



RESEARCH PAPER

A pentatricopeptide repeat protein restores *nap* cytoplasmic male sterility in *Brassica napus*

Zhi Liu^{1,2,3,*}, Faming Dong^{1,2,3,*}, Xiang Wang^{1,3}, Tao Wang^{1,3}, Rui Su^{1,3}, Dengfeng Hong^{1,2,3} and Guangsheng Yang^{1,2,3,†}

¹ National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, PR China

² National Research Center of Rapeseed Engineering and Technology, National Rapeseed Genetic Improvement Center (Wuhan Branch), Huazhong Agricultural University, Wuhan 430070, PR China

³ College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China

* These authors contributed equally to this work.

† Correspondence: gsyang@mail.hzau.edu.cn

Received 2 February 2017; Editorial decision 15 June 2017; Accepted 19 June 2017

Editor: Dabing Zhang, Shanghai Jiao Tong University

Abstract

Two forms of male-sterile cytoplasm, designated *nap* and *pol*, are found in the oilseed rape species, *Brassica napus*. The *nap* cytoplasm is observed in most *B. napus* varieties, and it confers male sterility on a limited number of cultivars that lack the corresponding restorer gene, *Rfn*. In the present study, using linkage analysis in combination with 5652 BC₁ progeny derived from a cross between a *nap* cytoplasmic male sterility (CMS) line 181A and a restorer line H5, we delimited the *Rfn* gene to a 10.5 kb region on chromosome A09, which contained three putative ORFs. Complementation by transformation rescue revealed that the introduction of *ORF2*, which encodes a pentatricopeptide repeat (PPR) protein, resulted in the recovery of fertility of *nap* CMS plants. Expression analysis suggested that the *Rfn* was highly expressed in flower buds and it was preferentially expressed in the tapetum and meiocytes during anther development. Further RNA gel blots and immunodetection suggested that the *Rfn* gene may play a complicated role in restoring the *nap* CMS. Our work laid the foundation for dissecting the molecular basis of CMS fertility restoration and the nuclear–mitochondrial interactions in CMS/*Rf* systems.

Key words: *Brassica napus*, cytoplasmic male sterility, map-based cloning, mitochondria, pentatricopeptide repeat protein, restorer of fertility (*Rf*).

Introduction

Cytoplasmic male sterility (CMS), a maternally inherited trait that fails to produce functional pollen, has been observed in >150 higher plant species. At the molecular level, CMS is typically associated with an aberrant chimeric gene in the mitochondria (Hanson and Bentolila, 2004). In most cases, the male sterility phenotype can be recovered by dominant

nuclear genes, termed restorer of fertility (*Rf*) genes, which can specifically reduce the accumulation of CMS-associated RNAs or proteins in F₁ hybrids. Thus, CMS/*Rf* systems provide both an ideal model for investigating nuclear–cytoplasmic interaction and a practical tool for facilitating hybrid seed production.

Abbreviations: BAC, bacterial artificial chromosome; CDS, coding sequence; CMS, cytoplasmic male sterility; DIG, digoxigenin; GFP, green fluorescent protein; PPR, pentatricopeptide repeat; qRT-PCR, quantitative reverse transcription-PCR; *Rf*, restorer of fertility.

© The Author 2017. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

In oilseed rape species, *Brassica napus*, there are two different endogenous male-sterile cytoplasm, *nap* and *pol*. The *nap* cytoplasm is the most prevalent cytoplasm in *B. napus* germplasm, but it confers male sterility only in a limited number of cultivars that lack the corresponding fertility restorer gene, *Rfn*. In the absence of the nuclear restorer gene *Rfn*, plants carrying the *nap* cytoplasm display male sterility when grown under relatively low temperature (22/16 °C; 16/8 h), while the sterility can be gradually repaired or even completely reversed with an increase of temperature (Fan and Stefansson, 1986). Because flowering occurs in spring or summer and the environmental temperature gradually goes up, this specific temperature-sensitive male sterility is hardly used for hybrid production. In contrast, the *pol* CMS is highly stable in most cases; thus it can be widely used for commercial hybrid seed production (Fan and Stefansson, 1986; Zhou and Fu, 2007).

Previous studies indicated that *pol* CMS is associated with expression of the novel ORF, *orf224*, which is located upstream of a normal mitochondrial gene, *atp6* (Singh and Brown, 1991, 1993), while *nap* CMS is associated with a different but related ORF, *orf222*, that is co-transcribed with an exon of a *trans*-spliced gene, *nad5c*, and another short ORF, *orf139* (L'Homme *et al.*, 1997). The nuclear fertility restorer genes for *nap* and *pol* CMS, *Rfn* and *Rfp*, represent different alleles or haplotypes of a single nuclear locus (Li *et al.*, 1998). For both the *nap* and *pol* CMS systems, the nuclear restorer genes may take the role of down-regulating the expression of their corresponding mitochondrial CMS-associated transcripts (Singh and Brown, 1991; L'Homme *et al.*, 1997; Liu *et al.*, 2016). To validate the allelism of *Rfn* and *Rfp* and better understand how *Rfn* regulates CMS-associated transcripts, it is necessary to clone *Rfn*.

To date, except for *Rf2* (encoding a mitochondrial aldehyde dehydrogenase) for T-CMS in maize (Cui *et al.*, 1996), *Rf17* (encoding an acyl-carrier protein synthase-like protein) for CW-CMS in rice (Fujii and Toriyama, 2009), *Rf2* (encoding a mitochondrial glycine-rich protein) for LD-CMS in rice (Itabashi *et al.*, 2011), and *Rf1* (encoding a putative peptidase of the M48 family) for Owen-CMS in sugar beet (Matsuhira *et al.*, 2012), the rest of the cloned *Rf* genes all encode pentatricopeptide repeat (PPR) proteins (Bentolila *et al.*, 2002; Brown *et al.*, 2003; Kazama and Toriyama, 2003; Koizuka *et al.*, 2003; Komori *et al.*, 2004; Klein *et al.*, 2005; Wang *et al.*, 2006, 2013; Hu *et al.*, 2012; Tang *et al.*, 2014; Huang *et al.*, 2015; Igarashi *et al.*, 2016; Liu *et al.*, 2016). Such proteins are characterized by the succession of tandem degenerate motifs of ~35 amino acids. Based on the structure of the PPR motifs, PPR proteins can be divided into P and PLS subfamilies. P-type PPR proteins contain only canonical 35 amino acid repeats (P), whereas PLS PPR proteins consist of sequential repeats of P, long (L), and short (S) PPR motifs (Lurin *et al.*, 2004). P-type PPR proteins were shown to participate in multiple aspects of organelle RNA processing steps such as cleavage, splicing, stabilization, and/or translation of their target transcripts, while PLS PPR proteins have been almost exclusively associated with C-to-U RNA editing (Barkan and Small, 2014). It has been verified that PPR proteins have the

ability to recognize ssRNA following a one-motif to one-base rule (Barkan *et al.*, 2012; Okuda *et al.*, 2014). Recent computational, biochemical, and structural studies indicated that the PPR motif first produces an antiparallel helix–turn–helix fold whose repetition then forms a solenoid-like structure (Barkan *et al.*, 2012; Yagi *et al.*, 2013; Yin *et al.*, 2013). Within each PPR repeat, the combinations of two amino acids, known as the ‘PPR code’ at two key positions (the 5th and the 35th), confer recognized RNA specificity (Barkan *et al.*, 2012; Barkan *et al.*, 2013; Ke *et al.*, 2013; Yagi *et al.*, 2013; Yin *et al.*, 2013; Gully *et al.*, 2015; Shen *et al.*, 2016).

In the present study, we finely mapped and cloned the nuclear restorer gene *Rfn* for *nap* CMS in *B. napus* by using a map-based cloning strategy and preliminarily investigated how it regulates the fertility of *nap* CMS. The *Rfn* encodes a mitochondria-targeted PPR protein, and it may play a much more complicated role in restoring the *nap* CMS.

Materials and methods

Genetic analysis of *nap* CMS restoration

A *nap* CMS male-sterile line 181A, a male-fertile line H5, and three *pol* CMS lines (1141A, 245A, and 7492A) were used in this study. The cross between 181A as female parent and H5 as male parent yielded male-fertile F₁ plants. Self-pollination of F₁ plants yielded an F₂ segregating population, which was used to analyze the genetic control of *nap* CMS restoration. The *pol* CMS lines 1141A, 245A, and 7492A as female were also each used to cross with H5 for genetic analysis. The individual fertility of each cross combination and self-pollination was assessed as previously described (Liu *et al.*, 2012) during the early flowering period.

Fine mapping of the *Rfn* gene

The *nap* CMS line 181A as female was crossed with H5 to create the F₁ plants, and the CMS plants were backcrossed with F₁ plants as pollen donor to produce a backcross segregation population containing 5652 individuals. Polymorphic markers around the *Rfn* locus were developed based on the synteny region of *B. napus* and other *Brassica* species, which were then used for fine mapping of the *Rfn* gene.

BAC screening, sequencing, and candidate gene prediction

A bacterial artificial chromosome (BAC) clone library described by Li *et al.* (2015) was screened with co-segregated markers through a two-stage PCR screening method (Crooijmans *et al.*, 2000). The target BAC clone was then sequenced by Illumina HiSeq 2000 according to the manufacturer's standard protocol. The online-based tool FGENESH (<http://www.softberry.com>) was used to predict putative ORFs in the delimited gene region. The genomic or coding sequences (CDSs) of predicted genes were submitted to BRAD (<http://brassicadb.org>; Cheng *et al.*, 2011) for homology search and basic functional annotation.

Complementation by transformation rescue

The genomic sequence of candidate gene (including the putative promoter, CDS, and 3'-downstream region) was PCR amplified with H5 as a template using specific primers (see Supplementary Table S1 at *JXB* online). The PCR product was inserted into the *Eco*RI and *Bam*HI sites of the binary vector pFGC5941 and then introduced into the CMS line 181A by *Agrobacterium*-mediated

transformation. Individual transgenic plants were raised to maturity and visually assessed for male fertility/sterility under relatively low temperature during flowering time. Self-pollination of the fertility-recovered T_0 plants produced T_1 plants, and the CMS parent line as female was backcrossed with T_0 plants to produce test cross progeny plants. Pollen grains from the anthers of the transgenic plants were collected before anthesis and dyed in 1% aceto-carmin staining solution. The stained pollen was photographed under a microscope.

RNA extraction and quantitative reverse transcription-PCR

Total RNA was extracted from various rapeseed tissues using an RNeasy plant mini kit (Qiagen, <http://www.qiagen.com/>). RNA samples of ~2 µg were converted into cDNA with GoScript™ Reverse Transcriptase following the manufacturer's instructions (Promega, USA), and the products were diluted 10-fold with distilled water for template. Quantitative reverse transcription-PCR (qRT-PCR) was performed with the Bio-Rad CFX96 Real-time system (Bio-Rad) and the GoTaq qPCR Master Mix (Promega, USA). Relative expression levels were calculated by the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). The relative amount of PCR product that was amplified using the designed primer sets (listed in Supplementary Table S1) was normalized to the internal control gene *BnActin* (GenBank: AF111812.1). The data are expressed as the mean \pm SD ($n=3$ biological replicates).

GUS staining assay

An ~2000 bp upstream region of the *Rfn* gene was amplified from the restorer line H5 (Primers Pro-F and Pro-R; Supplementary Table S1) and inserted into the cloning vector pMDC162. The Pro-*Rfn*-GUS (β -glucuronidase) construct was then introduced into wild-type Arabidopsis plants by floral-dip transformation (Clough and Bent, 1998). GUS activity was visualized by staining various tissues from the heterozygous transgenic plants and wild-type plants overnight at 37 °C in X-Gluc solution containing 50 mM sodium phosphate buffer of pH 7.0, 10 mM EDTA, 1 mM X-Gluc (5-bromo-4-chloro-3-indolyl-glucuronide), and 20% methanol (w/v). The stained samples were cleared in 75% (v/v) ethanol and then photographed under a stereomicroscope (Olympus SZX16, Japan).

RNA in situ hybridization analysis

Floral buds at different developmental stages from the restorer line H5 were freshly collected and fixed in FAA (50% ethanol, 5.0% acetic acid, and 3.7% formaldehyde). Fixed floral buds were dehydrated with a 50–100% ethanol series, cleaned by a series of xylenes from 25% to 100%, and embedded in Paraplast. Anther sections (8 µm thickness) were cut using a Leica RM2235 microtome and placed on RNase-free polylysine-coated slides. A 124 bp CDS fragment from the *Rfn* gene was amplified using the insitu-F and insitu-R primers (Supplementary Table S1) and inserted into the pGEM-T Easy vector (Promega, USA) for sequencing. The probes were labeled with digoxigenin (DIG) using an RNA labeling kit (Roche Applied Science). The subsequent hybridization and immunological detection procedures were performed as previously described by Wan *et al.* (2010).

Transient expression and subcellular localization of RFN

The coding sequence of the putative N-terminal 44 residue signal peptide was PCR amplified using primers *Rfn*-*Xba*-F and *Rfn*-*Xba*-R (Supplementary Table S1) and then inserted into the pM999-GFP vector to generate a green fluorescent protein (GFP) fusion product. The fusion construct was introduced into Arabidopsis protoplasts by polyethylene glycol (PEG)/calcium-mediated transformation (Yoo *et al.*, 2007). The mitochondria of protoplasts were marked using MitoTracker Red CMX-Ros (Molecular Probes, Invitrogen) staining solution after incubation at 23 °C for 14–18 h.

GFP expression and co-localization of GFP fusion protein to mitochondria were imaged using a confocal scanning microscope system (Olympus FV1200, Japan) with 488 nm laser light for fluorescence excitation of GFP and 580 nm laser light for excitation of MitoTracker Red CMX-Ros.

RNA gel blot analysis

For RNA gel blot analysis, total RNA was isolated from young flower buds using Trizol reagent (Invitrogen, <http://www.invitrogen.com/>) as described by the manufacturer. Approximately 20 µg of total RNA from each sample was size fractionated on 1.2% agarose/2.2 M formaldehyde horizontal gels and transferred to Hybond-N⁺ nylon membranes (Amersham) by capillary blotting with 3 M NaCl/0.3 M sodium citrate for 16–20 h. The *orf222* and *orf139* fragments were amplified from 181A using the primers listed in Supplementary Table S1. The PCR products were inserted into the pGEM-T Easy vector (Promega, USA) for sequencing. Antisense DIG-labeled RNA probes were made with the RNA labeling kit (Roche Applied Science). The hybridization and detection with DIG Easy Hyb and the DIG Nucleic Acid Detection Kit (Roche Applied Science) were performed following the manufacturer's instructions.

Antibody preparation and immunodetection

A peptide antigen (QSKLKLGGKDRTSK) corresponding to 14 residues of ORF222 was synthesized by a chemical synthesis method (GenScript, Nanjing, China) and used to immunize rabbits for antibody production. Total proteins were extracted from the young flower buds of CMS line, restorer line, and transgenic fertility-restored plants. The proteins were separated by 12% SDS-PAGE and transferred onto a membrane (PVDF type, Millipore). The following procedures were performed as previously described by Jing *et al.* (2012).

Results

Male fertility restoration is controlled by one dominant gene

To investigate the genetic control of fertility restoration of *nap* CMS, we prepared an F₂ population derived from the cross of a *nap* CMS line 181A as female and a male fertile line H5 (Supplementary Fig. S1). Fertility assessment showed that 715 of 940 plants had a fertile phenotype and the remaining plants had a sterile phenotype, fitting the 3:1 segregation ratio ($\chi^2=0.51$, $P<0.05$) of a single locus, whereas all the F₁ plants derived from the crosses between three *pol* CMS lines and H5 remained male sterile (Table 1). These results indicated that the male-fertile line H5 can specifically restore the *nap* CMS line 181A in a sporophytic manner and possesses only one dominant *Rfn* gene.

The *Rfn* locus was limited to a 10.5 kb region

A previous study indicated that the *Rfn* and *Rfp* genes represent different alleles or haplotypes of a single nuclear locus (Li *et al.*, 1998). To map the *Rfn* locus precisely, sequence information from the collinear region around the *Rfp/Rfn* locus was used for development of simple sequence repeat (SSR) or sequence characterized amplified region (SCAR) markers. Subsequently, 16 markers tightly linked to the *Rfn* gene were successfully obtained. These markers were further

used to assay each individual of the backcross segregation population. Among these plants, four and two recombinant individuals were identified using the closest flanking markers BnSR33 and BrIP77, respectively, and two markers (BrSC64 and BrSC65) co-segregated with the *Rfn* locus (Fig. 1A). To characterize precisely the genomic sequence around the *Rfn* locus in *B. napus*, we further screened a BAC clone library using the two co-segregated markers; a BAC clone (HBnB022L06) was also successfully identified. After sequencing the target BAC clone and sequence alignment of the closest flanking markers, the *Rfn* locus was narrowed down to a 10.5 kb physical region on chromosome A09 of *B. napus* (Fig. 1B). Gene prediction showed that this region contained two complete CDSs (*ORF1* and *ORF2*) and a truncated CDS (*ORF3*) (Fig. 1C). Homologous gene annotation

in *Arabidopsis* showed that *ORF1* and *ORF2* encode a zinc finger (AN1-like) family protein and a PPR-containing protein, respectively, while *ORF3* encodes a serine-type endopeptidase. Considering most *Rf* genes encoding PPR proteins, we took *ORF2* as the *Rfn* candidate.

Validation of the target gene of *Rfn* for *nap* CMS

To validate *ORF2* as the target of *Rfn*, we isolated the genomic sequence (including the putative promoter, ORF, and 3'-downstream region) of *ORF2* from the restorer line H5 and cloned it into binary vector pFGC5941. The construct was then introduced into the *nap* CMS line 181A by *Agrobacterium*-mediated transformation. Nine independent T₀ transgenic plants displayed a normal male fertile phenotype when grown

Table 1. Genetic analysis of the fertility restorer gene in H5

The male-sterile lines 1141A, 245A, and 7492A are *pol* cytoplasm; the male-sterile line 181A is *nap* cytoplasm

Cross combination	Progeny	Total plants	Fertile plants	Sterile plants	Expected ratio	χ^2 value ^a
1141A (<i>pol</i>)×H5	F ₁	115	0	115	–	–
245A (<i>pol</i>)×H5	F ₁	67	0	67	–	–
7492A (<i>pol</i>)×H5	F ₁	69	0	69	–	–
181A (<i>nap</i>)×H5	F ₁	83	83	0	–	–
	F ₂	940	715	225	3:1	0.51

^a χ^2 (0.05, 1)=3.84.

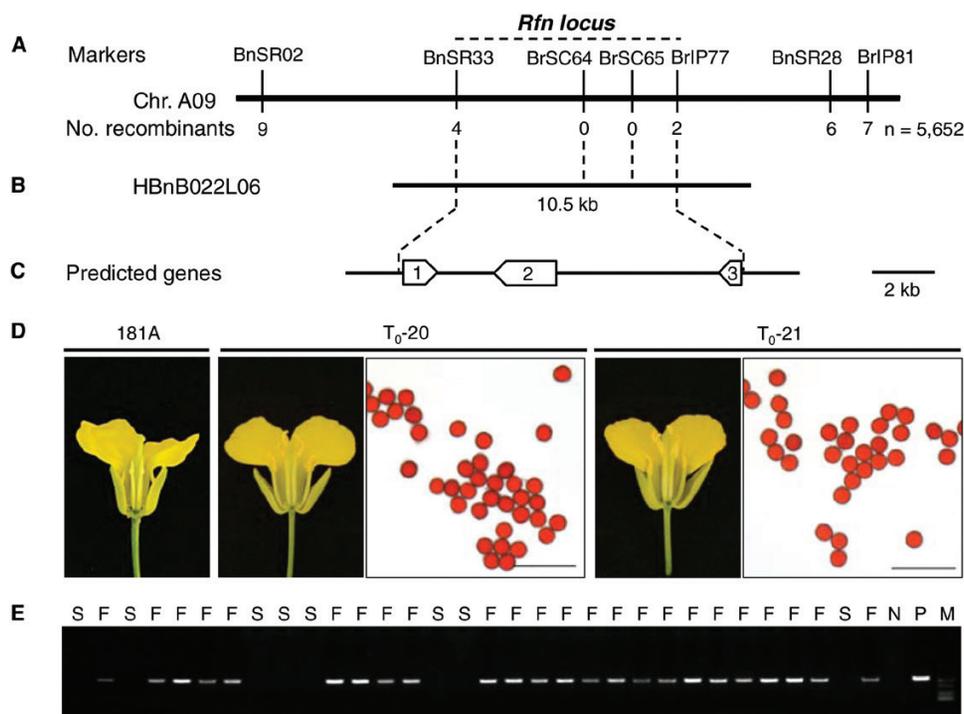


Fig. 1. Map-based cloning of the *Rfn* gene for *nap* CMS in *Brassica napus*. (A) Molecular mapping of the *Rfn* locus using a BC₁ population containing 5652 individuals. The numbers of recombinants between the markers and *Rfn* are shown. (B) Physical map based on a sequenced BAC. The *Rfn* locus was narrowed down to a 10.5 kb region. (C) The predicted genes from <http://www.softberry.com>. It contained two complete CDSs (*ORF1* and *ORF2*) and a partial CDS (*ORF3*). (D) Morphology of flowers and pollen grains of a *nap* CMS plant and two hemizygous *Rfn* candidate transgenic plants. Dissected flowers were photographed with a digital camera. Pollen grains were stained with 1% aceto-carmin solution. Scale bar=100 μ m. (E) Co-segregation analysis of the introduced DNA and fertile plants among the T₁-20 progeny. F, fertile plant; S, sterile plant; N, negative control; P, positive control; M, DL2000 molecular markers. (This figure is available in colour at JXB online.)

to maturity under relatively low temperature (Fig. 1D). The recovered fertility of T₁ progeny plants could be stably co-transmitted and co-segregated with the introduced DNA (Fig. 1E). Meanwhile, a 3:1 or 15:1 genetic segregation ratio was observed in the five independent transgenic T₁ progeny of the self-cross and a 1:1 or 3:1 genetic segregation ratio was observed in the corresponding test-crossed progeny (Supplementary Table S2). These results suggested that a single site or two independently heritable T-DNA insertions was introduced into 181A in the T₀ transgenic plants. Thus, our data proved that *ORF2* possesses the ability to restore the *nap* CMS in a sporophytic manner and it is the causal gene of *Rfn*.

Sequence analysis of the *Rfn* gene and its recessive allele

To characterize the recessive allele of *Rfn* existing in the CMS line, *ORF2* was amplified from the 181A genome and sequenced. Comparing the putative promoter region (~2.0 kb) and the full-length CDS of *Rfn* from H5 with that of the *rfn* allele from 181A, we found that only five single nucleotide polymorphisms (SNPs) were detected in the putative promoter region, while rich sequence variations including 184 SNPs were identified in the coding sequence. These SNPs have no effect on the deduced protein length, but result in 99 amino acids mutations (Supplementary Figs S2, S3). Sequence analysis indicated that both *Rfn* and *rfn* encode novel PPR proteins (RFN and rfn, respectively) of 629 amino acids with 16 PPR motifs (Supplementary Fig. S4), which contain a predicted mitochondrial transit signal at the N-terminus (<https://ihg.gsfc.de/ihg/mitoprot.html>; Claros and Vincens, 1996). Further analysis of the PPR motifs between RFN and rfn indicated that nine of the ‘code’ combinations in each PPR motif are different from each other (Fig. 2). This difference suggested that the recognized sequence or the binding ability of CMS-associated RNA target sequence may be different, and thus cause the functional divergence between RFN and rfn.

Expression pattern of the *Rfn* gene

qRT-PCR analysis showed that the *Rfn* and *rfn* alleles are constitutively expressed in the selected tissues, and both

maintained relatively high-level expression in the flower buds (Fig. 3A). Similar results were obtained by GUS staining assay in the *Rfn* promoter–GUS fusion transgenic Arabidopsis plants (Supplementary Fig. S5). To determine precisely the spatial and temporal expression pattern of *Rfn* during different anther development stages, we performed RNA *in situ* hybridization using an *Rfn*-derived probe. As shown in Fig. 3, faint hybridization signals were visible at the microspore mother cell (MMC) and meiotic stages (Fig. 3B, C). At the tetrad stage, the expression of the *Rfn* gene was detected in the tapetum and tetrads at its maximal level (Fig. 3D). The hybridization signals were slightly weakened in the subsequent stages and remained detectable in the tapetum but sharply decreased in the pollen grains (Fig. 3E, F). Subcellular localization analysis using a construct expressing a GFP fused with the N-terminal 44 residue signal peptide of RFN confirmed that this protein targets to mitochondria (Fig. 4), consistent with its function in restoring CMS.

Effects of *Rfn* on processing of *nap* CMS-associated transcripts or protein

In a previous study, northern blot analysis suggested that expression of the *orf222/nad5clorf139* region may be associated with the *nap* CMS (L’Homme *et al.*, 1997). To examine the mechanism through which RFN affects this CMS-associated transcript, RNA gel blot analysis was performed with the probes *orf222* and *orf139*, respectively. As shown in Fig. 5, the *orf222*-containing transcript levels in the restorer and transgenic fertility-restored lines were not significantly changed when compared with those in the *nap* CMS line (Fig. 5B). In contrast, an ~1.2 kb transcript was specifically observed in the restorer and transgenic fertility-restored lines when using *orf139* as a probe (Fig. 5B). Further immunodetection by the antibody against the *orf222*-encoded protein (ORF222) showed that the accumulation of ORF222 was not obviously changed in the restorer line and transgenic fertility-restored lines compared with the CMS line (Fig. 5C). These results indicate that the *Rfn* gene may not be responsible for processing the *orf222* transcripts but rather the *orf139* transcripts.

PPR motif	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
amino acid residue	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35	5 35
RFN	K T	N S	S D	S N	N D	G Q	N N	T N	N N	N D	N D	N S	S D	S E	N S	N D
rfn	K T	N S	S D	S N	N N	G D	N N	N T	N D	N D	N N	N D	K D	N D	T N	N D
Recognized RNA base																
RFN	N	C	G	A	U	N	Y	A	Y	U	U	C	G	R	C	U
rfn	N	C	G	A	Y	G	Y	Y	U	U	Y	U	N	U	A	U

Fig. 2. Combinations of amino acid residues at positions 5 and 35 of PPR motifs and the RNA sequence predicted to be recognized by these combinations for RFN and rfn proteins. ‘N’ represents that the recognized RNA base is unknown; ‘R’ represents that the recognized RNA base is A or G; ‘Y’ represents that the recognized RNA base is U or C. The same or different amino acid residue combinations in each PPR motif between RFN and rfn and their recognized RNA bases are indicated by different shading. (This figure is available in colour at *JXB* online.)

Discussion

Rfn and *Rfp* are tightly linked independent genes and highly homologous

A previous study indicated that the restorer genes *Rfn* for *nap* cytoplasm and *Rfp* for *pol* cytoplasm represent different alleles or haplotypes of a single nuclear genetic locus (Li *et al.*, 1998). Very recent research confirmed that *Rfn* indeed localizes to the

same genomic region as *Rfp* (Gaborieau and Brown, 2016). In this study, through a map-based cloning strategy, we finely mapped the *Rfn* gene on the end of chromosome A09 of *B. napus* and successfully isolated it; and it was shown to encode a novel mitochondria-localized PPR protein. According to the *B. napus* reference genome of Darmor-*bzh* (Chalhoub *et al.*, 2014), we found that the physical distance between *Rfn* and *Rfp* (Liu *et al.*, 2016) is ~125 kb. Sequence comparison showed that RFN and RFP are highly conserved, due to their sequence identity of 70% at the amino acid level (Supplementary Fig. S4). BlastP showed that both RFN and RFP have the highest similarity to RFL2 (At1g12300) in *Arabidopsis thaliana* (Supplementary Fig. S4). However, the gene *Rfp* which can restore the *nap* CMS, because some *pol* CMS restorer lines failed to restore the *nap* CMS of 181A in F₁ plants, and all the positive T₀ transformants remained male sterile phenotype when we introduced *Rfp* into 181A (data not shown). So, our results strongly suggested that *Rfn* and *Rfp* are tightly linked independent genes and they are likely to be evolutionarily homologous genes, but underwent a diversifying selection for adapting to newly emerging sterility-inducing genes (Geddy and Brown, 2007).

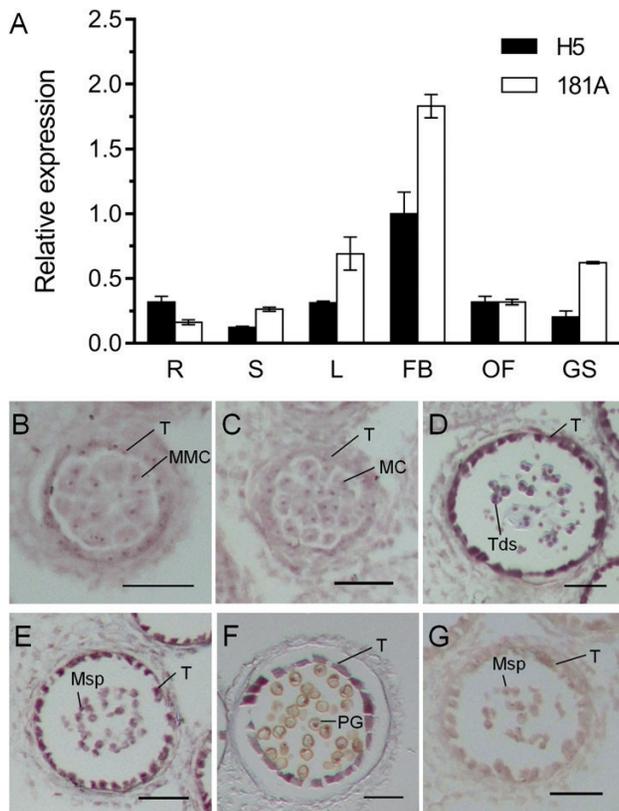


Fig. 3. The expression patterns of *Rfn* (*rfn*) by qRT-PCR and RNA *in situ* hybridization. (A) qRT-PCR analysis of *Rfn* and *rfn* expression in the selective tissues of parent lines H5 and 181A. R, root; S, stem; L, leaves; FB, flower buds; OF, opening flowers; GS, green siliques. The expression levels were normalized to *BnActin*. The values are presented as the means \pm SD ($n=3$ biological replicates). (B–G) The expression patterns of *Rfn* by RNA *in situ* hybridization in the anther of the fertile parent H5 at different developmental stages detected by the antisense probe (B–F) and sense probe (G). MC, meiotic cell; MMC, microspore mother cell; Msp, microspores; PG, pollen grains; T, tapetum; Tds, tetrads. Scale bars=50 μ m. (This figure is available in colour at JXB online.)

The possibility of functional divergence between RFN and rfn

PPR proteins constitute a large group of RNA-binding proteins in terrestrial plants, and function as sequence-specific ssRNA-binding proteins mainly in chloroplasts and mitochondria, where they are involved in various steps of organelle gene expression (Barkan and Small, 2014). Recent structural studies of PPR proteins indicated that one PPR motif has the ability to recognize a single nucleotide, and the amino acid combinations at the 5th and 35th residues of each PPR motif are very important for recognizing the specific RNA sequence (Yin *et al.*, 2013; Shen *et al.*, 2016). In *A. thaliana*, an amino acid mutation at position 35 of the third PPR motif results in RFL9 losing its function of processing the *rps3* and *orf240* transcript in the mitochondria (Arnal *et al.*, 2014). The comparison of coded amino acids combinations between RFN and *rfn* showed that nine of them are different. Furthermore, we found that *Rfn* and *rfn* are constitutively expressed in various tissues, and the expression level of *rfn* is even higher than that of *Rfn* in flower buds. So, it is suggested that the functional divergence between RFN and *rfn* may not due to their few sequence variations in

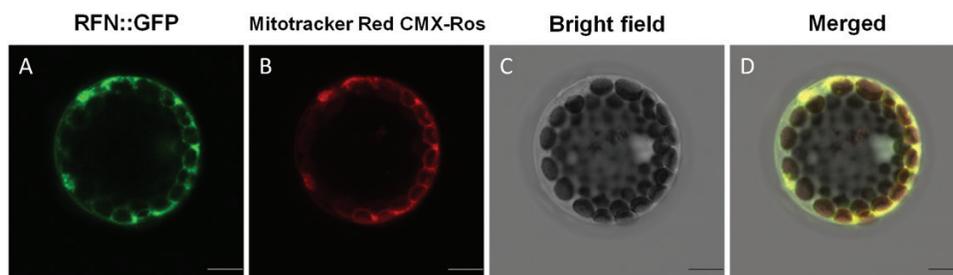


Fig. 4. Subcellular localization of the N-terminal 44 residue signal peptide of RFN fused with green fluorescent protein (GFP). (A) The protoplast showed a green fluorescent signal at 488 nm. (B) The same protoplast showed a red fluorescent signal (stained by MitoTracker Red CMX-Ros) at 580 nm. (C) The bright-field image. (D) The merged image of (A), (B), and (C). Scale bars=10 μ m. (This figure is available in colour at JXB online.)

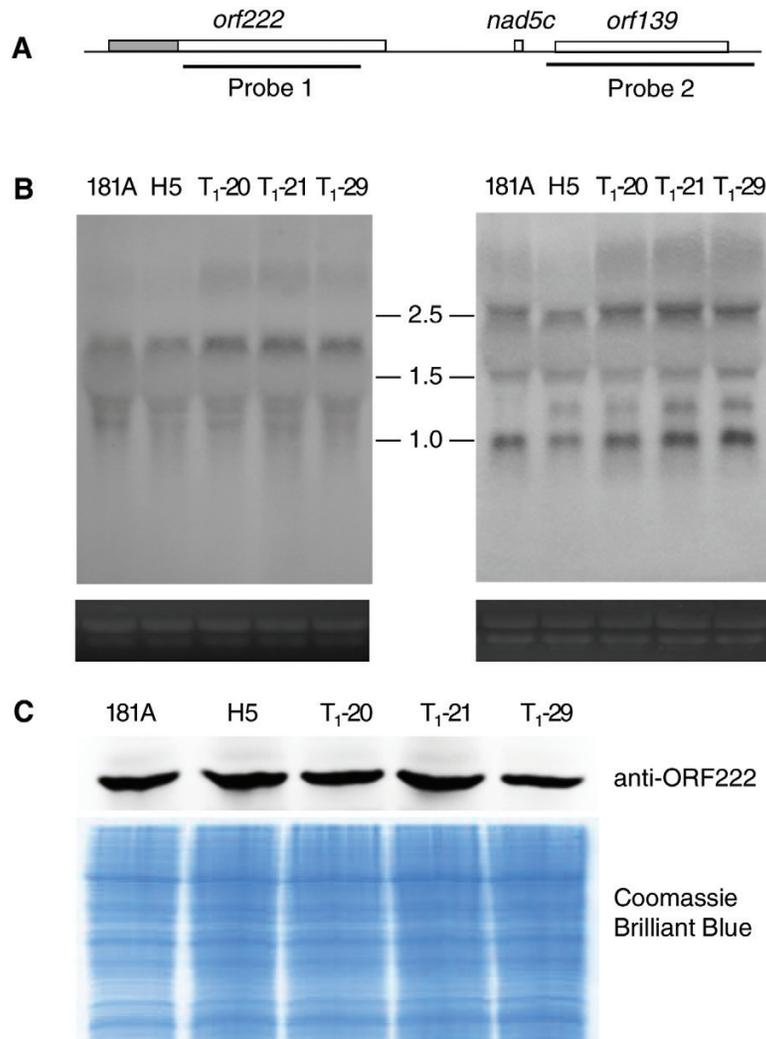


Fig. 5. RNA gel blot analysis of the *orf222/nad5c/orf139* transcripts with probes *orf222* and *orf139*, respectively, and immunodetection of ORF222 from CMS, restorer line, and transgenic fertility-restored plants. (A) The structure of *nap* CMS-associated transcripts *orf222/nad5c/orf139*. The shaded portion indicated the *orfB* homologous region. (B) RNA gel blots analysis with probes *orf222* (left panel) and *orf139* (right panel), respectively. The estimated sizes of the different transcripts are shown in kilobases. Ethidium bromide stain of the gel confirms equal RNA loading. (C) Immunodetection of CMS protein expression in T₁ progeny with an anti-ORF222 antibody. Equal amounts of total protein were loaded in each lane stained by Coomassie Brilliant Blue. (This figure is available in colour at *JXB* online.)

promoter regions, but have resulted from the changes in nine amino acids residing in the code positions.

The possible role of *Rfn*

The restoration of fertility in CMS/*Rf* systems may be achieved by various mechanisms at different molecular levels, such as the genomic level, post-transcriptional level, translational or post-translational level, and metabolic level (Chen and Liu, 2014). Most CMS-associated transcripts or proteins in various CMS crops are cleaved or degraded by the corresponding RF proteins. In BT-CMS rice, the CMS-associated transcripts *B-atp6/orf79* were cleaved by RF1A but degraded by RF1B (Wang *et al.*, 2006). In Ogu-CMS rapeseed, the RFO protein may impede *orf138* mRNA translation by preventing either association or progression of mitochondrial ribosomes with the *orf138* mRNA (Uyttewaal *et al.*, 2008). A previous study indicated that the *nap* CMS-associated transcripts

orf222/nad5c/orf139 were affected by the action of *Rfn* at the post-transcriptional level (L'Homme *et al.*, 1997). In the current work, we detected the CMS-associated transcripts in the *nap* CMS line, restorer line, and transgenic fertility-restored lines using the *orf222* and *orf139* probe, respectively. We found that the CMS-associated transcripts were not significantly changed when using *orf222* as a probe, but an ~1.2 kb transcript appeared in the restorer line and transgenic fertility-restored lines when *orf139* was used as a probe. The accumulation of ORF222 was also not obviously changed in the restorer line and transgenic fertility-restored lines compared with the *nap* CMS line. These findings suggested two possibilities: first, there is another CMS-associated gene but not *orf222*; secondly, RFN may process the *nap* CMS-associated *orf222* in a unique stage. Further investigation is needed to identify the target RNA of RFN or RNA blotting with a specific stage of the anther and thus illustrate the molecular mechanism of how RFN recovers the *nap* CMS in *B. napus*.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. Opening flower morphology of the parent plants 181A and H5 under relatively low temperature.

Fig. S2. The putative promoter region and coding sequence of *Rfn* and its deduced amino acids from the restorer line H5.

Fig. S3. Nucleotide sequence of recessive *rfn* allele from 181A and the deduced amino acids.

Fig. S4. Alignment of the amino acid sequences of RFN, RFP, and their homologous protein RFL2 in Arabidopsis.

Fig. S5. GUS assay from different tissues of Pro_{*Rfn*}:GUS-introduced Arabidopsis plants.

Table S1. The primers used in this study.

Table S2. Fertility of T₁ progeny derived from five independent T₀ transgenic plants by selfing and test cross.

Acknowledgements

We thank Qinghua Zhang and Liang Liu for sequencing and assembling the BAC clone. We are grateful to the anonymous reviewers for their constructive comments and suggestions. This work was supported by the National Natural Science Foundation of China (no. 31300160), National Key Research and Development Plan of China (2016YFD0101300), Technological innovation project of Hubei (2016ABA084), and Modern Agro-Industry Technology Research System (CARS-13).

References

- Arnal N, Quadrado M, Simon M, Mireau H. 2014. A restorer-of-fertility like pentatricopeptide repeat gene directs ribonucleolytic processing within the coding sequence of *rps3-rpl16* and *orf240a* mitochondrial transcripts in *Arabidopsis thaliana*. *The Plant Journal* **78**, 134–145.
- Ban T, Ke J, Chen R, *et al.* 2013. Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *Journal of Biological Chemistry* **288**, 31540–31548.
- Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. 2012. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genetics* **8**, e1002910.
- Barkan A, Small I. 2014. Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology* **65**, 415–442.
- Bentolila S, Alfonso AA, Hanson MR. 2002. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male sterile plants. *Proceedings of the National Academy of Sciences, USA* **99**, 10887–10892.
- Brown GG, Formanová N, Jin H, *et al.* 2003. The radish *Rfo* restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *The Plant Journal* **35**, 262–272.
- Chalhoub B, Denoeud F, Liu S, *et al.* 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* **345**, 950–953.
- Chen L, Liu YG. 2014. Male sterility and fertility restoration in crops. *Annual Review of Plant Biology* **65**, 579–606.
- Cheng F, Liu S, Wu J, Fang L, Sun S, Liu B, Li P, Hua W, Wang X. 2011. BRAD, the genetics and genomics database for *Brassica* plants. *BMC Plant Biology* **11**, 136.
- Claros MG, Vincens P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. *European Journal of Biochemistry* **241**, 779–786.
- Clough SJ, Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**, 735–743.
- Croijmans RP, Vrebalov J, Dijkhof RJ, van der Poel JJ, Groenen MA. 2000. Two-dimensional screening of the Wageningen chicken BAC library. *Mammalian Genome* **11**, 360–363.
- Cui X, Wise RP, Schnable PS. 1996. The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* **272**, 1334–1336.
- Fan Z, Stefansson BR. 1986. Influence of temperature on sterility of two cytoplasmic male-sterility systems in rape (*Brassica napus* L.). *Canadian Journal of Plant Science* **66**, 221–227.
- Fujii S, Toriyama K. 2009. Suppressed expression of RETROGRADE-REGULATED MALE STERILITY restores pollen fertility in cytoplasmic male sterile rice plants. *Proceedings of the National Academy of Sciences, USA* **106**, 9513–9518.
- Gaborieau L, Brown GG. 2016. Comparative genomic analysis of the compound *Brassica napus* *Rf* locus. *BMC Genomics* **17**, 834.
- Geddy R, Brown GG. 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* **8**, 130.
- Gully BS, Cowieson N, Stanley WA, Shearston K, Small ID, Barkan A, Bond CS. 2015. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding *atpH* RNA. *Nucleic Acids Research* **43**, 1918–1926.
- Hanson MR, Bentolila S. 2004. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell* **16** (Suppl), S154–S169.
- Hu J, Wang K, Huang W, *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. *The Plant Cell* **24**, 109–122.
- Huang W, Yu C, Hu J, *et al.* 2015. Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. *Proceedings of the National Academy of Sciences, USA* **112**, 14984–14989.
- Igarashi K, Kazama T, Toriyama K. 2016. A gene encoding pentatricopeptide repeat protein partially restores fertility in RT98-type cytoplasmic male-sterile rice. *Plant and Cell Physiology* **57**, 2187–2193.
- Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K. 2011. The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *The Plant Journal* **65**, 359–367.
- Jing B, Heng S, Tong D, *et al.* 2012. A male sterility-associated cytotoxic protein ORF288 in *Brassica juncea* causes aborted pollen development. *Journal of Experimental Botany* **63**, 1285–1295.
- Kazama T, Toriyama K. 2003. A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. *FEBS Letters* **544**, 99–102.
- Ke J, Chen RZ, Ban T, *et al.* 2013. Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nature Structural and Molecular Biology* **20**, 1377–1382.
- Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF. 2005. Fertility restorer locus *Rf1* [corrected] of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theoretical and Applied Genetics* **111**, 994–1012.
- Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J. 2003. Genetic characterization of a pentatricopeptide repeat protein gene, *orf687*, that restores fertility in the cytoplasmic male-sterile Kosena radish. *The Plant Journal* **34**, 407–415.
- Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N. 2004. Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). *The Plant Journal* **37**, 315–325.
- L'Homme Y, Stahl RJ, Li XQ, Hameed A, Brown GG. 1997. *Brassica nap* cytoplasmic male sterility is associated with expression of a mtDNA region containing a chimeric gene similar to the pol CMS-associated *orf224* gene. *Current Genetics* **31**, 325–335.
- Li S, Chen L, Zhang L, *et al.* 2015. *BnaC9.SMG7b* functions as a positive regulator of the number of seeds per silique in *Brassica napus* by regulating the formation of functional female gametophytes. *Plant Physiology* **169**, 2744–2760.
- Li XQ, Jean M, Landry BS, Brown GG. 1998. Restorer genes for different forms of *Brassica* cytoplasmic male sterility map to a single nuclear locus that modifies transcripts of several mitochondrial genes. *Proceedings of the National Academy of Sciences, USA* **95**, 10032–10037.

- Liu Z, Liu P, Long F, Hong D, He Q, Yang G.** 2012. Fine mapping and candidate gene analysis of the nuclear restorer gene *Rfp* for pol CMS in rapeseed (*Brassica napus* L.). *Theoretical and Applied Genetics* **125**, 773–779.
- Liu Z, Yang Z, Wang X, et al.** 2016. A mitochondria-targeted PPR protein restores pol cytoplasmic male sterility by reducing *orf224* transcript levels in oilseed rape. *Molecular Plant* **9**, 1082–1084.
- Livak KJ, Schmittgen TD.** 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- Lurin C, Andrés C, Aubourg S, et al.** 2004. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant Cell* **16**, 2089–2103.
- Matsuhira H, Kagami H, Kurata M, et al.** 2012. Unusual and typical features of a novel restorer-of-fertility gene of sugar beet (*Beta vulgaris* L.). *Genetics* **192**, 1347–1358.
- Okuda K, Shoki H, Arai M, Shikanai T, Small I, Nakamura T.** 2014. Quantitative analysis of motifs contributing to the interaction between PLS-subfamily members and their target RNA sequences in plastid RNA editing. *The Plant Journal* **80**, 870–882.
- Schmitz-Linneweber C, Small I.** 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science* **13**, 663–670.
- Shen C, Zhang D, Guan Z, et al.** 2016. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nature Communications* **7**, 11285.
- Singh M, Brown GG.** 1991. Suppression of cytoplasmic male sterility by nuclear genes alters expression of a novel mitochondrial gene region. *The Plant Cell* **3**, 1349–1362.
- Singh M, Brown GG.** 1993. Characterization of expression of a mitochondrial gene region associated with the *Brassica* 'Polima' CMS: developmental influences. *Current Genetics* **24**, 316–322.
- Tang H, Luo D, Zhou D, Zhang Q, Tian D, Zheng X, Chen L, Liu YG.** 2014. The rice restorer *Rf4* for wild-abortive cytoplasmic male sterility encodes a mitochondrial-localized PPR protein that functions in reduction of *WA352* transcripts. *Molecular Plant* **7**, 1497–1500.
- Uyttewaal M, Arnal N, Quadrado M, et al.** 2008. Characterization of *Raphanus sativus* pentatricopeptide repeat proteins encoded by the fertility restorer locus for Ogura cytoplasmic male sterility. *The Plant Cell* **20**, 3331–3345.
- Wan L, Xia X, Hong D, Yang G.** 2010. Molecular analysis and expression of a floral organ-specific polygalacturonase gene isolated from rapeseed (*Brassica napus* L.). *Molecular Biology Reports* **37**, 3851–3862.
- Wang Z, Zou Y, Li X, 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. *The Plant Cell* **18**, 676–687.
- Wang ZW, Wang C, Wang C, Gao L, Mei SY, Zhou Y, Xiang CP, Wang T.** 2013. Heterozygous alleles restore male fertility to cytoplasmic male-sterile radish (*Raphanus sativus* L.): a case of overdominance. *Journal of Experimental Botany* **64**, 2041–2048.
- Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T.** 2013. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* **8**, e57286.
- Yin P, Li Q, Yan C, et al.** 2013. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168–171.
- Yoo SD, Cho YH, Sheen J.** 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2**, 1565–1572.
- Zhou YM, Fu TD.** 2007. Genetic improvement of rapeseed in China. In: *Proceedings of the 12th International Rapeseed Congress*. Wuhan, China, 26–30.