



Characterization of genes specific to *sua*-CMS in *Nicotiana tabacum*

Yeqiang Zheng^{1,3} · Zhiwen Liu^{1,3} · Yuhe Sun^{1,2} · Guanshan Liu^{1,2} · Aiguo Yang^{1,2} · Fengxia Li^{1,2}

Received: 29 March 2018 / Accepted: 1 June 2018 / Published online: 29 June 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Key message Six unique ORFs were characterized in tobacco plants with *sua*-CMS sterile cytoplasm, identifying the mtDNA basis for pollen sterility.

Abstract *sua*-CMS (cytoplasmic male sterility), the most widely used sterile system in tobacco hybrids, is the only CMS type identified as having no negative effects on agronomic or quality traits in tobacco (*Nicotiana tabacum*) and as being fully male sterile. CMS is often associated with alterations of mitochondrial DNA (mtDNA), including novel chimeric open reading frames (ORFs), which result from rearrangement and recombination. Here, we obtained 34 mitochondrial ORFs in the *sua*-CMS line msZhongyan100 (sZY) by BLAST analysis. When we amplified these mitochondrial ORFs in seven tobacco CMS lines including *sua*-, *glu*-, *rep*-, *rus*-, *tab1*-, *tab2*-, and *tab3*-CMS types and in fertile tobacco, we found that six ORFs—*orf82*, *orf103*, *orf115^a*, *orf91*, *orf115^b*, and *orf100*—were located in three small regions (m-sr) of the mitochondrial genome of sZY and were unique to the *sua*-CMS line. We further amplified the m-sr fragments in three different backcross populations of the seven types of CMS, three F1 hybrids with *sua*-CMS sterile cytoplasm, two *sua*-CMS lines, and 284 fertile tobacco accessions. The ORFs were specific to plants with the *sua*-CMS background. All six unique ORFs were chimeric and had no homology with the mitochondrial genomes of fertile tobacco. Transcript analysis revealed that the ORFs were highly expressed in the anthers and floral buds of sZY. These six ORFs were specific to *sua*-CMS and could be used as molecular markers to identify *sua*-CMS lines, which is useful for improving breeding for heterosis in tobacco.

Keywords Tobacco · *sua*-CMS · Mitochondrial ORF · Molecular markers · Transcript analysis

Communicated by Carlos F. Quiros.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00299-018-2309-2>) contains supplementary material, which is available to authorized users.

✉ Aiguo Yang
yangaiguo@caas.cn

✉ Fengxia Li
lifengxia@caas.cn

¹ Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101, China

² Department of Key Laboratory for Tobacco Gene Resources, State Tobacco Monopoly Administration, Qingdao 266101, China

³ Department of Graduate School of Chinese Academy of Agricultural Science, Beijing 100081, China

Introduction

Cytoplasmic male sterility (CMS) is a trait in plants where viable pollen cannot be produced. CMS is maternally inherited and occurs in more than 150 higher plant species and is important in the evolution of gynodioecious species (Dufay et al. 2007). CMS has been extensively exploited for producing hybrids in a wide variety of crops (Touzet and Meyer 2014; Chen and Liu 2014) showing heterosis. Heterosis refers to performance of hybrids that is superior to that of the parental lines from which they are derived (Baranwal et al. 2012).

Stamen-defective CMS is determined by variation of mitochondrial genes, resulting in a lack of compatibility between the nucleus and the cytoplasm (Tang et al. 2017). Chimeric mitochondrial open reading frames (ORFs), which are generated by rearrangement and recombination of mitochondrial genomes, are the primary or only cause of CMS (Horn et al. 2014). In the mitochondrial genome, chimeric genes are often co-transcribed with genes encoding

respiratory chain components, producing aberrant transcripts or encoding toxic proteins that interfere with the normal functioning of mitochondria at critical stages of anther development, resulting in pollen sterility (Touzet and Meyer 2014).

In tobacco (*Nicotiana tabacum*), over 30 male sterile lines have been developed, all of which were derived from somatic protoplast fusion or the traditional breeding to result in interspecies hybridization between tobacco and other *Nicotiana* species, such as *N. debeni*, *N. megalosiphon*, *N. suaveolens*, *N. bigelovii*, *N. hesperis*, *N. rustica*, *N. glauca*, and *N. repanda*. *sua*-CMS is the only sterile cytoplasm thus far identified as having no negative effects on growth or quality (Tong 1997) and is, therefore, widely used to produce hybrid tobacco. At present, most of the CMS lines in China, America, Japan, and Brazil used in hybrid tobacco are *sua*-CMS (Cao and Yu 2017).

The sterile cytoplasm of the *sua*-CMS line was developed by somatic fusion between tobacco and *N. suaveolens* in the 1950s. Ma et al. (2005) confirmed that there were no significant differences between the *sua*-CMS lines and the homogeneous maintainer lines in terms of biological characteristics, agronomic traits, or disease resistance. The *sua*-CMS flower is similar to that of the maintainer lines in all traits (shape, sepal morphology, petal color, and pistil morphology) except male sterility. Cytological studies of *sua*-CMS lines showed a lack of male reproductive ability between primordial differentiation of the stamens and the formation of sporogenous cells (Tong 1997).

Development and identification of different CMS types are necessary for sustainability of heterosis in a crop species. Cytology, protein markers, and DNA markers are often used to identify CMS lines. Some cytological studies about anther development in tobacco CMS have been reported (Pollak 1992; Farbos et al. 2001; Fitter et al. 2005). However, anther development, especially in male sterile plants, is often affected by temperature, photoperiod, or moisture (Yuan et al. 1993; Xue et al. 2018). Molecular markers based on DNA polymorphisms are relatively stable and are widely used to identify CMS lines (Bonhomme et al. 1992). The associated genes for CMS are often used as specific markers to identify different sterile cytoplasm. In the *sua*-CMS line of tobacco, some SNPs located in mitochondrial genes were identified (Zhao et al. 2009). The mitochondrial genome of the *sua*-CMS line has been sequenced (Li et al. 2016), allowing for variations in ORFs associated with *sua*-CMS to be developed.

The *sua*-CMS lines of tobacco are used globally to produce male sterile plants and increase heterosis, yet a little work has been done to characterize the underlying genes. Description of the genetic basis of these lines could assist with breeding efforts and the development of new lines. In this study, six specific ORFs of the tobacco *sua*-CMS lines

were characterized and the expression patterns of the ORFs were analyzed.

Materials and methods

Plant materials

The plant materials in this study included seven types of CMS (*sua*-, *glu*-, *rep*-, *rus*-, *tab1*-, *tab2*-, and *tab3*-) and their maintainer lines. Each CMS type included a CMS line (BC5F1–BC10F1) and three BC1F1 offspring obtained from the male sterile line and the three common maintainer lines. In addition, two *sua*-CMS lines that were backcrossed more than ten generations with their corresponding maintainer lines and three F1 hybrids that harbored *sua*-sterile cytoplasm were also used to identify genes unique to *sua*-CMS. The material characters and types of CMS used in this study are listed in Table 1. All plant materials were grown in the experimental field of the Tobacco Research Institute in Qingdao, Shandong, China in 2017.

Nicotiana accessions (284) were obtained from the Chinese Tobacco Germplasm Repository (Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, Shandong, China; Table S1). These belonged to three subgenera: subgenus *Nicotiana* included 201 flue-cured, 42 sun-cured, 12 burley, three cigar, and nine oriental tobacco accessions, and two wild-type species; subgenus *Rustica* included nine rustica tobacco accessions and one wild-type species; and subgenus *Petunioides* included five wild-type species. These plant materials were grown in the experimental field of the Tobacco Research Institute in Qingdao, Shandong, China in 2014–2016.

Observation of stamen morphology and scanning electron microscopy of pollen

Around 10:00 AM in sunny weather in July 2017, emerging flowers were sampled from fully flowering plants of the *sua*-CMS line msZhongyan100 (sZY) and the maintainer line Zhongyan100 (ZY). Petals and sepals were carefully peeled back to observe the morphology of the stamens. Mature anthers of sZY and ZY were collected and fixed in a glutaraldehyde fixation solution. The pollen grains were sputter-coated with platinum particles and examined using a scanning electron microscope (JEOL: JSM-840, Tokyo, Japan).

Pollen vitality assay

To test pollen viability, mature anthers were stained with fluorescein diacetate (FDA) according to the methods of McCabe et al. (1997). Anthers were cut with a sharp blade

Table 1 Characters and types of tobacco plant materials evaluated in this study

No.	Name (code)	Character	Type of CMS
1	msZhongyan100	Sterile line, BC10F1	<i>sua</i> -CMS
2	86-6	Sterile line, BC10F1	<i>glu</i> -CMS
3	98-43	Sterile line, BC10F1	<i>rep</i> -CMS
4	200-18	Sterile line, BC10F1	<i>rus</i> -CMS
5	L1	Sterile line, BC6F1	<i>tab1</i> -CMS
6	4-1	Sterile line, BC6F1	<i>tab2</i> -CMS
7	CH03	Sterile line, BC5F1	<i>tab3</i> -CMS
8	Zhongyan100	Maintainer line of msZhongyan100	Fertile cytoplasm
9	MS2017-1	BC1F1, msZhongyan100×Speight G-28	<i>sua</i> -CMS
10	MS2017-2	BC1F1, msZhongyan100×K326	<i>sua</i> -CMS
11	MS2017-3	BC1F1, msZhongyan100×NC82	<i>sua</i> -CMS
12	MS2017-4	BC1F1, 86-6×Speight G-28	<i>glu</i> -CMS
13	MS2017-5	BC1F1, 86-6×K326	<i>glu</i> -CMS
14	MS2017-6	BC1F1, 86-6×NC82	<i>glu</i> -CMS
15	MS2017-7	BC1F1, 98-43×Speight G-28	<i>rep</i> -CMS
16	MS2017-8	BC1F1, 98-43×K326	<i>rep</i> -CMS
17	MS2017-9	BC1F1, 98-43×NC82	<i>rep</i> -CMS
18	MS2017-10	BC1F1, 200-18×Speight G-28	<i>rus</i> -CMS
19	MS2017-11	BC1F1, 200-18×K326	<i>rus</i> -CMS
20	MS2017-12	BC1F1, 200-18×NC82	<i>rus</i> -CMS
21	MS2017-13	BC1F1, L1×Speight G-28	<i>tab1</i> -CMS
22	MS2017-14	BC1F1, L1×K326	<i>tab1</i> -CMS
23	MS2017-15	BC1F1, L1×NC82	<i>tab1</i> -CMS
24	MS2017-16	BC1F1, 4-1×Speight G-28	<i>tab2</i> -CMS
25	MS2017-17	BC1F1, 4-1×K326	<i>tab2</i> -CMS
26	MS2017-18	BC1F1, 4-1×NC82	<i>tab2</i> -CMS
27	MS2017-19	BC1F1, CH03×Speight G-28	<i>tab3</i> -CMS
28	MS2017-20	BC1F1, CH03×K326	<i>tab3</i> -CMS
29	MS2017-21	BC1F1, CH03×NC82	<i>tab3</i> -CMS
30	MS2017-22	Speight G-28, maintainer line	Fertile cytoplasm
31	MS2017-23	K326, maintainer line	Fertile cytoplasm
32	MS2017-24	NC82, maintainer line	Fertile cytoplasm
33	msYunyan87	Sterile line, BC10F1	<i>sua</i> -CMS
34	msK326	Sterile line, BC10F1	<i>sua</i> -CMS
35	KRK26	F1 hybrid, popularized variety	<i>sua</i> -CMS
36	YUN203	F1 hybrid, popularized variety	<i>sua</i> -CMS
37	Zhongyan202	F1 hybrid, popularized variety	<i>sua</i> -CMS

on a glass slide in a few drops of FDA working solution and pollen grains were released. After incubation in the dark for 5 min, pollen grains were observed using a microscope (Nikon Eclipse 80i, Tokyo, Japan) equipped with a charge-coupled device (CCD) for photography.

Annotation and selection of ORFs

The potential ORFs in mitochondrial genomes were identified using the Artemis software (Rutherford et al. 2000). ORF-Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) and tRNAscan-SE (Lowe and Eddy 1997) were also used to further

define ORFs of at least 210 nt. Common ORFs with >99% shared identity between sZY and ZY were identified using blastn and blastx (Altschul et al. 1990), and were subsequently discarded. To improve identification accuracy, the unique ORFs in the mitochondrial genome of sZY were compared with additional plant mitochondrial genome annotations in NCBI using blastn and blastx.

Isolation of total DNA and PCR amplification of mitochondrial gene-specific primers

Total DNA was isolated from young leaves using the CTAB method (Doyle and Doyle 1990). We designed 27 primer pairs to amplify 34 unique ORFs of *sua*-CMS in *sua*-, *glu*-, *rep*-, *rus*-, *tab1*-, *tab2*-, and *tab3*-CMS types and in fertile tobacco. Each primer pair contained one, two, or three complete ORFs according to the positions of the ORFs. Primer pairs used in this study are listed in Table 2.

PCR was performed in 10 μ L reaction mixtures containing 100 ng total DNA, 1 unit *Taq* DNA polymerase (Thermo Scientific, Waltham, MA, USA), 0.2 mM dNTP mix, 1 μ L of 10 \times *Taq* buffer with $(\text{NH}_4)_2\text{SO}_4$, and 0.5 μ M of each primer (Tsingke Biotech, Beijing, China). The PCR protocol was as follows: 94 $^\circ\text{C}$ for 5 min; 30 cycles of 94 $^\circ\text{C}$ for 30 s; 50–56 $^\circ\text{C}$ for 30 s according to the annealing temperature of each primer pair, 72 $^\circ\text{C}$ for 45 s, and a final extension at 72 $^\circ\text{C}$ for 7 min. PCR products were detected by electrophoresis on 1% agarose gels, followed by staining with ethidium bromide.

Cloning and sequencing

The target DNA fragments were recovered using a PCR purification kit (TIANGEN Biotech, Beijing, China). The recovered DNA fragments were ligated into pMD18-T vectors (TaKaRa, Shiga, Japan), and the products were transformed into competent cells (DH5a). Single-colony cultures containing the inserted target fragments were verified by PCR and sequenced (Tsingke Biotech).

Screening and characterization of mitochondrial ORFs of *sua*-CMS

To assay the stability of the specific mitochondrial genes in *sua*-CMS, we amplified the mitochondrial small region (m-sr) fragments in three different backcross populations (BC1F1s), three F1 hybrids with *sua*-CMS sterile cytoplasm, and two *sua*-CMS lines (BC10F1). BC1F1s were sZY/Speight G-28//Speight G-28, sZY/K326//K326, and sZY/NC82//NC82, and the six different types of CMS (*glu*-CMS, *rep*-CMS, *rus*-CMS, *tab1*-CMS, *tab2*-CMS, and *tab3*-CMS)

Table 2 Primer sequences used for amplification of open reading frames (ORFs) in tobacco hybrids

No.	Primer sequences (5'–3')	ORFs
1	CTTCTTTCTCTCTCCGCT; CGCATTATCTTCATTCCAC	<i>orf95</i>
2	CGCTTAGAGTTCGGTGGA; GTGCAGTTGAAGTGGAGT	<i>orf77^a</i>
3	ATCCTCCTTTTTCAACT; GTGTGCCTACATTCTTAG	<i>orf76</i>
4	GCAGAGGTGACAAAGCAT;GAAACCAAACAAAGAGCG	<i>orf82</i>
5	CTATTGAATGCGTTGTCGG; TGAGCTCTATGCCTTTGCC	<i>orf77^b</i>
6	TACATGAGATGCGGTTGGA; TACTGGGTAGATCAGTTGG	<i>orf116</i>
7	GGTGCGAAATTCTATTCT; TATTCCCCTCGTGCTTAA	<i>orf71</i>
8	TCAGGTCCAGAGGGGA; CGGAGAAAGCGAGATA	<i>orf88^a</i>
9	TGATTAGAGTTCAGCTCAG; GTAGTCAAGTTCAAGGGTC	<i>orf142; orf149</i>
10	AGGAAGTCCACAGTAAGG; TGGAATATGTAGGCGTAG	<i>orf73</i>
11	TTTTTATTGATTTCATAC; TCCAGTCTTGTTTCATGCT	<i>orf93; orf73</i>
12	GCAGTGAGTGAGTCCCTG; GTCGTATTTGCTTTTGAA	<i>orf74^a</i>
13	TCCTAGTTGGAGTTGGCA; GAATCTATTTTCGGTCTTT	<i>orf83</i>
14	ATCGGCTTGACCTCTTTTC; AACTTGCTGGCTTGCTCC	<i>orf79</i>
15	ACATCTCGGATAGGGAGGG; GAGGGCGCTGCTGTTGAAC	<i>orf74^b</i>
16	GGAGTTCCCCCCTTGAAG; CGGTGCCGAAGTTGGTT	<i>orf103; orf115^a</i>
17	CAGGCGAGTAGAAGGATA;GTGTTGGATGAGAAGCAG	<i>orf111^a; orf71</i>
18	ATGAAGAGGCGCAGAAGC;TAACCGGAAAGGGCAAAA	<i>orf78</i>
19	CCGTGAGTGTAAGGGGTA; TCCAGAAATGAAAAAGG	<i>orf89</i>
20	CTTTGTGATTTCCATCT;TCTTCTCTTTCTTTGTC	<i>orf74^c</i>
21	GCATTGATTTATGGGACA;ACCTACTGAAGATGGACA	<i>orf111^b</i>
22	TCTTCTTTTCTTTGCCT; ATCTAGTTGTGATTGCCA	<i>orf91</i>
23	ATCAGGACGGTAGCACTT; GGTCTTTCCAAAACAAA	<i>orf91; orf115^b; orf109</i>
24	AAGTTCAAGACGAATAAG; GATCGAGTAAACAAAAAG	<i>orf72; orf115^c</i>
25	GAGGGAACGAGCGAAAGAT; GATGGACACTGGCTTGG	<i>orf94^a</i>
26	AGATCACATCTCCCGTTACTA; CTCCTCTCCACTTCCTTAT	<i>orf88^b</i>
27	GAGGGAACGAGCGAAAGAT; GATGGACACTGGCTTGGGG	<i>orf94^b</i>

had three similar nuclear backgrounds. The F1 hybrids were msK326×RW, msZY90×CT107 and ms317×kx10.

We further amplified m-sr fragments in fertile tobacco accessions. These accessions belonged to the three subgenera (see “Plant materials”) and included flue-cured, sun-cured, burley, cigar, oriental, and rustica tobacco types, as well as wild-type species.

Transcript analysis

Total RNA was isolated from tobacco leaves, floral buds, anthers, and roots using a total RNA reagent kit (Thermo Scientific). After removing any possible genomic DNA contaminants using RNase-Free DNase I (Promega, Madison, WI, USA), cDNA synthesis was performed using a Prime-Script first-strand cDNA synthesis kit (TaKaRa). The RT-PCR reaction was terminated after 24 cycles, and tobacco *Tublin* was used as the internal control. Primer sequences of the specific ORFs (Table 3) were designed using Premier Primer 5.0. All products were visualized using 1.5% agarose gels.

qRT-PCR was performed on an ABI 7500 (Thermo Scientific), using SYBR® Premix Ex Taq™ (TakaRa) according to the manufacturer’s instructions. The final volume of the PCR mixture was 20 µl, including 10 µl of the SYBR Green Master mix reagent, 7.8 µl of sterile water, 0.4 µl of DyeII, 1 µl of cDNA, and 0.4 µl (2 mM) of each real-time PCR primer. The following cycling conditions were used: 95 °C for 30 s followed by 40 cycles of 95 °C for 5 s, and then 60 °C for 34 s. The expression patterns of ORFs were determined relative to tobacco *Tublin*, and relative expression levels were calculated. All assays were repeated at least three times.

Results

Stamen morphology and pollen vitality of *sua*-CMS

ZY exhibited five full anthers, and the anthers were as long as the stigma, which is necessary for normal self-pollination (Fig. 1a). In contrast, the *sua*-CMS line sZY had shortened filaments and stigma-like or completely degraded anthers

(Fig. 1d). Unlike its maintainer line ZY, in which the anther dehiscence when the flower opened, the anther of sZY did not change in shape or color throughout the flowering process, even when the flower withered. The pollen grains of ZY were nearly round and had visible grooves based on scanning electron microscopy (SEM) (Fig. 1b) and they were viable (Fig. 1c). In sZY, no pollen grains were visible, other than some shards in the anther observed under SEM (Fig. 1e). sZY did not produce any viable pollens (Fig. 1f), and thus, it was considered fully male sterile.

Screening and characterization of mitochondrial sequences of *sua*-CMS

There were 393 and 417 mitochondrial ORFs (> 210 codons) in ZY and sZY, respectively. Thirty-four ORFs were unique to sZY according to blastn and blastx analyses. We designed 27 primer pairs to amplify all 34 unique ORFs across seven CMS types and fertile tobacco, which comprised sZY (*sua*-CMS), ZY (fertile tobacco), 86-6 (*glu*-CMS), 98-43 (*rep*-CMS), 200-18 (*rus*-CMS), L1 (*tab1*-CMS), 4-1 (*tab2*-CMS), and CH03 (*tab3*-CMS). The six ORFs—*orf82*, *orf103*, *orf115^a*, *orf91*, *orf115^b*, and *orf100*—located at 122,845–123,313 bp (m-sr1), 342,705–343,743 bp (m-sr2), and 414,430–415,454 bp (m-sr3) of the mitochondrial genome of the sZY (Accession no. KR071121), were only amplified in sZY. The other 28 ORFs were also amplified in some of the other six CMS types and therefore were not considered unique to *sua*-CMS (Table 4). The specific amplification products in sZY were cloned and sequenced. The sequences of the amplification products were completely consistent with the corresponding m-sr fragments. When the three m-sr fragments were analyzed via BLAST with the fertile tobacco mitochondrial genomes (Accession no. BA000042 and KR780036), we found that six ORFs in the three m-sr sequences were chimeric (Fig. 2). *orf82* in m-sr1 was 246 bp in length and contained six small fertile tobacco mitochondrial DNA fragments, accounting for 51.6% of *orf82*. The other sequences shared no homology to fertile tobacco mitochondrial genomes. *orf103*, *orf115^a*, *orf91*, *orf115^b*, and *orf100* were between 273 and 345 bp in length, and each ORF was also determined to contain 5–6 small fertile tobacco mitochondrial DNA fragments. The

Table 3 Primer sequences used in transcript analysis

Target ORFs	Primer sequences (5′–3′)
<i>orf82</i>	CACTAGCGCCCTTCTCT;CGACCTGGGTTCCATTCC
<i>orf103</i>	CTTCGGGTTTTGGAGGGC;GTAAGGCAGGGGAGTTG
<i>orf115^a</i>	ATGCTATTTCGTTTCTTCCCA;ACTCCCTCCCATCTCTTC
<i>orf91</i>	AAGTATGAGAATTTGATT;ATTTTTACATTGGTGAGG
<i>orf115^b</i> , <i>orf100</i>	CACTCCTTTCCTTCTCTTT;CTCTCTGGTGATGTTGCGCC
<i>Tublin</i>	CAAGGAAATCACCGCTTTG;AAGGGATGCGAGGATGGA

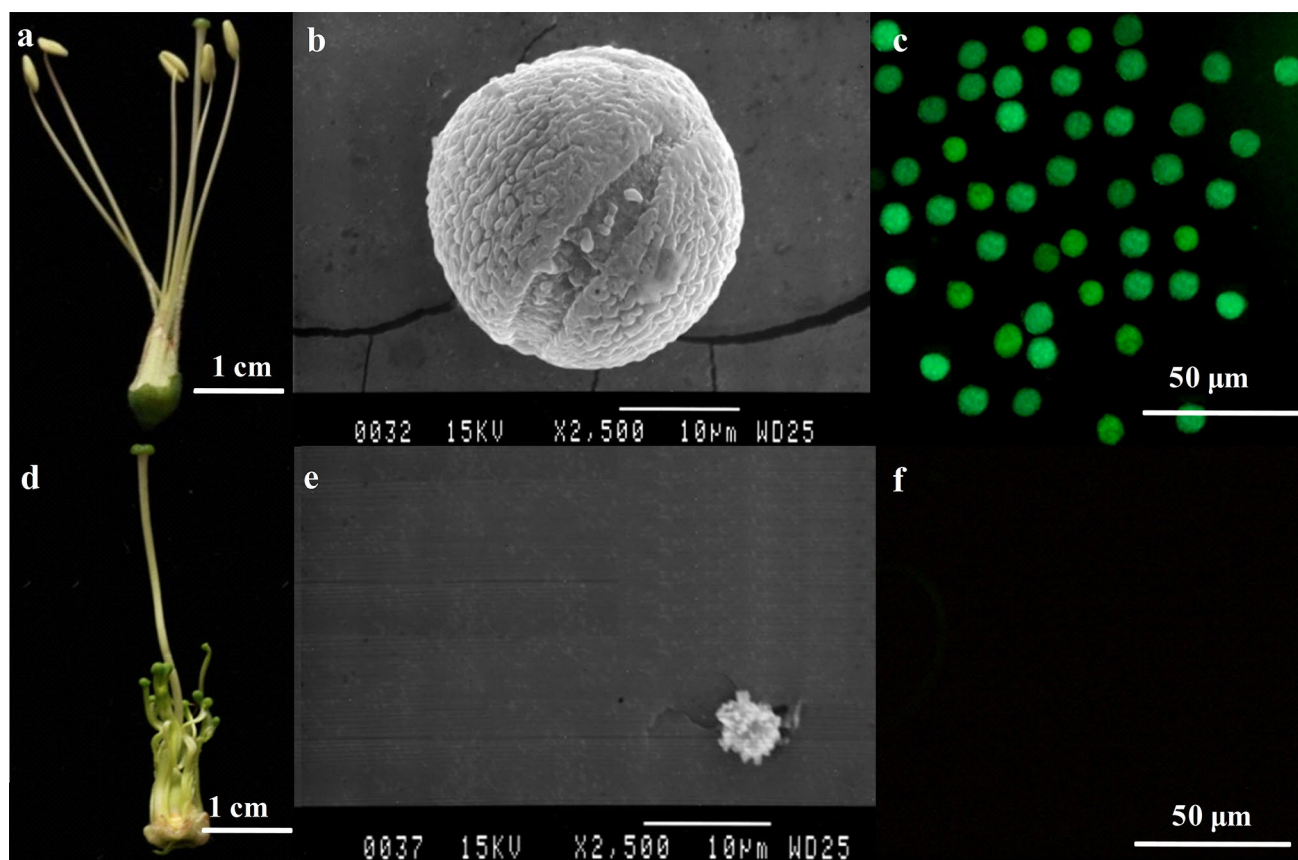


Fig. 1 Anther morphology, pollen structure, and pollen viability for fertile tobacco Zhongyan 100 (a–c) and the *sua*-CMS line msZhongyan100 (d–f) (a/d anther morphology; b/e pollen grain; c/f pollen viability)

similarity sequences accounted for 21.7, 24.9, 25.6, 39.9, and 27.4% of *orf103*, *orf115^a*, *orf91*, *orf115^b*, and *orf100*, respectively, and the other sequences shared no homology to fertile tobacco mitochondrial genomes.

Stability and specificity of the mitochondrial sequences

The amplification results showed that the specific mitochondrial fragments were consistently amplified in all plants with *sua*-CMS background, and amplification was consistent with changes in the nuclear genomes (Fig. 3). These m-sr fragments were not amplified in BC1F1s of the other six tobacco CMS types or in their backcrossed male parents (Speight G-28, K326 and NC82), suggesting that the ORFs were related to *sua*-CMS. The three m-sr fragments were not amplified in any of the fertile tobacco accessions (Fig. S1).

Transcript analysis of ORFs

In sZY, all six ORFs were expressed in the four plant tissues. They were expressed highly in floral buds and anthers and to a lesser extent in roots and leaves. In contrast, no

RT-PCR products of any ORFs were observed in the four tissues of the maintainer line ZY (Fig. 4). The qRT-PCR results validated those of RT-PCR (Fig. 5). Both methods revealed similar changes in the expression of ORFs in sZY.

Discussion

In this study, we identified six unique ORFs in the tobacco line *sua*-CMS, all of which were located in three small regions of the mitochondrial genome of *sua*-CMS. These ORFs were all chimeric and showed no homology with mitochondrial genomes of the tested fertile tobacco lines. We could not identify the origin of 48.4–78.3% of the sequences in these ORFs (Fig. 2). In plants with CMS, the formation of new ORFs is commonly due to recombination or rearrangement of the mitochondrial genome and new ORFs are often associated with CMS (Chen and Liu 2014). Mitochondrial genomes of different species vary greatly, because plant mitochondrial DNA (mtDNA) has extensive genomic variation and the non-coding sequences of mtDNA are not conserved across species (Gualberto and Newton 2017). Comparison of the mitochondrial genomes

Table 4 Open reading frames (ORFs) amplified in msZhongyan100 (*sua*-CMS), Zhongyan100, 86-6 (*glu*-CMS), 98-43 (*rep*-CMS), 200-18 (*rus*-CMS), L1 (*tab1*-CMS), 4-1 (*tab2*-CMS), and CH03 (*tab3*-CMS)

ORFs	sZY (<i>sua</i> - CMS)	ZY (fertile tobacco)	86-6 (<i>glu</i> - CMS)	98-43 (<i>rep</i> - CMS)	200-18 (<i>rus</i> - CMS)	L1 (<i>tab1</i> - CMS)	4-1 (<i>tab2</i> - CMS)	CH03 (<i>tab3</i> - CMS)
<i>orf95</i>	+	–	–	–	–	–	+	–
<i>orf77^a</i>	+	–	+	–	–	–	–	–
<i>orf76</i>	+	–	–	–	–	–	+	–
<i>orf82</i>	+	–	–	–	–	–	–	–
<i>orf77^b</i>	+	–	–	+	–	–	–	–
<i>orf116</i>	+	–	–	–	+	–	–	–
<i>orf71</i>	+	–	+	–	–	–	–	–
<i>orf88^a</i>	+	–	–	–	–	–	+	–
<i>orf142; orf149</i>	+	–	–	–	–	–	+	–
<i>orf73</i>	+	–	–	–	–	+	–	–
<i>orf93; orf73</i>	+	–	–	–	–	–	+	–
<i>orf74^a</i>	+	–	+	–	+	–	–	–
<i>orf83</i>	+	–	–	–	–	–	+	–
<i>orf79</i>	+	–	–	–	+	–	–	–
<i>orf74^b</i>	+	–	–	+	–	+	+	–
<i>orf103; orf115^a</i>	+	–	–	–	–	–	–	–
<i>orf111^a; orf71</i>	+	–	–	–	–	+	–	–
<i>orf78</i>	+	–	+	+	–	+	+	–
<i>orf89</i>	+	–	–	–	–	–	+	+
<i>orf74^c</i>	+	–	–	–	+	–	–	–
<i>orf111^b</i>	+	–	–	–	–	–	+	–
<i>orf91</i>	+	–	+	–	–	–	–	–
<i>orf91; orf115^b; orf100</i>	+	–	–	–	–	–	–	–
<i>orf72; orf115^c</i>	+	–	–	–	–	–	+	–
<i>orf94^a</i>	+	–	–	+	+	+	+	–
<i>orf88^b</i>	+	–	–	–	–	–	+	–
<i>orf94^b</i>	+	–	+	–	–	–	–	–

+ present; – absent

of *Brassica napus* and *Arabidopsis thaliana* (Unselde et al. 1997; Handa 2003) in the Brassicaceae family showed that one-third of the mtDNA of *B. napus* had no homology with the mtDNA of *A. thaliana*. Much of the mtDNA of *Silene noctiflora* and *S. conica*, which have many circular chromosomes (Sloan et al. 2012), lacked identifiable genes and could be rapidly gained or lost (Wu et al. 2015). When plants undergo interspecies somatic hybridization, mitochondrial genomes of the hybrids coexist with those of both parents and show recombination or rearrangement (Cardi et al. 1999; Zubko et al. 2001). The mitochondrial genomes differ significantly between CMS lines developed by interspecies hybridization and their parents. In the CMS line TK81-MS of sugar beet (*Beta vulgaris* L), a total of 68 kb (about 13.6% of the total genome) were missing compared to the normal sugar beet line, TK81-O (Satoh et al. 2006). The variation in the mitochondrial genome of *sua*-CMS observed in this and other studies is, therefore, not unusual. The *sua*-CMS line of tobacco was developed by somatic hybridization between

N. tabacum and *N. suaveolens*. Sequencing of the mitochondrial genomes confirmed that the mtDNA of the *sua*-CMS lines was 25% larger than that of the maintainer line or of other fertile tobacco lines, and most of this increase was due to non-coding repeat sequences (Sugiyama et al. 2005; Li et al. 2016).

Mitochondria are the location of cellular energy metabolism, in which the electron transport chain generates ATP through proton gradients and maintains the vital activity of cells (Millar et al. 2011). CMS in plants is associated with disturbance of the electron transport chain, impairing energy production. The energy impairment does not affect vegetative development or female flower development. The leading hypothesis to explain this phenomenon is that the developing microspores consume much more energy than any other tissues (Hanson and Bentolila 2004). Thus, genes encoding electron transport chain proteins and ORFs associated with CMS are characterized by high expression in floral tissues (Zhao et al. 2016; Yang et al. 2017). In addition to

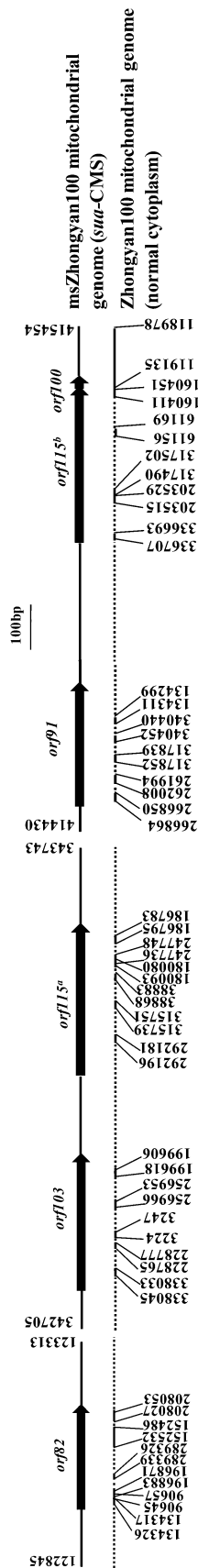


Fig. 2 Schematic structure of six ORFs and their flanking regions. Solid lines indicate homologous sequence; dotted lines indicate non-homologous sequence

mitochondrial energy deficiency, CMS protein cytotoxicity and premature tapetal programmed cell death (PCD) in the anther also explain some CMS systems (Chen and Liu 2014), which are encoded or induced by ORFs associated with CMS. These ORFs are specific or highly expressed in the anther. In this study, the six described ORFs were unique genes of *sua*-CMS tobacco and were highly expressed in the floral buds and anthers (Figs. 4, 5). These results indicated that the ORFs may be related to the mitochondrial function of *sua*-CMS. Further research out of our lab will be conducted to confirm this hypothesis.

The most notable feature of CMS is anther abortion. Morphological and cytological identification of CMS lines could be used to classify sterile cytoplasm (Gonzalez-Melendi et al. 2008; Yang et al. 2008). Variation in the mitochondrial genome is an important driver of plant CMS formation (Hanson and Bentolila 2004; Gualberto and Newton 2017). Based on the mitochondrial genomic changes of sterile cytoplasm, CMS types have been identified precisely and accurately at the molecular level. Kim et al. (2007) developed a combination of molecular markers based on mtDNA rearrangement of radish (*Raphanus sativus* L.), which is now used for the reliable classification of radish cytoplasm. A novel cytoplasmic male sterility in radish was subsequently identified using these molecular markers (Lee et al. 2008). More than 60 CMS lines have been developed from interspecies, intersubspecies, and intervarieties of *Oryza* plants with the AA genome in rice. Based on PCR amplification of 264 random primers in two different backcross populations of rice, a mitochondrial SCAR marker of the HL-CMS/Rf system was developed (Tan et al. 2015). Heng et al. (2017) identified mitotype-specific sequences (MSSs) by comparative analysis of six mitochondrial genomes in *B. napus* and developed 12 mitotype-specific markers that could distinguish between six different mitotypes. These markers can provide guidance for classifying different mitotypes in *B. napus* breeding, and can be used to improve the efficiency of cross-breeding and heterosis utilization in cruciferous vegetables. We identified six specific ORFs of tobacco *sua*-CMS by mitochondrial genome analysis and amplification in different tobacco CMS types and in fertile *Nicotiana* accessions. These specific genes will be useful for molecular markers to distinguish *sua*-CMS from other types of CMS.

Conclusion

In this study, we obtained 34 unique ORFs in tobacco sZY by BLAST analysis of the mitochondrial genome ORFs of the *sua*-CMS line sZY and its maintainer line ZY. Six ORFs were amplified in plants with *sua*-sterile cytoplasm, including the F1 hybrids, BC1F1s offspring, and CMS lines, and were not amplified in plants with *glu*-, *rep*-,

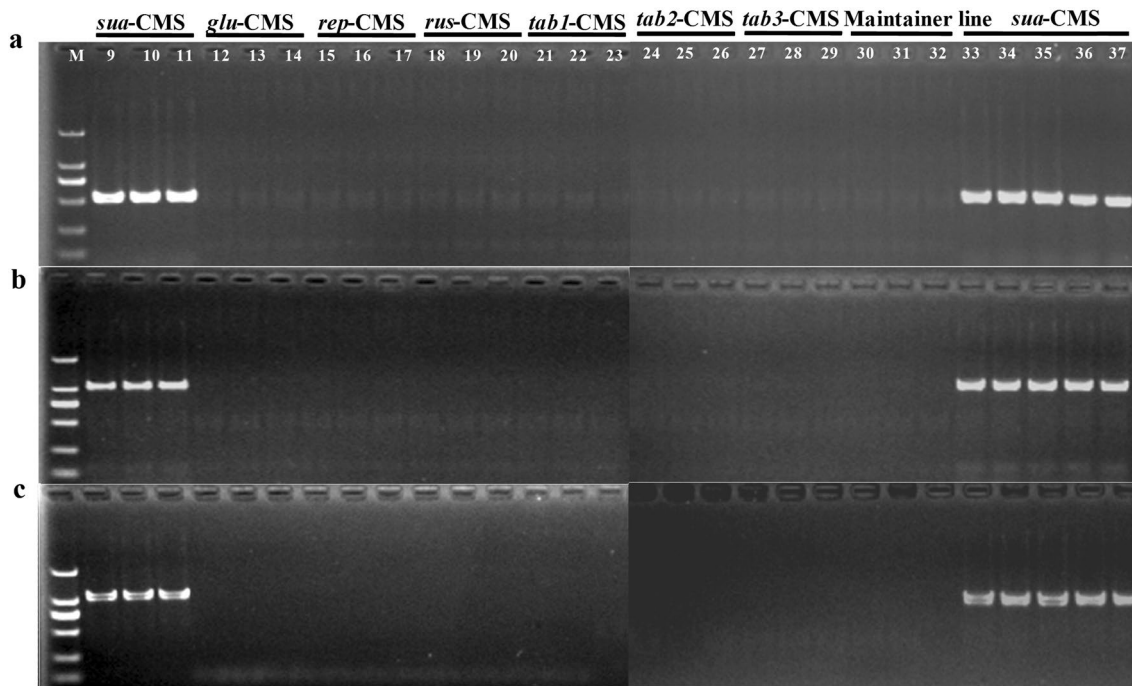
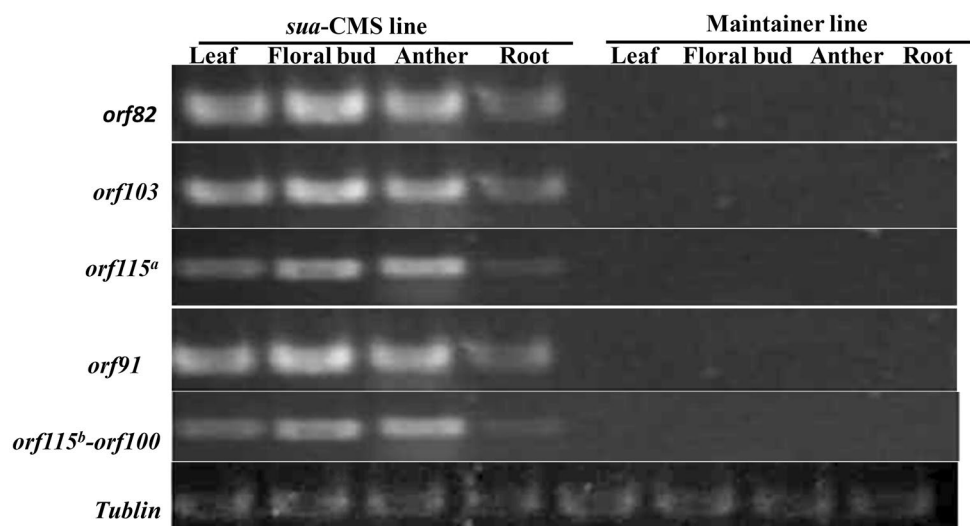


Fig. 3 Amplified m-sr fragments present in different types of CMS lines with the same nuclear background. **a** m-sr1 amplification; **b** m-sr2 amplification; **c** m-sr3 amplification. Number above each lane

matches the corresponding plant materials (listed in Table 1). *M* DNA marker DL2000

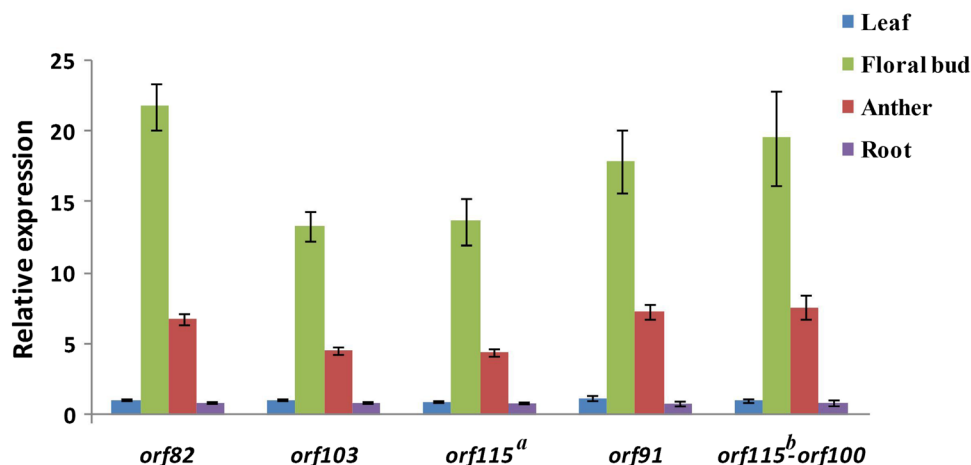
Fig. 4 Expression of the specific ORFs in the *sua*-CMS line msZhongyan100 and its maintainer line Zhongyan100 in four plant tissues



rus-, *tab1*-, *tab2*-, and *tab3*- tobacco sterile cytoplasm or the 284 fertile tobacco accessions. This result indicated that they were *sua*-CMS sterility-specific genes, which can be used to identify *sua*-CMS lines. Transcriptional analysis showed that the six ORFs were highly expressed in anthers and flower buds, but were expressed at low levels in vegetative organs, including leaves and roots, in sZY, confirming that these ORFs play a role in male sterility in *sua*-CMS.

Author contribution statement AY and FL conceived and designed the study. YZ and ZL carried out the experiments. YS and GL contributed materials. YZ, AY, and FL analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Fig. 5 Relative expression of the six specific ORFs in the *sua*-CMS line msZhongyan100 in four plant tissues



Acknowledgements We thank Min Ren and Xinwei Zhang for providing the genomic DNA of the 284 fertile tobacco accessions. The present investigation was financially supported by the Agricultural Science and Technology Innovation Program (ASTIP-TRIC02) and the Key Laboratory for Utilization of Gene Resources of Tobacco Program (110201603007).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 15:403–410
- Baranwal VK, Mikkilineni V, Zehr UB, Tyagi AK, Kapoor S (2012) Heterosis: emerging ideas about hybrid vigour. *J Exp Bot* 18:6309–6314
- Bonhomme S, Budar F, Lancelin D, Small I, Defrance MC, Pelletier G (1992) Sequence and transcript analysis of the *Nco2.5* ogura-specific fragment correlation with cytoplasmic male sterility in *Brassica* cybrids. *Mol Gen Genet* 235:340–348
- Cao JL, Yu J (2017) Utilization and insights on flue-cured tobacco hybrids in Zimbabwe. *Tob Sci Technol* 50:17–24
- Cardi T, Bastia T, Monti L, Earle ED (1999) Organelle DNA and male fertility variation in *Solanum* spp. and interspecific somatic hybrids. *Theor Appl Genet* 99:819–828
- Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579–606
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Dufay M, Touzet P, Maurice S, Cuguen J (2007) Modelling the maintenance of male-fertile cytoplasm in a gynodioecious population. *Heredity* 99:349–356
- Farbos I, Mouras AA, Glimelius K (2001) Defective cell proliferation in the floral meristem of alloplasmic plants of *Nicotiana tabacum* leads to abnormal floral organ development and male sterility. *Plant J* 26:131–142
- Fitter JT, Thomas MR, Niu C, Rose RJ (2005) Investigation of *Nicotiana tabacum* (+) *N. suaveolens* cybrids with carpelloid stamens. *J Plant Physiol* 162:225–235
- Gonzalez-Melendi P, Uyttewaal M, Morcillo CN, Hernandez Mora JR, Fajardo S, Budar F, Lucas MM (2008) A light and electron microscopy analysis of the events leading to male sterility in Ogu-INRA CMS of rapeseed (*Brassica napus*). *J Exp Bot* 59:827–838
- Gualberto JM, Newton KJ (2017) Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol* 68:225–252
- Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rape seed (*Brassica napus* L.) comparative analysis of the mitochondrial genomes of rape seed and *Arabidopsis thaliana*. *Nucleic Acids Res* 31:5907–5916
- Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* 16(Suppl 1):S154–S169
- Heng S, Chen F, Wei C, Hu K, Yang Z, Wen J, Yi B, Ma C, Tu J, Si P, Fu T, Shen J (2017) Identification of different cytoplasms based on newly developed mitotype-specific markers for marker-assisted selection breeding in *Brassica napus* L. *Plant Cell Rep* 36:901–909
- Horn R, Gupta KJ, Colombo N (2014) Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* 19:198–205
- Kim S, Lim H, Park S, Cho KH, Sung SK, Oh DG, Kim KT (2007) Identification of a novel mitochondrial genome type and development of molecular markers for cytoplasm classification in radish (*Raphanus sativus* L.). *Theor Appl Genet* 115:1137–1145
- Lee YP, Park S, Lim C, Kim H, Lim H, Ahn Y, Sung SK, Yoon MK, Kim S (2008) Discovery of a novel cytoplasmic male-sterility and its restorer lines in radish (*Raphanus sativus* L.). *Theor Appl Genet* 117:905–913
- Li F, Yang A, Lv J, Gong D, Sun Y (2016) The complete mitochondrial genome sequence of *sua*-type cytoplasmic male sterility of tobacco (*Nicotiana tabacum*). *Mitochondrial DNA Part A DNA Mapp Seq Anal* 27:2929–2930
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964
- Ma W, Zhang H, Bai Y, Li Y, Chen J (2005) Comparison of traits of the mail sterile lines and their maintainer lines of flue-cured tobacco. *J Hunan Agric Univ (Nat Sci)* 31:496–499
- McCabe PF, Levine A, Meijer PJ, Tapon NA, Pennell RI (1997) A programmed cell death pathway activated in carrot cells cultured at low cell density. *Plant J* 12:267–280
- Millar AH, Whelan J, Soole KL, Day DA (2011) Organization and regulation of mitochondrial respiration in plant. *Annu Rev Plant Biol* 62:79–104

- Pollak PE (1992) Cytological differences between a cytoplasmic male sterile tobacco hybrid and its fertile counterpart during early anther development. *Am J Bot* 79:937–945
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945
- Satoh M, Kubo T, Mikami T (2006) The Owen mitochondrial genome in sugar beet (*Beta vulgaris* L.): possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions. *Theor Appl Genet* 113:477–484
- Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol* 10:e1001241
- Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiyama M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Gen Genom* 272:603–615
- Tan Y, Xu X, Wang C, Cheng G, Li S, Liu X (2015) Molecular characterization and application of a novel cytoplasmic male sterility-associated mitochondrial sequence in rice. *BMC Genet* 16:45
- Tang H, Zheng X, Li C, Xie X, Chen Y, Chen L, Zhao X, Zheng H, Zhou J, Ye S, Guo J, Liu YG (2017) Multi-step formation, evolution, and functionalization of new cytoplasmic male sterility genes in the plant mitochondrial genomes. *Cell Res* 27:130–146
- Tong D (1997) Tobacco breeding. China Agricultural Press, Beijing, pp 279–294
- Touzet P, Meyer EH (2014) Cytoplasmic male sterility and mitochondrial metabolism in plants. *Mitochondrion* 19:166–171
- Unsold M, Marienfeld JR, Brandt P, Brennicke A, Handa H (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* 15:57–61
- Wu Z, Cuthbert JM, Taylor DR, Sloan DB (2015) The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc Natl Acad Sci* 112:10185–10191
- Xue Z, Xu X, Zhou Y, Wang X, Zhang Y, Liu D, Zhao B, Duan L, Qi X (2018) Deficiency of a triterpene pathway results in humidity-sensitive genic male sterility in rice. *Nat Commun* 9:604
- Yang S, Terachi T, Yamagishi H (2008) Inhibition of *chalcone synthase* expression in anthers of *Raphanus sativus* with Ogura male sterile cytoplasm. *Ann Bot* 102:483–489
- Yang W, Lou X, Li J, Pu M, Mirbahar AA, Liu D, Sun J, Zhan K, He L, Zhang A (2017) Cloning and functional analysis of MADS-box genes, TaAG-A and TaAG-B, from a wheat K-type cytoplasmic male sterile line. *Front Plant Sci* 8:1081
- Yuan SC, Zhang ZG, He HH, Zen HL, Lu KY, Lian JH, Wang BX (1993) Two photoperiodic-reactions in photoperiod-sensitive genic male-sterile rice. *Crop Sci* 33:651–660
- Zhao T, Zhu T, Liu Q, Zhang M, Jiang H (2009) SNP in tobacco mitochondrial gene *atp6* and its correlation with CMS. *Acta Agron Sin* 35:1655–1661
- Zhao N, Xu X, Wamboldt Y, Mackenzie SA, Yang X, Hu Z, Yang J, Zhang M (2016) MutS HOMOLOG1 silencing mediates *ORF220* substoichiometric shifting and causes male sterility in *Brassica juncea*. *J Exp Bot* 67:435–444
- Zubko MK, Zubko EI, Ruban AV, Adler K, Mock HP, Misera S, Gleba YY, Grimm B (2001) Extensive developmental and metabolic alterations in cybrids *Nicotiana tabacum* (+ *Hyoscyamus niger*) are caused by complex nucleo-cytoplasmic incompatibility. *Plant J* 25:627–639