1	<i>Cajanus cajanifolius</i>	ICPA 2039 (A4) [†]	ICPB 2039	BC ₁₅
2	<i>Cajanus cajanifolius</i>	ICPA 2199 (A4)	ICPB 2199	BC ₃
3	<i>Cajanus cajanifolius</i>	ICPA 2098 (A4)	ICPB 2098	BC ₈
4	<i>Cajanus cajanifolius</i>	ICPA 2092 (A4)	ICPB 2092	BC ₈
5	<i>Cajanus cajanifolius</i>	ICPA 2189 (A4)	ICPB 2189	BC ₅
6	<i>Cajanus cajanifolius</i>	ICPA 2166 (A4)	ICPB 2166	BC ₆
7	<i>Cajanus cajanifolius</i>	ICPA 2051 (A4)	ICPB 2051	BC ₉
8	<i>Cajanus cajanifolius</i>	ICPA 2048 (A4)	ICPB 2048	BC ₈
9	<i>Cajanus cajanifolius</i>	ICPA 2047 (A4)	ICPB 2047	BC ₉
10	<i>Cajanus cajanifolius</i>	ICPA 2043 (A4)	ICPB 2043	BC ₁₀
11	<i>Cajanus sericeus</i>	ICPA 2061 (A1)	ICPB 2061	BC ₆
12	<i>Cajanus scarabaeoides</i>	ICPA 2052 (A2)	ICPB 2052	BC ₆
13	<i>Cajanus lineatus</i>	ICPA 2209 (A6)	ICPB 2209	BC ₆
14	<i>Cajanus reticulatus</i>	CMS <i>reticulatus</i>	Vyshali	BC ₆

[†] Details in parentheses indicate the source of cytoplasm used to develop CMS lines.

Materials and Methods

Plant Materials

A CMS line (ICPA 2039) and its cognate maintainer line (ICPB 2039) derived from *C. cajanifolius* (A₄) were used for identification of mitochondrial gene variations through expressional and structural variation analysis. A set of 10 plants each from ICPA 2039 and ICPB 2039 lines were used to confirm the structural variation associated with the gene(s) during the structural variation analysis. Nine different A₄ derived CMS lines and their cognate maintainer lines were used for validation of the gene-based Indel marker (Table 1). To estimate the accuracy level of the gene-based Indel marker in detecting contamination of maintainer line, known levels of admixtures of maintainer in CMS were formed through DNA admixture. To check the utility of this marker at the commercial scale, 100 seeds of ICPA 2043, female parent of world's first CGMS based pigeonpea hybrid ICPH 2671 (Saxena et al., 2013) having A₄ cytoplasm, were tested. The gene-based Indel marker was also amplified on CMS and maintainer lines of four different cytoplasm of pigeonpea, namely, A₁ (*C. sericeus*), A₂ (*C. scarabaeoides*), A₆ (*C. lineatus*), and A₈ (*C. reticulatus*).

Ribonucleic Acid and Deoxyribonucleic Acid Extraction

Pool of 10 flower buds of 5 mm size from three randomly selected plants each from male sterile and male fertile plants as described in Dalvi et al. (2008) were harvested and used for RNA isolation. Further, the buds were fixed in Carnoy's fluid to confirm the pollen fertility. After microscopic conformation of sterility or fertility, total RNA was extracted from the selected buds using TRIzol (Invitrogen, Grand Island, NY) and purified through RNeasy Plant Mini kit (Qiagen, GmbH, Hilden, Germany). Further, three

micrograms of RNA was used to construct single strand cDNA using the SuperScript III RT enzyme (Invitrogen).

Genomic DNA was isolated from young leaves of single plant by following extraction method as described in Cuc et al. (2008). Quality and quantity of DNA was checked on 0.8% agarose gel and concentration was normalized to approximately 5ng/ μ L.

Primer Designing

A total of 55 exons present in 34 protein coding genes were targeted for amplification in ICPA 2039 and ICPB 2039 lines. Primer pairs were designed from the exonic regions which ranged from 1 to 5 exons per gene using the Primer3 software (Rozen and Skaletsky, 2000; Table S1). These primer pairs were used for both expressional and structural variation analysis. The criteria used for designing the primer pairs included annealing temperature range of 55 to 60°C with an average of 57°C, amplicon size 70 to 150 bp, primer length 20 ± 5 bp, and GC% 50 ± 5 .

Quantitative Real-Time Polymerase Chain Reaction

Quantitative real-time PCR (qRT-PCR) was performed using ABI SYBR GREEN PCR reaction on an ABI Fast 7500 System (Applied Biosystems, Foster City, CA). To calculate mean relative expression levels, each reaction was performed in three biological and two technical replicates, along with a no-template control. For qRT-PCR analysis, 1 μ L of cDNA was used as a template in a 10 μ L PCR reaction mixture. Cycle initiated with preincubation at 50°C for 2 min, denaturation at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing and extension at 60°C for 1 min. The real-time PCR efficiencies for each reaction was >95%. Also, a melting curve analysis was performed to determine the specificity of the reaction. *Actin* was used as an internal control for normalization of expression level, and quantity of each mRNA was calculated from the threshold points located in the log-linear range. The data were analyzed using the 7500 Sequence Detection Software (Applied Biosystems, USA) with default baseline and threshold. Student's *t* test was used for data analysis (at $P \leq 0.01$ and $P \leq 0.05$) to declare statistically significant values. OrganellarGenomeDraw (OGDRAW, <http://ogdraw.mpimp-golm.mpg.de/>, verified 15 May 2015) was used to create high-quality visual representations of expressional differences.

DNA Amplification through Polymerase Chain Reaction

Polymerase chain reactions were performed in a 5 μ L reaction volume [0.5 μ L of $10 \times$ PCR buffer, 1.0 μ L of 15 mM MgCl₂, 0.25 μ L of 2 mM dNTPs, 0.50 μ L of 2 pM/ μ L primer (MWG-Biotech AG, Bangalore, India), 0.1 U of Taq polymerase (Bioline, London, UK), and 1.0 μ L (5 ng/ μ L) of template DNA in 96-well micro liter plate (ABgene, Rockford, IL) using thermal cycler GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA).

The PCR program used to amplify the DNA fragments was: initial denaturation was for 5 min at 94°C, followed by 35 cycles of denaturation for 20 s at 94°C, annealing for 20 s at 60°C, and extension for 30 s at 72°C, and final extension for 7 min at 72°C. The amplified PCR products were separated on 3.5% Metaphor gels stained with ethidium bromide, visualized under UV light, and photographed on gel documentation system.

Sequence Analysis and Homology Search

Polymerase chain reaction amplicons showing expression difference and length variation were used for sequencing at both ends using corresponding primers by deploying Sanger sequencing methodology. Sequences obtained were merged and contigs were developed for each amplicon using DNA baser software (<http://www.dnabaser.com/>, verified 15 May 2015). Contigs obtained from DNA baser software were subjected for homology search using BLASTn (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, verified 15 May 2015), and further multiple sequences were aligned by using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>, verified 15 May 2015) software.

Protein Structure Analysis

To identify the conformational changes at protein level, protein structure evaluation of identified associated genes was performed between ICPA 2039 and ICPB 2039. Sequence data were subjected for ORF prediction using the getorf program of EMBOSS suite (Olson, 2002). One-dimensional property of the protein sequences were predicted using ProtParam (<http://www.expasy.ch/tools/protparam.html>, verified 15 May 2015). Secondary protein structures were predicted using Jpred3 (Cole et al., 2008). To predict protein disorders, DisEMBL (<http://dis.embl.de>, verified 15 May 2015) was used. Homology-based modeling was performed by submitting the sequences to automated homology prediction servers such as Phyre2 (Kelley and Sternberg, 2009), SWISS-MODEL (<http://swissmodel.expasy.org>, verified 15 May 2015), ModWeb (<http://salilab.org/modweb>, verified 15 May 2015), HHPredB (Söding et al., 2005), and M4T (Fernandez-Fuentes et al., 2007).

Results

Relative Expression Profiling of Mitochondrial Genes

Expression profiling of selected 34 protein coding genes revealed a range of expressional variation ranging from a minimum of 0.04-fold (*nad1e* and *nad5d*) to a maximum of 13.66-fold (*nad4L*) differences between the CMS (ICPA 2039) and the maintainer line (ICPB 2039). The differential gene expression values for each gene between the two lines are illustrated in Fig. 1. In summary, the genes of mitochondrial respiratory chain/electron transport chain of Complex I showed a wide range of differential expression between ICPA 2039 and ICPB 2039. The different exonic regions of the *nad1* showed differential

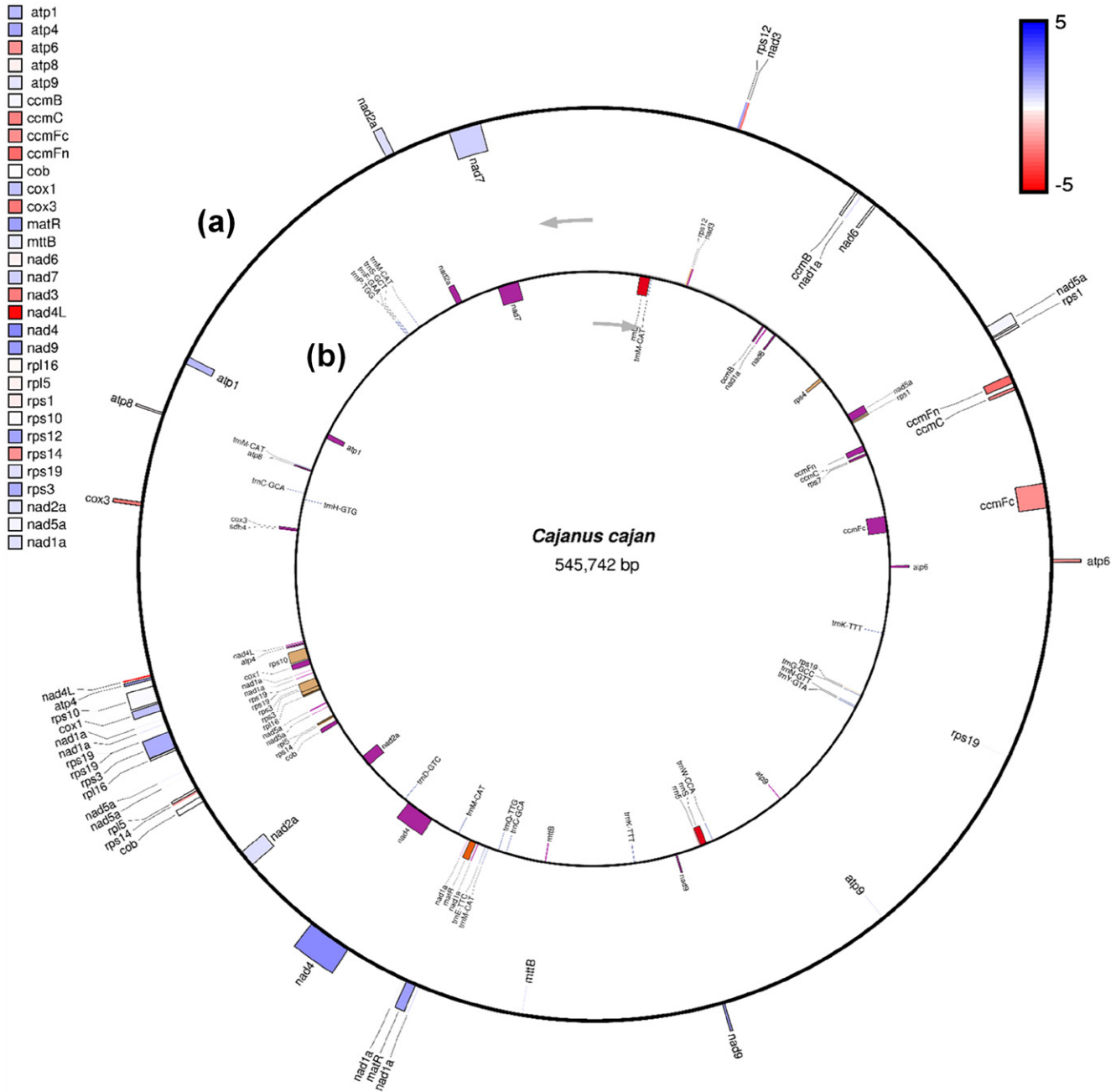


Figure 1. A genome-wide map of mitochondria showing relative expression profiling of 34 genes between ICPA 2039 (cytoplasmic male sterile A) and ICPB 2039 (maintainer line B) of pigeonpea. (a) Genomic-wide transcript profiling of 34 mitochondrial genes belonging to respiratory chain/electron transport chain Complex I to V for identification of differentially expressed genes between ICPA 2039 and ICPB 2039 using quantitative real-time polymerase chain reaction (qRT-PCR). Expression profiling of both lines were analyzed using RNA isolated from young flower bud of both lines. Transcription direction is down for the genes presented in the inner circle and up for those genes presented toward the outer circle of genome. The color of different bars represents the differential gene expression between both of the lines. The genes with blue color bars showed upregulation, genes with red color showed downregulation, and the bars with intermediate color showed average gene regulation between both of the lines. (b) Reference mitochondrial genome of pigeonpea adapted from Tuteja et al. (2013) for understanding the organization of the genome.

expressions ranging from 0.04 to 2.22, for *nad2* from 0.47 to 0.80. However, *nad3* and *nad4L* possessing one exon each showed fold differences of 2.90 and 13.66, respectively. Similarly, the exonic region of *nad4* ranged from 0.28 to 4.60, *nad5* from 0.04 to 2.16, and *nad7* from 0.21 to 1.43. The genes, *nad6* and *nad9* having one exon each, showed 1.22 and 0.26-fold differences, respectively.

sdh4 gene in the Complex II and *cob* gene in the Complex III showed expression differences of 0.56 and 0.95-fold, respectively, between the two lines. Further, in the Complex IV, which has two genes, *cox1* and *cox3* displayed expression differences of 0.44 and 6.21 between the two lines, respectively. The Complex V comprises of five genes namely *atp1*, *atp4*, *atp6*, *atp8*, and *atp9*. These

Table 2. Mitochondrial genes with significant fold expression difference between ICPA 2039 and ICPB 2039 lines of pigeonpea.

Serial No.	Gene	Primer name	Fold difference
I. Complex I			
1	<i>nad1</i>	<i>nad1a</i>	0.20*
		<i>nad1e</i>	0.04**
2	<i>nad3</i>	<i>nad3</i>	2.90*
3	<i>nad4L</i>	<i>nad4L</i>	13.66**
4	<i>nad4</i>	<i>nad4a</i>	4.60**
		<i>nad4d</i>	0.28**
5	<i>nad5</i>	<i>nad5a</i>	0.22**
		<i>nad5b</i>	2.16*
		<i>nad5c</i>	2.07*
		<i>nad5d</i>	0.04**
6	<i>nad7</i>	<i>nad7c</i>	0.27*
		<i>nad7e</i>	0.21**
IV. Complex IV			
7	<i>cox3</i>	<i>cox3</i>	6.21**
V. Complex V			
8	<i>atp6</i>	<i>atp6</i>	2.20**
VI. Cytochrome c biogenesis			
9	<i>ccmFCb</i>	<i>ccmFcb</i>	2.18*
10	<i>ccmFn</i>	<i>ccmFn</i>	3.71**
VII. Ribosomal protein			
11	<i>rps3</i>	<i>rps3b</i>	0.11*
12	<i>rps12</i>	<i>rps12</i>	0.20*
13	<i>rps14</i>	<i>rps14</i>	2.18*
14	<i>rps19</i>	<i>rps19a</i>	0.36*
VIII. Other protein coding			
15	<i>matR</i>	<i>matR</i>	0.14**

* $P \leq 0.05$, unpaired student's *t* test.

** $P \leq 0.01$, unpaired student's *t* test.

genes showed expression variation from 0.39 to 2.20 between ICPA 2039 and ICPB 2039. Similarly, five genes, *ccmB*, *ccmC*, *ccmFCa*, *ccmFCb*, and *ccmFn*, belonging to Cytochrome c biogenesis, showed differential gene expression of 0.88, 2.72, 0.61, 2.18, and 3.71 between these lines, respectively. In the ribosomal protein gene complex, which has 10 genes, the expression difference ranged from 0.11 for *rps3* to 2.18 for *rps14* between the lines. The other protein coding genes complex has two genes, *matR* and *mttB*, and for these genes, differential gene expression was observed as 0.14 and 0.72, respectively, between the A and B lines. Out of 34 protein coding genes, nine genes (subunits of nine genes: *nad1e*, *nad4L*, *nad4a*, *nad4d*, *nad5a*, *nad5d*, *nad7e*, *cox3*, *atp6*, *ccmFn*, and *matR*) showed significant fold difference at $P \leq 0.01$ and 15 showed significant fold difference at $P \leq 0.05$. However, out of nine genes showed significantly different at $P \leq 0.01$ between the A and B lines, *nad4L* showed a very high level of significant fold difference between these lines (13.66-fold). The list of genes with significant fold difference between A and B lines are presented in Table 2.

Structural Variations in Mitochondrial Genes

To further detect the possible structural variations in the mitochondrial genomes of ICPA 2039 and ICPB 2039, DNA polymorphism was analyzed for all 34 mitochondrial genes between A (ICPA 2039) and B (ICPB 2039) lines utilizing 55 exon based primers (Table S1). Only *nad7a*, fragment of *nad7* gene, showed length variation between these lines (Fig. 2a). For the *nad7a* gene, the amplicons observed in ICPA 2039 and ICPB 2039 were of 150 and 160 bp sizes, respectively (Fig. 2b). Further, to confirm the identified structural variation, a set of 10 plants each for ICPA 2039 and ICPB 2039 were screened with primers for *nad7a* gene, which clearly differentiated the male sterile and fertile lines (Fig. 2c).

Based on expression profiling and length variation analysis between ICPA 2039 and ICPB 2039, the two genes, *nad4L* (13.66-fold gene expression in ICPA 2039 line as compared with ICPB 2039) and *nad7a* (10 bp deletion in ICPA 2039) were selected for Sanger sequencing from ICPA 2039 and ICPB 2039 lines. To amplify the full-length gene sequences, two primer pairs for *nad4L* and one primer pair for *nad7a* were designed. Polymerase chain reaction amplified products were sequenced, and the data generated were used for further analysis (Table 3). Comparative sequence analysis of amplicons for *nad4L* did not show any difference in the coding region of the gene between ICPA 2039 and ICPB 2039. However, two nucleotide substitutions were observed in the upstream region at 129 bp (A in ICPA 2039 to C in ICPB 2039) and 130 bp (C in ICPA 2039 to G in ICPB 2039) of the gene (Fig. 3a). Similarly, comparative sequence analysis of *nad7a* gene revealed 10 bp deletion in ICPA 2039 when compared with ICPB 2039 at 180 to 189 bp (Fig. 3b). Although, as per our earlier study (Tuteja et al., 2013), this deletion was present in the upstream of the coding region of *nad7a* gene, our annotation analysis of amplified sequence of *nad7a* gene in this study showed occurrence of the deletion in coding region.

Conformational Protein Changes for CMS Associated Genes

In addition to analyzing variation at DNA level, for understanding conformational protein changes, ORF prediction was performed for *nad4L* and *nad7a* gene sequences for ICPA 2039 and ICPB 2039 lines. While no change was observed in amino acids between ICPA 2039 and ICPB 2039 for *nad4L* gene, in the case of *nad7a* gene, 110 and 103 amino acids residues were predicted in ICPA 2039 and ICPB 2039 lines, respectively. In *nad7a*, because of the 10 bp deletion in ICPA 2039, the predicted ORF start site shifted 20 bases upstream of the deletion adding six amino acids. The seventh amino acid is formed from the triplet codon derived from the two residual bases upstream and one base downstream of the deletion. The deletion causes a frame-shift mutation thereby adding a seventh amino acid at N terminal region in ICPA 2039 resulting in a total of 110 amino acids. However in ICPB 2039 the ORF starts after the

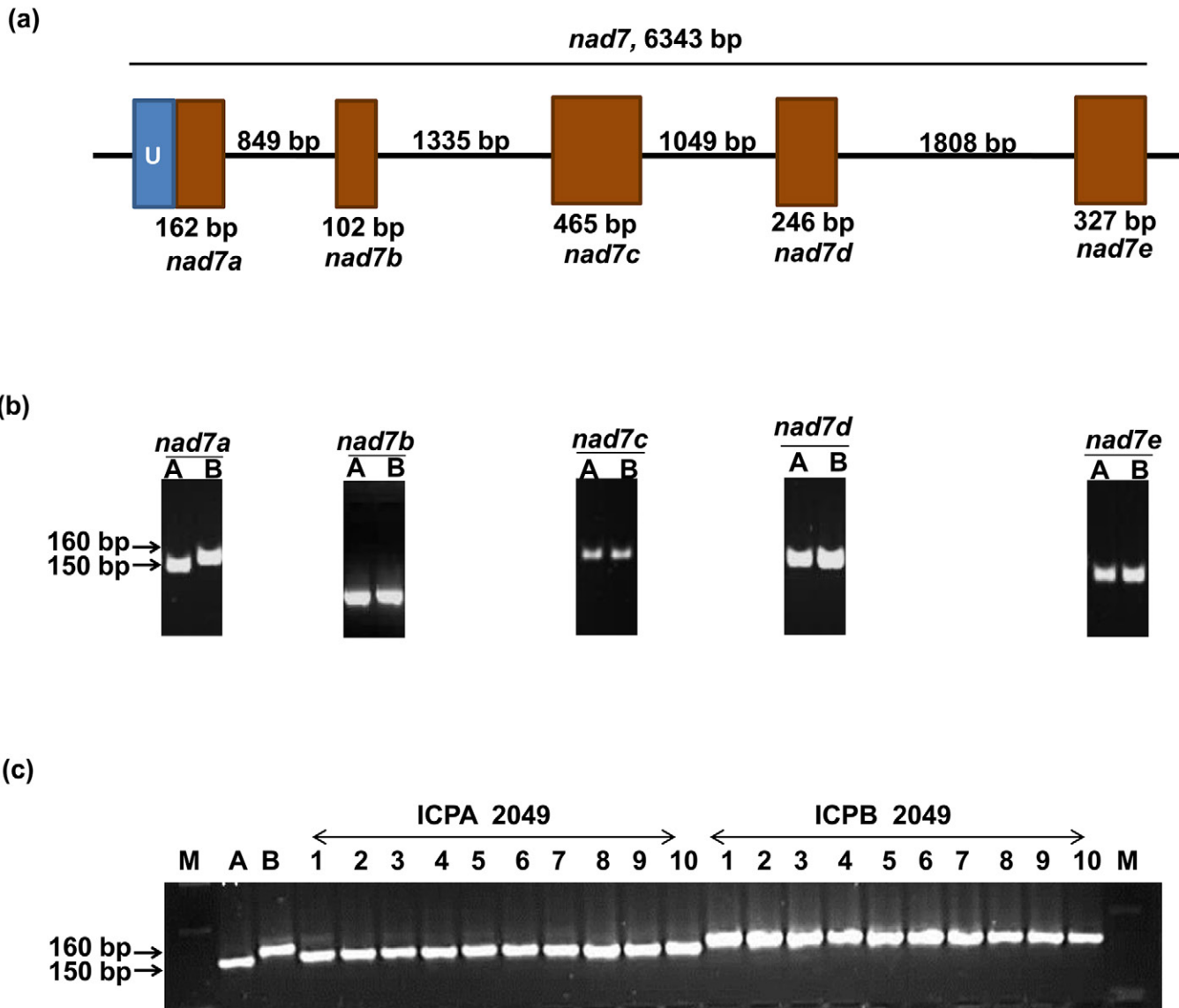


Figure 2. Genomic organization of *nad7* gene belonging to the respiratory chain Complex I. (a) Structure of the pigeonpea *nad7* gene with lengths of exons (brown bars) and introns (lines) shown in base pairs. Chimeric genic regions associated with the *nad7* region are presented in blue color bars. The position of introns and exons were determined by the published mitochondrial genome sequence. (b) Structural variation analysis of five subunits of *nad7* showed presence of 10 bp length variation between ICPA 2039 and ICPB 2039 in the first subunit, *nad7a*. (c) Confirmation of identified 10 bp deletion in *nad7a* on a set of 10 plants each of ICPA 2039 and ICPB 2039.

Table 3. Single nucleotide polymorphisms (SNPs) and Indels in *nad4L* and *nad7a* region of ICPA 2039 and ICPB 2039 by Sanger sequencing.

Gene	Nucleotide position bp	SNPs/Indels		Type
		ICPA 2039	ICPB 2039	
<i>nad4L</i>	129	A	C	Transversion
	130	C	G	Transversion
<i>nad7a</i>	180–189	–	TTTTTAGTG	Deletion

same 10 bp region resulting in 103 amino acids (Fig. 3b). As a result, the molecular weight of NAD7 protein sequence was estimated as 12 kDa in ICPA 2039 as compared to 11.3 kDa in ICPB 2039. Additionally, changes in

the number of negatively (Glu, Asp) and positively (Arg, Lys) charged residues were also observed in ICPA 2039 (8 negatively and 12 positively) and in ICPB 2039 (8 negatively and 11 positively). Further, the secondary protein structure analysis of ICPA 2039 and ICPB 2039 for *nad7a* showed significant differences in β sheet and helix with a predominant N-terminal loop regions (Supplemental Fig. S1). Preliminary analysis with predicted secondary protein structures of both the lines revealed significant disordered regions in NAD7A protein. Three different disordered loop regions in the predicted proteins were identified in ICPA 2039 (at 1 to 22, 42 to 74, and 83 to 110 amino acids) and ICPB 2039 (at 1 to 15, 35 to 67, and 76 to 103 amino acids). Tertiary protein structure of NAD7A from both the lines showed superimposed

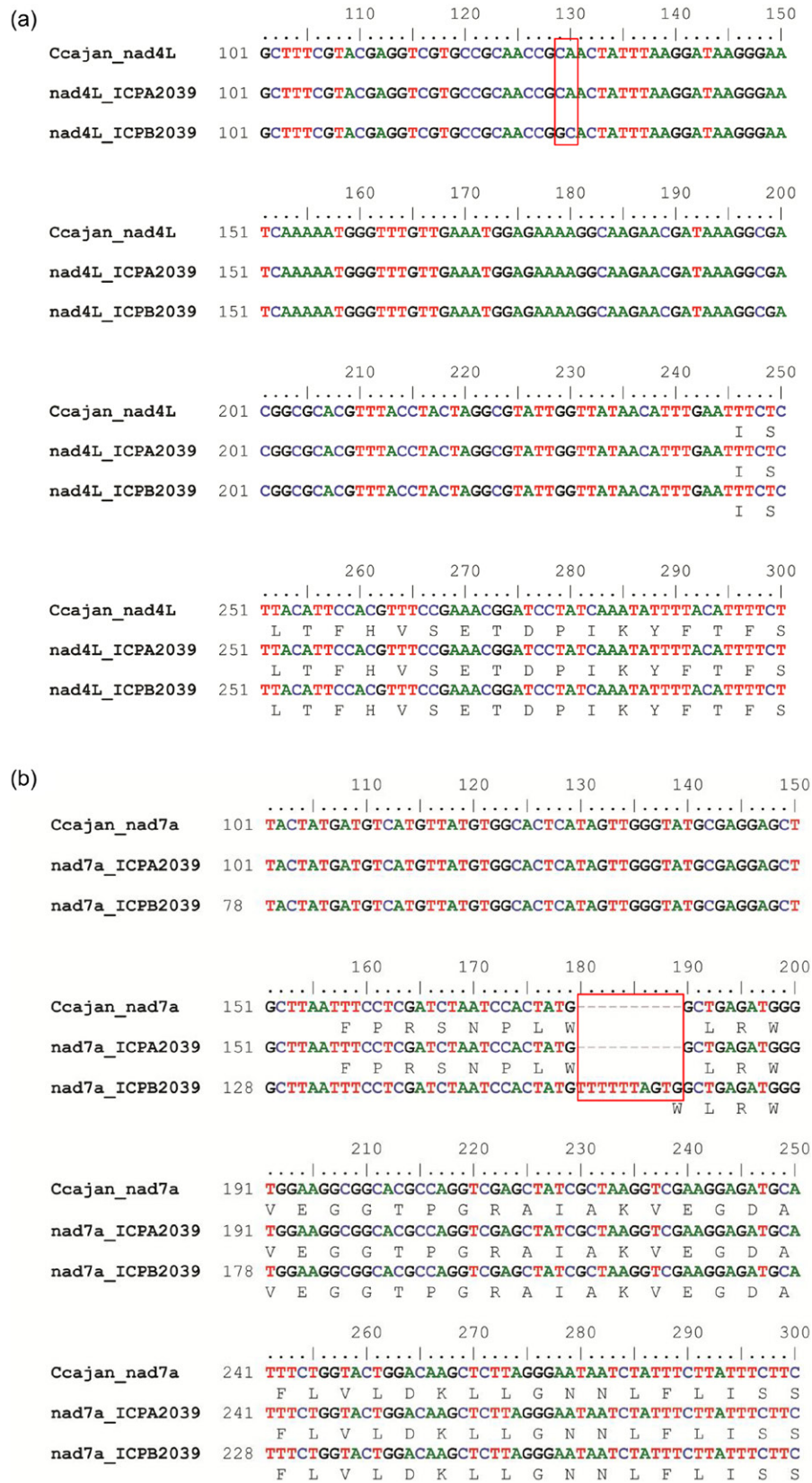


Figure 3. Sequence alignment of amplicons for *nad4L* and *nad7a* from cytoplasmic male sterile line ICPA 2039 and its cognate maintainer ICPB 2039. (a) Alignment of 360 bp *nad4L* amplified fragment that showed 13.66-fold differential expression between ICPA 2039 and ICPB 2039 provides two single nucleotide polymorphisms (SNPs) in the promoter region. (b) Sequence alignment of ICPA 2039 and ICPB 2039 for another gene *nad7a* of 619 bp shows a 10 bp deletion from position 180 to 189 in ICPA 2039 in comparison with its maintainer line.

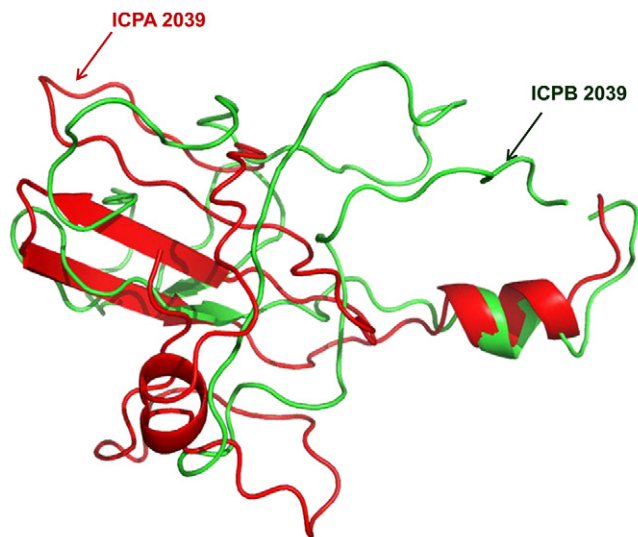


Figure 4. Protein structure analyses for NAD7A in a cytoplasmic male sterile (CMS) line (ICPA 2039) and its cognate maintainer lines, ICPB 2039. Homology based modeling between ICPA 2039 and ICPB 2039 using automated homology prediction servers predicted difference between two lines that is a putative indicator for the male sterility of CMS line.

structure onto each other with matching features and conformational differences between ICPA 2039 and ICPB 2039 that may account for the instability of the subunit (Fig. 4).

Gene Organization and Synteny of *nad7* Gene

Based on Sanger sequencing analysis, the *nad7* gene is comprised of five exons, which are ordered as *nad7a* (162 bp) along with unknown origin of sequence, *nad7b* (102 bp), *nad7c* (465 bp), *nad7d* (246 bp), and *nad7e* (327 bp), and separated by introns of 849, 1335, 1049, and 1808 bp, respectively (Fig. 2a). Further, to understand the synteny of *nad7* gene of pigeonpea, mitochondrial genomes for select species such as rice, Arabidopsis [*Arabidopsis thaliana* (L.) Heynh.], lotus [*Lotus japonicus* (Regel) K. Larsen (= *L. corniculatus* var. *japonicus* Regel)], mung bean [*Vigna radiata* (L.) R. Wilczek], and soybean [*Glycine max* (L.) Merr.] were analyzed for homology search. Multiple sequence alignment showed that the *nad7a* subunit was highly conserved in the selected five plant species. For instance, the *nad7a* gene of pigeonpea showed 100% homology with the gene sequence of mung bean, soybean, and lotus. With the Arabidopsis sequence, 96.5% homology was observed, whereas least (48%) homology was observed with rice (Supplemental Fig. S2a). Subsequently, the upstream region of *nad7a* where the 10 bp deletion was observed in ICPA 2039, was searched for homology with other crops in NCBI database using BLASTN search. As a result, the sequence with 73% query coverage showed 88% nucleotide identity with *orf124* of *Beta vulgaris* subsp. *maritima* (L.) Arcang. genotype male-sterile E mitochondrion (accession No. FQ014226.1; Supplemental Fig. S2b).

Table 4. Details for deletion based genic marker for seed purity testing analysis.

Primer name	Nature	Nucleotide sequence (5'–3')	ICPA 2039	ICPB 2039
<i>nad7a_del</i>	Codominant	F CTGACGGATGCCTACTAT R CATCTCCTTCGACCTTAG	150	160

Development and Validation of *nad7a* Gene-Based Marker

The marker analysis of *nad7a* between CMS and maintainer lines clearly differentiated the male sterile line (ICPA 2039) from the fertile line (ICPB 2039) and the marker was designated as *nad7a_del* (Table 4). Screening of a set of nine different A₄ derived CMS and their maintainer lines along with ICPA 2039 and ICPB 2039 confirmed the marker ability to differentiate sterile and fertile lines (Fig. 5). The marker, *nad7a_del* was also used to detect the contaminants in ICPA 2039 with ICPB 2039 in different known admixtures (2, 5, 10, and 15%). The marker was able to detect as low as 2% admixtures level of ICPB 2039 specific fragment in DNA of ICPA 2039. The intensity of contaminant specific allele was more prominent and clear with increasing admixture level from 5 to 15% (Fig. 6). Further, to check the utility of the marker, DNA analysis of 100 individual seeds of ICPA 2043 revealed 97.7% purity (Supplemental Fig. S3).

Screening of the *nad7a_del* marker on the other four available pigeonpea cytoplasms also demonstrated its utility and specificity for A₄ cytoplasm. All four different CMS lines, ICPA 2061 (A₁), ICPA 2052 (A₂), ICPA 2206 (A₆) and CMS *reticulatus* (A₈) and their corresponding maintainer lines (ICPB 2061, ICPB 2052, ICPB 2206 and Vyshali) did not show any structural variation with the marker *nad7a_del*. Due to amplification of a similar amplicon (150 bp) in both, CMS and its maintainer line, the marker could not differentiate the CMS line from its cognate maintainers for other four cytoplasms. Therefore, it was concluded from our study that the marker *nad7a_del* is specific for A₄ derived CMS lines and can be utilized for purity testing of A₄ derived CMS lines from their maintainers (Supplemental Fig. S4).

Discussion

Expression Variation of Mitochondrial Genes

To understand the molecular basis of A₄ based CMS in pigeonpea, the CMS line, ICPA 2039, along with its maintainer, ICPB 2039, were compared for 34 protein coding genes, identified earlier through pigeonpea mitochondrial genome sequencing (Tuteja et al., 2013). Four genes, *cox3*, *nad7*, *mttB*, and *ccmFc*, showed chimeric gene sequence between A and B lines during our previous studies (Tuteja et al., 2013). All except *mttB* showed significant fold differences between the two lines in expression profiling in the present study. Most of the

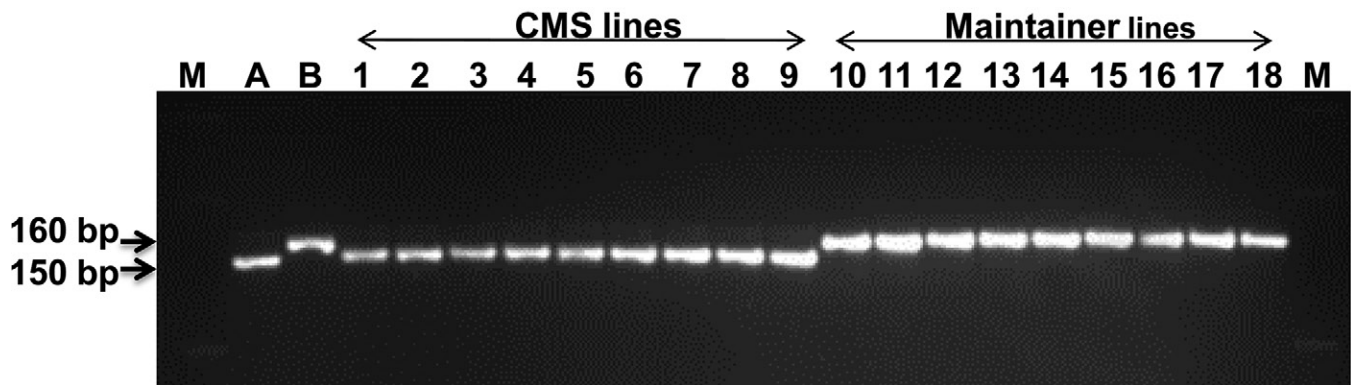


Figure 5. Validation of *nad7* gene based marker *nad7a_del* in different A_4 derived cytoplasmic male sterile (CMS) and its cognate maintainer lines. Lane M, 100 bp standard DNA ladder; Lane A, CMS line ICPA 2039; Lane B, Maintainer line ICPB 2039. Lanes 1 to 9 (different CMS line): Lane 1, ICPA 2039; Lane 2, ICPA 2098; Lane 3, ICPA 2092; Lane 4, ICPA 2189; Lane 5, ICPA 2166; Lane 6, ICPA 2051; Lane 7, ICPA 2048; Lane 8, ICPA 2047; Lane 9, ICPA 2043. Lane 10–18 (cognate maintainer lines of CMS lines): Lane 10, ICPB 2039; Lane 11, ICPB 2098; Lane 12, ICPB 2092; Lane 13, ICPB 2189; Lane 14, ICPB 2166; Lane 15, ICPB 2051; Lane 16, ICPB 2048; Lane 17, ICPB 2047; Lane 18, ICPB 2043. Lane M, 100 bp standard DNA ladder. Fragments were resolved on 3.5% Metaphore agarose gel and the expected product sizes are mentioned with the arrow.

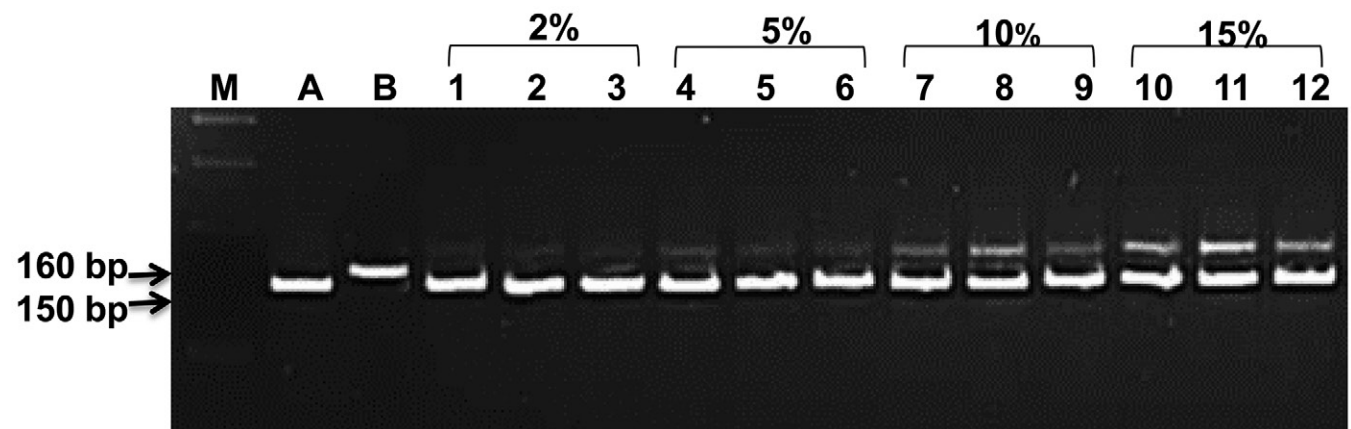


Figure 6. Evaluation of efficiency of newly developed marker *nad7a_del* for identification of admixtures of known concentration between ICPA 2039 and ICPB 2039. Lane M, 100 bp standard DNA ladder; Lane A, CMS line ICPA 2039; Lane B, maintainer line ICPB 2039. Lanes 1 to 12 represent the amplification profile of known admixtures of DNA with 2, 5, 10, and 15% of ICPB 2039 in ICPA 2039 (2:98, 5:95; 10:90, and 15:85 by volume of 100 μ L). The amplified fragments were resolved on 3.5% Metaphore agarose gel. The intensity of the identification of admixtures was increased with the increasing level of admixtures. The fragment corresponding to the CMS (150 bp) and the maintainer line (160 bp) is indicated by the arrow to the left of the image.

genes belonging to the Complex I of mitochondrial respiratory electron transport system (ETS) showed significant fold expression differences between ICPA 2039 and ICPB 2039 lines. The Complex I of respiratory ETS is one of the major pathways for the entry of electrons into cyanide-sensitive electron transfer chain (Marienfeld and Newton, 1994). Mutations in the mitochondrial DNA of this complex were found associated with CMS in many other plant species like *Arabidopsis* (Brandt et al., 1992), tobacco (*Nicotiana tabacum* L.; Gutierrez et al., 1997), maize (Marienfeld and Newton, 1994), and wheat (Bonet et al., 1990). Interestingly, our study also showed a high level of expression difference between the CMS and its maintainer line in one of the component genes (*nad4L*) of Complex I. Furthermore, sequence analysis of *nad4L* in ICPA 2039 revealed two single nucleotide substitutions in comparison with maintainer lines in upstream

regions of the gene. Altered transcription levels of mitochondrial genes was found to be associated with CMS in some plant species (Dieterich et al., 2003; Hanson and Bentolila, 2004; Wang et al., 2006). For instance, a comparative study of *atp9* gene between CMS and maintainer lines identified RNA editing with two nucleotide sites (C to U) in the conserved region of the gene associated with the male sterility (Jiang et al., 2011). However, the two nucleotide substitutions observed in ICPA 2039 at the upstream region of *nad4L* did not show any conformational changes at the protein structure level, which needs further functional confirmation.

Structural Variation in Mitochondrial Genes

The structural variation analysis of 34 mitochondrial protein coding genes at length level showed 10 bp deletion occurred at the upstream region of *nad7a* in ICPA

2039. The CMS-associated genes are often chimeric and are derived due to rearrangement in the mitochondrial genome from the time of evolution. These chimeric ORFs often include novel sequences of unknown origin, which lead to male sterility through altering the biological function of known candidate genes (see Schnable and Wise, 1998; and Chen and Liu, 2014). For instance, absence of the last two exons of *nad7* in *Nicotiana sylvestris* Speg. & Comes CMS mutant, though not lethal to plant cells, was associated with the abnormal CMS phenotype due to deficiency of Complex I (Pla et al., 1995; Sabar et al., 2000). Classical studies on the presence of abnormality in *nad7* gene in many crops revealed that it leads to respiratory deficiencies and causes abnormal respiration (Nehls et al., 1992; Connett and Hanson, 1990; Pla et al., 1995; Akçay et al., 2012). Recently, DNA methylation of sterile and fertile phenotypes of photoperiod-thermosensitive male sterile line of rice (PA64S) revealed *nad7* as one of the candidate genes responsible for male sterile phenotype (Chen et al., 2014). However, to confirm the association of the above mentioned two genes (*nad7* and *nad4L*) with CMS, the expression of the genes need to be further tested in the presence of fertility restorer genes in F₁ plants.

Implication of Structural Variation on NAD7A Protein

To understand the conformational changes in protein structure of male sterile and fertile lines, prediction of longest ORFs between ICPA 2039 and ICPB 2039 was the initial and challenging step due to the presence of 10 bp deletion in the ICPA 2039 lines. This amino acids length variation between the two lines causes significant difference in protein molecular weights. Protein modeling of the *nad7a* gene between ICPA 2039 and ICPB 2039 lines confirmed the presence of high levels of disorder between these lines due to presence of predominant random coil (loop) regions. Protein modeling with various models revealed that a SWISS-MODEL-derived protein data bank was the most accurate structural alignment (root-mean square deviation near zero) with Chain 4 of the 4HEA crystal structure.

The presence of 10 bp length variation was associated with the conformational changes in the secondary and tertiary protein structure of *nad7* subunit (NAD7A) between ICPA 2039 and ICPB 2039. In earlier studies, aberration in *nad7* gene or its associated subunits was not lethal for plants, but was associated with severe developmental defects including CMS due to reduced respiration rate (Sabar et al., 2000). Therefore, the conformational differences for NAD7A between the male sterile and fertile line might be associated with CMS in A₄ cytoplasm of pigeonpea, which needs further functional confirmation. The 10 bp deletion observed in ICPA 2039 that caused changes in the ORFs in the CMS line might have affected the stability and functions of protein. It was evident from the multiple sequence alignment results that the *nad7a* region is highly conserved in legumes in comparison with other model crops (rice and Arabidopsis).

The upstream region having deletion showed 88% sequence similarity to the chimeric gene *orf124* known candidate gene for CMS in *Beta vulgaris* (L.).

Application of CMS Based Marker for Purity Testing

To maintain the genetic purity of CMS lines, we have developed the mitochondrial gene based marker, *nad7a_{del}*, that can differentiate ICPA 2039 with its maintainer ICPB 2039. Further validation of the marker on different set of nine A₄ cytoplasm derived CMS lines, along with their maintainers, clearly demonstrated that the marker can be successfully applied to distinguish and identify the off-type plants from A₄ CMS seed lot. Contrary to the traditional, time consuming grow-out test for CMS seed purity testing, mitochondrial gene based markers provide breeders a quick and accurate tool for screening the admixtures (Yashitola et al., 2004; Ngangkham et al., 2010; Suzuki et al., 2013). Additionally, the efficiency of *nad7a_{del}* for testing the purity of ICPA 2039 up to 2% level of contamination demonstrated their immediate use in marker based seed purity of parental line. Earlier, marker based hybrid purity tests were also developed for genetic purity analysis of A₄ based F₁ hybrids (ICPH 2671 and ICPH 2438) in pigeonpea (Saxena et al., 2010b; Bohra et al., 2011). Our findings will further supplement the pigeonpea hybrid breeding program by selecting pure CMS seed lot for successful production of hybrid seeds.

To assist the hybrid pigeonpea breeding process, the marker developed for A₄ cytoplasm (derived from *C. cajanifolius*) was also checked across other four promising cytoplasm. The finding revealed that the marker was not able to differentiate the CMS from their cognate maintainers. This clearly showed that the deletion in *nad7* associated region was specific for A₄ cytoplasm. Therefore, in the near future, such types of analyses will be required for other pigeonpea cytoplasm to understand the molecular mechanism involved in male sterility and to develop their associated markers for CMS seed purity.

Conclusions

Based on variation analysis of 34 proteins coding genes, two putative genes, *nad4L* and *nad7a*, were found associated with CMS in ICPA 2039. Further protein structure analysis predicted conformation changes in NAD7A between the CMS and maintainer lines due to 10 bp deletion in ICPA 2039, which needs further functional validation. To confirm the association of *nad7a* with A₄ CMS, expression of the gene needs to be verified in the presence of a fertility restorer gene, in hybrids derived from male sterile and fertility restorer lines. The *nad7a_{del}* marker developed in this study is very effective to differentiate A₄-derived CMS lines with their cognate maintainers. In summary, this study will assist with deciphering the molecular mechanism involved in male sterility for A₄ cytoplasm, as well as for enhancing hybrid breeding in pigeonpea.

Acknowledgments

Authors thank United States Agency for International Development (USAID) for financial support for the research work. This work has been undertaken as part of the CGIAR Research Program on Grain Legumes. ICRISAT is a member of CGIAR Consortium.

References

- Akçay, N., M. Bor, T. Karabudak, F. Özdemir, and I. Türkan. 2012. Contribution of gamma amino butyric acid (GABA) to salt stress responses of *Nicotiana glauca* CMSII mutant and wild type plants. *J. Plant Physiol.* 169:452–458. doi:10.1016/j.jplph.2011.11.006
- Bohra, A., A. Dubey, R.K. Saxena, R.V. Penmettsa, K.N. Poornima, N. Kumar, A.D. Farmer, G. Srivani, H.D. Upadhyaya, S.R. Ramesh, D. Singh, K.B. Saxena, P.B. KaviKishor, N.K. Singh, C.D. Town, G.D. May, D.R. Cook, and R.K. Varshney. 2011. Analysis of BAC-end sequences (BESs) and development of BES-SSR markers for genetic mapping and hybrid purity assessment in pigeonpea (*Cajanus* spp.). *BMC Plant Biol.* 11:56–70. doi:10.1186/1471-2229-11-56
- Bonen, L., S. Bird, and L. Belanger. 1990. Characterization of the wheat mitochondrial *orf25* gene. *Plant Mol. Biol.* 15:793–795. doi:10.1007/BF00016131
- Brandt, P., S. Sunkel, M. Unseld, A. Brennicke, and V. Knoop. 1992. The *nad4L* gene is encoded between exon c of *nad5* and *orf25* in the *Arabidopsis* mitochondrial genome. *Mol. Gen. Genet.* 236:33–38.
- Chase, C.D. 2007. Cytoplasmic male sterility: A window to the world of plant mitochondrial-nuclear interactions. *Trends Genet.* 23:81–90. doi:10.1016/j.tig.2006.12.004
- Chen, X.J., J.H. Hu, H.Y. Zhang, and Y. Ding. 2014. DNA methylation changes in photoperiod-thermo-sensitive male sterile rice PA64S under two different conditions. *Gene* 537:143–148. doi:10.1016/j.gene.2013.12.015
- Chen, L., and Y.G. Liu. 2014. Male sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* 65:579–606. doi:10.1146/annurev-arplant-050213-040119
- Cheng, S.H., L.Y. Cao, S.H. Yang, and H.Q. Zhai. 2004. Forty years' development of hybrid rice: China's experience. *Rice Sci.* 11:225–230.
- Cole, C., J.D. Barber, and G.J. Barton. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* 36:W197–W201. doi:10.1093/nar/gkn238
- Connett, M.B., and M.R. Hanson. 1990. Differential mitochondrial electron transport through the cyanide-sensitive and cyanide-insensitive pathways in isonuclear lines of cytoplasmic male sterile, male fertile and restored Petunia. *Plant Physiol.* 93:1634–1640. doi:10.1104/pp.93.4.1634
- Crow, J.F. 1998. 90 years ago: The beginning of hybrid maize. *Genetics* 148:923–928.
- Cuc, L.M., E.S. Mace, J.H. Crouch, V.D. Quang, T.D. Long, and R.K. Varshney. 2008. Isolation and characterization of novel micro-satellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). *BMC Plant Biol.* 8:55. doi:10.1186/1471-2229-8-55
- Dalvi, V.A., K.B. Saxena, I.A. Madrap, and R.V. Kumar. 2008. Cytogenetic studies in A4 cytoplasmic-nuclear male-sterility system of pigeonpea. *J. Hered.* 99:667–670. doi:10.1093/jhered/esn056
- Dieterich, J.H., H.P. Braun, and U.K. Schmitz. 2003. Alloplasmic male sterility in *Brassica napus* (CMS 'tournefortii-stiewe') is associated with a special gene arrangement around a novel *atp9* gene. *Mol. Gen. Genet.* 269:723–731. doi:10.1007/s00438-003-0886-3
- Duvick, D.N., J.S.C. Smith, and M. Cooper. 2004. Long-term selection in a commercial hybrid maize breeding program. *Plant Breed. Rev.* 24:109–151.
- Eckardt, N.A. 2006. Cytoplasmic male sterility and fertility restoration. *Plant Cell* 18:515–517. doi:10.1105/tpc.106.041830
- Ellur, R.K., A.K. Singh, and H.S. Gupta. 2013. Enhancing Rice Productivity in India: Aspects and Prospects. In: P.K. Shetty, S. Ayyappan, and M.S. Swaminathan, editors, *Climate Change and Sustainable Food Security*. National Institute of Advanced Studies, Bangalore and Indian Council of Agricultural Research, New Delhi. p. 99–132.
- Fernandez-Fuentes, N., C.J. Madrid-Aliste, B.K. Rai, J.E. Fajardo, and A. Fiser. 2007. M4T: A comparative protein structure modeling server. *Nucleic Acids Res.* 35:W363–368. doi:10.1093/nar/gkm341
- Gutierrez, S., M. Sabar, C. Lelandais, P. Chetrit, P. Diolet, H. Degand, M. Boutry, F. Vedel, Y. deKouchkovsky, and R. DePaepe. 1997. Lack of mitochondrial and nuclear-encoded subunits of complex I and alteration of the respiratory chain in *Nicotiana glauca* mitochondrial deletion mutants. *Proc. Natl. Acad. Sci. USA* 94:3436–3441. doi:10.1073/pnas.94.7.3436
- Hanson, M.R., and S. Bentolila. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16:S154–S169. doi:10.1105/tpc.015966
- Hu, J., K. Wang, W. Huang, G. Liu, Y. Gao, J. Wang, Q. Huang, Y. Ji, X. Qin, L. Wan, R. Zhu, S. Li, D. Yang, and Y. Zhu. 2012. The rice pentatricopeptide repeat protein RF5 restores fertility in Hong-Lian cytoplasmic male-sterile lines via a complex with the glycine-rich protein GRP162. *Plant Cell* 24:109–122. doi:10.1105/tpc.111.093211
- Igarashi, K., T. Kazama, K. Motomura, and K. Toriyama. 2013. Whole genomic sequencing of RT98 mitochondria derived from *Oryza rufipogon* and northern blot analysis to uncover a cytoplasmic male sterility-associated gene. *Plant Cell Physiol.* 54:237–243. doi:10.1093/pcp/pcs177
- Jiang, W., S. Yang, D. Yu, and J. Gai. 2011. A comparative study of ATPase subunit 9 (*Atp9*) gene between cytoplasmic male sterile line and its maintainer line in soybeans. *Afr. J. Biotechnol.* 10:10,387–10,392. doi:10.5897/AJB11.2342
- Kelley, L.A., and M.J. Sternberg. 2009. Protein structure prediction on the web: A case study using the phyre server. *Nat. Protoc.* 4:363–371. doi:10.1038/nprot.2009.2
- Longin, C.F., M. Gowda, J. Mühleisen, E. Ebmeyer, E. Kazman, R. Schachschneider, J. Schacht, M. Kirchoff, Y. Zhao, and J.C. Reif. 2013. Hybrid wheat: Quantitative genetic parameters and consequences for the design of breeding programs. *Theor. Appl. Genet.* 126:2791–2801. doi:10.1007/s00122-013-2172-z
- Luo, D.P., H. Xu, Z.L. Liu, J.X. Guo, H.Y. Li, L.T. Chen, C. Fang, Q.Y. Zhang, M. Bai, N. Yao, H. Wu, H. Wu, C.H. Ji, H.Q. Zheng, Y.L. Chen, S. Ye, X.Y. Li, X.C. Zhao, R.Q. Li, and Y.G. Liu. 2013. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45:573–577. doi:10.1038/ng.2570
- Marienfeld, J.R., and K.J. Newton. 1994. The maize NCS2 abnormal growth mutant has a chimeric *nad4-nad7* mitochondrial gene and is associated with reduced complex I function. *Genetics* 138:855–863.
- Mühleisen, J., H.P. Maurer, G. Stiewe, P. Bury, and J.C. Reif. 2013. Hybrid breeding in barley. *Crop Sci.* 53:819–824. doi:10.2135/cropsci2012.07.0411
- Narayanan, K.K., P. Senthikumar, S. Venmadhi, G. Thomas, and J. Thomas. 1996. Molecular genetics studies on the rice mitochondrial genome. In: G.S. Khush, editor, *Rice Genetics. Proc. of 3rd Int. Rice Genetic Symp., International Rice Research Institute, Los Baños, Manila, the Philippines*. p. 689–695.
- Nehls, U., T. Friedich, A. Schmiede, T. Onishi, and H. Weiss. 1992. Characterization of assembly intermediates of NADH: Ubiquinone oxidoreductase (complex I) accumulated in *Neurospora* mitochondria by gene disruption. *J. Mol. Biol.* 227:1032–1042. doi:10.1016/0022-2836(92)90519-P
- Ngangkham, U., S.K. Parida, S. De, K.A.R. Kumar, A.K. Singh, N.K. Singh, and T. Mohapatra. 2010. Genic markers for wild abortive (WA) cytoplasm based male sterility and its fertility restoration in rice. *Mol. Breed.* 26:275–292. doi:10.1007/s11032-010-9397-1
- Olson, S.A. 2002. EMBOSS opens up sequence analysis. *European Molecular Biology Open Software Suite. Brief. Bioinform.* 3:87–91. doi:10.1093/bib/3.1.87
- Pla, M., C. Mathieu, R. DePaepe, P. Chetrit, and F. Vedel. 1995. Deletion of the last 2 exons of the mitochondrial *Nad7* gene results in lack of the *nad7* polypeptide in a *Nicotiana glauca* CMS mutant. *Mol. Gen. Genet.* 248:79–88. doi:10.1007/BF02456616
- Rai, K.N., D.S. Murty, D.J. Andrews, and P.J. Bramel-Cox. 1999. Genetic enhancement of pearl millet and sorghum for the semi-arid tropics of Asia and Africa. *Genome* 42:617–628. doi:10.1139/g99-040

- Rajendrakumar, P., A.K. Biswal, S.M. Balachandran, M.S. Ramesha, B.C. Viraktamath, and R.M. Sundaram. 2007. A mitochondrial repeat specific marker for distinguishing wild abortive types cytoplasmic male sterile rice lines from their cognate isogenic maintainer lines. *Crop Sci.* 47:207–211. doi:10.2135/cropsci2006.06.0365
- Reddy, B.V.S., A.A. Kumar, C.R. Reddy, and V.A. Tonapi. 2013. Sorghum hybrid technology. In: B. Roy, A.K. Basu, and A.B. Mandal, editors, *Breeding. Biotechnology and seed production of field crops*, p. 367.
- Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: S. Krawetz and S. Misener, editors, *Bioinformatics methods and protocols: Methods in molecular biology*. Humana Press, Totowa, NJ. p. 365–386.
- Sabar, M., R. De Paepe, and Y. de Kouchkovsky. 2000. Complex I impairment, respiratory compensations, and photosynthetic decrease in nuclear and mitochondrial male sterile mutants of *Nicotianasyvestris*. *Plant Physiol.* 124:1239–1250. doi:10.1104/pp.124.3.1239
- Sane, A.P., P. Seth, S.A. Ranade, P. Nath, and P.V. Sane. 1997. RAPD analysis of isolated mitochondrial DNA reveals heterogeneity in elite wild abortive (WA) cytoplasm in rice. *Theor. Appl. Genet.* 95:1098–1103. doi:10.1007/s001220050668
- Sawargaonkar, S.L. 2011. Study of Heterosis, combining ability, stability and quality parameters in CGMS based pigeonpea [*Cajanus cajan* (L.) Millsp.] hybrids. PhD Thesis, Marathwada Agricultural University, Parbhani, Maharashtra. 151 p.
- Saxena, K.B., R.V. Kumar, N. Srivastava, and B. Shiyong. 2005. A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. *Euphytica* 145:289–294. doi:10.1007/s10681-005-1647-7
- Saxena, K.B., R.V. Kumar, A.N. Tikle, M.K. Saxena, V.S. Gautam, S.K. Rao, D. Khare, Y.S. Chauhan, R.K. Saxena, R.K. Varshney, B.V.S. Reddy, D. Sharma, L.J. Reddy, J.M. Green, D.G. Faris, M. Mula, R. Sultana, R.K. Srivastava, C.L.L. Gowda, and S. Sawargaonkar. 2013. ICPH 2671—The world's first commercial food legume hybrid. *Plant Breed.* 132:479–485.
- Saxena, R.K., K.B. Saxena, and R.K. Varshney. 2010b. Application of SSR markers for molecular characterization of hybrid parents and purity assessment of ICPH 2438 hybrid of pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *Mol. Breed.* 26:371–380. doi:10.1007/s11032-010-9459-4
- Saxena, K.B., R. Sultana, N. Mallikarjuna, R.K. Saxena, R. Kumar, S.L. Sawargaonkar, and R.K. Varshney. 2010a. Male-sterility systems in pigeonpea and their role in enhancing yield. *Plant Breed.* 129:125–134. doi:10.1111/j.1439-0523.2009.01752.x
- Schnable, P.S., and R.P. Wise. 1998. The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends Plant Sci.* 3:175–180. doi:10.1016/S1360-1385(98)01235-7
- Söding, J., A. Biegert, and A.N. Lupas. 2005. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* 33:W244–W248. doi:10.1093/nar/gki408
- Suzuki, H., J. Yu, F. Wang, and J. Zhang. 2013. Identification of mitochondrial DNA sequence variation and development of single nucleotide polymorphic markers for CMS-D8 in cotton. *Theor. Appl. Genet.* 126:1521–1529. doi:10.1007/s00122-013-2070-4
- Tang, H., D. Luo, D. Zhou, Q. Zhang, D. Tian, X. Zheng, L. Chen, and Y.G. Liu. 2014. The Rice Restorer *Rf4* for wild-abortive cytoplasmic male sterility encodes a PPR protein that functions in reduction of WA352 transcripts. *Mol. Plant* 7:1497–1500. doi:10.1093/mp/ssu047
- Tuteja, R., R.K. Saxena, J. Davila, T. Shah, W. Chen, Y.L. Xiao, G. Fan, K.B. Saxena, A.J. Alverson, C. Spillane, C. Town, and R.K. Varshney. 2013. Cytoplasmic male sterility-associated chimeric open reading frames identified by mitochondrial genome sequencing of four *Cajanus* genotypes. *DNA Res.* 20:485–495. doi:10.1093/dnares/dst025
- Wang, Z.H., Y.J. Zou, X.Y. Li, Q.Y. Zhang, L. Chen, H. Wu, D.H. Su, Y.L. Chen, J.X. Guo, D. Luo, Y. Long, Y. Zhong, and Y.G. Li. 2006. Cytoplasmic male sterility of rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. *Plant Cell* 18:676–687. doi:10.1105/tpc.105.038240
- Yashitola, J., R.M. Sundaram, S.K. Biradar, T. Thirumurugan, M.R. Vishnupriya, R. Rajeshwari, B.C. Viraktamath, N.P. Sarma, and R.V. Sonti. 2004. A sequence specific PCR marker for distinguishing rice lines on the basis of wild abortive cytoplasm from their cognate maintainer lines. *Crop Sci.* 44:920–924. doi:10.2135/cropsci2004.0920