Genome-wide identification of the restorer-of-fertility-like (RFL) gene family in Brassica napus and expression analysis in Shaan2A cytoplasmic male sterility

Luyun Ning, Hao Wang, Dianrong Li, Yonghong Li, Kang Chen, Hongbo Chao, Huaxin Li, Jianjie He and Maoteng Li

Abstract

Background: Cytoplasmic male sterility (CMS) is very important in hybrid breeding. The restorer-of-fertility (Rf) nuclear genes rescue the sterile phenotype. Most of the Rf genes encode pentatricopeptide repeat (PPR) proteins.

Results: We investigated the restorer-of-fertility-like (RFL) gene family in Brassica napus. A total of 53 BnRFL genes were identified. While most of the BnRFL genes were distributed on 10 of the 19 chromosomes, gene clusters were identified on chromosomes A9 and C8. The number of PPR motifs in the BnRFL proteins varied from 2 to 19, and the majority of BnRFL proteins harbored more than 10 PPR motifs. An interaction network analysis was performed to predict the interacting partners of RFL proteins. Tissue-specific expression and RNA-seq analyses between the restorer line KC01 and the sterile line Shaan2A indicated that BnRFL1, BnRFL5, BnRFL6, BnRFL8, BnRFL11, BnRFL13 and BnRFL42 located in gene clusters on chromosomes A9 and C8 were highly expressed in KC01.

Conclusions: In the present study, identification and gene expression analysis of RFL gene family in the CMS system were conducted, and seven BnRFL genes were identified as candidates for the restorer genes in Shaan2A CMS. Taken together, this method might provide new insight into the study of Rf genes in other CMS systems.

Keywords: CMS, Rf, RFL, PPR, RNA-sequencing
mitochondrial proteins [10]. To date, many other Rf genes have been identified in different CMS systems. Most of these Rf genes encode pentatricopeptide repeat (PPR) proteins. Examples of such Rf genes include Rf-PPR592 in petunia (Petunia hybrida) [11], Rfo [12] and orf687 [13] in radish (Raphanus sativus), Rf4 [14], Rf5 [15] and Rf6 [16] in rice (Oryza sativa) and Rfp [17] and Rfn [18] in rapeseed. The PPR proteins were first identified as tandem repeats of degenerate 35-amino-acid motifs (PPR motifs) in Arabidopsis thaliana [19] and were classified into PLS and P subfamilies, according to the PPR motif structure [20]. The PPR gene family is a large family comprising 441 members in Arabidopsis [21], 491 members in rice [22] and 626 members in poplar (Populus alba) [23]. Except for the PPR13 in sorghum (Sorghum bicolor), most of the RF-related PPR proteins belong to the P subfamily and lack the catalytic sites for RNA editing or binding [15]. Two partner proteins of the RF-related PPR proteins have been reported, including GRP162, which associates with RF5 [15], and hexokinase 6 (HXK6), which functions together with RF6 [16] to rescue CMS in rice.

To date, all reported Rf genes have been identified via genetic mapping, which is a time-consuming method and takes several years to narrow down the genomic region of interest. However, the concept of restorer-of-fertility-like (RFL) gene was put forward in 2011, and 212 RFL genes were identified based on BLAST searches using the Rf-PPR592 and Rf5 sequences against the genomes of 13 different dicot and monocot species, including Arabidopsis, soybean (Glycine max) and sorghum [24]. AtRFL2, together with RNase P, regulates the processing of mitochondrial orf291 RNA [25]. AtRFL4 is needed for processing the 5′-end of nad4 mRNA in mitochondria [26]. AtRFL9, also known as RNA PROCESSING FACTOR 4 (RPF4), participates in the generation of extra 5′ termini of ccmB transcripts in Arabidopsis [27]. These results enhanced our understanding of mitochondrial RNA processing in plants and provided novel insights into the function of RFL proteins.

In the present study, we performed BLAST searches using the sequences of Rf-PPR592 and AtRFLs against the genome of rapeseed and identified 53 BnRFL genes. Based on these 53 BnRFL genes, candidate Rf genes were analyzed in the Shaan2A CMS system by RNA-seq and tissue-specific expression analyses. Our data provide a strong foundation for the study of Rf genes in other CMS systems.

**Results**

**Identification of BnRFL genes**

A total of 53 BnRFLs were identified in this study, based on the homology with the RFL genes in Arabidopsis (AtRFL1–26) and petunia (Rf-PPR592) (Table 1). First, sequences of all 26 AtRFLs were searched in the database one at a time. Of the 26 AtRFL genes, nine showed no homologs in B. napus, including AtRFL5, AtRFL6, AtRFL9, AtRFL10, AtRFL14, AtRFL15, AtRFL16, AtRFL25 and AtRFL26. Then, using Rf-PPR592 as a reference [24], 26 BnRFLs were identified (E-value < 1e−100). Taken together, there should be a total of 53 BnRFLs genes in B. napus. We also identified two known restorer genes, BnRFL6 (Rfn) and BnRFL13 (Rfp) (previously identified in the nap and pol CMS systems) and four candidate restorer genes (BnRFL2, BnRFL10, BnRFL11 and BnRFL42; previously identified in B. napus by fine genetic mapping) [31].

The number of PPR motifs in the BnRFL proteins varied from 2 to 19, although most of the BnRFL proteins contained at least 10 PPR motifs and the average number of PPR motifs was 12 (Table 1). Approximately one-fifth of the BnRFLs showed relatively low pl (< 6), whereas nearly half of the BnRFLs showed relatively high pl (> 8). The molecular weight of these RFL proteins ranged from 11.7–92.4 kDa. Additionally, the GRAVY value of nearly two-fifth BnRFLs and most of the selected Rf genes was less than 0, indicating that these RFL proteins were hydrophilic. Most of the BnRFLs were predicted to localize to the mitochondria, which is consistent with the subcellular localization of RF proteins (Table 1).

**Chromosomal location and structural analysis of BnRFL genes**

First, we downloaded the chromosomal distribution of AtRFLs from The Arabidopsis Information Resource (TAIR) (Fig. 1a). All 26 AtRFLs were located in a cluster on chromosome 1. Of the 53 BnRFLs identified in this study, 46 BnRFLs were distributed unevenly on 10 of the 19 chromosomes, and 18 and 10 BnRFLs formed highly dense clusters on chromosomes A9 and C8, respectively. The remaining seven BnRFLs were located on the unmapped scaffold (Fig. 1b and c).

Next, we determined the exon-intron structure of the BnRFL genes and a few known restorer genes (Additional file 1). Most of the BnRFLs were intron-less, similar to the restorer genes, such as Rf4, Rf5 and Rf6, in rice CMS line. Of the 53 BnRFL genes, 10 contained a single intron, similar to the Rf genes, Rfk1, Rfob and orf687, in radish. Notably, the intron in BnRFL23 was more than 4 kb in length, unlike other BnRFLs.

Because PPR proteins generally contain tandem repeats of PPR motifs, we searched for the PPR motifs in the BnRFL proteins and a few known restorer proteins (Additional file 2). To investigate whether BnRFLs contained additional motifs, 53 BnRFLs and 9 known restorer proteins from other species were submitted to MEME. The results showed 20 motifs in the BnRFL proteins (Fig. 2). Interestingly, all of the identified RFL
| A. thaliana gene | A. thaliana ID | B. napus gene | Chr. | Gene position | PPR number | Protein Length (AA) | pI | Molecular weight (Da) | GRAVY | Subcellular location |
|-----------------|----------------|----------------|------|----------------|-------------|---------------------|-----|----------------------|-------|---------------------|
| AtRFL1          | At1G06580      | BnRFL1         | C8   | 45,301,508     | 11          | 497                 | 8.55 | 56,608.61            | 0.183 | M<sup>j</sup>         |
| AtRFL2          | At1G12300      | BnRFL2         | A9   | 42,781,901     | 15          | 637                 | 5.97 | 70,948.87            | 0.026 | M                   |
|                 |                | BnRFL3         | A9   | 42,776,628     | 15          | 647                 | 6.31 | 71,868.61            | 0.038 | M                   |
|                 |                | BnRFL4         | C8   | 42,874,876     | 14          | 640                 | 5.61 | 71,088.91            | 0.032 | M                   |
|                 |                | BnRFL5         | C8   | 42,856,820     | 15          | 626                 | 6.2  | 69,734.49            | 0.001 | M                   |
|                 |                | BnRFL6         | A9   | 42,851,654     | 15          | 629                 | 8.33 | 69,640.90            | 0.16  | M                   |
|                 |                | BnRFL7         | A1   | 25,159,038     | 15          | 626                 | 6.16 | 70,426.2             | 0.001 | M                   |
|                 |                | BnRFL8         | C8   | 42,947,165     | 15          | 621                 | 7.34 | 68,638.38            | 0.016 | M                   |
| AtRFL3          | At1G12620      | BnRFL9         | A8   | 25,819,298     | 14          | 612                 | 6.65 | 68,324.47            | –0.043| M                   |
|                 |                | BnRFL10        | Un†  | 20,135         | 15          | 648                 | 8.21 | 72,194.66            | 0.043 | M                   |
|                 |                | BnRFL11        | A9   | 42,353,345     | 15          | 648                 | 8.2  | 72,022.51            | 0.007 | –                   |
|                 |                | BnRFL12        | C8   | 17,918,175     | 15          | 608                 | 7.37 | 67,610.72            | –0.023| M                   |
|                 |                | BnRFL13        | A9   | 42,707,709     | 15          | 650                 | 8.41 | 73,399.39            | –0.054| M                   |
|                 |                | BnRFL14        | C8   | 17,950,498     | 8           | 323                 | 4.67 | 35,505.17            | 0.006 | –                   |
|                 |                | BnRFL15        | C8   | 42,022,999     | 5           | 215                 | 5.29 | 23,581.51            | 0.006 | M                   |
| AtRFL4          | At1G12700      | BnRFL16        | C8   | 42,872,137     | 14          | 604                 | 5.47 | 67,048.93            | 0.002 | M                   |
|                 |                | BnRFL17        | A9   | 42,773,730     | 14          | 603                 | 6.57 | 66,969.91            | –0.018| M                   |
| AtRFL7          | At1G62680      | BnRFL18        | Un   | 30,312         | 12          | 538                 | 8.45 | 60,152.19            | 0.016 | M                   |
|                 |                | BnRFL19        | A9   | 7,486,280      | 12          | 535                 | 8.56 | 59,967.03            | 0.012 | M                   |
|                 |                | BnRFL20        | A9   | 9,401,981      | 12          | 523                 | 8.33 | 58,558.11            | –0.002| M                   |
|                 |                | BnRFL21        | A9   | 9,446,842      | 11          | 440                 | 4.81 | 48,891.56            | 0.127 | –                   |
| AtRFL8          | At1G62720      | BnRFL22        | A9   | 7,519,673      | 13          | 485                 | 8.39 | 55,092.25            | –0.012| M                   |
| AtRFL11         | At1G62930      | BnRFL23        | C1   | 45,473,259     | 15          | 487                 | 6.36 | 55,189.36            | –0.013| S                   |
|                 |                | BnRFL24        | A9   | 9,204,272      | 4           | 178                 | 5.29 | 20,006.32            | 0.001 | –                   |
| AtRFL12         | At1G63070      | BnRFL25        | A9   | 9,203,511      | 4           | 188                 | 5.47 | 21,000.35            | –0.047| –                   |
| AtRFL13         | At1G63080      | BnRFL26        | C4   | 53,751,551     | 2           | 107                 | 6.26 | 11,658.37            | –0.133| –                   |
|                 |                | BnRFL27        | C4   | 53,761,784     | 2           | 107                 | 6.26 | 11,658.37            | –0.133| –                   |
| AtRFL17         | At1G63400      | BnRFL28        | C1   | 24,931,745     | 19          | 796                 | 6.17 | 89,097.00            | –0.119| M                   |
|                 |                | BnRFL29        | A1   | 15,383,421     | 18          | 796                 | 6.8  | 88,897.86            | –0.12 | M                   |
| AtRFL18         | At1G64100      | BnRFL30        | A9   | 9,036,004      | 17          | 681                 | 5.94 | 76,309.56            | 0.017 | M                   |
|                 |                | BnRFL31        | A9   | 9,033,346      | 17          | 681                 | 5.94 | 76,309.56            | 0.017 | M                   |
| A. thaliana gene | A. thaliana ID | B. napus gene | Chr. | Gene position | PPR number | Protein Length (AA) | pl [28] | Molecular weight (Da) [28] | GRAVY [28] | Subcellular location |
|-----------------|----------------|---------------|------|---------------|-------------|---------------------|--------|--------------------------|-----------|----------------------|
|                 |                | BnRFL32       | C1   | 45,499,113    | 17          | 681                 | 6.24   | 76,10647                 | 0.022     | M                    |
| AtRFL19         | At1G64580      | BnRFL33       | A9   | 8,665,312     | 12          | 507                 | 9.1    | 57,06397                 | −0.001    | M                    |
|                 |                | BnRFL34       | C5   | 40,329,839    | 10          | 445                 | 9.16   | 50,07412                 | 0.009     | −                    |
|                 |                | BnRFL35       | Un   | 47,833        | 12          | 491                 | 9.05   | 55,20681                 | 0.062     | M                    |
|                 |                | BnRFL36       | A9   | 9,201,254     | 12          | 491                 | 9.04   | 55,22177                 | 0.036     | M                    |
|                 |                | BnRFL37       | Un   | 21,883        | 10          | 445                 | 9.16   | 50,07412                 | 0.009     | −                    |
| AtRFL20         | At3G16710      | BnRFL38       | A1   | 29,342,555    | 12          | 506                 | 8.64   | 57,35677                 | 0.032     | M                    |
|                 |                | BnRFL39       | Un   | 67,690        | 10          | 453                 | 8.94   | 51,20474                 | 0.017     | M                    |
| AtRFL21         | At3G22470      | BnRFL40       | A8   | 25,799,162    | 13          | 631                 | 6.99   | 70,63148                 | 0.008     | M                    |
|                 |                | BnRFL41       | C8   | 42,220,823    | 15          | 671                 | 8.44   | 75,3904                  | −0.004    | M                    |
|                 |                | BnRFL42       | A9   | 42,760,108    | 15          | 652                 | 6.73   | 73,14836                 | −0.005    | M                    |
|                 |                | BnRFL43       | C1   | 14,072,487    | 11          | 423                 | 5.4    | 46,85141                 | −0.008    | −                    |
|                 |                | BnRFL44       | A8   | 25,805,814    | 5           | 229                 | 4.63   | 25,24808                 | −0.006    | −                    |
|                 |                | A4           | C8   | 17,950,054    | 4           | 165                 | 8.12   | 18,1614                  | 0.0335    | S                    |
| AtRFL22         | A14G26800      | BnRFL46       | A1   | 11,953,788    | 10          | 502                 | 8.86   | 56,4999                  | −0.017    | M                    |
|                 |                | BnRFL47       | C1   | 20,149,199    | 11          | 501                 | 8.94   | 56,44183                 | 0         | M                    |
| AtRFL23         | At5G16640      | BnRFL48       | Un   | 18,562        | 12          | 501                 | 8.46   | 56,33467                 | 0.011     | M                    |
|                 |                | BnRFL49       | A10  | 15,694,475    | 11          | 498                 | 8.73   | 56,12871                 | 0.018     | C                    |
| AtRFL24         | At5G41170      | BnRFL50       | Un   | 83,037        | 11          | 481                 | 8.76   | 54,11004                 | 0.023     | M                    |
|                 |                | BnRFL51       | A4   | 11,026,724    | 11          | 478                 | 8.84   | 53,924                   | 0.064     | M                    |
|                 |                | BnRFL52       | A9   | 27,726,100    | 16          | 811                 | 7.06   | 92,42964                 | −0.036    | M                    |
|                 |                | BnRFL53       | C7   | 55,839,984    | 14          | 821                 | 7.89   | 93,22571                 | −0.038    | M                    |
| RF-PPR592       | [11]           |               |      |               | 14          | 592                 | 7.81   | 67,34037                 | −0.071    | M                    |
| RF4             | [14]           |               |      |               | 18          | 782                 | 6.56   | 86,28274                 | −0.037    | M                    |
| RF5             | [15]           |               |      |               | 17          | 791                 | 6.1    | 87,61443                 | −0.013    | M                    |
| RF6             | [16]           |               |      |               | 15          | 798                 | 8.4    | 88,61748                 | −0.096    | M                    |
| RF6             | [12]           |               |      |               | 17          | 687                 | 4.96   | 76,50042                 | 0.022     | M                    |

| Subcellular location position | Prowler [29] | TargetP [30] |
|-----------------------------|-------------|-------------|
| M                           | M          | M          |
| C                           | C          | C          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| C                           | C          | C          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| C                           | C          | C          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |
| M                           | M          | M          |

References:

a Unplaced Scaffold
b Mitochondrion
c Secretory pathway
d Any other location
e Chloroplast
proteins contained motif 1, which contained 80 amino acids.

**Phylogenetic and Syntenic analysis**

To identify the homologs of BnRFLs in different monocot and dicot species, multiple sequence alignments were performed and sequence similarity was determined. The rice RF5 protein was used for BLAST searches against the rice and maize genomes, and Rf-PPR592 was used for BLAST searches against the radish genome. An additional 16 OsRFLs (including 4 reported restorer genes), 9 ZmRFLs, and 22 RsRFLs (including 4 known restorer genes) were identified (E-value <1e^{-100}). Phylogenetic analysis revealed that RFLs mainly formed two separate clusters, and RFLs in monocot and dicot species were clustered together, respectively (Fig. 3a). Additionally, two reported restorer genes (BnRFL6 and BnRFL13) and four candidate restorer genes (BnRFL2, BnRFL10, BnRFL11 and BnRFL42) clustered together. Six additional BnRFLs (BnRFL3, BnRFL4, BnRFL5, BnRFL8, BnRFL15 and BnRFL41) clustered together with the reported and candidate restorer genes. These 12 BnRFLs have been deeply investigated in the following study.

Next, we examined the synteny of BnRFLs with their homologs in *Arabidopsis*, *B. rapa* and *B. oleracea* (Fig. 3b). The results showed syntenic relationships between *AtRFL* genes on chromosome 1 and *RFL* genes on chromosomes BraA8, BraA9, BolC8, BnaA8, BnaA9, BnaC3, BnaC8 and BnaC9. The *AtRFL* gene on chromosome 3 showed synteny with *RFL* genes on BraA1 and BnaA1. The *AtRFL* genes on chromosome 4 showed synteny with *RFL* genes on BraA1, BolC1, BnaA1 and BnaC1, and the *AtRFL* genes on chromosome 5 showed synteny with *RFL* genes on BraA4, BnaA4, BnaC4 and BnaC9.

The Ks and Ka values indicate the evolutionary pressure on species. A Ka/Ks ratio < 1 indicates functional constraint, whereas Ka/Ks ratio > 1 indicates positive selection [32]. To explore the selection pressure on
BnRFLs, we calculated the Ka/Ks ratios (Additional file 3). All BnRFL genes showed a Ka/Ks ratio of 0.1–0.7. The Ka/Ks ratio of most of the BnRFLs was relatively low (<0.4). However, BnRFL46 and BnRFL47 showed relatively high Ka/Ks ratios (>0.6).

**Interaction analysis of AtRFL proteins**

Most of the RFL proteins belong to the P subfamily and need to interact with other proteins to perform RNA processing [15]. To predict the interacting partners of RFL proteins (no B. napus information in STRING database), an interaction network for AtRFLs were constructed based on STRING 10.5 and Cytoscape 3.6.1. Except AtRFL10, which did not have interaction information in the STRING database, 25 AtRFLs and their predicted partners are shown in Fig. 4 and Additional file 4. Interestingly, AtRFL11, AtRFL12 and AtRFL13 and three HXKs, including HXK1, HXK2 and HXK3, shared a common interacting protein, namely replication factor C2 (RFC2), a multi-subunit complex critical for high-speed ATP-dependent DNA synthesis [33]. No homologs of AtRFL25 were identified in B. napus. Approximately one-quarter of the BnRFL genes were homologous to AtRFL2 and AtRFL3. Further analysis revealed that AtRFL2 and AtRFL3 interact with AtRFL25, which showed interaction with the glycine-rich proteins, GRP7 and GR-RBP2 (Fig. 4). Moreover, AT1G48510, SURF1, COX15 and COX11 were predicted to interact with atp6−1, AT3G48810, NAD9, CCMH and most of the AtRFLs.

**Expression analysis of BnRFL genes**

Based on the results of phylogenetic analysis, 12 BnRFL genes, which were mentioned in the phylogenetic analysis, were selected for tissue-specific expression analysis in the sterile line Shaan2A, the maintainer line Shaan2B and the restorer line KC01 by qRT-PCR. Although BnRFL10 and BnRFL11 were located on different
Fig. 3 Phylogenetic and syntenic analysis of RFL genes. 

a Phylogenetic analysis of RFLs in rapeseed (*Brassica napus*), Arabidopsis thaliana, rice (*Oryza sativa*), maize (*Zea mays*) and radish (*Raphanus sativus*). Monocot RFLs are indicated in light blue. Dicot RFLs are indicated in light purple. RFLs in the cluster of known RF proteins are indicated in pink. Red solid circles, known RF proteins; red diamonds, candidate RF proteins in a previous study. Yellow, purple, light blue and dark blue circles and green triangles represent the RFL genes in *R. sativus*, *B. napus*, *O. sativa*, *Z. mays* and *A. thaliana*, respectively. Numbers at nodes (range 0–100) indicate the reliability of the corresponding branch; higher the number, higher the reliability of the branch.

b Synteny analysis of the RFL genes in *A. thaliana*, *B. napus*, *B. oleracea* and *B. rapa*. At Chr1–5, *A. thaliana* chromosomes 1–5; Bra A1-A10, *B. rapa* chromosomes 1–10; Bol C1-C9, *B. oleracea* chromosomes 1–9; Bna A1-A10, *B. napus* chromosomes A1-A10; Bna C1-C9, *B. napus* chromosomes C1-C9.

Fig. 4 Interaction analysis of AtRFLs. Purple nodes represent AtRFLs. Proteins interacting with AtRFLs, shown in other colors, were searched in the STRING database.
chromosomes, the coding sequence (CDSs) of these genes were highly similar (identity = 1926/1947; 99%), and it was difficult to distinguish them by qRT-PCR. Therefore, we finally analyzed the expression of 11 BnRFL genes. In the restorer line, the expression of BnRFL6, BnRFL13 and BnRFL42 was lower in leaves than in the perianth (Fig. 5a, Additional file 5). The majority of the selected BnRFLs showed higher expression level in MA when compared with leaves, except for BnRFL2, BnRFL3 and BnRFL4. However, the expression of the BnRFL genes, except BnRFL41, was lower in the gynoecium and LA when compared with leaves. Compared with Shaan2A tissues, the expression of 11 BnRFL genes was higher in KC01 tissues, especially in MA (Fig. 5b, Additional file 5). What’s more, the expression of these genes in Shaan2B LA was higher than those in Shaan2A (Fig. 5b, Additional file 5). However, the expression of most of these BnRFLs was lower in the gynoecium of the restorer line than in that of Shaan2A.

To compare the expression level of all 53 BnRFL genes in Shaan2A vs. KC01, three biological replicates of RNA-seq were performed using RNA isolated from young buds (YB, < 1 mm, representing pre-meiosis) and small anthers (SA, sampled from buds 1–2 mm in length (representing tetrad stage to microspore release stage). A total of 320,892,232 raw sequence reads were generated, with approximately 50 million raw reads representing each tissue sample (SRA number: PRJNA511929). Additionally, to conduct comparative transcriptome analysis of the three lines in the Shaan2A CMS system, raw transcriptome reads representing the same stages of Shaan2A and Shaan2B were downloaded (SRA number: PRJNA502996). RNA-seq data analysis of Shaan2A and KC01 revealed only nine BnRFLs exhibited differential expression (Fig. 5c). These results provide important clues for analyzing the candidate restorer genes in the Shaan2A CMS system.

Transcriptomic analysis between Shaan2A and KC01
To investigate the differences between Shaan2A and KC01, possibly caused by the male sterile genes and restorer genes, we also identified the differentially expressed genes (DEGs) between Shaan2B and KC01, as Shaan2A and Shaan2B share the same nuclear genetic background. Thus, common DEGs identified based on Shaan2A vs. KC01 comparison and Shaan2B vs. KC01 comparison would represent the DEGs identified between different genetic backgrounds, i.e., Shaan2A (or Shaan2B) and KC01. A total of 2980 and 8243 DEGs were identified in YB and SA stage, respectively, based on the comparison between Shaan2A and KC01 (|log2 Ratio| > 1; Additional file 6).

A

| BnRFL | MA | LA |
|-------|----|----|
| BnRFL2 | 0.58 | -3.67 | -3.68 | -2.71 |
| BnRFL3 | 0.29 | -0.79 | -0.53 | -1.83 |
| BnRFL4 | 1.22 | -8.87 | -0.12 | -8.29 |
| BnRFL5 | 0.05 | -9.19 | 1.10 | -9.63 |
| BnRFL6 | -3.84 | -3.74 | 2.64 | -3.01 |
| BnRFL6 | 0.54 | -0.06 | 5.39 | -0.75 |
| BnRFL10 | 0.63 | -0.22 | 3.19 | -1.00 |
| BnRFL11 | -0.84 | -0.72 | 2.58 | -0.49 |
| BnRFL11 | 1.82 | -1.82 | 4.39 | -1.66 |
| BnRFL41 | 4.62 | 3.19 | 5.43 | 1.37 |
| BnRFL42 | -4.11 | -3.93 | 1.68 | -3.88 |

B

| BnRFL | MA | LA |
|-------|----|----|
| BnRFL2 | 0.98 | 1.54 |
| BnRFL3 | 1.00 | 1.23 | 2.23 | 1.00 | 0.80 | 3.34 | 1.00 | 1.65 | 0.65 | 1.00 | 5.01 | 2.53 | 1.00 | 1.26 | 0.54 |
| BnRFL4 | 1.00 | 1.26 | 2.42 | 1.00 | 1.44 | 4.23 | 1.00 | 1.01 | 0.66 | 1.00 | 1.33 | 3.43 | 1.00 | 2.85 | 1.00 |
| BnRFL5 | 1.00 | 1.24 | 5.84 | 1.00 | 0.79 | 2.11 | 1.00 | 0.69 | 0.72 | 1.00 | 1.15 | 5.03 | 1.00 | 7.71 | 7.37 |
| BnRFL6 | 1.00 | 2.28 | 4.96 | 1.00 | 1.31 | 2.09 | 1.00 | 1.26 | 0.98 | 1.00 | 0.50 | 3.38 | 1.00 | 2.18 | 1.12 |
| BnRFL9 | 1.00 | 1.60 | 0.59 | 1.00 | 0.90 | 0.46 | 1.00 | 1.08 | 1.13 | 1.00 | 0.89 | 2.06 | 1.00 | 1.43 | 0.14 |
| BnRFL10 | 1.00 | 1.03 | 2.37 | 1.00 | 0.43 | 0.99 | 1.00 | 0.71 | 0.56 | 1.00 | 3.05 | 15.35 | 1.00 | 4.79 | 2.91 |
| BnRFL11 | 1.00 | 2.51 | 5.69 | 1.00 | 1.52 | 15.48 | 1.00 | 0.71 | 0.20 | 1.00 | 3.44 | 6.60 | 1.00 | 11.86 | 6.38 |
| BnRFL41 | 1.00 | 1.86 | 6.89 | 1.00 | 0.91 | 1.72 | 1.00 | 0.66 | 0.50 | 1.00 | 1.05 | 1.55 | 1.00 | 5.64 | 1.60 |

C

| BnRFL | YB | SA |
|-------|----|----|
| BnRFL1 | -2.4 | -1.72 |
| BnRFL6 | -1.58 |
| BnRFL9 | 1.42 | 1.9 |
| BnRFL10 | -5.76 | -3.51 |
| BnRFL13 | -2.33 |
| BnRFL15 | -2.7 |
| BnRFL16 | 4.77 |
| BnRFL38 | -1.33 |
| BnRFL48 | -5.1 |

Fig. 5 Expression profiles of BnRFL genes. a Expression profiles of BnRFL genes in different tissues of KC01. Gene expression in other tissues was calculated relative to that in leaves. The log2-normalized ratios are shown. b Comparative expression analysis of BnRFL genes in different tissues of KC01, Shaan2B and Shaan2A. Gene expression levels in KC01 and Shaan2B was calculated relative to those in Shaan2A. c RNA-seq analysis of the differentially expressed BnRFLs at YB and SA stage. Gene expression levels in Shaan2A was calculated relative to those in KC01. Red indicated higher expression levels. Green represented the lower expression levels. * indicates no significant difference in expression.
Based on GO analysis, only one GO term in the molecular function category, ‘sequence-specific DNA binding’, was significant at YB stage (Fig. 6a). By contrast, at SA stage, 8243 DEGs identified between Shaan2A and KC01 were categorized in three main categories, including molecular function, cellular component and biological process, which were further classified into many functional sub-categories (Additional file 7). The top 30 sub-categories, including ‘disaccharide metabolic process’, ‘regulation of RNA biosynthetic process’, ‘regulation of RNA metabolic process’ and ‘regulation of transcription, DNA-dependent’, are shown in Fig. 6b.

To validate the RNA-seq data, the expression of 11 DEGs, potentially involved in anther development, was analyzed by qRT-PCR (Additional file 8). The results of qRT-PCR analysis for most of these DEGs at the two

---

**Fig. 6** Top 30 gene ontology (GO) sub-categories of the DEGs identified between Shaan2A and KC01. Asterisk indicates the corrected p-value < 0.05. **a** YB stage. **b** SA stage
stages were consistent with those of RNA-seq analysis, indicating that the reliability of our RNA-seq data.

**Discussion**

Many CMS systems have been used in *B. napus*, including *pol*, *nap*, *Ogu* and Shaan2A [4–7]. To date, *Rf* genes have been identified via genetic mapping only in *pol* and *nap* CMS systems [17, 18]. In the present study, 53 *BnRFL* genes were identified in the Shaan 2A CMS system. Most of these genes contained more than 10 PPR motifs, which is consistent with the previously reported restorer proteins such as *Rf-PPR592* (14 PPR motifs), *Rf4* (18 PPR motifs) and *Rfo* (17 PPR motifs) [11, 17, 24]. Moreover, most of the *BnRFLs* identified in this study were predicted to localize to mitochondria, similar to the known restorer genes [15–18]. Since the *Rf* genes function with the toxic chimeric genes in mitochondria to rescue male sterility [14–16], the mitochondrial localization of the proteins seems appropriate.More importantly, we also identified *BnRFL6* (*Rfn*) and *BnRFL13* (*Rfp*), previously confirmed as restorer genes in *nap* and *pol* CMS systems, respectively [17, 18], and four candidate restorer genes (*BnRFL2, BnRFL10, BnRFL11* and *BnRFL42*), previously identified in *B. napus* via genetic mapping [31]. Taken together, analysis of the *RFL* gene family for the identification of candidate restorer genes were viable, which would also provide a new way to analyze the restorer genes in other CMS systems as one supplementary method except for the traditional genetic mapping to locate the candidate genes.

Nearly 7500 years ago, *B. napus* originated from the hybridization of *B. rapa* and *B. oleracea* [34], and the Brassica plants experienced the extra whole genome triplication (WGT) event when compared with *Arabidopsis* [35]. The *Arabidopsis* genome contains 26 *RFL* gene family members, so considering the WGT event there should be over 78 *RFL* genes in *B. oleracea* or *B. rapa* genome, and finally generate even more *RFL* genes in *B. napus*. While only 53 *BnRFLs* were identified in the present study, which implied that nearly 50% *RFL* genes were lost after the WGT event.

Most of the *BnRFLs* were unevenly distributed on 10 of the 19 chromosomes of *B. napus*, while a few formed gene clusters on chromosomes A9 and C8, similar to the gene cluster in *Arabidopsis* (chromosome 1; Table 1, Fig. 1), rice and barley (*Hordeum vulgare*) [36, 37]. Gene clustering has also been observed in many other gene families, such as the *LEA* gene family in *B. napus* [38] and *Phyllostachys edulis* [39] and laccase gene family in *Citrus sinensis* [40]. The *LEA* gene clusters on *B. napus* chromosomes A9 and C4 probably resulted from chromosomal rearrangement during the evolution of *Brassica* species [38]. The *RFL* genes on *Arabidopsis* chromosome 1 showed synteny with the *RFL* genes on BraA9, BolC8, BnaA9 and BnaC8. Additionally, *BnRFLs* maintained a syntenic relationship with *RFL* genes in *B. rapa* and *B. oleracea*, suggesting that a conserved role of *BnRFLs* located on chromosomes A9 and C8. Moreover, *AtRFL2, AtRFL4* and *AtRFL9* were located within the gene cluster on *Arabidopsis* chromosome 1 and participated in the processing of the mitochondrial RNA [25–27]. The phylogenetic analysis revealed that the *BnRFLs* have the closer phylogenetic relationship with *AtRFLs* and *RsRFLs* (Fig. 3), and the structural analysis showed that all of the *BnRFLs* and the known restorer genes in radish share a conserved motif (Fig. 2), and all *BnRFL* genes showed a Ka/Ks ratio < 1 (Additional file 3), which indicated that there was no positive selection on the *BnRFL* genes during the evolution. What’s more, *BnRFL6* (*Rfn*) and *BnRFL13* (*Rfp*) were located within the gene cluster on chromosome A9. These data suggest that the *RFL* genes within gene clusters on chromosomes A9 and C8 represent the restorer genes in the CMS system, as these likely exhibit a conserved role in mitochondrial RNA processing.

Tandem repeats of a degenerate 35-amino-acid PPR motif are the most prominent feature of the PPR family, and all of the 53 *BnRFL* proteins showed this trait. Although 212 *RFL* genes in 13 different species [24] and 26 *RFL* genes in barley [37] have been identified previously, the conserved domain of the *RFL* proteins has not yet been analyzied. Therefore, we investigated motifs other than PPR in the *RFL* proteins, revealing 20 motifs among the 53 *BnRFLs* and a few known *RF* proteins. Interestingly, all *RFLs* contained motif 1, comprising 80 amino acids. We propose motif 1 as the conserved motif in the *RFL* protein family. This motif will serve as a reference for *RFL* family analysis in other species.

Because *RF*-related PPR proteins belong to the P subfamily and do not exhibit endonuclease activity, these proteins form functional complexes with other proteins [20]. To date, only two *RFL*-interacting partner proteins have been identified, including GRP162 and HXK6 in the rice CMS system [15, 16]. In the present study, we constructed an interaction network for *AtRFLs* (Fig. 4). Interestingly, *AtRFL11, AtRFL12* and *AtRFL13* and HXKs (HXK1, HXK2 and HXK3) shared a common interacting partner, RFC2, which was critical for high-speed DNA synthesis [22], whereas *AtRFL25* showed interaction with GRP7 and GR-RBP2. Moreover, AT1G48510, SURF1, COX11 and COX15 were predicted to interact with most of the *AtRFLs*. AT1G48510 is a surfeit locus 1 cytochrome c oxidase biogenesis protein. SURF1 is associated with cytochrome c oxidase assembly [41]. Both COX11 and COX15 are mitochondrial proteins and belong to the cytochrome c oxidase protein family [42]. COX11 likely plays a key role as a mitochondrial chaperone in the assembly of the COX complex
and regulates pollen germination and plant growth [43]. Overall, the interaction network indicates possible partner proteins of RFL proteins in Arabidopsis. These data provide important clues for the identification of interaction factors of RF proteins in other species.

Previously, it has been shown that Rf4 is constitutively expressed in different rice organs at relatively low levels [14]. Although Rf6 expression is detectable in various rice tissues, it is expressed to a higher level in the panicle than in other tissues [16]. In the pol CMS system, Rfp shows relatively high expression in flower buds and weak expression in opening flowers, leaves, stems and roots [17]. In the phylogenetic tree, two previously reported restorer genes (BnRFL6 and BnRFL13), four candidate restorer genes (BnRFL2, BnRFL10, BnRFL11 and BnRFL42) and another six BnRFLs (BnRFL3, BnRFL4, BnRFL5, BnRFL8, BnRFL15 and BnRFL41) clustered together, suggesting these genes as the more probable candidates of restorer genes in the rapeseed CMS system. Analysis of expression patterns revealed that most of these genes, except for BnRFL2, BnRFL3 and BnRFL4, were expressed to relatively higher levels in MA than in leaves in KC01. Additionally, these BnRFLs showed higher expression in KC01 tissues, especially MA, than in Shaan2A tissues. Expression profiling of BnRFL genes in Shaan2A vs. KC01 showed that BnRFL1, BnRFL6, BnRFL10, BnRFL13, BnRFL15, BnRFL38 and BnRFL49 were down-regulated in Shaan2A. However, BnRFL15 only harbored five PPR motifs, which was much lower than the number of PPR motifs in the known RF proteins. While BnRFL38 and BnRFL49 are located on chromosomes A1 and A10, respectively, BnRFL1, BnRFL5 and BnRFL8, BnRFL6, BnRFL11, BnRFL13 and BnRFL42 are located in gene clusters on chromosomes C8 and A9. Interestingly, almost all of the known rice Rf genes are located in the RFL gene cluster on chromosome 10 [36]. These data suggest BnRFL1, BnRFL5, BnRFL6, BnRFL8, BnRFL11, BnRFL13 and BnRFL42 as the more likely candidates of restorer genes in the Shaan2A CMS system. In O. sativa, RF6 with a characteristic duplication of PPR motifs in the restorer line of Honglian CMS can restore sterility, while the duplicated motifs are absent in rf6 of sterile line [16]. In the next steps, we will clone these candidate restorer genes in the restorer line and sterile line of Shaan2A CMS respectively, and compare the difference of sequences between them, for we wonder if there is the similar motif difference between these candidate restorer genes. Then we will narrow down the list of candidate genes, and conduct the transgenic work in sterile line to investigate their function.

Furthermore, DEGs identified in small anthers of Shaan2A vs. KC01 were annotated as involved in the ‘regulation of RNA biosynthetic process’, ‘regulation of RNA metabolic process’ and ‘regulation of transcription, DNA-dependent’. The RF-related PPR proteins interact with their partner proteins to bind or to edit RNA [15]. Here, the regulation of RNA biosynthetic, RNA metabolic process and transcription was different between the sterile line and restorer line, which might be caused by the sterile genes in Shaan2A and restorer genes in KC01. However, the detailed mechanism needs further investigation.

Conclusions

In CMS, the Rf nuclear genes rescue the sterile phenotype and most of the Rf genes encode pentatricopeptide repeat (PPR) proteins. In the present study, a total of 53 BnRFL genes were identified in B. napus. Most of the BnRFL genes were distributed on 10 of the 19 chromosomes, and gene clusters were identified on chromosomes A9 and C8. The interaction network analysis was performed to predict the interacting partners of RFL proteins. Tissue-specific expression and RNA-seq analyses between the restorer line KC01 and the sterile line Shaan2A indicated that BnRFL1, BnRFL5, BnRFL6, BnRFL8, BnRFL11, BnRFL13 and BnRFL42 located in gene clusters on chromosomes A9 and C8 were highly expressed in KC01, which suggest these seven BnRFL genes as strong candidates for the restorer genes in Shaan2A CMS. Our results would provide new insight into the study of Rf genes in other CMS systems.

Methods

Plant materials

The sterile line Shaan2A, maintainer line Shaan2B and restorer line KC01 of B. napus, gifted by Professor Dianrong Li at the Hybrid Rape Research Center of Shaanxi Province, were used in this study. Plants were cultivated on the experimental field of the Huazhong University of Science and Technology (Wuhan, Hubei province, China). After harvest, plant samples were immediately frozen in liquid nitrogen and stored at −80°C until needed for total RNA isolation.

Identification of the RFL gene family in B. napus and other species

The RFL genes were identified as described previously [24]. Briefly, AtRFL1–26 [24] and RF-PPR592 [11] sequences were used for BLAST searches against the genome of the rapeseed cultivar ‘ZS11’ [44]. The sequence of rice RF5 (also known as RF1a) [15, 45] was used for BLAST searches against the genome sequences of monocots (E-value <1e−100), including O. sativa (RGAP, [http://rice.plantbiology.msu.edu/]) and Z. mays [46]. The RF-PPR592 sequence was used for BLAST searches against the genome sequences of dicots (E-value <1e−100), including B. rapa [47], B. oleracea [48] and R. sativus [49].
The grand average of hydropathy (GRAVY) value, isoelectric point (pI) and molecular weight of RFL proteins were calculated using ExPASy (http://www.expasy.org/tools/) [28]. The subcellular location of RFL proteins was predicted using the Protein Prowler Subcellular Localization Predictor version 1.2 (http://bioinf.scmb.uq.edu.au:8080/prowler_webapp_1-2/) [29] and TargetP1.1 server (http://www.cbs.dtu.dk/services/TargetP/) [30].

Structural analysis of RFL genes
The exon-intron structure of BnRFL genes and a few known Rf genes were based on the alignments of the CDS with the corresponding genomic sequences, and the diagram was conducted using the Gene structure display server (GSDS, http://gsds.cbi.pku.edu.cn/) [50]. The PPR motifs in all BnRFL proteins and a few known RF proteins were analyzed using the NCBI Conserved Domain Search tool (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [51]. Conserved motifs in RFL proteins were analyzed using Multiple Expectation Maximization for Motif Elicitation (MEME, http://alternate.meme-suite.org/) [52].

Phylogenetic and Syntenic analysis of RFL genes
Multiple sequence alignment of the predicted amino acid sequences of BnRFLs, AtRFLs, RsRFLs, OsRFLs and ZmRFLs was performed using ClustalX [53]. A phylogenetic tree of these RFL proteins was constructed with MEGA 7 using the Neighbor Joining (NJ) method [54]. Analysis of synteny among BnRFL, AtRFL, BoRFL and BrRFL genes was performed using the syntenic gene tool in the Brassica database (BRAD, http://brassicadb.org/brad/) [55]. The non-synonymous to synonymous nucleotide substitution ratio (Ka/Ks) was calculated using TBtools [56].

Interaction analysis
The interaction analysis of AtRFLs was based on the STRING 10.5 database, which included the known and predicted protein–protein interactions. First, the interaction proteins of AtRFLs were searched. After deleting the repeat proteins, the interaction network was visualized using Cytoscape 3.6.1.

RNA extraction, RNA-seq and qRT-PCR analysis
Gene expression was analyzed in various tissues of the sterile line Shaan2A and restorer line KC01 including leaves, perianths, gynoecium, medium anthers (MA) and large anthers (LA). MA were harvested from buds 2–4.5 mm in length and represented the uninnuclear microspore stage, whereas LA were harvested from buds 4.5 mm in length and represented the mature pollen formation stage.

Total RNA extraction, RNA-seq analysis and qRT-PCR were conducted according to our previous protocols [57], with minor modifications. Briefly, approximately 100 mg plant samples were used for total RNA extraction using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA), according to the manufacturer’s instructions. Then, cDNA sequencing libraries were constructed using TruSeq™ RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). RNA-seq was performed on the Illumina NovaSeq 6000 platform. The raw data were filtered using the NGSQC toolkit (v2.2.3), and the clean reads were mapped to the reference genome of the rape seed cultivar ‘ZS11’. The differentially expressed genes (DEGs) were evaluated using DESeq2, with normalized fold-change ≥2 and p-value < 0.05. Gene Ontology (GO) annotation was using the Web Gene Ontology Annotation Plot (WEGO) software.

To perform qRT-PCR analysis, RNA was reverse transcribed using the TaKaRa PrimeScript™ RT Reagent Kit with gDNA Eraser, according to the manufacturer’s instructions. Actin was used as the internal reference gene [58]. The qRT-PCR experiments and transcript quantification were performed as described previously [57]. Primers used in this study are listed in Additional file 9.

Supplementary Information
Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-020-07163-z.

Additional file 1: Exon-intron structure of the BnRFL genes and known Rf genes.
Additional file 2: Distribution of PPR motifs in the identified RFL proteins.
Additional file 3: Non-synonymous (Ka) and synonymous (Ks) nucleotide substitution rates of the coding sequence of RFL genes in A. thaliana and B. napus.
Additional file 4: List of the interaction proteins.
Additional file 5: Original data of qRT-PCR.
Additional file 6: RNA-seq analysis of DEGs identified between Shaan2A and KC01 at YB and SA stage, respectively.
Additional file 7: List of GO sub-categories of the DEGs identified between Shaan2A and KC01.
Additional file 8: Validation of the expression of selected DEGs by qRT-PCR. (A) Results of qRT-PCR analysis. (B) Results of RNA-seq analysis. The numbers indicate log2 X-normalized ratios. Red indicated higher expression levels. Green represented the lower expression levels. “–” indicates no significant difference in RNA-seq data.

Additional file 9: List of primers used in this study.

Abbreviations
RFL: Restorer-of-fertility-like; CMS: Cytoplasmic male sterility; Rf: Restorer-of-fertility; PPR: Pentatricopeptide repeat; CMS: Chemical induced male sterility; GMS: Generic male sterility; HXK6: Hexokinase 6; RPF4: RNA PROCESSING FACTOR 4; TAIR: The Arabidopsis Information Resource; RFC2: Replication factor C2; GRAVY: Grand average of hydropathy; CDS: Coding sequence; GSDS: Gene structure display server; NJ: Neighbor Joining; DEG: Differentially expressed gene; GO: Gene Ontology.
Acknowledgements
Not applicable.

Authors’ contributions
LYN, DRL and MTL conceived and designed the study. LYN performed the experiments, analyzed the experimental data and wrote the manuscript. LYN, HW, YHL and KC participated in the identification, chromosome location and structural analysis of BnRFL genes. LYN, HBC, HXL and JH contributed to the phylogenetic analysis, syntenic analysis and interaction analysis. MTL reviewed the manuscript. All authors reviewed and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (31871666), the National Key Research Project of China (2016YFD0101200) and New Century Talents Support Program of the Ministry of Education of China (NCET110172).

Availability of data and materials
Raw RNA-seq data of RFL proteins was predicted using the Protein Prowler Subcellular localization genome of

References
1. Fu TD. Breeding and utilization of rapeseed hybrids. Hubei Sci Technol. 2000;167–9 (in Chinese).
2. Chen L, Liu YG. Male sterility and fertility restoration in crops. Annu Rev Plant Biol. 2014;65:579–606.
3. Horn R, Gupta KJ, Colombo N. Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion. 2014;19:198–205.
4. Fu TD. Production and research of rapeseed in the people’s republic of China. Eucarpia Cruciferaceae Newsletter. 1981;6–8.
5. Shiga T, Baba S. Cytoplasmic male sterility in rape plant, Brassica napus L. Jap J Breed. 1971;21:16–7 (in Japanese).
6. Ogura H. Studies on the male sterility in Japanese radish with special reference to the utilization of this sterility towards the practical raising of hybrid seed. Mem Fac AgricKogoshima Univ. 1968;0:9–78 (in Japanese).
7. Li DR. Report on three-lines breeding in Brassica napus. Shannan J Agricultural Sci. 1989;1:26–9 (in Chinese).
8. Fu TD. Breeding and utilization of rapeseed hybrids. Huhan: Hubei Science and Technology Press. 1995. (in Chinese).
9. Zhao HX, Li ZJ, Hu SW, Sun GL, Chang JJ, Zhang ZH. Identification of cytoplasm types in rapeseed (Brassica napus L.) accessions by a multiplex PCR assay. Theor. Appl. Genet. 2010;121:643–50.
10. Cui X, Wise RP, Schnable P. The r22 nuclear restorer gene of male-sterile T-cytoplasm maize. Science. 1996;272:1334–6.
11. Bentolilla S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci U S A. 2002;99:10887–92.
12. Brown GG, Formanová N, Jin H, Wargachuk R, Dendy C, Patil P, et al. The radish Rfo restorer gene of Ogura cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J. 2003;35:262–72.
13. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kono-Murase J, et al. Genetic characterization of a pentatricopeptide repeat gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J. 2003;34:407–15.
14. Tang H, Luo D, Zhou D, Zhang Q, Tian D, Zheng X, et al. The Rice Pentatricopeptide repeat protein RFS restores fertility in Hong-Lian cytoplasmic male sterile lines via a complex with the glycine-rich protein GPR162. Plant Cell. 2012;24:109–22.
15. Huang W, Yu C, Hu J, Wang L, Dan Z, Zhou W, et al. The Rice Pentatricopeptide repeat protein RFS restores fertility in Hong-Lian cytoplasmic male sterility lines via a complex with the glycine-rich protein GPR162. Plant Cell. 2012;24:109–22.
16. Wu H, Yu C, Hu J, Wang L, Dan Z, Zhou W, et al. Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. Proc Natl Acad Sci U S A. 2015;112:14984–9.
17. Liu Z, Wang Z, Wang H, Liu G, Gao Y, Wang J, et al. The Rice Pentatricopeptide repeat protein RFS restores fertility in Hong-Lian cytoplasmic male sterility lines via a complex with the glycine-rich protein GPR162. Plant Cell. 2012;24:109–22.
18. Huang W, Yu C, Hu J, Wang L, Dan Z, Zhou W, et al. Pentatricopeptide-repeat protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. Proc Natl Acad Sci U S A. 2015;112:14984–9.
19. Liu Z, Zhang Y, Wang X, Li K, An H, Liu J, et al. A mitochondria-targeted PPR protein restores pol cytoplasmic male sterility by reducing orf224 transcript levels in oilseed rape. Mol Plant. 2016;9:1082–4.
20. Liu Z, Dong F, Wang X, Wang T, Su R, Hong D, et al. A pentatricopeptide-repeat protein restores nap cytoplasmic male sterility in Brassica napus. J Exp Bot. 2017;68:1115–23.
21. Small ID, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organelar proteins. Trends Biochem Sci. 2000;25:46–7.
22. Schmitz-Linneweber C, Small ID. Pentatricopeptide repeat proteins: a scaffold set for organelle gene expression. Trends Plant Sci. 2008;13:663–70.
23. Lurin C. Genome-wide analysis of Arabidopsis Pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell. 2004;16:2089–103.
24. Chen G, Zou Y, Hu J, Ding Y. Genome-wide analysis of the rice PPR gene family and their expression profiles under different stress treatments. BMC Genomics. 2018;19:720.
25. Xing H, Fu X, Yang C, Tang X, Guo L, Li C, et al. Genome-wide investigation of pentatricopeptide repeat gene family in poplar and their expression analysis in response to biotic and abiotic stresses. Sci Rep. 2018;8:2817.
26. Fuji S, Bond CS, Small ID. Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. Proc Natl Acad Sci U S A. 2011;108:1723–8.
27. Fuji S, Suzuki T, Gégo P, Higashiyama T, Koizuka N, Shikanai T. The restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291 RNA in Arabidopsis. Plant J. 2016;86:104–13.
28. Höhler A, Jonietz C, Törjek O, Altmann T, Binder S, Forner J. A RESTORER OF FERTILITY-like PPR gene is required for 5′-end processing of the nad4 mRNA in mitochondria of Arabidopsis thaliana. Plant J. 2011;65:737–44.
27. Stoll K, Jonietz C, Schleicher S, des Francs-CC, Small ID, Binder S. In Arabidopsis thaliana distinct alleles encoding mitochondrial RNA PROCESSING FACTOR 4 support the generation of additional 5' termini of ccmR transcripts. Plant Mol Biol. 2017;93(6):639–68.

28. Gasteiger E. ExPaSY: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31:3794–8.

29. Boden M, Hawkins J. Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics. 2005;21:2279–86.

30. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their n-terminal amino acid sequence. J Mol Biol. 2000;300:1005–16.

31. Gabonelle A, Brown GG. Comparative genomic analysis of the compound Brassica napus RF locus. BMC Genomics. 2016;17:834.

32. Anton N, Katenya DM, Li WH. The KA/KS ratio test for assessing the protein-coding potential of genomic regions: An empirical and simulation study. Methods. 2001;12:198–202.

33. Chen Y, Qian J, You L, Zhang X, Jiao J, Liu Y, et al. Subunit interaction differences between the replication factor C complexes in Arabidopsis and rice. Front Plant Sci. 2018;9:777.

34. Chahloub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014;345:950–3.

35. Cheng F, Wu J, Wang X. Genome triplication drove the diversification of Brassica plants. Hortic Res. 2014;1:14024.

36. Melonek J, Stone JD, Small ID. Evolutionary plasticity of restorer-of-fertility-like proteins in rice. Sci Rep. 2016;6:3552.

37. Melonek J, Zhou R, Bayer PE, Edwards D, Stein N, Small D. High intraspecific diversity of restorer-of-fertility-like genes in barley. Plant J. 2018;97:281–95.

38. Liang Y, Xiong Z, Zheng J, Xu D, Zhu Z, Xiang J, et al. Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in Brassica napus. Sci Rep. 2016;6:2426.

39. Huang Z, Zhong X, He J, Jin S, Guo H, Yu X, et al. Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding LEA (late embryogenesis abundant) proteins in Moso bamboo (Phyllostachys edulis). PLoS One. 2016;11:e165953.

40. Xu Z, Zhou Y, Wang B, Ding L, Wang Y, Luo L, et al. Genome-wide identification and characterization of laccase gene family in Citrus sinensis. Gene. 2019;689:114–23.

41. Poyau A, Buchet K, Godinot C. Sequence conservation from human to prokaryotes of Surf1, a protein involved in cytochrome c oxidase assembly, deficient in Leigh syndrome. FEBS Lett. 1999;462:416–20.

42. Vishwakarma A, Tetali SD, Selinski J, Scheibe R, Padmasee K. Importance of the alternative oxidative (AOX) pathway in regulating cellular redox and ROS homeostasis to optimize photosynthesis during restriction of the cytochrome oxidase pathway in Arabidopsis thaliana. Ann Bot. 2015;116:555–69.

43. Radin I, Mansilla N, Rödel G, Steinbreuner I. The Arabidopsis COX11 homolog is essential for cytochrome c oxidase activity. Front Plant Sci. 2018;9:11091.

44. Sun F, Fan G, Hu Q, Zhou Y, Guan M, Tong C, et al. The high-quality genome of Brassica napus cultivar ‘ZS1’ reveals the introgression history in semi-winter morphotype. Plant J. 2017;92:452–68.

45. Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, et al. 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. 2006;18:676–87.

46. Jiao Y, Peluso P, Shi J, Liang T, Sitzker MC, Wang B, et al. Improved maize reference genome with single-molecule technologies. Nature. 2017;546:524–7.

47. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet. 2011;43:1035–9.

48. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 2014;5:3930.

49. Kitashiba H, Li F, Hirakawa H, Kawanabe T, Zou Z, Hasegawa Y, et al. Draft sequences of the radish (Raphanus sativus L.) genome. DNA Res. 2014;21:481–90.

50. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7.

51. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017;45:D200–3.

52. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202–8.

53. Higgins DG, Sharp PM. CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene. 1988;73:237–44.

54. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.

55. Cheng F, Wu J, Fang L, Wang X. Systenic gene analysis between Brassica rapa and other Brassicaceae species. Front Plant Sci. 2012;3:198.

56. Chen CJ, Chen H, Zhang Y, Thomas HR, Frank MH, Ye Y, et al. TBTools, a toolkit for biologists integrating various HTS-data handling tools with a user-friendly interface. Mol Biol. 2020;13:1194–202.

57. Ning LY, Lin ZW, Gu JW, Gan L, Li Y, Wang H, et al. The initial deficiency of protein processing and flavonoids biosynthesis were the main mechanisms for the male sterility induced by SX-1 in Brassica napus. BMC Genomics. 2018;19:896.

58. An H, Yang Z, Yi B, Wen J, Shen J, Tu J, et al. Comparative transcript profiling of the fertile and sterile flower buds of pol CMS in B. napus. BMC Genomics. 2014;15:258.

59. Ning LY, Wang H, Li DR, Lin ZW, Li YH, Zhao WG, et al. Transcriptomic and proteomic analysis of Shaan2A24 cytoplasmic male sterility and its maintainer line in Brassica napus. Front Plant Sci. 2019;10:252.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.