Plant Molecular Biology (2018) 97:131–147
https://doi.org/10.1007/s11103-018-0728-7

A novel mitochondrial *orf147* causes cytoplasmic male sterility in pigeonpea by modulating aberrant anther dehiscence

Pooja Bhatnagar-Mathur[1] · Ranadheer Gupta[1] · Palakolanu Sudhakar Reddy[1] · Bommineni Pradeep Reddy[1] · Dumbala Srinivas Reddy[1] · C. V. Sameerkumar[1] · Rachit Kumar Saxena[1] · Kiran K. Sharma[1]

Received: 13 April 2017 / Accepted: 6 April 2018 / Published online: 17 April 2018

Abstract

Key message **A novel open reading frame (ORF) identified and cloned from the A4 cytoplasm of *Cajanus cajanifolius* induced partial to complete male sterility when introduced into *Arabidopsis* and tobacco.**

Abstract Pigeonpea (*Cajanus cajan* L. Millsp.) is the only legume known to have commercial hybrid seed technology based on cytoplasmic male sterility (CMS). We identified a novel ORF (*orf147*) from the A4 cytoplasm of *C. cajanifolius* that was created via rearrangements in the CMS line and co-transcribes with the known and unknown sequences. The bi/polycistronic transcripts cause gain-of-function variants in the mitochondrial genome of CMS pigeonpea lines having distinct processing mechanisms and transcription start sites. In presence of *orf147*, significant repression of *Escherichia coli* growth indicated its toxicity to the host cells and induced partial to complete male sterility in transgenic progenies of *Arabidopsis thaliana* and *Nicotiana tabacum* where phenotype co-segregated with the transgene. The male sterile plants showed aberrant floral development and reduced lignin content in the anthers. Gene expression studies in male sterile pigeonpea, *Arabidopsis* and tobacco plants confirmed down-regulation of several anther biogenesis genes and key genes involved in monolignol biosynthesis, indicative of regulation of retrograde signaling. Besides providing evidence for the involvement of *orf147* in pigeonpea CMS, this study provides valuable insights into its function. Cytotoxicity and aberrant programmed cell death induced by *orf147* could be important for mechanism underlying male sterility that offers opportunities for possible translation for these findings for exploiting hybrid vigor in other recalcitrant crops as well.

Keywords Arabidopsis · *Cajanus cajan* · Cytoplasmic male sterility · Hybrid vigor · Pigeonpea · Tobacco

Introduction

Pigeonpea [*Cajanus cajan* (L.) Millsp.] is an important high protein (20–22%) food legume grown in the rainfed tropics and sub-tropics of Asia, Africa and South America, cultivated by smallholder farmers. Pigeonpea has a unique advantage of being partially out-crossing (20–50%) and thus being potentially amenable to hybrid breeding. Indeed, a genic male sterility (GMS) and four cytoplasmic male sterility (CMS; A_1 through A_4) systems have been developed (Dalvi et al. 2008), but only the A4 CMS system (Saxena et al. 2005) has been suitable for commercial hybrid production with yield improvement of 30–40% (Saxena 2015; Saxena et al. 2016). This is a remarkable feat in pulse breeding in view of food and nutritional security. Further improvements in hybrid pigeonpea breeding required an understanding of the basis of male sterility in this system.

The A_4 male sterile pigeonpea line ICPA2039 is based on (*Cajanus cajanifolius*) cytoplasm and the CMS hybrid system comprises in addition a maintainer line ICPB2039. The molecular basis of A_4 CMS, as explained fully later, remained unknown despite the attempts to decipher it. CMS in other plants has been characterized. It is an outcome of incompatible interactions between the mitochondrial and nuclear genomes. Plant mitochondrial genome constitutes

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11103-018-0728-7) contains supplementary material, which is available to authorized users.

✉ Pooja Bhatnagar-Mathur
p.bhatnagar@cgiar.org

✉ Kiran K. Sharma
k.sharma@cgiar.org

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

many sequences that actively recombine among themselves either by dividing into sub-genomic molecules or by joining different molecules into new ones. These recombination processes in mitochondrial genomes may create new molecular structures and novel open reading frames (ORFs), including CMS genes (Mackenzie and McIntosh 1999). Several mutations associated with the CMS trait include *T-urf13* in *Zea mays* (Dewey et al. 1986), *pcf* in *Petunia* (Young and Hanson 1987), *cox1* in *Oryza sativa* (Wang et al. 2006), and mutations in *ATPase* subunits in *Helianthus annuus* (Laver et al. 1991) and *Brassica napus* (Landgren et al. 1996). While most of the identified CMS-related chimeric genes have been linked to *ATP synthetase* or *cytochrome C oxidase* (Hanson and Bentolila 2004), variations in DNA sequence within the genes, as well as in their up- or down-stream regions have been shown to be closely linked to CMS in many crop species (Young and Hanson 1987; Köhler et al. 1991).

The mitochondrial genome sequences of the male sterile line (ICPA 2039), its maintainer line (ICPB 2039), a wild progenitor (ICPW 29) and a hybrid (ICPH 2433) have been determined. These indicate, for instance several re-arrangements putatively associated with CMS in pigeonpea (Tuteja et al. 2013). A subsequent study comparing variations between the CMS and maintainer line indicated a 10 bp indel in the coding region of *nad7* gene, that had been deduced to cause a chimeric ORF and resultant perturbations to the nad7 protein structures (Sinha et al. 2015). This study attributed the male sterility phenotype to these predicated structural aberrations. Here, we show that these are not the cause but a novel ORF of 444 nucleotides, we refer to as *orf147*, is responsible for CMS in pigeonpea. Transgenic expression of the pigeonpea *orf147* in *Arabidopsis* and tobacco affected anther functions and caused male sterility, thus generating valuable information for adapting this system further for hybrid breeding.

Results

Since male sterility is maternally inherited through the mitochondria, there must be a change in mitochondrial genome of the male sterile line of pigeonpea (ICPA 2039) that is responsible for CMS. To explore the CMS causing genes, previously reported rearrangement sites unique to the line ICPA 2039 were initially compared within the pigeonpea mitotypes ICPA 2039 (male sterile) and ICPB 2039 (maintainer) lines (data not shown). Interestingly, upon comparing the flanking sequences of *nad7* gene from the male sterile line with those from the maintainer line, a variable fragment was found to be located 5′ to the *nad7* subunit of complex I (the main dehydrogenase of the mitochondrial respiratory chain) in the male sterile line of pigeonpea. While genome walking revealed sequence variations in the 5′ upstream region of *nad7* from the male fertile and sterile lines, the coding as well as the 3′ regions were observed to be identical (Fig. S1). Sequence divergence was observed in the upstream region starting from −259 bp of the *nad7* initiation site (Fig. 3d).

ORF prediction and validation by RT-PCR

Predictive analyses of nucleotide sequences of *nad7* region in the fertile (ICPB 2039) and CMS (ICPA 2039) lines revealed two ORFs based on a threshold of 85 amino acids, with reasonably high level of variability. Reverse transcription-PCR (RT-PCR) analysis using different sets of primers resulted in amplification of various regions including ORF sequences upstream of the *nad7* gene. cDNAs of both fertile and sterile pigeonpea lines amplified a 402 bp fragment referred to as *orf133* using primer set 14 (Fig. 1a) that revealed difference in three amino acids between the fertile and sterile lines (data not shown). However, the primer set 12 resulted in an amplification of 444 bp fragment specific only to the male sterile line (ICPA 2039), hereby referred to as *orf147* (Fig. 1b). Database search for the similarity of *orf147* gene fragment from the sterile line to known ORFs using BLASTX (http://www.ncbi.nlm.nih.gov/BLAST) did not detected any significant sequence homology. However,

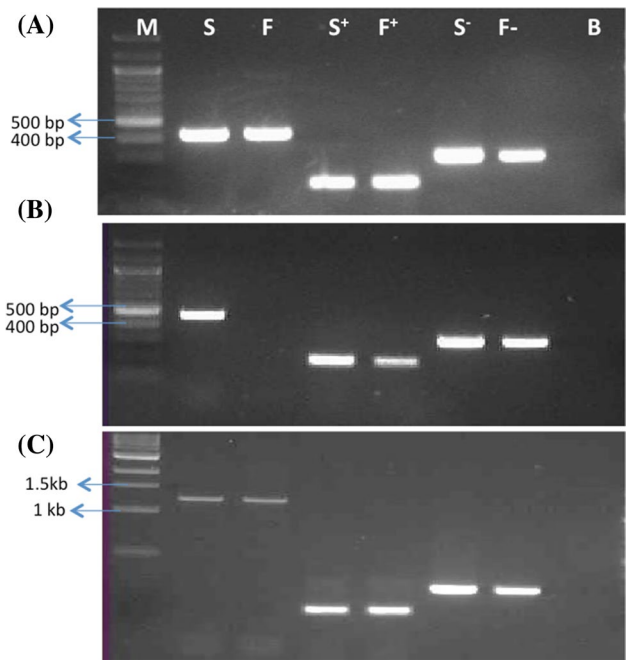


Fig. 1 RT-PCR analysis of *orf133, orf147* and *nad7* genes in male sterile (S) and male fertile (F) pigeonpea lines. **a** Amplification of *orf133* transcript (402 bp). **b** Amplification of the unique *orf147* transcript (444 bp). **c** Amplification of *nad7* transcript (1.2 kb) in both male fertile and sterile parents. (+) Depicts the cDNA of respective lines amplifying the GAPDH cds and (−) reflect the genomic DNA amplification of GAPDH

its deduced amino acid sequence showed partial homology to "*orf124*" of *Beta vulgaris* subsp. maritime genotype male-sterile E mitochondrion (accession # FQ014226.1).

Transcription of *orf147* is polycistronic

RT-PCR carried out using primer combinations specific to different internal regions, viz. orf147F/nad7P-GER, orf133F/nad7P-1R, and orf1331F/nad7GER resulted in amplification of 1741, 1327, and 2143 bp, respectively. These overlapping amplicons indicated presence of a single 2455 bp polycistronic transcript in the mitochondria from the sterile cytoplasm encompassing *orf133, orf147* and *nad7* and spacer sequences (Fig. 2a). In contrast, the fertile maintainer line did not show any amplification with any of these primer sets, indicating the monocistronic nature of the *nad7* transcript (Fig. 2b, c). Nevertheless, *orf133* amplification with gene-specific primers (orf133-F/orf133-R) in the maintainer line indicated its existence as a separate transcript (Fig. 1b). Three mitochondrial gene-specific probes, nad7, orf133 and orf147 were used to analyse the expression differences of these gene transcripts between the male-sterile and maintainer lines of pigeonpea. The *nad7* transcripts showed different banding patterns in Northern blots, while two bands were observed in the male-sterile line, only a low intensity shorter band of 402 bp was present as expected (Fig. 2a) in the maintainer line (Fig. 2d). RNA blotting experiments with the *orf147* probe also confirmed the polycistronic transcript in the male-sterile line only, while no signal was detected in the male fertile line. Nonetheless, when probed with *orf133* the maintainer line did not pick signal, apparently due to a very low expression of the transcript as observed in the expression studies using qPCR (data not shown).

These results were further confirmed by identification of transcription start sites (TSS) in the male sterile and fertile pigeonpea lines carried out using two different sets of RACE primers. While PCR with primers RACE F1 and RACE R1 following cDNA circularization resulted in amplification in both male sterile (ICPA 2039) and maintainer (ICPB 2039) lines, the primers RACE F1 and RACE R2 showed amplification with the male sterile line (ICPA 2039) only, thereby indicating absence of this region in the cDNA transcript (Fig. 3a). For the male sterile line (ICPA 2039), the sequence results using RACE F1 and RACE R2 (referred to as ICPA "transcript 1") identified T residue located 1681 bp upstream to the *nad7* start codon as the TSS (Fig. 3a), while the product of RACE F1 and RACE R1 (referred to as "transcript 2") showed the G at 656 bp upstream of the *nad7A* start codon as the other functional TSS (Fig. 3b), with "*orf147*" in common. For the male fertile maintainer line (ICPB 2039), sequence analysis revealed the TSS at T (-556) with primers RACE F1 and RACE R1 (Fig. 3c). These results confirmed that *orf147* transcripts in the male sterile line existed with more than one cistron, while the male fertile line had a monocistronic transcript (Fig. 3d).

RNA editing and secondary structure of *orf147*

To study the occurrence of RNA editing, the cDNA sequence of 5′ upstream region of the *nad7* gene in both male sterile and fertile lines were compared with their respective mitochondrial genome sequences and the genome-walked PCR products. This identified differential RNA editing pattern in the sequenced clones for each line. Several consistent edited sites among independent clones were detected for each line. The male sterile line (ICPA 2039) exhibited ten edited changes in this region in contrast to 22 observed in the maintainer line. There were four edited events in the CMS line in the *orf133*: a glycine residue at 59 aa position was edited to serine; glutamine at 119 was edited to leucine; serine at 124 was changed to arginine, and isoleucine at 125 was edited to lysine. However, no edited event was observed in *orf147*.

The secondary structure of *orf147* transcripts of the CMS line revealed a perfect hairpin loop at the 5′ end (Fig. 4a). In silico analysis using a homology-based modeling program (http://www.expasy.org) suggested that the product of *orf147* does not contain any transmembrane domain and might be a soluble protein (Fig. 4b).

orf147 encodes a cytotoxic peptide

To examine the function of *orf147* and *orf133* transcripts, their coding sequences were cloned into the expression region of PET32a vector, followed by IPTG-induced expression in *Escherichia coli*. While the growth curve analysis of *orf133* expressing cells had no apparent effect on the growth upon IPTG induction, *orf147* resulted in cytotoxicity to the *E. coli* cells (Fig. 5a, b). The bacterial cell growth stalled resulting in decrease in cell density upon IPTG induction of *orf147*, when compared to un-induced control and plasmid control (Fig. 5c).

Expression of *orf147* in *Arabidopsis* and tobacco results in male sterility

Transformation with AP3::CoxIV-Orf147 gene cassette (Fig. 6a) carrying *coxIV-orf147* gene fusion driven by *AtAP3* promoter resulted in 24 primary transgenic events in *Arabidopsis*, and over 20 in tobacco that were grown to maturity and T1 seed collected. Vegetative growth of the transgenic plants (i.e., growth rate, plant morphology) was uniform and similar to the untransformed counterparts in both the tested species.

About 80% of the T1 and T2 progenies from selected 12 independent *Arabidopsis* events at flowering exhibited

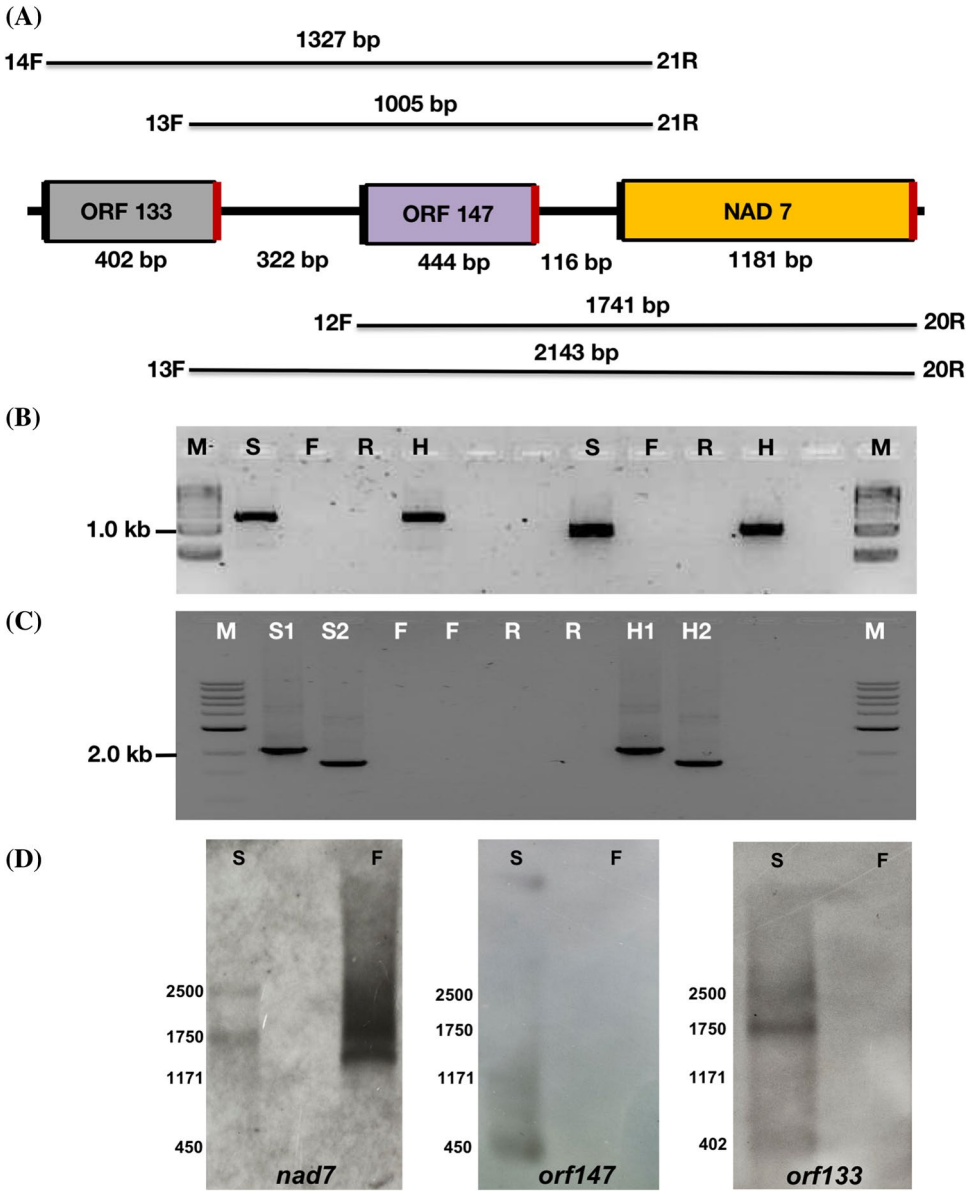


Fig. 2 Identification and analyses of *orf147* gene. **a** Organization of mitochondrial genomic regions associated with the *orf147* gene in male sterile pigeonpea line (ICPA 2039). Boxes represent coding sequences and the horizontal lines indicate flanking regions of the ORFs. The amplification of different regions along the transcript are depicted. **b, c** RT-PCR analysis of the 5′ region of *nad7* in the CMS (S; ICPA 2039), maintainer (F; ICPA 2039), restorer (R; ICPR 2438) and the restored hybrid (H; ICPH 2438) lines revealing amplification of different regions of the transcript. The black and the red filled box represent the N and the C terminus of respective ORFs. **b** Amplifications of 1327 bp (left) and 1005 bp (right) regions spanning *orf133, orf147* and part of *nad7* cds in male sterile line and hybrid. **c** Amplification of a 2143 bp fragment spanning across *orf133, orf147* and *nad7* in male sterile (S1) and hybrid (H1) lines using primers (13F & 20R); a 1741 bp amplicon is obtained in male sterile line (S2) and hybrid (H2) regions using another set of primers (12F & 20R), revealed co-transcription of *orf147* with upstream *orf133*, and downstream located *nad7* gene in these mitotypes. **d** Northern blot analysis of total RNA from flower buds of the male-sterile pigeonpea line (S) and the maintainer line (F) for three mitochondrial probes (*nad7, orf147* and *orf133*). The male sterile line (S) showed polymorphic band patterns of RNA transcripts; the RNA blots could not detect any bands for orf147 in the male fertile line (F) aligning with the results obtained by RT-PCR and cRT-PCR

semi-sterile or sterile phenotype, resulting in poor seed set when compared to the wild type (WT) (Fig. 6a, b). At the dehiscence stage, the sterile transgenic events did not produce any pollen grains or normal siliques. The sterile plants had flowers with smaller sepals and petals than their WT counterparts, with protruding pistil, shortened stamen filaments, and impaired anther dehiscence. The semi-sterile plants bore two kinds of siliques, one shorter and with no or fewer seeds than WT, and the other normal siliques like the WT. The male sterile plants had very short siliques with

no seeds (Fig. 6c–e). Similarly, out of 20 primary transgenic events of tobacco, four confirmed positive events showed complete male sterility. The flowers of male sterile transgenic tobacco plants expressing *orf147* were relatively smaller with shortened filaments and either produced very small fruits exhibiting partial sterility, or detached collapsed capsules in the completely sterile ones (Fig. 6f–i). qPCR analysis of several selected transgenic *Arabidopsis* and tobacco plants revealed variation in the *orf147* transcript levels with male sterile phenotypes showing strong *orf147* expression (Fig. 7b, c).

Expression of anther development-related genes

To detect the expression of anther development-related genes that act after tapetal specification, quantitative RT-PCR analyses for *orf147* expression (Fig. 7a) as well as key genes involved in anther development were carried out in male sterile and fertile pigeonpea lines (ICPA 2039 and ICPB 2039) and transgenic *Arabidopsis* plants along with their WT controls.

Interestingly, in CMS pigeonpea line, while the transcripts of *Defective in tapetal development and function* (*TDF1/MYB35*), *Dysfunctional tapetum1* (*DYT1*) *and Male sterlity1* were significantly down-regulated compared with those in the fertile maintainer line, an increased accumulation of transcripts of *Aborted microspore* (*AMS*) *gene* was observed (Fig. 8a). Similarly, the ectopic expression of *orf147* in *Arabidopsis* transgenics resulted in down-regulation of the transcripts of *DYT1, AMS* and *MS1* that are required for normal tapetal function and pollen wall development (Fig. 8b). This data suggested that expression of *orf147* in pigeonpea male sterile line induced male sterility in transgenic *Arabidopsis* plants, possibly by regulating the transcriptional expression of key genes specific to anther development.

Male sterile anthers have reduced endothecial secondary wall lignification

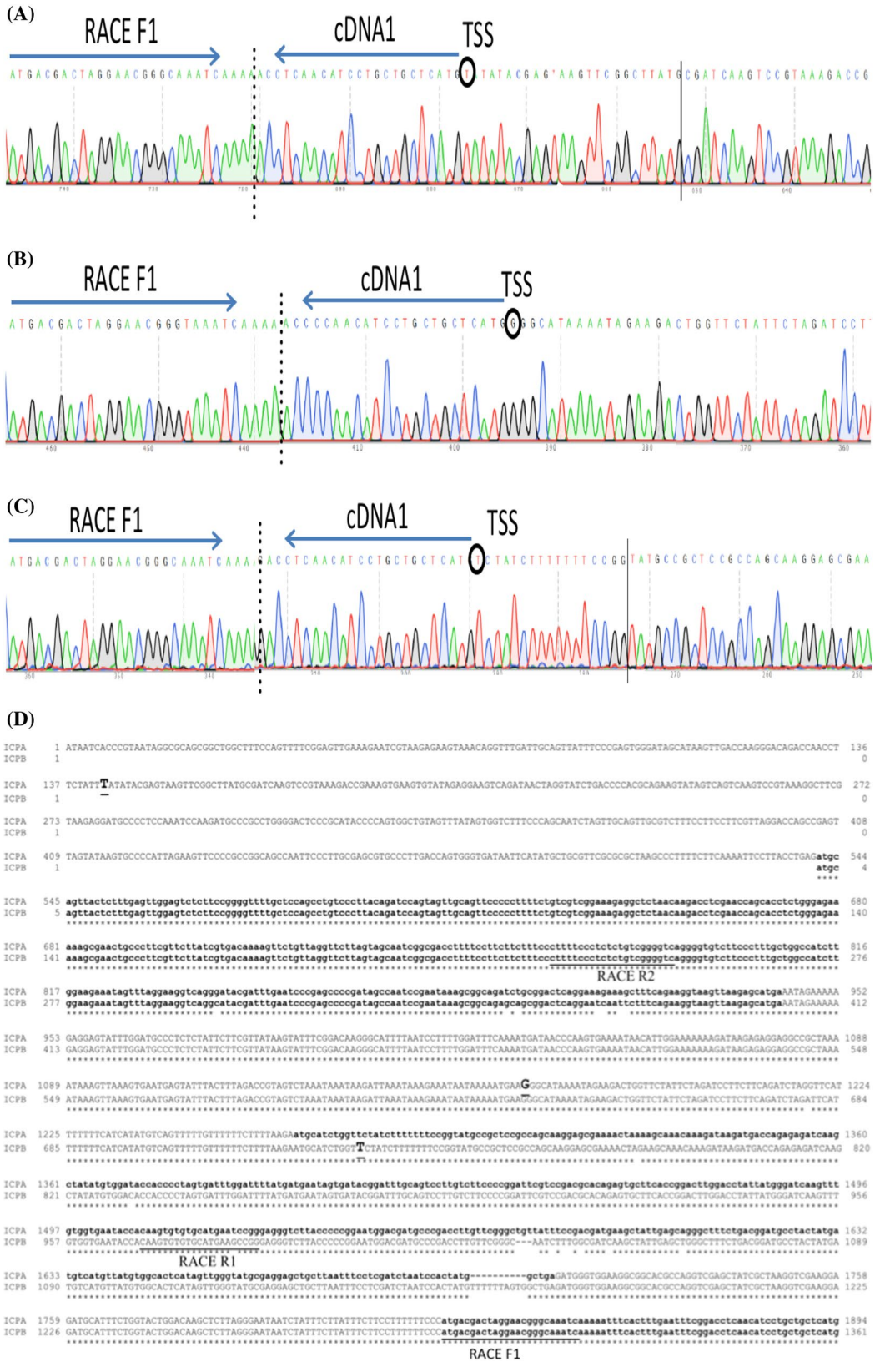
Investigation of lignification patterns using phloroglucinol staining of anthers of both transgenic male sterile and WT flowers of *Arabidopsis* and tobacco revealed a high degree of phloroglucinol stain accumulation in the WT anthers at all stages of development, when compared to the anthers of male sterile transgenic lines (Fig. 9a–f). Further, these observations correlated well with the gene expression profiles of key genes that are involved in lignin biosynthesis like *4CL* (*4 Coumarate:CoAligase*), *CCoAOMT* (*Caffeoyl CoA O-methyltransferase*), and *C3H* (*Cinnamic acid 3-hydroxylase*). Clearly, these genes expressed at significantly lower levels in the flowers of male sterile lines when compared to the WT. The relative expression of *4CL, C3H, CCoAOMT* were 0.85, 0.55 and 0.8, respectively in the male sterile plant, as compared to 1.18, 1.7 and 1.25, respectively in the WT plants (Fig. 9g).

Discussion

Heterosis in CMS-based hybrid breeding system in pigeonpea, world's first and only pulse, has achieved increased grain yield (> 30%) and disease resistance as compared to the pure line varieties (Saxena et al. 2010). Commercial success of the pigeonpea hybrid system using the only available A4 cytoplasm-based CMS, not only depends on a stable, non-reverting source of male sterility, but also on the identification of robust genotypes that reversibly suppresses the male sterility trait (Saxena and Hingane 2015). While the CMS trait is inherited maternally, very little is known about the underlying molecular mechanism to efficiently control pollination for hybrid seed production. This provides a good opportunity to study the regulation of mitochondrial gene expression in this important pulse crop. While, 12.29% of the mitochondrial genome of pigeonpea has been reported to comprise of the coding region, the remaining is non-coding (Tuteja et al. 2013). Although, a few genomic segments, particularly a gene-based indel marker was recently detected in several A4 cytoplasm-based CMS, maintainer lines and commercial hybrids (Sinha et al. 2015), functional studies linking the mitochondrial changes to the CMS trait have been lacking.

Several studies have confirmed the role of chimeric ORFs in male sterility by disrupting their function through insertion or deletion of a few base pairs (Hanson and Bentolila 2004). To determine whether a candidate ORF responsible for the male sterility phenotype could be identified in the CMS line of pigeonpea, we relied on the underlined sequence divergence in the mitochondrial genomic region of ICPA 2039 and ICPB 2039 that could have resulted in novel/chimeric ORFs causing CMS. Based on the available data, transcription and translation patterns of predicted ORFs in the contigs spanning a 10 kb region upstream and downstream of the known *nad7* gene were comprehensively evaluated in the mitochondrial genomes of pigeonpea genotypes ICPA 2039 and ICPB 2039. Our data on expression of an aberrant mitochondrial gene leading to CMS in pigeonpea is consistent with earlier reports in several species where unique ORFs created by aberrant recombination events in the mitochondrial genome have been shown to encode unique polypeptides associated with CMS phenotypes (Schnable and Wise 1998).

Our study showed that a 444 bp long unique CMS-associated novel *orf147* detected upstream of and co-transcribing with the known *nad7* gene in the mitochondrial genome of pigeonpea male sterile ICPA 2039 that derived its A_4


(A)
RACE F1
cDNA1
TSS
(B)
RACE F1
cDNA1
TSS
(C)
RACE F1
cDNA1
TSS
(D)
RACE R2
RACE R1
RACE F1

◂**Fig. 3** Sequence divergence and TSS analyses in male sterile and fertile pigeonpea line. DNA sequencing profile used to identify the transcription start sites (TSSs) in pigeonpea male sterile line (**a, b**) and maintainer fertile line (**c**). Dashed vertical lines represent DNA sequences from the TA vector. Arrows depict the primer annealing regions, and main TSS is encircled. **d** Comparison of nucleotide sequences of 5′ flanking region of *nad7* in CMS (top line) and male fertile (lower line) lines. The partial *nad7* coding sequences are included to show the primer region. *Sequence divergence. The *orfs* are indicated in small bold characters. The continuous underlined stretch indicates the primer region for cRT-PCR. The underlined single nucleotide shows TSS in both lines relative to the start codon of *nad7* validating the non-existence of *orf147* transcript in the fertile line

cytoplasm from *C. cajanifolius* is very likely to be responsible for mitochondrial dysfunction. The *orf147* did not contain components of any known functional mitochondrial genes and its transcripts were detected in both, CMS (ICPA 2039) and restored hybrid lines that carried the CMS-inducing cytoplasm, but not in the ICPB 2039 (fertile maintainer) mitochondrial genome. The restorer line (ICPR 2438) used in this study derives its cytoplasm from a different source i.e., *Cajanus cajan* and hence the absence of an *orf147* in the restorer line is due to the absence of ICPA 2039-CMS mitochondrial genome in these lines.

The RNA editing sites in the CMS line were less frequent when compared to the fertile maintainer line. Several reports have indicated such diversity in RNA editing patterns and its association with the CMS trait (Araya et al. 1992; Kurek et al. 1997; Tang et al. 1999; Howad et al. 1999; Gallagher et al. 2002). Incomplete RNA editing events producing dysfunctional proteins have also been shown to be associated with fertility restoration in rice (Iwabuchi et al. 1993).

Occurrence of a novel or chimeric transcript near or disrupting a functional gene has been reported in several CMS plants [*atp1* in *Solanum melongena* (Yoshimi et al. 2013), *atp6* in *Z. mays* (Dewey et al. 1986), *Brassica tournefortii* (Landgren et al. 1996), *Capsicum annuum* (Sabar et al. 2003), *atp9* in *Petunia* (Young and Hanson 1987), *B. napus* (Dieterich et al. 2003), *Sorghum bicolor* (Tang et al. 1996), *C. annuum* (Li et al. 2013), *Oryza rufipogon* (Igarashi et al. 2013), *Arabidopsis thaliana* (Geisler et al. 2012), *T. aestivum* (Xu et al. 2008), *Z. mays* (Wen et al. 2003) and *Nicotiana tabacum* (Bergman et al. 2000)]. However, our observation with respect to non-disruption of any known functional mitochondrial essential gene in CMS pigeonpea is distinct from most reports. Moreover, there was no indication of the novel transcript resulting in a reduced/loss or gain of function change in the *nad7* gene *per se*, thereby indicating no apparent bearing on the oxidative phosphorylation. Our results contradict and diverge from the recent report (Sinha et al. 2015), where a frame-shift mutation in the *nad7* gene and the resulting disordered predicted protein structure was suggested to be the cause of CMS in pigeonpea, especially in the absence of any in situ function validation data. Our results on the structural and functional variations in the male sterile and fertile lines, comprehensive characterization of the causal ORF, and its functional validation in both prokaryotic and eukaryotic biological systems are not in agreement with the conformational differences and interpretations drawn from protein prediction analyses reported previously (Sinha et al. 2015).

We performed Northern blotting, RT-PCR and circular RT-PCR analyses to detect transcriptional differences between fertile and sterile lines, with two longer transcripts found only in the male-sterile line. Distinct TSSs reflecting transcriptional differences also detected in sterile and fertile lines of pigeonpea, where the male sterile line (ICPA 20139) exhibited two transcripts, while the fertile line revealed only a single transcript. Sequencing six independent Circular RT-PCR products of each of the above transcripts resulted in two slightly scattering 3′ ends in the CMS line. The multiple active TSSs in the male sterile line could have been responsible for the altered post-processed bi-cistronic transcript involving *orf147-nad7* and a longer tri-cistronic transcript comprising of *orf133-orf147-nad7*. This type of co-transcription has also been observed in several other CMS systems (Bonhomme et al. 1992; Krishnasamy and Makaroff 1993; Wang et al. 2006). Single promoters with multiple TSS, rather than multiple promoters have been reported to exist throughout the mitochondrial genome where multiple products arise as a consequence of extensive processing and stability differences (Holec et al. 2006). Moreover, the secondary structure of *orf147* transcripts from the male sterile pigeonpea line revealed a perfect hairpin loop at the 5′ end, that has been suggested to provide stability to the male sterility-associated transcripts in maize (Xiao et al. 2006).

Previous reports on CMS-associated proteins such as PCF in *Petunia*, ORF125 in Kosena radish, *Raphanus sativus* cv. Kosena (Nivison and Hanson 1989; Iwabuchi et al. 1999) and expression of *orf79* in BT-type CMS rice have been shown to be cytotoxic (Duroc et al. 2005; Wang et al. 2006). In the case of pigeonpea, we also observed that ORF147 is a soluble protein that is cytotoxic to *E. coli*, and its recombinant transgene leads to male sterility in two tested model plant species (*Arabidopsis* and tobacco). To eliminate possibility of the observed toxic effects resulting from overloading of the protein synthesis machinery of overexpressed heterologous ORF147 proteins in *E. coli* cells, a similar expression study was carried out with *orf133*, another existing upstream ORF in the CMS line, whose accumulation had no adverse effect on the growth of *E. coli*. This characteristic of *orf147* might also affect the development of floral organs by weakening the mitochondria as has been reported for several other CMS-associated genes (Jing et al. 2011).

Functional validation of the *orf147* effect on male sterility through stamen/petal-specific promoter of the APETALA3

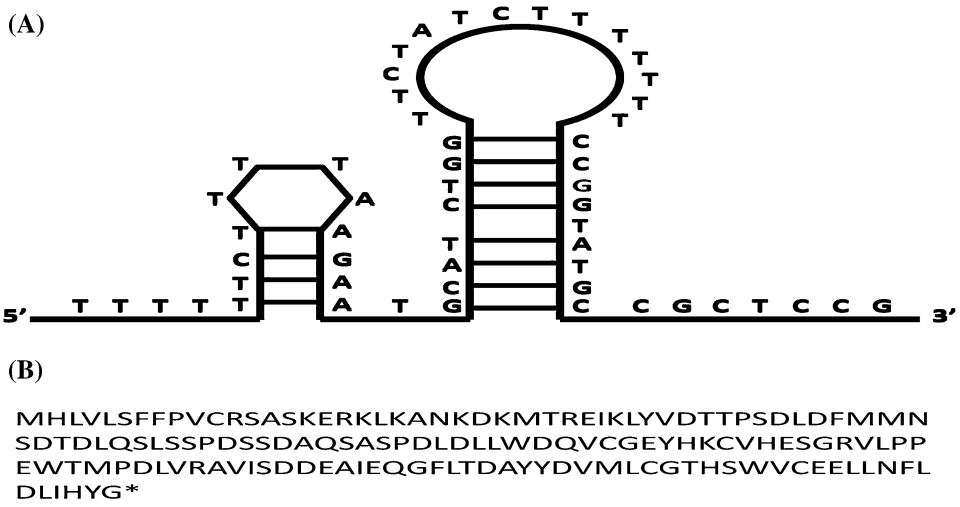


Fig. 4 Predicted amino acid sequence and secondary structure of ORF147. **a** Secondary structure of ORF147 from the male sterile line using *S*fold 2.2 software (http://sfold.wadsworth.org/cgi-bin/srna.pl). **b** Predicted amino acid sequence of *orf147*

in *A. thaliana* and *N. tabacum* transgenic plants resulted in partial to complete male sterility, thereby suggesting that the encoded cytotoxic protein results in disruption of the development of male sporophytic and/or gametophytic cells. The *Arabidopsis* APETALA3 gene is expressed specifically in stamens and petals during most floral development, and its promoter has been successfully used to identify the CMS-associated gene *orf129* in sugarbeet (Yamamoto et al. 2008). In our study, the *orf147* was fused to the mitochondrial transit peptide sequence of the nuclear *coxIV* gene of yeast for mitochondrial targeting and for investigating its effects on floral organs. The promoter of *AtAPETALA3* gene (Hill et al. 1998) and the pre-sequence of *coxIV* for mitochondrial localization used in our study have also been used successfully to identify CMS-associated genes in several previous studies (Köhler et al. 1997; Yamamoto et al. 2008). The functional analyses of CMS genes such as *orf79, orfH79*, and *WA352* in *O. sativa* (Wang et al. 2006; Peng et al. 2010; Hu et al. 2012; Luo et al. 2013), *orf129* in *B. vulgaris* (Yamamoto et al. 2008), *orf288* in *B. napus* (Jing et al. 2011), *orf220* in *B. juncea* (Yang et al. 2010), and *orf456* in *C. annuum* (Kim et al. 2007) have been successfully carried out by fusing 5′ mitochondrial targeting signal sequences.

Clearly, the tissue-specific expression of *orf147* not only disturbed the differentiation of stamens, but also affected the development of petals in the transgenic *Arabidopsis* plants. Nevertheless, no morphological differences were observed in the transgenic plants of tobacco that could possibly be due to a relatively weaker expression in flowers under the influence of a heterologous *Arabidopsis* promoter used for transformation experiments. The male sterile or semi-sterile *Arabidopsis* as well as tobacco transgenic plants formed shortened stamen filaments. Genetic transformation with an empty vector construct did not affect the fertility status of the null plants, thereby indicating that the expression of *orf147* in *Arabidopsis* and tobacco plants could lead to male sterility in the presence of an external mitochondrial-targeting peptide, and an appropriate promoter only. Interestingly, there were no notable differences in vegetative growth of the transgenic plants and their wild type counterparts in both the tested plant species, which could be attributed to the specific interaction of CMS-associated genes with floral organs (Jing et al. 2011). The male sterile transgenic phenotype in both *Arabidopsis* and tobacco was heritable, and strong *orf147* expression in the T1 and T2 progenies indicated complete penetrance, an important finding in terms of research on the CMS mechanism in this pulse crop. Typically reported transmembrane nature of cytotoxic CMS proteins having hydrophobic region does not hold true for *orf147*. This is in line with the observations in male sterile rice where transgenic plants expressing the truncated WA352 proteins contain no transmembrane segments (Luo et al. 2013), thereby indicating that CMS induction is not dependent on the transmembrane domains.

Proteins involved with CMS have been reported to interact with nuclear-encoded mitochondrial factors to induce abnormal programmed cell death (PCD) in tapetum and microspores (Wilson et al. 2001; Sorensen et al. 2003; Zhang et al. 2006; Ito et al. 2007; Zhu et al. 2008; Wilson and Zhang 2009; Chen and Liu 2014). To reveal the biological functions and relationships of key regulators involved in CMS process in the male sterile pigeonpea and transgenic *Arabidopsis* phenotypes expressing *orf147*, we monitored gene expression of key transcriptional factors involved in the initiation of tapetal development and genes having roles in tapetal and microspore maturation and PCD. PCD being an apoptosis-like cellular process is controlled by mitochondria-driven signals (Chen and Liu 2014). Obvious reduction in the expression levels of transcripts of key anther biogenesis genes in male sterile pigeonpea and transgenic *Arabidopsis* plants highlighted the underlying mechanism by which *orf147* induces

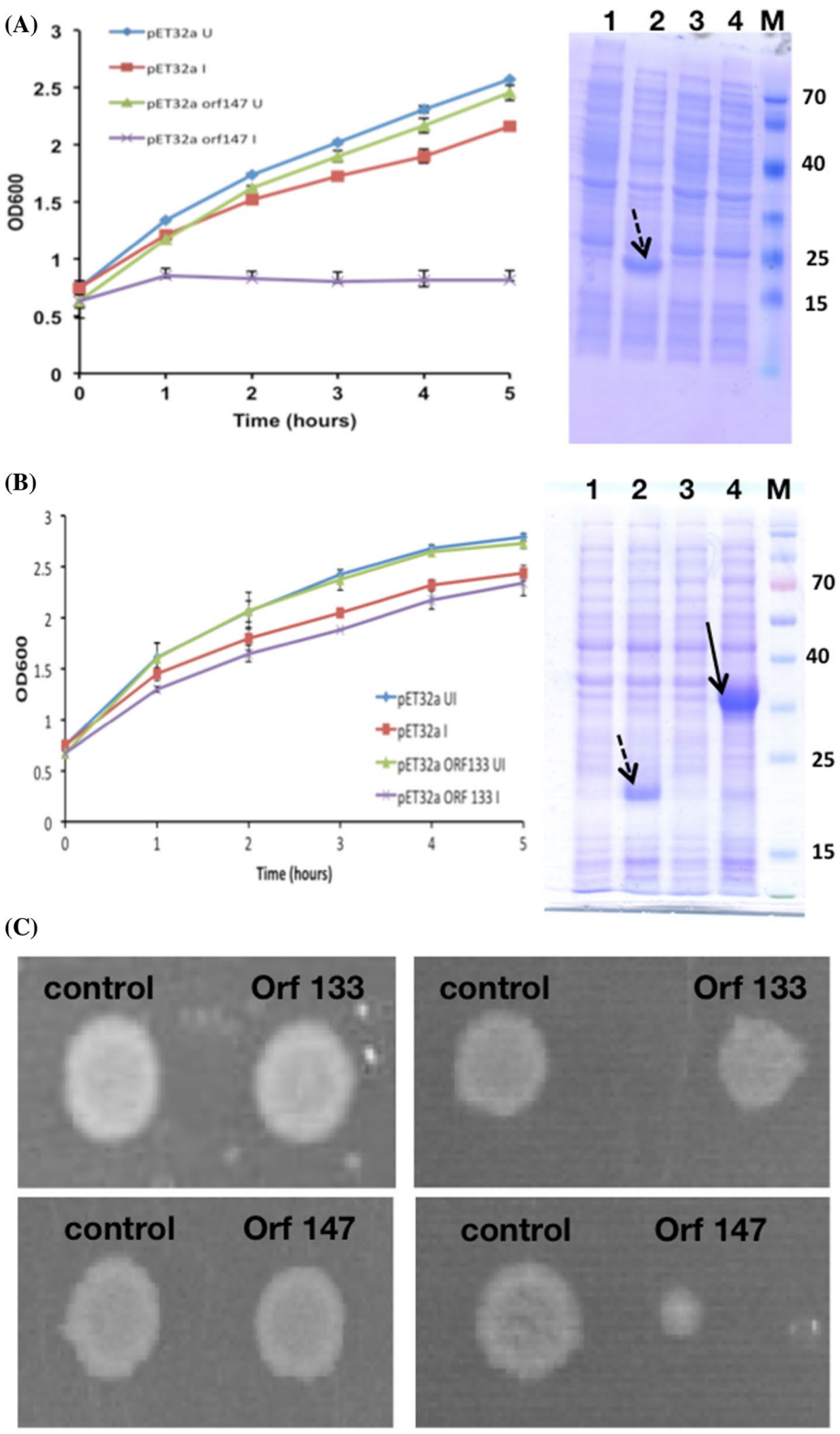


Fig. 5 Effect of *orf147* and *orf133* expression on the growth of *Escherichia coli*. **a** *E. coli* cells in liquid cultures with or without IPTG (pET32aUI), the control expression vector not induced by IPTG; pET32aI—the control expression vector induced by IPTG; pET32aORF147UI—*orf147*-containing vector not induced by IPTG; pET32aORF147 I—orf147-containing vector induced with IPTG. Induction of ORF147-Trx recombinant fusion proteins resolved on 12% SDS-PAGE gel (1. pET 32b uninduced, 2. pET 32b induced, 3. ORF147 pET 32b uninduced, 4. ORF147 pET 32b induced, M. pre-stained protein ladder. The Trx protein is depicted by dashed arrow; No ORF147-Trx recombinant protein detected). **b** *E. coli* cells in liquid cultures with or without IPTG (pET32aUI—the control expression vector not induced by IPTG, pET32aI—the control expression vector induced by IPTG, pET32aORF133UI—orf133-containing vector not induced by IPTG, pET32aORF133 I—orf133-containing vector induced with IPTG). The induction of ORF133-His recombinant fusion proteins resolved on 12% SDS-PAGE gel (1. pET 32b uninduced; 2. pET 32b induced 3. ORF133 pET 32b uninduced; 4. ORF133 pET 32b induced; M. pre-stained protein ladder. While the Trx protein is depicted by dashed arrow; the induced ORF133-Trx recombinant protein is indicated by solid arrow). **c** Upon IPTG induction of *orf147*, the bacterial cell growth stalled resulting in decreased cell density compared to un-induced sample and plasmid control; no adverse effects were observed in the bacterial growth upon induction of *orf133*

aberrant PCD-controlled cellular degeneration of the tapetum, indicating a retrograde regulation mechanism, resulting in gametophytic and/or sporophytic sterility.

Expression of *AMS, TEK* and *MS1* involved in pollen wall formation are regulated by *DYT1* primarily via *TDF1*, a putative *myb* transcription factor, acting downstream of

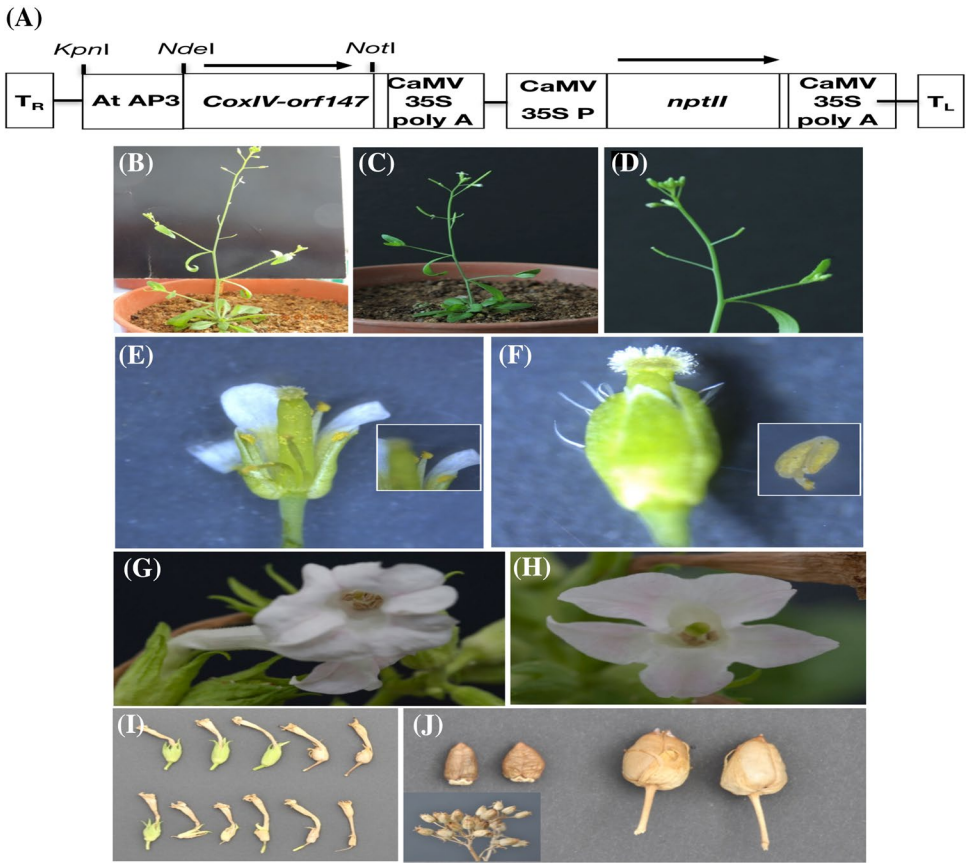


Fig. 6 Transformation vector construct and evidence of *orf147* induced CMS in *Arabidopsis* and tobacco transgenic plants. **a** Schematic representation of T-DNA region of plant transformation vector carrying *orf147* fused with a yeast mitochondrial targeted peptide (CoxIV), and cloned under *AP3* promoter from *Arabidopsis thaliana*. **b** Male sterile transgenic *Arabidopsis* plant showing normal growth and development. **c** Wild type plant with primary branches showing normal siliques. **d** Male sterile transgenic plant with short siliques with no developing seeds. **e** Normal mature flowers of wilt type plants (inset shows normal anther dehiscence). **f** Flower of male sterile line with fused carpels, protruding pistil and short filaments (inset non-dehiscent anther in the transgenic flower). **g** Flower size, color, and structure in the wild type tobacco plant. **h** Flower of male sterile tobacco plant with anthers below the stigma. **i** Flowers with developing seed capsules from wild type tobacco plants (top row), and from sterile progeny (bottom row). **j** Mature seed capsules of wild type tobacco plants (right), collapsed and detached seed capsules in partially sterile transgenic phenotypes (left), (*inset*) wild type (WT) floral branches bearing seed capsules

DYT1 and upstream of *AMS*, a bHLH protein expressed in the tapetum involved in the transcriptional regulatory networks that regulate tapetal development (Sorensen et al. 2003; Zhang et al. 2008, 2011; Li et al. 2006; Xu et al. 2010; Niu et al. 2013). *Arabidopsis* mutant for plant homeodomain (PHD) transcription factor failed to produce viable pollen, indicating the role of *MS1* in regulating the late tapetal gene expression and pollen wall deposition (Wilson et al. 2001; Ito et al. 2007; Yang et al. 2007) and alterations in PCD (Vizcay-Barrena and Wilson 2006). The detrimental interactions of CMS protein with genes involved in tapetal development have also been previously reported in rice CMS-WA lines, triggering premature tapetal PCD much earlier than that reported in male fertile rice (Luo et al. 2013), CMS-HL rice and *H. annuus* tapetal cells (Balk and Leaver 2001; Sabar et al. 2003; Li et al. 2004) by invoking oxidative stress responses (Pring et al. 2006; Fujii and Toriyama 2008, 2009).

Our study also suggested that besides having a role in abnormal PCD, *orf147* also disrupted the secondary thickening of endothecium that is essential for providing the mechanical force for anther dehiscence, causing male sterile phenotype (Bonner and Dickinson 1989; Dawson et al. 1999; Steiner-Lange et al. 2003; Mitsuda et al. 2007; Yang et al. 2007). Studies using Phloroglucinol staining in unopened flower buds of *Arabidopsis* and tobacco transgenic plants indicated reduced lignin in anther walls of male sterile phenotypes. Reduced gene expression of lignin biosynthesis genes such as *4CL* (*4 Coumarate:CoAligase*), *CCoAOMT* (*Caffeoyl CoA O-methyltransferase*), and C3H (*Cinnamic acid 3-hydroxylase*) in sterile plants also reflected the disruption of monolignol biosynthesis pathway contributing to

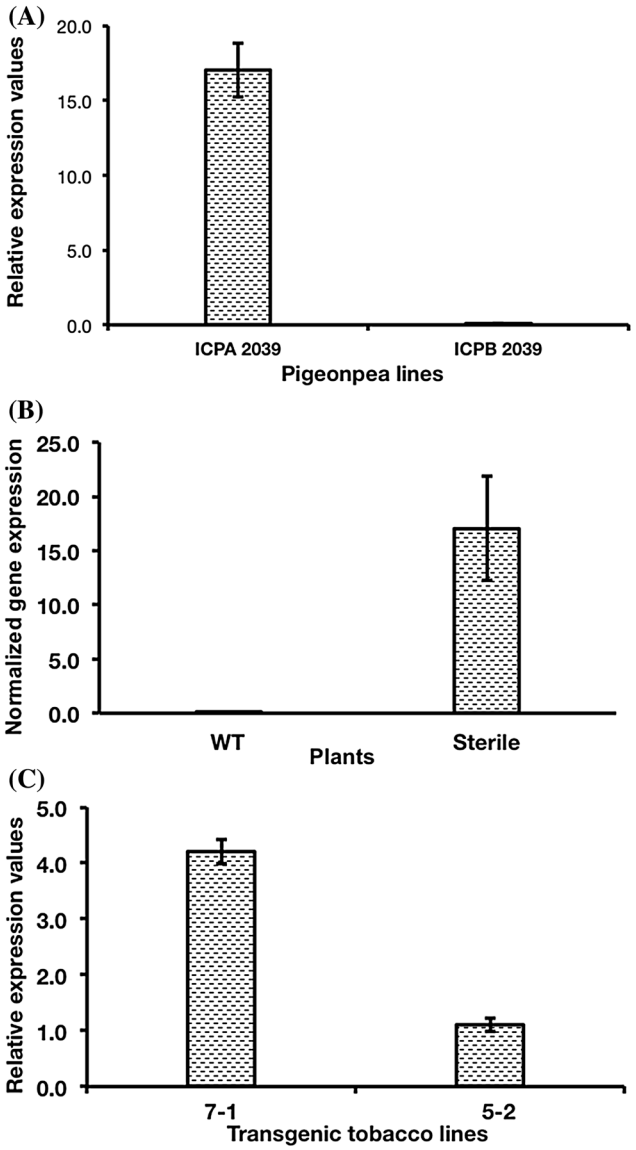


Fig. 7 Relative gene expression of mitochondrial *orf147* in inflorescences. **a** Male sterile pigeonpea line (ICPA 2039), fertile maintainer line (ICPB 2039). **b** Wild type and transgenic male sterile *Arabidopsis* plants. **c** Transgenic tobacco plants with complete (left) and partial sterility (right) were evaluated for *orf147* expression. Values are mean ± SE over triplicate reactions; gene expression was normalized with internal reference gene expression

abnormal endothecial secondary thickening (Thevenin et al. 2011). Clearly, *orf147* expression triggered transcriptional responses that affected the PCD-controlled cellular degeneration of the tapetum as well as the endothecium secondary thickening for anther maturation and pollen release.

In conclusion, our study provides evidence of direct link between a novel mitochondrial *orf147* and cytoplasmic male sterility in pigeonpea. This can lead to a greater understanding of the molecular processes underlying this phenomenon, besides possibilities to identify, develop, and deploy robust gene-specific markers for tracking CMS responsible genes. Future studies on the elucidation of restoration mechanisms and role of these nuclear factors in suppressing aberrant *orf147* expression in male sterile parent would be critical for success. Potentially, engineering CMS by utilizing this strategy could circumvent some of the issues pertaining to compromised heterosis, which currently is being encountered as a major bottleneck in pigeonpea hybrid technology. Besides, there would be unique opportunities in developing CMS systems in crops where this trait has been difficult to achieve.

Materials and methods

Plant material and growth conditions

Seeds of an A4 cytoplasm (*C. cajanifolius*) containing CMS line (ICPA 2039), the corresponding maintainer line (ICPB 2039), restorer line (ICPR 2438) and a restored hybrid (ICPH 2438) from medium maturity group were procured from Pigeonpea Breeding Unit of the International Crops Research Institute for Semi-Arid Tropics (ICRISAT) in Hyderabad, India. Details of the cytoplasmic nuclear/genetic male sterility have been described earlier (Saxena and Hingane 2015). Plants were grown in pots containing autoclaved sand and soil (1:1) mixture and maintained in glasshouse under a 16 h light, 8 h dark photoperiod at 25/23 °C (light/dark) with 70–80% relative humidity.

Genomic DNA extraction and genome walking

Total genomic DNA was prepared from fresh leaves of 1–2 weeks-old seedlings using NucleoSpin Plant II DNA isolation kit (Macherey-Nagel, Germany). Genomic DNA (500 ng) was used for genome walking as reported earlier (Reddy et al. 2008) for identifying the flanking sequences. Contigs from published mitochondrial genome of pigeonpea (Tuteja et al. 2013) of ICPA 2039 (male sterile) and ICPB 2039 (maintainer) lines were compared in silico. Based on unique rearrangement sites, a 1.29 kb region of genomic DNA upstream of the *nad7* gene in both sterile and fertile lines was amplified and cloned using PCR-based directional genome walking using cDNA sequence-specific antisense primer (5′-AATTCAAAGTGAAATTTTTG-3′).

RNA isolation and cloning of ORFs

Total RNA was isolated from leaves of 3-weeks old plants and unopened flower buds using Trizol (Invitrogen, USA) and RNeasy Plant Mini Kit (Qiagen, Germany), and used for cDNA synthesis using First-Strand cDNA Synthesis Kit (Invitrogen, USA). cDNAs from ICPA 2039 and ICPB

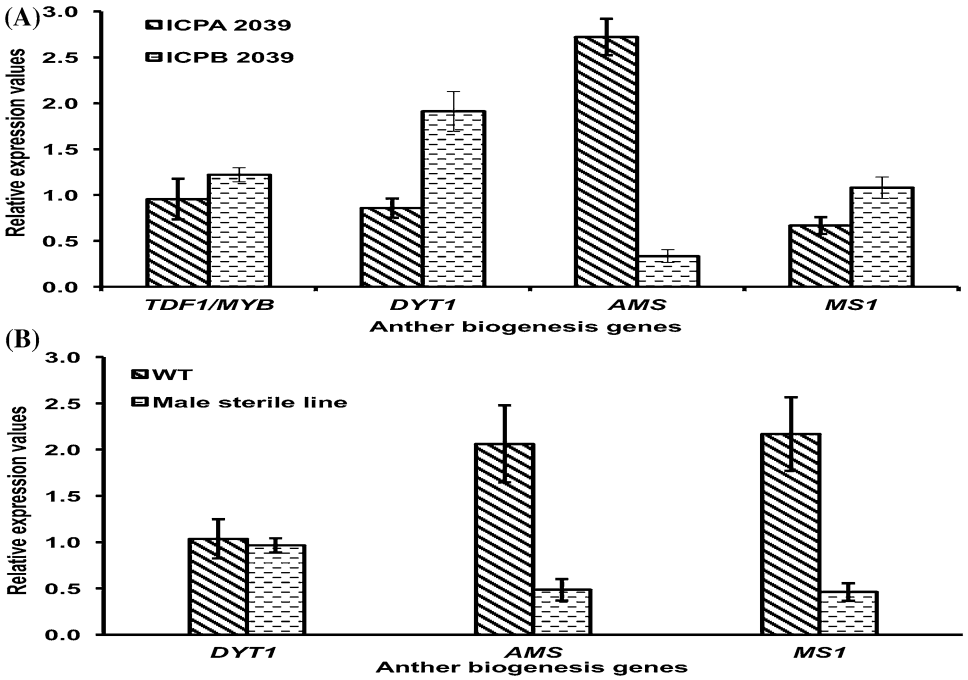


Fig. 8 Expressions of anther biogenesis genes. **a** qRT-PCR showing reduced gene expression of *TDF1, DYT and MS1* in flower buds of the male sterile pigeonpea line (ICPA 2039) when compared to the fertile maintainer line (ICPB 2039). **b** Male sterile transgenic *Arabidopsis* showing decreased expressions of key genes like *DYT1, AMS* and *MS1*. Values are mean ± SE and represent technical variation (over triplicates) from two pooled tissue samples normalized with internal reference gene expression

2039 were used for the amplification of orf sequences using specific primers (#11 and 14 respectively; Table S1). PCR fragments were used as templates for re-PCR with primers (sets 12 and 15 respectively; Table S1) containing restriction sites at the 5′ ends for cloning into bacterial expression plasmid. PCR fragments were cloned into "pJET 2.1 Blunt" plasmid (Thermo Fisher Scientific, USA) and confirmed by sequencing.

Northern blot analysis

Total RNA was fractionated on a 1.5% denaturing agarose gel containing formaldehyde and transferred onto Hybond N+ membranes (Amersham, UK). Hybridizations were carried out at 68 °C overnight in hybridization solution (Roche). PCR-amplified fragments of *orf147, orf133* and *nad7* gene fragments were used as a probe after labeling with the nonradioactive DIG Northern Starter Kit (Roche, Germany). Labeling, hybridization and detection were performed according to the manufacturer's instructions.

Identification of transcription start sites

The identification of TSS was carried out using the ARF-TSS strategy based on cDNA generation, circularization and PCR amplification, following a modified strategy based on Wang et al. (2006). Total RNAs from flower buds were subjected to cDNA synthesis with gene-specific primer (cDNA1 primer) followed by treatment with RNAse H. The cDNAs were self-ligated to undergo circularization using CircLigase™ (Epicentre, USA), and the ligated products were used as templates for PCR amplification using primers RACE F1 and RACE R1/RACE R2. The eluted PCR fragments were cloned and sequenced following agarose gel electrophoresis. The nucleotide base immediately next to the 5′ end of the gene-specific primer (cDNA1 primer) was identified as the TSS.

Sequence analysis

ORFs were predicted using Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/gorf/gorf.html) and MacVector assembly programs (V14.5.2). All predicted ORFs were verified against the publicly available mitochondrial nucleotide and protein sequence database as well as with the previously sequenced plant mitochondrial genomes. RNA-editing events were analyzed using PREP-Mt. The secondary structure of *orf147* transcripts of the CMS line was determined using *S*fold2.2-Ding RNA Bioinformatics Lab (Ding et al. 2004; http://sfold.wadsworth.org/cgi-bin/srna.pl). All the predicted protein sequences were aligned using ClustalW of MacVector software (V14.5.2). The transmembrane domains were predicted using Philius program (Reynolds et al. 2008).

Cloning of *orf 133* and *orf147* into an *E. coli* expression vector

The *orf147* from ICPA 2039 (male sterile line) and *orf133* from both sterile and fertile (ICPB 2039) lines were obtained by restriction digestion using *Nde*I and *Sal*I followed by cloning into the bacterial expression plasmid pET19b (Novagen, USA). Both fragments from the pET19b plasmid were

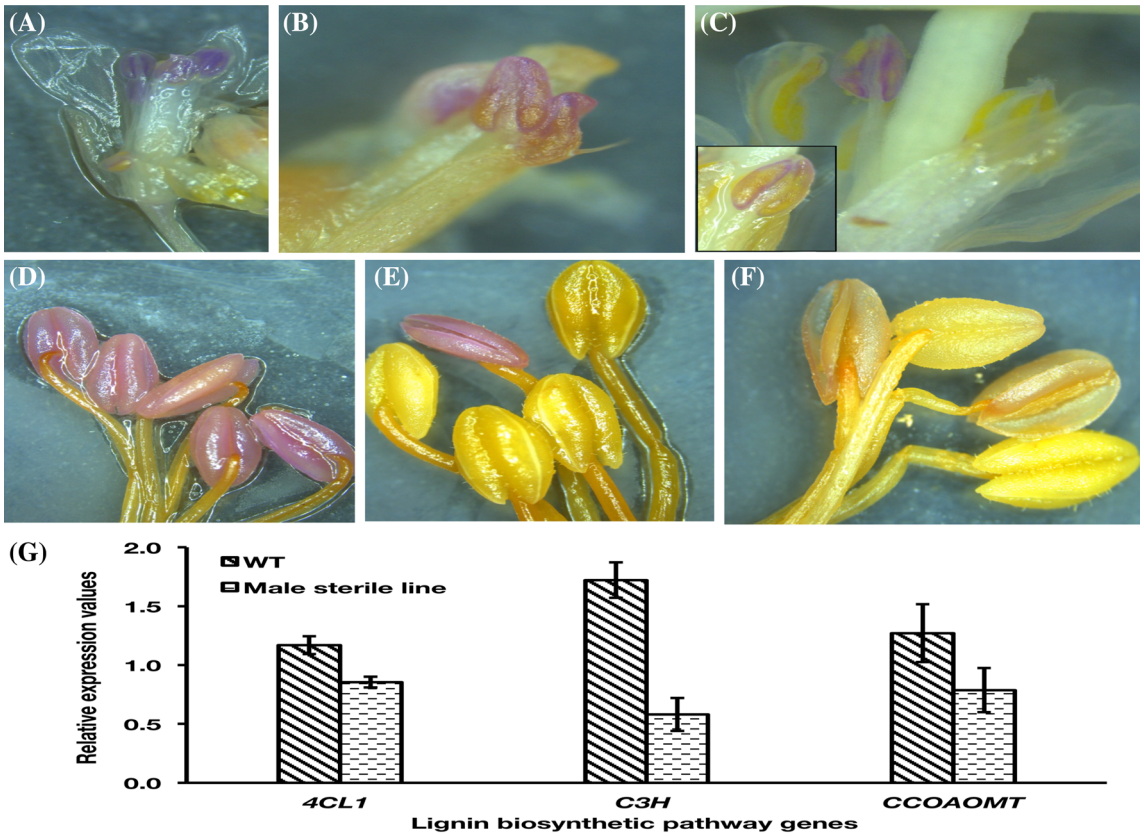


Fig. 9 Histochemical evidence of reduced lignification in phloroglucinol stained flower buds of *Arabidopsis* (**a**–**c**) and tobacco (**d**–**f**) transgenic plants expressing the *orf147*. **a** Anthers from wild type (WT) plants. **b** Close-up of deeply stained anthers from WT plants. **c** Anthers from male sterile transgenic plant with reduced staining indicating reduced lignification; (inset) partial staining of anther. **d** WT transgenic tobacco anthers showing intense staining. **e** Anthers from partial sterile tobacco transgenic plants. **f** Anthers from complete male sterile plant with negligible staining. **g** Reduced expression of lignin biosynthesis genes such as *4CL* (4 Coumarate:CoAligase), *CCoAOMT* (Caffeoyl CoA O-methyltransferase), and C3H (Cinnamic acid 3-hydroxylase) in male sterile transgenic *Arabidopsis* plants; values are mean ± SE and represent technical variation over triplicate reactions, from three pooled tissue samples normalized with internal reference gene expression

obtained by *Nco*I and *Bam*H1 digestion and cloned into pET32a adjacent to *Trx-tag*. Plasmid sequences were confirmed by sequencing and subjected to bacterial transformation into BL21 (DE3) *PlysS* strain for expression studies.

Monitoring of *E. coli* growth

Overnight-grown cultures of BL21(DE3) pLysS (Novagen, USA) harboring pET147, pET(133) or pET 32a control plasmids were obtained by using a single colony to inoculate 1.5 ml of LB medium supplemented with 500 mg l^{-1} carbenicillin and incubated overnight at 37 °C. 200 ml of LB medium supplemented with the same antibiotic was inoculated with 200 µl of the pre-culture and incubated at 37 °C with shaking. At an OD600 of 0.3, the culture was separated into two equal subcultures, induced with 0.5 mM IPTG and their growth monitored on hourly basis at 600 nm using a spectrophotometer (Eppendorf Bio Photometer plus; Eppendorf, Germany).

Plasmid construction

The 886 bp fragment of the flower-specific *AP3* promoter from *A. thaliana* was amplified using the primers AtAP3_Pro_KpnIF and AtAP3_Pro_NdeIR (set 15; Table S1), followed by cloning in pJET blunt 2.1 plasmid. The presequence mitochondrial transit peptide of the cytochrome oxidase subunit IV (CoxIV) from yeast (Köhler et al. 1997) was amplified using the TSPF and TSPR primers (Table S1) from *S. cerevisiae* cDNA. The *orf147* fragment was amplified using primers Orf 147 NdeI F and Orf 147 Not1R (Table S1), and subsequently fused to CoxIV by overlap extension PCR using the primers OE 147F and OE 147R (sets 16 and 17 respectively; Table S1). PCR amplified *AP3* promoter fragment (*Kpn*I, *Nde*I) and Cox-orf147 fusion fragment (*Nde*I, *Not*I) were together sub-cloned into a modified pL12R34H plasmid at *Kpn*I, *Not*I site, subsequently into pMDC100 (Curtis and Grossniklaus 2003),

followed by mobilization into *Agrobacterium tumefaciens* strain C58 for transformation of *Arabidopsis* and tobacco.

Transformation and confirmation of transgenic plants

Arabidopsis thaliana (Col-1) plants were transformed using floral dip protocol (Clough and Bent 1998), with inoculations repeated twice at 3 days intervals and seeds collected at maturity. Tobacco (*N. tabacum* L., var. Xanthi) seedlings were transformed using standard leaf disc method (Sunkara et al. 2013). Transgenic plants were grown in pots containing autoclaved sand and soil (1:1) in containment glasshouse until flowering and seed set at 23:20 °C (day:night) and 16 h light, 8 h dark photoperiod at 65–70% relative humidity for *Arabidopsis* and at 25:23 °C (day:night) and 16 h light, 8 h dark photoperiod at 70–80% relative humidity for tobacco. Genomic DNA from kanamycin resistant *Arabidopsis* and tobacco plants was isolated using NucleoSpin Plant II DNA isolation kit (Machery-Nagel, Germany), and subjected to PCR using *orf147* specific primers. PCR conditions included an initial denaturation cycle of 5 min at 94 °C, followed by 35 cycles of denaturation for 30 s at 94 °C, annealing for 1 min at 58 °C with an extension for 1 min at 72 °C with final extension for 10 min at 72 °C.

Candidate gene selection and primer design

Genome walking and cDNAs amplification in male sterile and fertile lines of pigeonpea was carried out using specific primers (Table S1). Seven candidate genes including *Defective in tapetal development and function* (*TDF1/MYB35*), *Aborted Microspore* (*AMS*), *Dysfunctional Tapetum1* (*DYT1*), and *Male Sterlity 1* (*MS1*), *4 Coumarate:CoAligase* (4CL), *Caffeoyl CoA O-Methyltransferase* (*CCoAOMT*), and *Cinnamic acid 3-hydroxylase* (C3H) were selected for qPCR analysis. Three reference genes *SAND, TIP41, and UNK*, showing highly stable expression (Czechowski et al. 2005) were selected as reference genes for this study. The retrieved *A. thaliana* sequences were used to design PCR primers using Primer 3 Plus software with GC content of 50%, primer length of 22 nucleotides and an expected product size of 80–150 base pairs (Table S1).

Quantitative real-time PCR analysis

All qRT-PCR reactions were carried out in Realplex (Eppendorf, Germany) Real Time PCR system using SYBR Green in 96 well optical reaction plates (Axygen, USA) sealed with ultra-clear sealing film (Platemax, USA). The PCR reaction was performed in a total volume of 10 µl, containing 1 µl of RNA (100 ng), 400 nM of each primer, 5 µl of 2× one step SYBR RT-PCR buffer 4 and 0.4 µl of prime script one step Enzyme Mix 2 (Takara, Japan) and made to 10 µl with Rnase-free H_2O. The qRT-PCR cycling conditions were as follows: 42 °C for 5 min and 95 °C for 10 s (reverse transcription) followed by 40 cycles of 15 s at 95 °C, 15 s at 62 °C with fluorescent signal recording and 15 s at 72 °C. All samples were collected from three independent plants and tested in three technical replicates. The raw Cq values of each gene were taken as the input data for estimating relative and average expression of candidate gene using qBase plus software (ver: 2.4; Biogazelle) (Hellemans et al. 2007).

Histochemical studies

Lignin content in *A. thaliana* and *N. tabacum* flower buds was analyzed histochemically using phloroglucinol-HCl staining (Liljegren 2012). Flowers were fixed in FAA (formaldehyde:acetic acid:alcohol) solution overnight and decolorized by using ethanol 25–85% series. The samples were subsequently stained with 2% (w/v) phloroglucinol in 92.5% ethanol for 1 h at room temperature, mounted on glass slides with 18.5% (v/v) HCl, and red coloration monitored immediately using Leica M125 microscope (Leica Microsystems, UK).

Acknowledgements This work was undertaken as part of the CGIAR Research Program on Grain Legumes. Thanks to Rahul Nitnavare, Chavvi Srivastava, Divya and Kedarinath for their technical assistance with the transformation and gene expression studies and PS Rao for photography. Critical reviews of the manuscript by Drs. Gopalan Selvaraj, Rajeev Gupta and Damaris Odeny are gratefully acknowledged.

Author contributions PBM and KKS conceptualized, designed and analyzed all experimental data. RG conducted expression studies in prokaryotic system, Sequence analysis was done by PSR, BPR assisted in cloning and transformation; DSR was involved in qPCR and Northern blot studies. RKS provided inputs on mitochondrial genomic sequence information and analysis. CVSK provided pigeonpea seed material. PBM and KKS conducted histochemical studies. PBM, KKS, RG and PSR contributed to manuscript preparation.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. PBM, RG and KKS are inventors on the patent applications of this work and are current employees of ICRISAT who owns the IP.

References

Araya A, Domec C, Begu D, Litvak S (1992) An in vitro system for the editing of ATP synthase subunit 9 mRNA using wheat mitochondrial extracts. Proc Natl Acad Sci USA 89:1040–1044

Balk J, Leaver CJ (2001) The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome c release. Plant Cell 13:1803–1818

Bergman P, Edqvist J, Farbos I, Glimelius K (2000) Male-sterile tobacco displays abnormal mitochondrial atp1 transcript

accumulation and reduced floral ATP/ADP ratio. Plant Mol Biol 42:531–544
Bonhomme S, Budar F, Lancelin D, Small I, Defrance M, Pelletier G (1992) Sequence and transcript analysis of the Nco2.5 Ogura-specific fragment correlated with cytoplasmic male sterility in *Brassica* cybrids. Mol Gen Genet 235:340–348
Bonner L, Dickinson H (1989) Anther dehiscence in *Lycopersicon esculentum* Mill. New Phytol 113:97–115
Chen L, Liu Y-G (2014) Male sterility and fertility restoration in crops. Annu Rev Plant Biol 65:579–606
Clough SJ, Bent AF (1998) Floral dip, a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16:735–743
Curtis M, Grossniklaus U (2003) A Gateway TM cloning vector set for high-throughput functional analysis of genes in plants. Plant Physiol 133:462–469
Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. Plant Physiol 139:5–17
Dalvi VA, Saxena KB, Madrap IA (2008) Fertility restoration in cytoplasmic-nuclear male-sterile lines derived from 3 wild relatives of pigeonpea. J Hered 99:671–673
Dawson J, SÖzen E, Vizir I, Van Waeyenberge S, Wilson ZA, Mulligan BJ (1999) Characterization and genetic mapping of a mutation (ms35) which prevents anther dehiscence in *Arabidopsis thaliana* by affecting secondary wall thickening in the endothecium. New Phytol 144:213–222
Dewey RE, Levings CS III, Timothy DH (1986) Novel recombinations in the maize mitochondrial genome produce a unique transcriptional unit in the Texas male-sterile cytoplasm. Cell 44:439–449
Dieterich JH, Braun HP, Schmitz UK (2003) Alloplasmic male sterility in *Brassica napus* (CMS "Tournefortii-Stiewe") is associated with a special gene arrangement around a novel *atp9* gene. Mol Genet Genomics 269:723–731
Ding Y, Chan CY, Lawrence CE (2004) Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res 32(Issue supplement 2):W135–W141
Duroc Y, Gaillard C, Hiard S, Defrance M, Pelletier G, Budar F (2005) Biochemical and functional characterization of ORF138, a mitochondrial protein responsible for Ogura cytoplasmic male sterility in Brassiceae. Biochimie 87:1089–1100
Fujii S, Toriyama K (2008) DCW11, Down-regulated gene 11 in CW-type cytoplasmic male sterile rice, encoding mitochondrial protein phosphatase 2C is related to cytoplasmic male sterility. Plant Cell Physiol 49:633–640
Fujii S, Toriyama K (2009) Suppressed expression of RETROGRADE-REGULATED MALE STERILITY restores pollen fertility in cytoplasmic male sterile rice plants. Proc Natl Acad Sci USA 106:9513–9518
Gallagher LJ, Betz SK, Chase CD (2002) Mitochondrial RNA editing truncates a chimeric open reading frame associated with S male-sterility in maize. Curr Genet 42:179–184
Geisler DA, Päpke C, Obata T, Nunes-Nesi A, Matthes A, Schneitz K, Maximova E et al (2012) Down regulation of the δ-subunit reduces mitochondrial ATP synthase levels, alters respiration, and restricts growth and gametophyte development in *Arabidopsis*. Plant Cell 24:2792–2811
Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16:S154–S169
Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 8:R19
Hill TA, Day CD, Zondlo SC, Thackeray AG, Irish VF (1998) Discrete spatial and temporal cis-acting elements regulate transcription of the *Arabidopsis* floral homeotic gene APETALA3. Development 125:1711–1721
Holec S, Lange H, Kuhn K, Alioua M, Borner T, Gagliardi D (2006) Relaxed transcription in *Arabidopsis* mitochondria is counter balanced by RNA stability control mediated by poly-adenylation and polynucleotide phosphorylase. Mol Cell Biol 26:2869–2876
Howad W, Tang HV, Pring DR, Kempken F (1999) Nuclear genes from T× CMS maintainer lines are unable to maintain atp6 editing in any anther cell-type in the *Sorghum bicolor* A3 cytoplasm. Curr Genet 36:62–68
Hu J, Wang K, Huang W, Liu G, Gao Y, Wang J, Huang Q et al (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
Igarashi K, Kazama T, Motomura K, Toriyama K (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
Ito T, Ng KH, Lim TS, Yu H, Meyerowitz EM (2007) The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis*. Plant Cell 19:3516–3529
Iwabuchi M, Kyozuka J, Shimamoto K (1993) Processing followed by complete editing of an altered mitochondrial apt6 RNA restores fertility of cytoplasmic male sterile rice. EMBO J 12:1437–1446
Iwabuchi M, Koizuka N, Fujimoto H, Sakai T, Imamura J (1999) Identification and expression of the kosena radish (*Raphanus sativus* cv. Kosena) homologue of the ogura radish CMS-associated gene, orf138. Plant Mol Biol 39:183–188
Jing B, Heng S, Tong D, Wan Z, Fu T, Tu J, Ma C et al (2011) A male sterility associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. J Exp Bot 63:1285–1295
Kim DH, Kang JG, Kim B-D (2007) Isolation and characterization of the cytoplasmic male sterility-associated orf456 gene of chili pepper (*Capsicum annuum* L.). Plant Mol Biol 63:519–532
Köhler RH, Horn R, Lössl A, Zetsche K (1991) Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. Mol Gen Genet 227:369–376
Köhler RH, Cao J, Zipfel WR, Webb WW, Hanson MR (1997) Exchange of protein molecules through connections between higher plant plastids. Science 276:2039–2042
Krishnasamy S, Makaroff CA (1993) Characterization of the radish mitochondrial *orfB* locus, possible relationship with male sterility in Ogura radish. Curr Genet 24:156–163
Kurek I, Ezra D, Begu D, Erel N, Litvak S, Breiman A (1997) Studies on the effects of nuclear background and tissue specificity on RNA editing of the mitochondrial ATP synthase subunits α, 6 and 9 in fertile and cytoplasmic male-sterile (CMS) wheat. Theor Appl Genet 95:1305–1311
Landgren M, Zetterstrand M, Sundberg E, Glimelius K (1996) Alloplasmic male-sterile *Brassica* lines containing *B. tournefortii* mitochondria express an ORF 3′ of the *atp6* gene and a 32 kDa protein. Plant Mol Biol 32:879–890
Laver HK, Reynolds SJ, Moneger F, Leaver CJ (1991) Mitochondrial genome organization and expression associated with cytoplasmic male sterility in sunflower (*Helianthus annuus*). Plant J 1:185–193
Li S, Wan C, Kong J, Zhang Z, Li Y, Zhu Y (2004) Programmed cell death during microgenesis in a Honglian CMS line of rice is correlated with oxidative stress in mitochondria. Funct Plant Biol 31:369–376
Li N, Zhang D-S, Liu H-S, Yin C-S, Li X-X, Liang W-Q, Yuan Z et al (2006) The rice tapetum degeneration retardation gene is

required for tapetum degradation and anther development. Plant Cell 18:2999–3014
Li J, Pandeya D, Jo Y-D, Liu W-Y, Kang B-C (2013) Reduced activity of ATP synthase in mitochondria causes cytoplasmic male sterility in chili pepper. Planta 237:1097–1109
Liljegren S (2012) Phloroglucinol stain for lignin. Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot4954
Luo D, Xu H, Liu Z, Guo J, Li H, Chen L, Fang C et al (2013) A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat Genet 45:573–577
Mackenzie SA, McIntosh L (1999) Higher plant mitochondria. Plant Cell 11:571–585
Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M (2007) NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. Plant Cell 19:270–280
Niu NN, Liang WQ, Yang XJ, Jin WL, Wilson ZA, Hu JP, Zhang DB (2013) EAT1 promote tapetal cell death by regulating aspartic proteases during male reproductive development in rice. Nat Commun 4:1445
Nivison HT, Hanson MR (1989) Identification of a mitochondrial protein associated with cytoplasmic male sterility in petunia. Plant Cell 1:1121–1130
Peng X, Wang K, Hu C, Zhu Y, Wang T, Yang J, Tong J, Li S, Zhu Y (2010) The mitochondrial gene *orfH79* plays a critical role in impairing both male gametophyte development and root growth in CMS-Honglian rice. BMC Plant Biol 10:125. https://doi.org/10.1186/1471-2229-10-125
Pring DR, Tang HV, Chase CD, Siripant MN (2006) Microspore gene expression associated with cytoplasmic male sterility and fertility restoration in sorghum. Sex Plant Reprod 19:25–35
Reddy PS, Mahanty S, Kaul T, Nair S, Sopory SK, Reddy MK (2008) A high-throughput genome-walking method and its use for cloning unknown flanking sequences. Anal Biochem 381:248–253
Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS (2008) Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. PLoS Comput Biol 4:e1000213. https://doi.org/10.1371/journal.pcbi.1000213
Sabar M, Gagliardi D, Balk J, Leaver CJ (2003) ORFB is a subunit of F1F (O)-ATP synthase, insight into the basis of cytoplasmic male sterility in sunflower. EMBO Rep 4:381–386
Saxena KB (2015) From concept to field: evolution of hybrid pigeonpea technology in India. Ind J Genet 75:279–293
Saxena KB, Hingane AJ (2015) Male sterility systems in major field crops and their potential role in crop improvement. In: Plant biology and biotechnology, plant diversity, organization, function and improvement. Springer, New Delhi, pp 639–656
Saxena KB, Kumar RV, Srivastava N, Shiying B (2005) A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. Euphytica 145:289–294
Saxena KB, Ravikoti VK, Dalvi VA, Pandey LB, Gaddikeri G (2010) Development of cytoplasmic nuclear male sterility, its inheritance, and potential use in hybrid pigeonpea breeding. J Hered 101:497–503
Saxena KB, Sameerkumar CV, Hingane AJ, Nagesh Kumar MV, Vijaykumar RA, Saxena RK, Patil S, Varshney RK (2016) Hybrid ICPH 2740 assures quantum jump in pigeonpea productivity in peninsular India. J Food Legum 29:142–144
Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3:175–180
Sinha P, Saxena KB, Saxena RK, Singh VK, Suryanarayana V, Sameerkumar CV, Katta MAVS. et al (2015) Association of *nad7a* gene with cytoplasmic male sterility in pigeonpea (*Cajanus cajan*). Plant Genome 8:1–12
Sorensen AM, Kröber S, Unte US, Huijser P, Dekker K, Saedler H (2003) The *Arabidopsis ABORTED MICROSPORES* (*AMS*) gene encodes a MYC class transcription factor. Plant J 33:413–423
Steiner-Lange S, Unte US, Eckstein L, Yang C, Wilson ZA, Schmelzer E, Dekker K, Saedler H (2003) Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. Plant J 34:519–528
Sunkara S, Bhatnagar-Mathur P, Sharma KK (2013) Isolation and functional characterization of two novel seed specific promoters from legumes. Appl Biochem Biotechnol 172:325–339
Tang HV, Pring DR, Shaw LC, Salazar RA, Muza FR, Yan B, Schertz KF (1996) Transcript processing internal to a mitochondrial open reading frame is correlated with fertility restoration in male-sterile sorghum. Plant J 10:123–133
Tang HV, Chen W, Pring DR (1999) Mitochondrial *orf107* transcription, editing, and nucleolytic cleavage conferred by the gene *Rf3* are expressed in sorghum pollen. Sex Plant Reprod 12:53–59
Thevenin J, Pollet B, Letarnec B, Saulnier L, Gissot L, Maia-Grondard A, Lapierre C, Jouanin L (2011) The simultaneous repression of CCR and CAD, two enzymes of the lignin biosynthetic pathway, results in sterility and dwarfism in *Arabidopsis thaliana*. Mol Plant 4:70–82
Tuteja R, Saxena RK, Davila J, Shah T, Chen W, Xiao Y-L, Fan G et al (2013) Cytoplasmic male sterility-associated chimeric open reading frames identified by mitochondrial genome sequencing of four *Cajanus* genotypes. DNA Res 20:485–495
Vizcay-Barrena G, Wilson ZA (2006) Altered tapetal PCD and pollen wall development in the *Arabidopsis* ms1 mutant. J Exp Bot 57:2709–2717
Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D et al (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
Wen L, Ruesch KL, Ortega VM, Kamps TL, Gabay-Laughnan S, Chase CD (2003) A nuclear restorer-of-fertility mutation disrupts accumulation of mitochondrial ATP synthase subunit alpha in developing pollen of S male-sterile maize. Genetics 165:771–779
Wilson ZA, Zhang DB (2009) From *Arabidopsis* to rice: pathways in pollen development. J Exp Bot 60:1479–1492
Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ (2001) The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J 28:27–39
Xiao H, Zhang F, Zheng Y (2006) The 5′ stem-loop and its role in mRNA stability in maize S cytoplasmic male sterility. Plant J 47:864–872
Xu P, Yang Y, Zhang Z, Chen W, Zhang C, Zhang L, Zou S, Ma Z (2008) Expression of the nuclear gene TaF(A)d is under mitochondrial retrograde regulation in anthers of male sterile wheat plants with timopheevii cytoplasm. J Exp Bot 59:1375–1381
Xu J, Yang CY, Yuan Z, Zhang DS, Gondwe MY, Ding ZW, Liang W, Zhang D, Wilson ZA (2010) The ABORTED MICROSPORES regulatory network is required for post-meiotic male reproductive development in *Arabidopsis thaliana*. Plant Cell 22:91–107
Yamamoto MP, Shinada H, Onodera Y, Komaki C, Mikami T, Kubo T (2008) A male sterility-associated mitochondrial protein in wild beets causes pollen disruption in transgenic plants. Plant J 54:1027–1036
Yang C, Xu Z, Song J, Conner K, Vizcay-Barrena G, Wilson ZA (2007) *Arabidopsis* MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. Plant Cell 19:534–548
Yang J, Liu X, Yan X, Zhang M (2010) Mitochondrially-targeted expression of a cytoplasmic male sterility-associated *orf220* gene causes male sterility in *Brassica juncea*. BMC Plant Biol 10:231

Yoshimi M, Kitamura Y, Isshiki S, Saito T, Yasumoto K, Terachi T, Yamagishi H (2013) Variations in the structure and transcription of the mitochondrial *atp* and *cox* genes in wild Solanum species that induce male sterility in eggplant (*S. melongina*). Theor Appl Genet 126:1851–1859

Young EG, Hanson MR (1987) A fused mitochondrial gene associated with cytoplasmic male sterility is developmentally regulated. Cell 50:41–49

Zhang W, Sun YJ, Timofejeva L, Chen CB, Grossniklaus U, Ma H (2006) Regulation of *Arabidopsis* tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development 133:3085–3095

Zhang D-S, Liang W-Q, Yuan Z, Li N, Shi J, Wang J, Liu Y-M, Zhang D-B (2008) Tapetum degeneration retardation is critical for aliphatic metabolism and gene regulation during rice pollen development. Mol Plant 1:599–610

Zhang DB, Luo X, Zhu L (2011) Cytological analysis and genetic control of rice anther development. J Genet Genomics 38:379–390

Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, Guan YF, Yang ZN (2008) Defective in tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. Plant J 55:266–277