Homology-based cloning and expression analysis of *Rf* genes encoding PPR-containing proteins in tobacco

A.M. Ding[1,2], F.X. Li[1], Y.Q. Chen[1,2], P. Zong[1,2], X. Qu[1,2], D.P. Gong[1], G.S. Liu[1] and Y.H. Sun[1]

[1]Key Laboratory for Tobacco Gene Resources,
Tobacco Research Institute of Chinese Academy of Agricultural Sciences,
Qingdao, China
[2]Graduate School of Chinese Academy of Agricultural Sciences,
Beijing, China

Corresponding author: Y.H. Sun
E-mail: yhsun@163.com

ABSTRACT. As a model plant, mechanisms of the cytoplasmic male sterility/restoration of fertility (CMS/*Rf*) system in tobacco are seldom studied. Using *Rf* gene sequences from other Solanaceae plants and the draft genome of *Nicotiana benthamiana*, degenerate primers were designed to amplify the cDNA pool of *N. tomentosiformis*. In total, six possible *Rf* sequences were identified, two of which contained base-deletion mutations. The other four were intact open reading frames, of which NtomPPR5 harbored a 3-pentatricopeptide repeat (PPR) motif deletion. Structure analysis revealed that they all encoded a PPR-containing protein with putative mitochondrial targeting signals at their N-terminus, and they all belong to the P subfamily. Phylogenetic analysis showed that all of the *Rf*-coding PPRs clustered together, and recent duplication events might have occurred in tobacco after the divergence of the species. Quantitative reverse transcription polymerase

chain reaction analysis demonstrated that the *NtomRfs* were expressed in all tissues of *N. tomentosiformis* and (CMS) K326, although the expression levels varied with gene, organ, and developmental stage. Furthermore, the expression levels of *Rf* sequences in K326 were lower than those in CMS K326. The molecular basis of the CMS/*Rf* system in tobacco requires further investigation.

Key words: *Rf* gene; Pentatricopeptide repeat protein family; Quantitative reverse transcription polymerase chain reaction; Sequence structure analysis; Phylogenetic tree

INTRODUCTION

The pentatricopeptide repeat (PPR) protein family was first defined and given this name in *Arabidopsis thaliana* by Small and Peeters (2000), and then it was realized that it is particularly prevalent in terrestrial plants as compared with other eukaryotes. With the accomplishment of whole-genome sequencing projects of many eukaryotic organisms, the number of PPR protein sequences they contain was uncovered. For example, the *Drosophila* and *Caenorhabditis elegans* genomes each contain 2 PPR proteins, and the *Homo sapiens* genome has 6 PPRs. In contrast, the number of PPRs in *A. thaliana* and *Oryza sativa* exceeds 440 and 650, respectively (Lurin et al., 2004). The PPR proteins have a classical structure of a variable length of organelle-targeting sequence at the N-terminus followed by 2-27 PPR repeats in a tandem array, with each motif containing a degenerate 35-amino acid sequence (Lurin et al., 2004; Saha et al., 2007). Because of the close resemblance in structure to tetratricopeptide repeats (TPRs), they were given the name PPR to briefly describe the large PPR-containing protein family.

Although the PPR proteins have only been recognized less than 15 years, it is already clear that this protein family is mainly targeted to mitochondria and chloroplasts and has broad essential functions including mitochondrial and chloroplast RNA processing such as transla-tion, RNA editing, RNA splicing, and RNA degradation (Barkan and Goldschmidt-Clermont, 2000; Shikanai, 2006; Wang et al., 2006); embryo development (Lurin et al., 2004; Cushing et al., 2005); plant development (Pring et al., 1995; Aida and Tasaka, 2006); and fertility restoration against cytoplasmic male sterility (CMS). CMS often causes the production of flowers with non-functional pollen, and its corresponding nuclear restoration of fertility (*Rf*) genes can suppress the expression of CMS inducer genes in mitochondria. The CMS and fertility-restoration system has been used for hybrid seed production in rice, radish, pepper, and sorghum. For many years, researchers have been trying to determine how the CMS occurs and the mechanism that allows *Rf* genes to compensate for their defects and restore fertility. To date, several genes have been proven to act as CMS restorers, including *Rf2* of maize (Cui et al., 1996); *Rf-PPR591/592* of petunia (Bentolila et al., 2002); *Rfs* of rice (Akagi et al., 2004; Huang et al., 2012); *Rfo*, *Rfob*, and *Rft* of radish (Desloire et al., 2003; Wang et al., 2008; Yasumoto et al., 2009); *Rf1* and *Rf2* of sorghum (Klein et al., 2005; Jordan et al., 2010); and *PePPR1* of pepper (Yeong et al., 2010). Fine-mapping of the *Rf* gene in soybean was also executed (Dong et al., 2012). These *Rf* genes all encode PPR-containing proteins except *Rf2* in maize and *Rf17* and *Rf2* in rice (Fujii and Toriyama, 2009; Itabashi et al., 2011). All the *Rf*

genes that were identified have a mitochondrial targeting sequence in their N-terminus, and they can alter the expression profile of CMS-associated genes in mitochondria except *Rf2*, which encodes acetaldehyde dehydrogenase.

Tobacco is not only an important economic crop worldwide but also an ideal model plant in scientific research. However, the CMS/*Rf* mechanisms in this plant are seldom studied. For years, we have obtained its CMS lines and maintainer lines, but we cannot identify any cultivar materials that can restore the fertility of the CMS lines. As a result, no mapping populations can be constructed. Thinking that the wild diploid type may carry this kind of genes, which can restore the fertility of the cultivar CMS lines, the wild diploid species *Nicotiana tomentosiformis* was used as plant material in this study to clone *Rf* genes. This study identified six *Rf* sequences in the wild tobacco species, two of which containing base deletions in their nucleotide sequences that lead to early terminations in their translation products. The other four genes encoding intact PPR-containing proteins may act as CMS restorers. To our knowledge, this is the first study to emphasize genes associated with fertility in the tobacco species. Further study is underway to determine their functions and other useful information.

MATERIAL AND METHODS

Plant materials

The wild diploid species *N. tomentosiformis*, the tobacco cultivated variety K326 and its corresponding cytoplasmic male sterility line (CMS K326) were planted in a greenhouse under natural conditions. During the vegetative period, roots, stem (stem apex), and leaves were collected from the plants; in the reproductive stage, leaves, flower buds, and blooming flowers were additionally sampled from *N. tomentosiformis*. After storing samples in liquid nitrogen, total RNA was extracted from each organ and stored at -70°C after purity and integrity detection.

Cloning of *Rf* genes in *N. tomentosiformis*

The GeneJET Plant RNA Purification Mini Kit (Thermo Fisher Scientific Inc., USA) was used to extract total RNA from the samples according to manufacturer directions. Then, DNase I (TIANGEN Biotech Co., Ltd., Beijing, China) was applied to eliminate genomic DNA contaminants. Reverse transcription was conducted using PrimeScript™ reverse transcription polymerase chain reaction (RT-PCR) kit (Takara Biotechnology Co., Ltd., Dalian, China) to synthesize the first cDNA strand, which was then kept at -20°C for further analysis.

Lurin et al. (2004) discovered that the PPR-coding genes are usually short and devoid of introns, making homology-based cloning relatively easy. Using known *Rf* genes in the Solanaceae plant pepper (Jo et al., 2010) and *Petunia hybrida* (Bentolila et al., 2002), as well as predicted PPR protein sequences in tomato, basic local alignment search tool (BLAST) was conducted against the draft genome sequence of *N. benthamiana* with the online SGN BLAST tool (http://solgenomics.net/organism/Nicotiana_benthamiana/genome). After this step, we obtained a series of *N. benthamiana*-specific PPR-containing sequences, most of which were short fragments. From these, we chose the long sequences with high e values (>1e-100), especially the ones matching the 5' or 3' end of the query sequences, which were

used as seeds for a second round of BLAST analysis. In total, more than 100 relevant sequences (fragments) were identified, from which we chose the 14 most probable sequences with feasibly long coding regions for further analysis. Each sequence was used in BLAST against the draft genome database to detect their most similar sequences in order to divide them into groups (Group1: Scf25162767, Scf25087176, Scf25037889, Scf25274342, and Scf25221533; Group2: Scf25037889, Scf24955369, Scf25256860, and Scf25225937; Group3: Scf25188490 and Scf25225937; and Group4: Scf25188490, Scf25280843, Scf24980057, and Scf25254137). Then, predicted open reading frames (ORFs) of each group were aligned using the DNAMAN software. We designed degenerate primers (Table 1) to PCR amplify the first strand cDNA in *N. tomentosiformis* to obtain the expressed ORFs. The PCR products were retrieved with the MiniBEST Agarose Gel DNA Extraction Kit (TaKaRa Biotechnology Co., Ltd., Dalian, China) and then cloned into the PMD® 18-T vector (TaKaRa Biotechnology Co., Ltd., Dalian, China) with matched *Escherichia coli* DH5α competent cells according to user guides. Five positive clones of each PCR product were sent to BGI Tech Solutions Co., Ltd. for sequencing.

Table 1. Sequences of the degenerate primers.

Primers[a]	Forward sequences	Reverse sequences
Rf1	ATGAGKARAMTTTSTCSCAATGGT	TCAMTKCTTATYTTYCAAGTGAAG
Rf2	ATGGCGAGAAGAATTTCTCTGTGT	TCACTTCTTAKTTTTCRAGTGAAA
Rf3	ATGACGAGAAYTTCTMTGCTKCGT	TCACTTWTTACTTCCCGAGTGAAA
Rf4	ATGATGAGARYTGCMGTGCAY	TCACYTCTTRTTTTCCAAGTG

[a]Only primer pairs which can get PCR products were listed. All the PCRs were in a 25-μL volume, including 1 μL cDNA template, 1 μL forward primer, 1 μL reverse primer, 2 μL 2.5 mM dNTPmix, 2.5 μL buffer, 0.5 μL Taq DNA polymerase, and 17 μL ddH$_2$O.

Sequence analysis

Pfam 27.0 (Punta et al., 2012; http://pfam.sanger.ac.uk/) as well as the online programs SMART (http://smart.embl-heidelberg.de/) and MEME (http://meme.nbcr.net/meme/cgi-bin/meme.cgi) were used to detect the conserved motifs of the predicted PPR-containing proteins in *N. tomentosiformis* as well as the *Brassica napus* Rfo and *Arabidopsis* CRR2. The online programs TargetP v1.1 (Emanuelsson et al., 2000), Predotar v.1.03 (Small et al., 2004), and MitoProt II v1.101 (Claros and Vincens, 1996) were used to perform the prediction of subcellular localization of the PPR proteins. Multiple sequence alignment was carried out with the MEGA 5.0 software (Tamura et al., 2011) using the sequences obtained and known *Rf* genes as well as partial PPRs from *Arabidopsis*. A phylogenetic tree was constructed with the neighbor-joining method.

Quantitative RT-PCR q(RT-PCR) analysis

Relative qRT-PCR was carried out to uncover the expression levels of the *NtomRf* genes. Comparisons proceeded between the vegetative period and reproductive stage of *N. tomentosiformis* and between K326 and CMS K326 at the vegetative period. Total RNA extracted from individual organs was first diluted to an equal concentration of 100 ng/μL. Then, samples were reverse transcribed into cDNA using the PrimeScript™ RT Reagent Kit with gDNA Eraser (Perfect Real Time, TaKaRa Biotechnology Co., Ltd., Dalian, China) according

to the product manual. The cDNA templates were then diluted five times. The SYBR® Premix Ex Taq™ II (Tli RNaseH Plus, TaKaRa Biotechnology Co., Ltd., Dalian, China) was used for qRT-PCR according to the user guide with an Applied Biosystems 7500 machine (ABI7500). The gene-specific primers for qRT-PCR were designed against the unique regions of each *NtomRf* sequence, and the tobacco ribosomal protein-coding gene *NtL25* (Schmidt and Delaney, 2010) was used as a reference gene (Table 2).

Table 2. Sequences of the gene-specific primers and primers for *NtL25*.

Primers	Forward sequences	Reverse sequences
GSP1[a]	TCCAATGGTATTATTCCCTTTA	AACAGAAGGAAGAGACTGAGAT
GSP4	TACTTTCGCAACTGTACTCC	GAAAAGGATGTGTTCTGACC
GSP5	GATTACTCATTCTGCTTCGG	AAGAGAAACAACAGCAGAGT
GSP6	CAAGGGAACACTAAGCCGAACA	CTCAGAGAACAAAGTCCTAACC
NtL25	CAAAAGTTACATTCCACCG	TTTCTTCGTCCCATCAGGC

[a]Because the sequences of *NtomRf1*, *2*, *3* are too similar to each other to design gene-specific primers (with an identity of 94.82%), one pair of primers were used to detect the expression levels of these three sequences.

RESULTS

Cloning of *NtomRf* genes

With each of the four pairs of degenerated primers listed in Table 1, we obtained a single and clear DNA band in agarose gel electrophoresis analysis (data not shown). For individual products, we chose five positive clones for sequencing after they were linked to the PMD® 18-T vector and transformed to *E. coli* DH5α. The sequencing results showed that the PCR products of primers Rf2, Rf3, and Rf4 were monoclones, which were termed *NtomRf4*, *NtomRf5*, and *NtomRf6*, whereas the products of Rf1 were polyclones, from which we obtained three different sequences named *NtomRf1*, *NtomRf2*, and *NtomRf3*. Details of the nucleotide and predicted protein sequences are listed in Table 3. It is worth noting that compared with *NtomRf1* and *NtomRf3*, a G deletion occurred at the site 1471 of *NtomRf2*, causing a frame-shift mutation and a premature transcription termination at the 496th codon (Figure 1). Similar events also happened in *NtomRf6*. These two genes may have lost their functions.
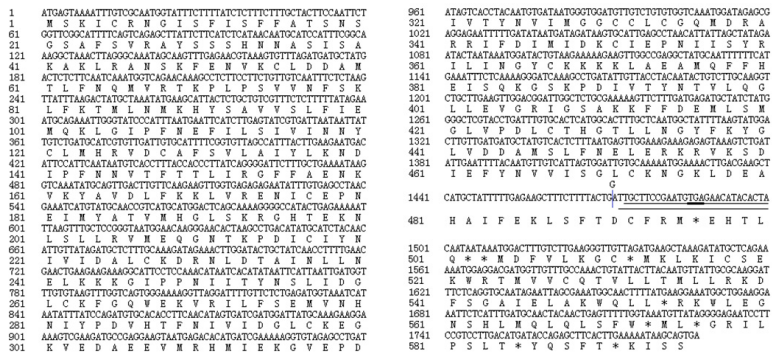

Figure 1. Sequences of *NtomRf2* and NtomPPR2. There was a G deletion at the site 1471, which caused a premature transcription termination at the 496th codon.

Features of the PPR-containing sequences

Figure 2 shows the PPR motifs for the four intact NtomPPRs as well as *B. napus* Rfo and *Arabidopsis* CRR2 detected by Pfam 27.0. It has been confirmed that Rfo belongs to the P subfamily, whereas AtCRR2 resides in the PLS subfamily. According to their structure profiles, all of the NtomPPRs resemble Rfo with no C-terminal domains, suggesting that they all belong to the P subfamily. As to the number of PPR repeats contained in individual sequences, different programs provide different results. Compared with Pfam 27.0, large differences were detected when SMART was used (Table 3). The program MitoProt II assigned all NtomPPRs except NtomPPR5 to mitochondria, Predotar predicted that NtomPPR2 and NtomPPR6 did not have a targeting sequence, and TargetP assigned NtomPPR2 and NtomPPR3 to chloroplasts rather than mitochondria (Table 3).
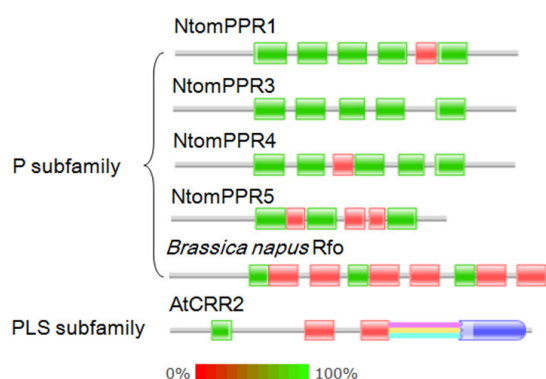

Figure 2. Schematic diagram showing PPR motifs distribution in the four NtomPPRs, the *Brassica napus* Rfo and the *Arabidopsis* CRR2 as obtained from Pfam 27.0. NtomPPR2 and 6 were rejected because of lacking C-terminus. The subfamilies that Rfo (belonging to P class) and AtCRR2 (belonging to PLS class) belong to are clear (Schmitz-Linneweber and Small, 2008). The green and red columns are tandem PPR repeats, which were colored according to the posterior probability. The bar with three colors and the blue one represent the E and DYW motifs, respectively.

Table 3. Details of the *NtomRfs* and NtomPPRs.

Gene name	Protein name	ORF length (bp)	No. of amino acid	No. of PPR repeats (with MEME)	Subcellular location[a]	Subfamily[b]
NtomRf1	NtomPPR1	1788	595	15	M, M, M	P
NtomRf2	NtomPPR2	1787	495 (fragment)	13	M, N, C	-
NtomRf3	NtomPPR3	1788	595	15	M, M, C	P
NtomRf4	NtomPPR4	1776	591	14	M, M, M	P
NtomRf5	NtomPPR5	1440	479	12	N, M, M	P
NtomRf6	NtomPPR6	1785	134 (fragment)	2	M, N, M	-

[a]For subcellular location, the three results in order are predicted with MitoProt II, Predotar, and TargetP, respectively. M = mitochondria; N = no targeting signals; C = chloroplasts. [b]P stands for the P subfamily. We cannot assign NtomPPR2 and NtomPPR6 to any subfamily because of C-terminus deficiency.

Although the PPR and TPR motifs have significant and characteristic differences, they can be aligned relatively easily (Figure 3). Das et al. (1998) proved that TPR motifs contain two anti-parallel alpha helices, with which the TPR repeats are expected to form a superhelix

including a groove to act as a protein-binding site. The amino acid composition of the key sites in PPR motifs is similar to that in TPR, and they also contain two alpha helices, according to which they were predicted to be nucleic acid-binding proteins (Small and Peeters, 2000). Recently, it has been confirmed in some cases that the PPRs are single-strand RNA-binding proteins. For example, Lahmy et al. (2000) found that the PPR protein P67 from radish seedlings is a chloroplastic RNA-binding protein, which could be involved in chloroplast RNA processing; Kotera et al. (2005) reported that PPR proteins in *Arabidopsis* participate in editing of chloroplast genes; Wang et al. (2006) uncovered that *Rf1a* and *Rf1b* in rice BoroII cytoplasm encode proteins cleaving or degrading mRNA of CMS inducer genes. Importantly, the mutants of some PPR-coding genes have apparent phenotypes (Barkan et al., 1994), ensuring the conjecture that they are sequence-specific-binding proteins with seldom complementary functions.
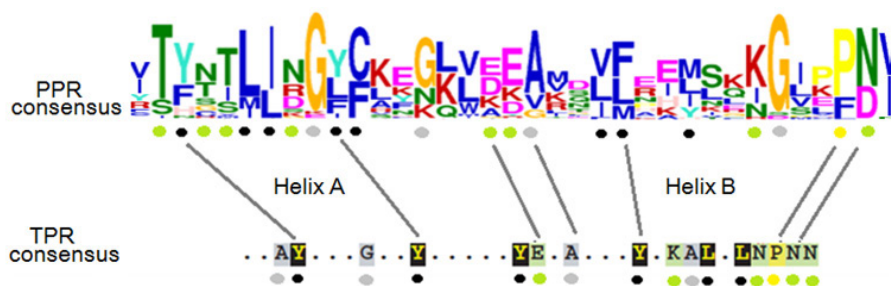

Figure 3. Alignment of PPR and TPR motifs. The PPR consensus was derived with MEME from the four intact NtomPPRs. The TPR consensus was detected by Small and Peeters (2000). The amino acids are marked with different colors of dots according to their properties: hydrophobic = black; hydrophilic = green; small side-chain = grey, and proline = yellow. The MEME-derived profile shows the degenetacy of the PPR repeats. Comparisons of the two motifs have been detailed by Saha et al. (2007).

Phylogenetic analysis of *Rf* genes in plants

Genome-wide analysis of *Arabidopsis* PPR proteins (Lurin et al., 2004) revealed that a 1-Mb region in the long arm of chromosome 1 harbored 19 PPR-coding genes in a cluster, several of which are closely related to *Rf* genes of radish, petunia, and rice. In this study, using the NtomPPRs, BLASTP was performed against the UniProt database by choosing "*Arabidopsis*" in the taxonomy filter. We then chose the ten most similar sequences to construct a neighbor-joining tree together with the *Rf* genes and eight other randomly chosen AtPPRs (Figure 4). The ten AtPPRs are all from chromosome 1. Compared with randomly chosen AtPPRs, all predicted PPR proteins from *Arabidopsis* chromosome 1 were clustered together with *Rf* genes, suggesting their evolutionarily closer relationships than other randomly selected AtPPRs. For the *Rf* genes, the PPRs from Solanaceae plants gathered in one cluster, whereas the ones from Cruciferae plants clustered in another group, indicating orthologous relationships between species. The *Rfs* from Gramineae plants, including rice and maize, are much more distantly related in genetic relationships with other *Rf* genes. In *N. tomentosiformis*, NtomPPR1, 2, 3, and 4 are more similar to each other than to NtomPPR5 and 6 and other *Rf* genes, indicating that duplication events occurred after the divergence of the species.
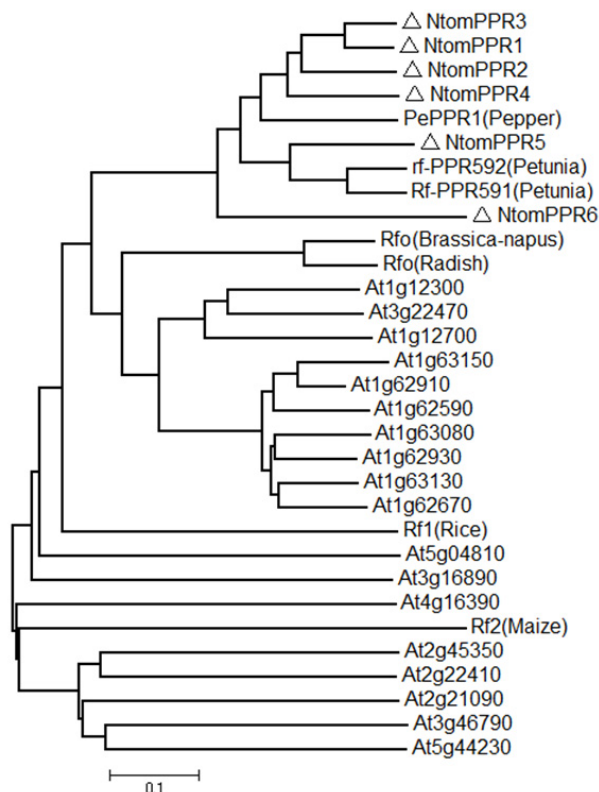
Figure 4. Phylogenetic tree of the *Rf* genes and partial AtPPR proteins. The predicted NtomPPRs are marked with triangle. NtomPPR2 and NtomPPR6 are protein fragments of 495 and 134 amino acids because of early termination.

Expression analysis of the *Rf* genes in *N. tomentosiformis* and (CMS) K326

Genome-wide analysis of the *Arabidopsis* PPR family by Lurin et al. (2004) demonstrated that 40-50% of the PPR-coding genes are expressed at low levels. The expression profiles of individual *Rf* genes may vary from different organs as well as different developmental stages. Expression analysis in specific organs of *N. tomentosiformis* at different developmental stages was shown in Figure 5A. In roots, all of the *NtomRfs* have stable and relatively high levels of transcription. The expression levels of *NtomRf1*, *2*, and *3* were enhanced mainly in flower buds and booming flowers; *NtomRf4* was expressed mainly in leaves of reproductive stage and generative organs; *NtomRf5* was expressed mainly in stem apex and leaves; and *NtomRf6* was expressed mainly in burgeoning tissues such as stem apex and flower buds. In common tobacco, it is interesting to find that the expression levels of all *NtomRfs* were lower in K326 than in its corresponding CMS line with some exceptions (Figure 5B). *NtomRf1*, *2*, *3*, *NtomRf4*, *NtomRf5*, and *NtomRf6* transcriptional differences were detected mainly in roots, stem and leaves, stem, and roots, respectively. In general, *NtomRfs* were expressed in all tissues of *N. tomentosiformis* and (CMS) K326, although the expression level varied with genes, organ, and developmental stage.
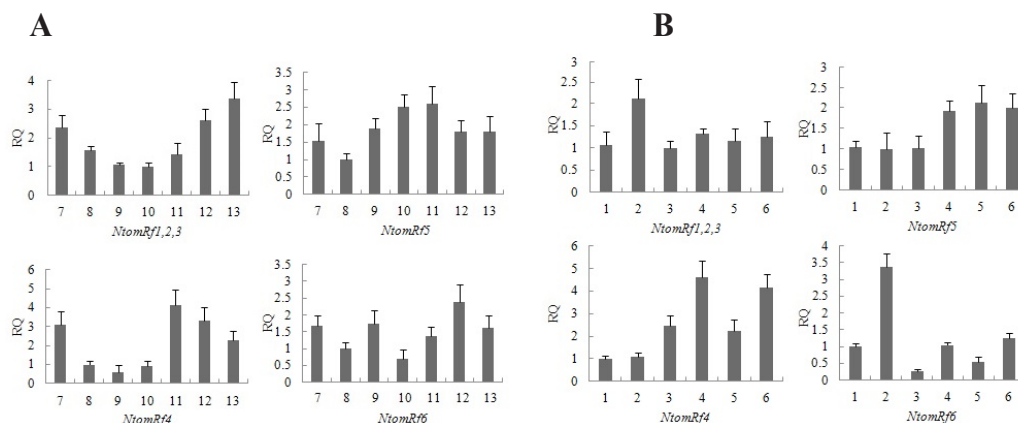
A   **B**

Figure 5. qRT-PCR analysis of *NtomRf* genes. **A. B.** Expression analysis in *Nicotiana tomentosiformis* and K326/ CMS K326, respectively. Arabic numbers stand for specific tissues: 1, 3 and 5 are roots, stem and leaves of K326 in vegetative stage; 2, 4 and 6 are roots, stem and leaves of CMS K326 in vegetative stage; 7, 8, 9, and 10 are roots, stem, stem apex, and leaves of *N. tomentosiformis* in vegetative stage; 11, 12 and 13 are leaves, flower buds and booming flowers of *N. tomentosiformis* in reproductive stage. RQ = relative quantity.

DISCUSSION

The candidate gene approach has been successfully applied to isolate genes of interest as well as develop molecular markers linked to target genes using conserved sequences. Often, cloning target genes in one Solanaceae plant using homologous sequences isolated from other Solanaceae plants is a crucial and successful strategy (Kang et al., 2005; Tomita et al., 2008). *Rf* genes contain highly conserved PPR motifs, although the conservative property of their N- or C-terminal sequences is much lower (Jo et al., 2010). In this study, because of the lack of information on *Rf* genes in the tobacco species, we chose the homology-based cloning method according to known and predicted *Rf* genes of other Solanaceae plants and the draft genome sequence of *N. benthamiana*. After the online BLAST analysis, we selected the *N. bethamiana*-specific *Rf*-related sequences, especially ones with matching 5' and 3' ends to the query sequences. With BLAST, we grouped the sequences and the degenerated primers were developed.

With these primers, six *NtomRf* genes were separated from *N. tomentosiformis*, two of which were detected with base deletions leading to premature termination in transcription, suggesting that they may be non-functional pseudogenes. Sequencing analysis showed that the PCR products of primer Rf1 were polyclones, from which we obtained three different sequences (*NtomRf1*, *2*, and *3*) with an average identity of 94.82% between sequences. It also uncovered that *NtomRf2* contained a G-deletion at the site 1471, causing a premature stop codon. Therefore, it could not be translated into functional protein. Similar events might also happen with *NtomRf6* because several early stop codons appeared in its predicted protein sequence, generating a fragment of only 134 amino acids. Lurin et al. (2004) proposed that most PPR-coding genes in *Arabidopsis* do not contain introns. In this study, PCR amplifications of *N. tomentosiformis* genomic DNA were also conducted with primers designed for *NtomRfs*. Sequencing results showed that the PCR products from cDNA and genomic DNA are the same; therefore, the genes are devoid of introns.

It has been proven that all *Rf* genes encode PPR-containing proteins except *Rf2* from maize and *Rf2* and *Rf17* from rice. Often, the PPR tandem repeats account for about 2/3 of the sequence itself. Polymorphisms within this protein family are mainly because of duplication or deletion of the tandem repeats, making the analysis of these sequences ambiguous. In this study, all of the NtomPPRs are predicted to contain PPR motifs, although the number varied according to the program that was used. The length of *NtomRf5* is about 340-bp shorter than that of the other *NtomRfs* as well as *Rf-PPR591/592* (orthologous genes of *NtomRf5*) from petunia. Sequence alignment demonstrated that 122 amino acid deletions occurred in NtomPPR5 (Figure 6), leading to the loss of three PPR motifs, which might be helpful in the functional diversification of *Rf* genes. Most of the known PPR proteins have been proven or predicted to be targeted to mitochondria or chloroplasts of plant cells (Small and Peeters, 2000; Lurin et al., 2004). Given this, we checked the proteins for predicted targeting signal peptides in the N-terminal region using online programs. Generally, all of the NtomPPRs are predicted to be located to mitochondria besides some exceptions, which might be attributed to algorithmic errors of the programs because the percentage of false negative results with Predotar or TargetP is about 20-30% (Emanuelsson et al., 2000; Small et al., 2004). The localization of NtomPPRs to mitochondria further assured their possible functions in interacting with CMS-related genes.
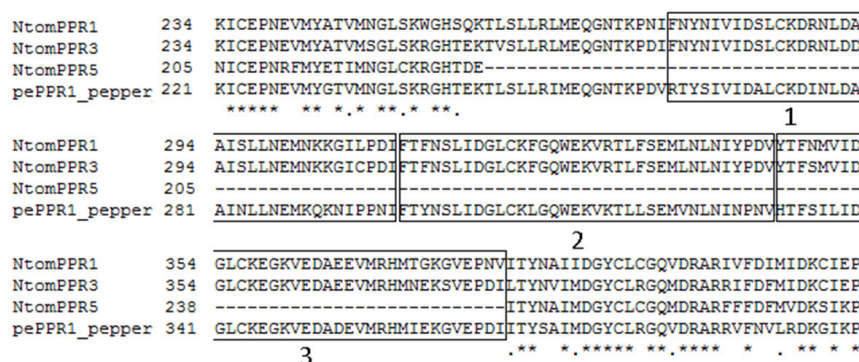

Figure 6. Alignment of predicted protein sequences of NtomPPR1, 3, 5 and pePPR1 from pepper. The 122 amino acid deletions in NtomPPR5 caused a missing of three PPR motifs (named 1, 2 and 3 in the figure). The deletions are indicated by dashes, and the missing PPRs are marked with box. Asterisks indicate amino acid identity; dots indicate low levels of amino acid similarity.

As it has been pointed out, plants have hundreds of PPR proteins with little or no redundant functions. It was suggested by Li et al. (1998) that the *Rf* loci might resemble the disease resistance loci, which were considered a result of gene duplication and modification events in order to cope with new pathogens (Richter and Ronald, 2000). Accordingly, Richter and Ronald (2000) proposed that the radish *Rfo* locus harbors tightly linked genes of related functions, which might have arisen through recent duplication events, and Brown et al. (2003) presented an interesting idea that multi-genes in the *Rf* loci arose in response to the appearance of new forms of CMS. It was proven in petunia (Bentolila et al., 2002) and radish (Wang et al., 2008) that the *Rf* loci contain closely linked *Rf* genes. In this study, the sequences of *NtomRf1*, *2*, *3*, and *4* are highly similar to each other, indicating that recent duplication events occurred in *N. tomentosiformis*, but whether they assemble in a cluster requires further investigation.

As a gene family, each member of the *Rf* genes may have a special role in different plant tissues or cells. The investigation of expression profiles in special tissues may help elucidate the specific mechanism of each *Rf* gene in suppressing CMS. It was verified that *Rf-PPR592* of petunia was only expressed in floral buds (Bentolila et al., 2002). In rice, studies by Akagi et al. (2004) showed that *Rf1* was expressed during booting stage in panicles (with pollen) and green leaves as well as in the vegetative organ calli. In this study, the *NtomRfs* that were identified were expressed in all tissues that were examined of both the wild species, *N. tomentosiformis*, and the common tobacco cultivar, K326, as well as its CMS line, although apparent transcriptional differences were detected. Because the nucleotide sequences of *NtomRf1*, *2*, and *3* are too similar to design gene-specific primers, one pair of primers was used to detect their expression levels. Therefore, the expression profile of each gene and whether they have differentiated in functions were unclear. *NtomRf6* was predicted to be a pseudogene, but it was still expressed in all plant organs, indicating that a recent mutation event occurred. We can also conjecture that functional copies of this gene still exist. qRT-PCR showed us that each *NtomRf* gene has a specific expression profile in *N. tomentosiformis*. However, it is interesting to find that their transcription levels were lower in K326 than in CMS K326, which was hard to explain. One conjecture was that the factors that caused infertility in K326 had nothing to do with these *Rf* genes but could induce their expression.

In this study, we identified six possible *Rf* sequences in *N. tomentosiformis*, four of which are intact ORFs and are homologous to the known *Rf* genes in pepper and petunia, indicating that they might be functional CMS suppression genes. Further studies of these four genes are needed in order to determine if they are real *Rf* genes and their mechanism; homologous sequences in the K326 and CMS K326 cultivars and why restorers for CMS K326 were not identified.

ACKNOWLEDGMENTS

REFERENCES

Aida M and Tasaka M (2006). Morphogenesis and patterning at the organ boundaries in the higher plant shoot apex. *Plant Mol. Biol.* 60: 915-928.

Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, et al. (2004). Positional cloning of the rice *Rf-1* gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. *Theor. Appl. Genet.* 108: 1449-1457.

Barkan A and Goldschmidt-Clermont M (2000). Participation of nuclear genes in chloroplast gene expression. *Biochimie* 82: 559-572.

Barkan A, Walker M, Nolasco M and Johnson D (1994). A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* 13: 3170-3181.

Bentolila S, Alfonso AA and Hanson MR (2002). A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. U. S. A.* 99: 10887-10892.

Brown GG, Formanova N, Jin H, Wargachuk R, et al. (2003). The radish *Rfo* restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. *Plant J.* 35: 262-272.

Claros MG and Vincens P (1996). Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* 241: 779-786.

Cui X, Wise RP and Schnable PS (1996). The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. *Science* 272: 1334-1336.

Cushing DA, Forsthoefel NR, Gestaut DR and Vernon DM (2005). *Arabidopsis emb175* and other *ppr* knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* 221: 424-436.

Das AK, Cohen PW and Barford D (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* 17: 1192-1199.

Desloire S, Gherbi H, Laloui W, Marhadour S, et al. (2003). Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep.* 4: 588-594.

Dong DK, Li Z, Yuan FJ, Zhu SL, et al. (2012). Inheritance and fine mapping of a restorer of fertility (Rf) gene for the cytoplasmic male sterility in soybean. *Plant Sci.* 188-189: 36-40.

Emanuelsson O, Nielsen H, Brunak S and von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300: 1005-1016.

Fujii S and Toriyama K (2009). Suppressed expression of retrograde-regulated male sterility restores pollen fertility in cytoplasmic male sterile rice plants. *Proc. Natl. Acad. Sci. U. S. A.* 106: 9513-9518.

Huang W, Hu J, Yu C, Huang Q, et al. (2012). Two non-allelic nuclear genes restore fertility in a gametophytic pattern and enhance abiotic stress tolerance in the hybrid rice plant. *Theor. Appl. Genet.* 124: 799-807.

Itabashi E, Iwata N, Fujii S, Kazama T, et al. (2011). The fertility restorer gene, Rf2, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *Plant J.* 65: 359-367.

Jo YD, Kim YM, Park MN, Yoo JH, et al. (2010). Development and evaluation of broadly applicable markers for restorer-of-fertility in pepper. *Mol. Breed.* 25: 187-201.

Jordan DR, Mace ES, Henzell RG, Klein PE, et al. (2010). Molecular mapping and candidate gene identification of the *Rf2* gene for pollen fertility restoration in sorghum [*Sorghum bicolor* (*L.*) Moench]. *Theor. Appl. Genet.* 120: 1279-1287.

Kang BC, Yeam I, Frantz JD, Murphy JF, et al. (2005). The *pvr1* locus in *Capsicum* encodes a translation initiation factor eIF4E that interacts with tobacco etch virus VPg. *Plant J.* 42: 392-405.

Klein RR, Klein PE, Mullet JE, Minx P, et al. (2005). Fertility restorer locus Rf1 of sorghum (*Sorghum bicolor* L.) encodes a pentatricopeptide repeat protein not present in the colinear region of rice chromosome 12. *Theor. Appl. Genet*. 111: 994-1012.

Kotera E, Tasaka M and Shikanai T (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433: 326-330.

Lahmy S, Barneche F, Derancourt J, Filipowicz W, et al. (2000). A chloroplastic RNA-binding protein is a new member of the PPR family. *FEBS Lett.* 480: 255-260.

Li XQ, Jean M, Landry BS and 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. *Proc. Natl. Acad. Sci. U. S. A.* 95: 10032-10037.

Lurin C, Andres C, Aubourg S, Bellaoui M, et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089-2103.

Pring DR, Van Tang H and Schertz KF (1995). Cytoplasmic Male Sterility and Organelle DNAs of Sorghum. In: Advances in Cellular and Molecular Biology of Plants (CS Levings III and IK Vasil, eds.). Kluwer Academic, 461-495.

Punta M, Coggill PC, Eberhardt RY, Mistry J, et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40: D290-D301.

Richter TE and Ronald PC (2000). The evolution of disease resistance genes. *Plant Mol. Biol.* 42: 195-204.

Saha D, Prasad AM and Srinivasan R (2007). Pentatricopeptide repeat proteins and their emerging roles in plants. *Plant Physiol. Biochem.* 45: 521-534.

Schmidt GW and Delaney SK (2010). Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Mol. Genet. Genomics* 283: 233-241.

Schmitz-Linneweber C and Small I (2008). Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* 13: 663-670.

Shikanai T (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell Mol. Life Sci.* 63: 698-708.

Small ID and Peeters N (2000). The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* 25: 46-47.

Small I, Peeters N, Legeai F and Lurin C (2004). Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4: 1581-1590.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

Tomita R, Murai J, Miura Y, Ishihara H, et al. (2008). Fine mapping and DNA fiber FISH analysis locates the tobamovirus resistance gene L3 of Capsicum chinense in a 400-kb region of R-like genes cluster embedded in highly repetitive sequences. *Theor. Appl. Genet.* 117: 1107-1118.

Wang Z, Zou Y, Li X, Zhang Q, 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.

Wang ZW, Zhang YJ, Xiang CP, Mei SY, et al. (2008). A new fertility restorer locus linked closely to the Rfo locus for cytoplasmic male sterility in radish. *Theor. Appl. Genet.* 117: 313-320.

Yasumoto K, Terachi T and Yamagishi H (2009). A novel *Rf* gene controlling fertility restoration of Ogura male sterility by RNA processing of *orf138* found in Japanese wild radish and its STS markers. *Genome* 52: 495-504.

Yeong DJ, Young-Min K, Mi-Na P, Jae-Hyoung Y, et al. (2010). Development and evaluation of broadly applicable markers for restore-of-fertility in pepper. *Mol. Breed.* 25: 187-201.