

Pentatricopeptide repeat proteins in maize

Kaifa Wei · Ping Han

Received: 7 August 2016 / Accepted: 14 November 2016 / Published online: 20 December 2016
© Springer Science+Business Media Dordrecht 2016

Abstract Nucleus-encoded pentatricopeptide repeat proteins (PPRs) are a superfamily of ubiquitous RNA-binding factors involved in organellar gene expression. To date, a large number of *PPR* genes have been identified and some of them have been functionally analyzed in higher plants, while the family in maize [*Zea mays* (Zm)] has long been unknown. In this study, a genome-scale analysis was conducted to identify all *ZmPPR* gene members, and exon/intron structure, chromosomal mapping and gene duplication events were analyzed. Then, homology modeling was used to explore structure-function relationships. Finally, publically transcriptome data and real-time PCR were used for studying gene expression. A total of 521 *PPR*-encoding genes were predicted to occur in the maize genome and unevenly distributed on ten chromosomes. More than 66% *PPR* genes rarely contain introns and similar exon/intron structural patterns were observed in the same classes/subclasses. Three-dimensional (3D) structural modeling of *ZmPPR479* provided insights into the mechanism of RNA recognition. The transcript abundance of 471 genes detected in 60 different developmental stages

revealed that some members exhibit tissue-specific expression. Under salt stress, five significantly differentially expressed genes were found in crown root (CR) and ten detected in primary root (PR) and seminal root (SR), respectively. For drought stress, the expressions of seven *ZmPPRs* showed obvious upregulation in leaf, whereas ten genes were significantly downregulated. Additionally, the expression of 25 and 145 genes were evidently increased and decreased in cob, respectively. Furthermore, the expression patterns of 18 *ZmPPRs* were confirmed using quantitative real-time PCR.

Keywords Maize · PPR · CMS/Rf · Expression profile · Development · Stress response

Introduction

The pentatricopeptide repeat (PPR) protein family, one of the largest and most enigmatic protein families in higher plants, was discovered by Small and Peeters during a bioinformatic screening of the Arabidopsis proteins targeted to mitochondria or plastid (Small and Peeters 2000). The PPR protein family is characterized by the signature motif of a degenerate 35-amino-acid repeat arranged in tandem arrays of 2–27 repeats per peptide and subdivided into two main classes: P and PLS, each accounting for about half of the family. Structurally, PPR motifs are highly similar to another repeat motif known as the tetratricopeptide repeat (TPR), a 34-amino acid repeating motif, which is generally found in eukaryotic proteins, especially yeast and *Drosophila* (Saha et al.

These authors have contributed equally to this work

Electronic supplementary material The online version of this article (doi:10.1007/s11032-016-0596-2) contains supplementary material, which is available to authorized users.

K. Wei (✉) · P. Han
School of Biological Sciences and Biotechnology, Minnan Normal University, 36 Xian-Qian-Zhi Street, Zhangzhou, Fujian 363000, China
e-mail: kaifa-wei@163.com

2007). PPR proteins are specific RNA-binding proteins involved in the RNA metabolism of organellar genes, whereas TPR proteins mediates protein-protein interactions and the assembly of multiprotein complexes within the cell (Desloire et al. 2003).

PPR proteins are mostly targeted to mitochondria or chloroplast and function extensively in RNA processing, fertility restoration in CMS plants, embryogenesis, and plant development. Previous studies have identified a large number of *PPR* genes, while only a few have been functionally characterized in detail (Wang et al. 2006). The first PPR protein to be functionally described was the *Saccharomyces cerevisiae* mitochondrial protein Pet309, which plays an important role in the translation of the mitochondrial *COX1* gene (Manthey and McEwen 1995). In a landmark study, a maize nuclear gene *crp1* (GRMZM2G083950), one of the earliest evidences of PPR proteins regulating messenger RNA (mRNA) metabolism in chloroplast, is required for the processing and accumulation of *petB* and *petD* mRNAs from a polycistronic precursor and translation of *petA* mRNA (Saha et al. 2007). *crp4* (At2g45350), a *PPR* gene in Arabidopsis, was the first gene found in plants to be directly implicated in RNA editing event in chloroplast (Kotera et al. 2005). However, the Mitochondria RNA Editing Factor 1 (*MEF1*) identified in 2009 was the first transfactor for RNA editing in mitochondria (Takenaka 2010). Some *PPR* genes function in fertility restoration in cytoplasmic male sterility (CMS) plants. A restorer-of-fertility-like (RFL) *PPR* gene, *RFL9* from Arabidopsis, directs ribonucleolytic processing within the coding sequences of *rps3-rpl16* and *orf240a* mitochondrial transcripts in the Col-0 accession (Arnal et al. 2014). In rice, two mitochondria-targeting proteins, RF1A and RF1B, can restore male fertility of BT-type CMS plants by blocking ORF79 production through endonucleolytic cleavage (RF1A) or degradation (RF1B) of dicistronic transcript *B-atp6/orf79* (Wang et al. 2006). In Petunia, *Rf-PPR592* gene, which encodes a 592-aa protein and is able to restore male fertility to CMS plant, was the first restoration of fertility (Rf) genes encoding mitochondria-targeting PPR proteins (Bentolila et al. 2002). PPR mutants in plants may display embryonic lethality or other spectacular phenotypes (Mootha et al. 2003). Arabidopsis *MEF7* involves in the mitochondria RNA editing, whereas loss-of-function mutations would affect plant growth, disable mitochondrial functions, and lead to embryo lethality (Zehrmann et al. 2012). As another example, *RIP1* is required for chloroplast and

mitochondrial RNA editing, whose mutants display a dwarf phenotype in Arabidopsis (Bentolila et al. 2012). There is still a great deal of work to be done to determine the functions of the other PPR proteins in plant development. Besides, PPR proteins targeted to mitochondrial or chloroplast act as diverse and important roles in responses to abiotic stresses. In Arabidopsis, six mitochondria-targeting PPR proteins (PPR40, ABO5, AHG11, SLG1, PGN, and SLO2) have been reported to regulate ABA signaling and salt or drought stress responses. A cytosol-nucleus dual-localized PPR protein SOAR1 in Arabidopsis is probably useful for improvement of crops through transgenic manipulation to enhance crop productivity in stressful conditions (Jiang et al. 2015). SVR7 (SUPPRESSOR OF VARIATION 7, AT4G16390), a homolog of ATP4 (ZmPPR017), is involved in photosynthesis and oxidative stress tolerance in Arabidopsis (Lv et al. 2014). WSL (white stripe leaf), a PPR protein that contains 14 PPR motifs and belongs to the PLS class, is required for the splicing of chloroplast transcript *rps12* in rice. However, the mutant *wsl* shows white stripes on the leaves as well as altered responses to abiotic stresses (Tan et al. 2014). Interestingly, no nucleus- or cytosol-localized *PPR* protein and almost no PPR genes in maize has been reported to regulate plant responses to abiotic stresses.

PPR genes have been reported in all sequenced eukaryotic genomes, but only few numbers are found in both animal and fungal. For example, only 28, 5, 2, and 6 *PPR* genes are recorded in the parasitic protozoan *Trypanosoma brucei*, yeast, drosophila, and human, respectively. More than 100 *PPR* genes have been found in *Physcomitrella*. The size of the gene family is greatly expanded in higher plants with 466 members in Arabidopsis and 477 in rice during the evolution of the land plants. Maize is a globally important cereal crop and has become an important model organism for the study of genetics, evolution, and other basic biological research, whereas the whole complement of PPR in maize has long been unknown. The availability of maize genome sequences provided an excellent opportunity for genome-wide analysis of PPR protein family. In this study, our analyses mainly focused on the identification, exon/intron structure, duplication events, and expression patterns in maize growth, development, and abiotic and biotic stresses of each member of the ZmPPR family. The results of this work try to provide a useful reference for further studies to better understand the potential functions of PPR proteins.

Materials and methods

PPR protein sequence retrieval and subcellular localization prediction

Maize genome sequence (ZmB73_5b_FGS_genes.fasta.gz) and the proteome sequence (ZmB73_5b_FGS_translations.fasta.gz) were downloaded from the Maize Genome Sequence Project (<http://ftp.maizesequence.org/release-5b/filtered-set/>). The Hidden Markov (HMM) profile of PPR domain (PF01535) downloaded from the Pfam data base (<http://pfam.sanger.ac.uk/>) was exploited to identify the PPR genes from the maize genome by HMMER 3.0 software (Finn et al. 2011) with a permissive E-value cut-off of 1. Moreover, sequences of *Arabidopsis thaliana* (dicotyledon) and *Oryza sativa* (monocotyledon) PPR proteins were used as queries to search against the maize proteome database. After that, the redundant sequences were then manually removed and all of the candidate sequences were further confirmed using Prosite (<http://prosite.expasy.org/>) and SMART (<http://smart.embl-heidelberg.de/>) for detecting the PPR protein domain. In addition, maize databases were used in searches for PPR homologs both in *japonica* rice and Arabidopsis proteome. The program of ExPaSy (http://web.expasy.org/compute_pi/) was used to predict the isoelectric point (pI) values and molecular weights (MWs) of the proteins. Finally, Predotar version 1.03 (<https://urgi.versailles.inra.fr/predotar/predotar.html>), TargetP version 1.1 (<http://www.cbs.dtu.dk/services/TargetP/>), MITOPROT (<https://ihg.gsf.de/ihg/mitoprot.html>), and PCLR Chloroplast Localization Prediction (<http://www.andrewschein.com/cgi-bin/pclr/pclr.cgi>) were used for prediction of organelle targeting from ZmPPR protein sequences. Multiple subcellular localization prediction programs were performed in this study because each uses different prediction methods that could cause targeting prediction variation. A Venn diagram was also created by Calculate and draw custom Venn diagrams (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to display the results of subcellular localization predictions.

Gene structure analysis of *ZmPPR* genes

To investigate the exon/intron structures of the PPR genes from maize, the Gene Structure Display Server

software (<http://gsds.cbi.pku.edu.cn>) was used to draw gene structure schematic diagrams by comparing the coding sequences (CDS) of maize PPR genes with their corresponding genomic sequences. For a better visualization and comparison, the 5' untranslated region (UTR) sequences were removed beforehand.

Chromosomal localization and gene duplication of *ZmPPR* genes

The chromosomal location data of the PPR-encoding genes were retrieved from the Maize Genome Sequence Project. Then, *ZmPPR* genes were directly mapped on maize chromosomes by using MapDraw 2.2. In this study, tandem repeats were defined as previously reported (Wei and Pan 2014). In addition, the CoGe SynMap program (<http://genomevolution.org/CoGe/SynMap.pl>) was applied to detect segmentally duplicated regions. The positions of syntenic regions between rice and maize were collected from the Synteny Mapping and Analysis Program (SyMAP) v4.0 (Soderlund et al. 2011). The segmental duplications and syntenic regions were finally visualized using Circos 0.67 (<http://circos.ca>). Moreover, synonymous (Ks) and non-synonymous substitution (Ka) rates of duplicated genes were calculated by DnaSP5.0 (Librado and Rozas 2009) according to methods used in previous studies. To deduce the selection mode and estimate the times of duplicated *ZmPPR* genes, the ratios (Ka/Ks, or ω) of non-synonymous to synonymous nucleotide substitutions for all duplicated pairs were analyzed. Generally, a Ka/Ks ratio >1 indicated accelerated evolution with positive selection; a ratio = 1 signified neutral selection that selection cannot be inferred; while a ratio <1 stood for negative selection or purifying selection against changing protein sequence. The duplication time (million years ago, Mya) was calculated by the eq. $T = Ks/2\lambda \times 10^{-6}$ Mya ($\lambda = 6.5 \times 10^{-9}$) (Yang et al. 2008).

ZmPPR479 protein homology modeling

The program of Swiss-model was performed to model the three-dimensional (3D) structure of *ZmPPR479* selected as a representative of the whole family through homology modeling. The critical amino acid sites were labeled on the structure. The crystal structure of the synthetic consensus PPR protein is available from the Protein Data Bank (PDB, <http://www.rcsb.org/pdb/home/home.do>) under accession number

4WSL. All the molecular graphics were generated using PyMOL (<http://www.pymol.org>).

Microarray-based expression analysis of *ZmPPR* genes

To understand the expression patterns of *ZmPPR* genes during different organs and developmental stages, the transcriptome data of genome-wide gene expression atlas of maize inbred line B73 were downloaded from the available database PLEXdb (<http://www.plexdb.org/>) with accession number ZM37. Pathogen infection and drought stress gene expression data were obtained from NCBI Gene Expression Omnibus (GEO) database with accession numbers GSE10023 and GSE16567, respectively. The Affymetrix Netaffx website (<http://www.affymetrix.com/estore/>) was used to identify probe sets for *ZmPPR* genes. The expression data were normalized using a robust multi-chip average (RMA) algorithm. Then, the expression values which were \log_2 -transformed beforehand were loaded into R and Bioconductor for further analysis (<http://www.bioconductor.org/>). Limma package was used to process raw data, and gplots package was applied to generate the heat maps.

Expression pattern investigation of *ZmPPR* genes based on RNA-seq

To observe and analyze the difference in *ZmPPR* gene expression profile along leaf developmental gradient (fifteen 1 cm segments), the RNA-seq data were downloaded from the ArrayExpress database with accession number E-GEOD-54274. The transcript abundance of each gene was estimated by fragments per kilobase of exon per million fragments mapped (FPKM) and the \log_2 -transformed FPKM values were used to draw a heat map. RNA-seq data from the ArrayExpress database under accession number E-MTAB-964 was downloaded to detect organ- and tissue-specific gene expression. In order to measure mRNA abundance of the stress-related *ZmPPR* genes, RNA-seq data of three root types [primary root (PR), seminal root (SR) and crown root (CR)] were obtained from the ArrayExpress database under the corresponding accession number E-GEOD-53995. A fold change of gene expression ≥ 2.0 ($|\log_2 \text{ ratio}| \geq 1$) and $P \leq 0.05$ were regarded as significantly differentially expressed. For drought stress, previous transcriptome profiles in drought-treated and well-

watered fertilized reproductive (ovaries) and vegetative tissue (basal leaf meristem) were used to detect drought-responsive *ZmPPR* genes (Kakumanu et al. 2012). The transcript abundance of each gene was estimated by FPKM.

Plant material and real-time PCR analysis

Plants of maize inbred line B73 were grown in experimental plots with a 1:1:1 mix of peat/vermiculite/perlite. The controlled environmental conditions (15 h light/25 °C, 9 h dark/20 °C, relative humidity of 55%) were provided as previously described (Wei et al. 2014). The seedlings were hand irrigated daily for about 2 weeks, and the nutrients were also supplied on a weekly basis according to the general purpose fertilizer. In this study, salt and drought treatments were performed as described previously (Zhang et al. 2015; Kakumanu et al. 2012). The real-time PCR was performed to confirm the RNA-seq data for selected genes under abiotic stresses. Real-time PCR was performed on an iCycler iQ5 Multicolor real-time PCR detection system (Bio-Rad) by using the Power SYBR Green PCR Master Mix (APPLIED Biosystems). Two biological replicates per sample were used and three technical replicates were performed for each biological replicate. Parameters of cycling conditions were determined as described previously (Wei and Pan 2014). The relative quantification method $2^{-\Delta\Delta C_t}$ was used to evaluate quantitative variation. For salt and drought stresses, maize *ubiquilin-1* (*UBQ1*) gene was used as internal control for normalization.

Cis-acting regulatory elements and miRNA-target prediction

With the aim of predicting *cis*-acting regulatory DNA elements (*cis*-elements) in promoter regions of *ZmPPR* genes, the PlantCARE database (Available online: <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was adopted to identify putative *cis*-elements in the 1500-bp genomic DNA sequences upstream of the initiation codon (ATG). MicroRNAs (miRNAs) play important roles in plant posttranscriptional gene regulation by targeting mRNAs for cleavage or translational repression. Target-align (Xie and Zhang 2010), a tool for plant miRNA target identification, was used to predict putative miRNA target genes based on previously parameter sets (Wei et al. 2014). A total of 321 mature miRNA sequences downloaded from miRBase database

(release 20) were reverse complemented and matched against the indexed maize transcript database (Griffiths-Jones et al. 2006).

Results and discussion

Gene family identification

As PPR protein contain 2–27 tandem repeats, the protein with orphan PPR motif was removed in this study. The results showed that a total of 521 maize genes were predicted to encode PPR proteins which far exceeds the number of genes in Arabidopsis and rice. As a matter of convenience, the *ZmPPR* genes were renamed from *ZmPPR1* to *ZmPP520* according to their exact positions on maize chromosomes 1–10 and from top to bottom (Wei et al. 2012). The exact position of gene GRMZM2G337701 was unknown, thus we named it *ZmPPR521*. We found that more than 95% proteins contained 2 to 19 repeats with 9 repeats being the most common number, and *ZmPPR063* had the maximum number of repeats in maize. ExPaSy analysis showed that *ZmPPR* proteins had large variations both in isoelectric point (pI) values (ranging from 4.59 to 10.08) and molecular weights (ranging from 11.71 to 212.66 kDa) (Table S1). According to the subtypes of PPR motifs, PPR proteins are divided into two classes: P and PLS. P-class PPR contain tandem arrays of 35-amino acid PPR motifs. PLS-class proteins contain alternating canonical P-type motifs and variant “long” (L)- and short (S)-type motifs and function mainly in RNA editing. Based on the C-terminal domain structure, the PLS class is subdivided into four major subclasses PLS, E, E+, and DYW (Fig. 1a) (Lurin et al. 2004). Of these genes, 283 *ZmPPRs* were grouped into P class, and 238 genes were classified as members of PLS class (31, 76, 49, and 82 members belong to PLS, E, E+, and DYW subclass respectively) (Table 1).

Subcellular localization of *ZmPPR* proteins

Subcellular localization, determining the environments in which protein operates, is important to elucidate protein function. In this study, more than 73% (381/521) *ZmPPR* proteins were predicted to target the mitochondrion or chloroplast using the TargetP 1.1, and the similar results were also obtained by the Predotar program. MITOPROT and PCLR Chloroplast Localization

Prediction analysis showed that 344 *ZmPPRs* are considered to be mitochondrially targeted and 282 belong to chloroplast-targeting proteins (Table S1). Interestingly, none of the *ZmPPR* proteins was predicted to target the nucleus. Up to now, only one PPR protein, Arabidopsis GRP23 (glutamine-rich protein 23), was reported to localize to the nucleus (Ding et al. 2006). The results of subcellular localizations were displayed in the Venn diagram (Fig. 1b).

CMS and fertility restoration gene *ZmPPR*

Cytoplasmic male sterility (CMS) caused by mitochondrial genes with coupled nuclear genes, a condition under which a plant fails to produce functional pollen, affect only male development and is widespread in flowering plants. Fortunately, nuclear restoration of fertility (*Rf*) genes can make CMS lines fertile. *Rf* genes in maize are dominant alleles, most of which encode PPR proteins. In maize, the CMS lines can be classified into three categories (cms-T, cms-S, and cms-C) based on their response to specific restorer genes (Bosacchi et al. 2015). However, according to the genetic model for pollen sterility, CMS lines in maize can be divided into two types: gametophytic and sporophytic. *Rf* genes in maize act sporophytically in cms-T and cms-C lines while function gametophytically in cms-S. Two restorer genes, *Rf1* and *Rf2*, are required for fertility restoration of cms-T (Laughnan and Gabay-Laughnan 1983). A single restorer gene, designated *Rf3*, is known to participate in the restoration of cms-S (Laughnan and Gabay-Laughnan 1983). The full restoration of fertility in cms-C lines is controlled by the dominant restorer gene *Rf4*. The mitochondrial chimeric gene *T-urf13*, encoding a 13-kDa protein termed URF13, is responsible for male sterility in CMS-T (Wise et al. 1987). Previous studies reported that the *Rf1* gene was mapped on chromosome 3, while the exact coordinates were not found till now (Duvick et al. 1961). The aldehyde dehydrogenase (ALDH) superfamily plays an important role in endogenous and exogenous aldehyde metabolism. The plant ALDH superfamily contains 13 distinct families, in which ALDH2 family members was initially identified as *Rf* genes (Skibbe et al. 2002). The maize *Rf* genes include *RF2A* (also called *Rf2*), *RF2B*, *RF2C*, *RF2D*, *RF2E*, and *RF2F*, also known as *ALDH2B2*, *ALDH2B5*, *ALDH2C1*, *ALDH2C2*, *ALDH2C4*, and *ALDH2C5*, respectively (Jimenez-Lopez et al. 2010). Two dominant allele genes, *Rf1* and *Rf2*, are

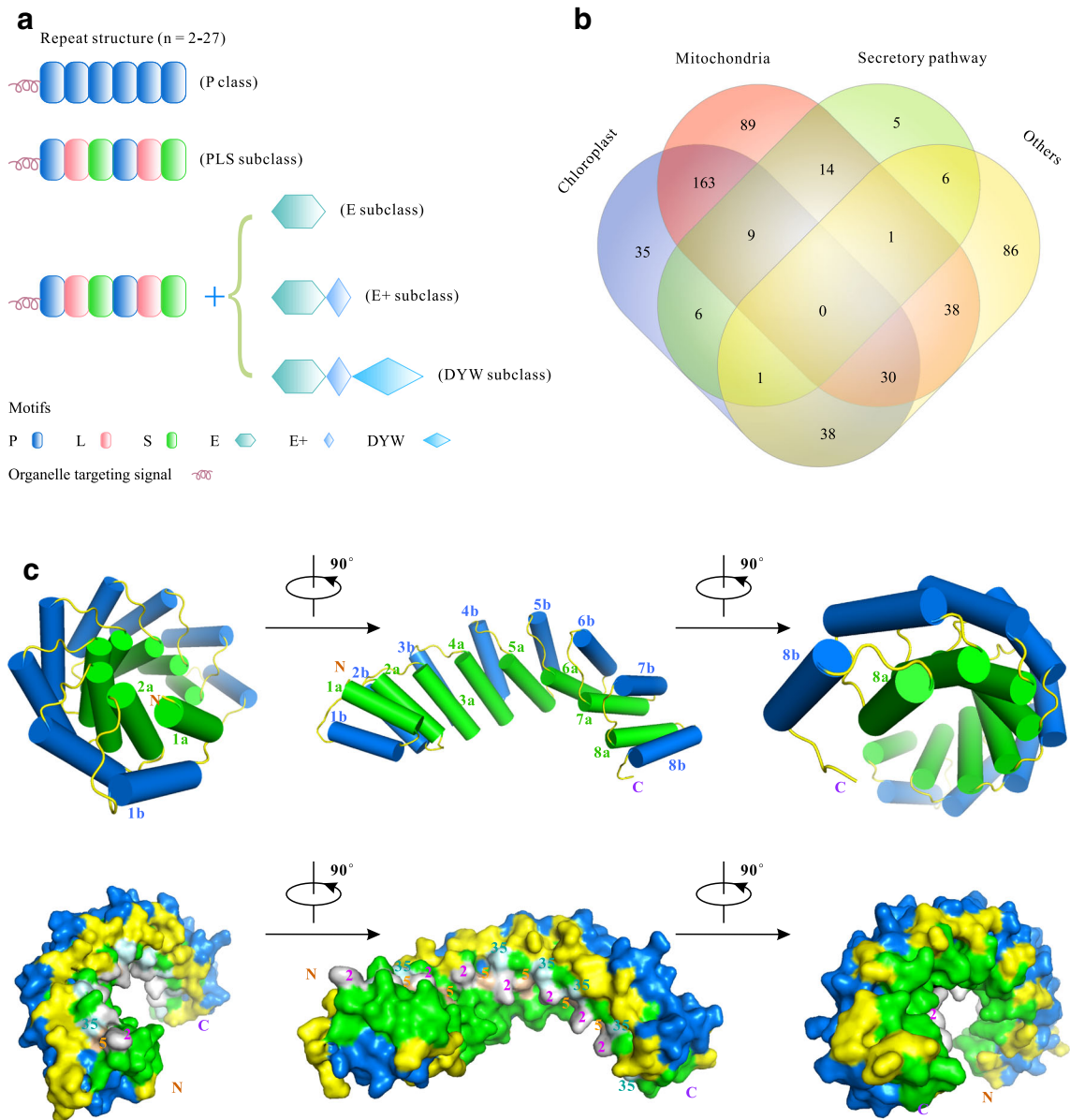


Fig. 1 Typical structures and subcellular localization of ZmPPRs. **a** Typical structures of PPR proteins in plant. **b** The Venn diagram displayed subcellular localization. **c** Structure of RNA-free ZmPPR479. The two helices (helix a and helix b) within each

repeats are colored *green* and *blue*, respectively. The predicted amino acid residues at the 2nd, 5th, and 35th position responsible for RNA recognition are labeled

complementary and necessary for restoring the URF13-mediated sterility. The *Rf3* allele was reported to cosegregates with internal processing and decreased accumulation of these transcripts, indicating an RNA editing function (Wen and Chase 1999). The expression of a chimeric mitochondrial gene region designated *orf355-orf77* contributes to CMS-S *rf3* pollen collapses (Zabala et al. 1997). All previous studies agreed that *Rf3*

gene belongs to the PPR family, and the locus has been localized to the long arm of chromosome 2. Even though *Rf3* gene has yet to be identified, five promising *Rf3* candidates (*ZmPPR142*, *ZmPPR143*, *ZmPPR145*, *ZmPPR148*, and *ZmPPR150*) were selected through different experimental approaches (Langewisch 2012). As can be seen from Table S1, all five of the candidates encode P-class PPR proteins, while only two

Table 1 Distribution of PPR genes, by chromosome, in maize

Chr.	P class	PLS class				Genes	Chr. size (Mb)	Genes per Mb
		PLS	E	E+	DYW			
1	51	6	10	5	12	84	301.48	0.28
2	50	4	5	3	10	72	237.92	0.30
3	31	2	10	8	9	60	232.25	0.26
4	24	5	10	3	5	47	242.06	0.19
5	33	4	12	9	10	68	217.96	0.31
6	11	3	9	3	14	40	169.41	0.24
7	29	2	8	3	6	48	176.83	0.27
8	28	0	3	2	5	38	175.38	0.22
9	12	3	4	4	6	29	157.04	0.18
10	13	2	5	9	5	34	149.63	0.23
Unknown	1	0	0	0	0	1	–	–
Total	283	31	76	49	82	521	2059.94	0.25

(*ZmPPR142* and *ZmPPR150*) were predicted to be targeted mitochondrion. The *Rf4* gene in maize has been identified as a basic helix-loop-helix (bHLH) transcription factor, and further mapping analysis revealed that *ZmbHLH165* (GRMZM2G021276) corresponds to the *Rf4* gene (Dow AgroSciences LLC, patent number: WO 2012/047,595 A2). In maize, the *Rf4* gene does not affect the steady-state level of *atp6-C* mRNA, indicating that restoration probability act at the protein level. The CMS phenotype in flowering plants can be reversed by the action of nuclear-encoded Rf proteins that reduce the expression of the aberrant mitochondrial gene using posttranscriptional mechanisms. In addition, CMS was not only an ideal model system to study the interaction between mitochondrial and nuclear genomes but also a useful genetic tool for breeding to exploit hybrid vigor in crops (Chen and Liu 2014). Although the *PPR* gene family is involved in the restoration of male fertility, the functions of most *ZmPPR* genes are still obscure. Therefore, exploring *PPR* gene functions will contribute to understanding another development, exploring CMS mechanism and improving molecular breeding in maize.

Exon/intron organization

A detailed illustration of the exon/intron structure is shown in Fig. S2 and different colors represent different subclasses. Gene structure analysis revealed that the number of introns of *ZmPPR* genes varied from 0 to 20, which is similar to intron number (in the range 0 to

21) in *japonica* rice (Fig. S3). About 66.8% (348/521) of *PPR* genes contain a single exon, and only 33% (173/521) contain more than one intron in maize. In addition, 60 and 79% of *ZmPPR* genes lack introns in P and E+ subclass, respectively. On average, a *PPR* gene contains about 0.6 introns per kb protein coding region, while the average number of introns per gene is about 1.13. Two genes, *ZmPPR215* and *ZmPPR267*, have the maximum number of introns in maize *PPR* gene family. Although *ZmPPR044* contains only five introns, the number of introns per kb protein coding region reaches the maximum (about 15.29 per kb). To further understand the structure diversity, we also analyzed the intron position and phase. We conclude that the phases of the splicing sites differ from each other, and the intron positions tend to be non-conserved.

Chromosomal locations and gene duplication

A total of 520 *ZmPPR* genes were mapped on ten chromosomes, while only one gene (*ZmPPR521*) was located on unmapped scaffolds. As depicted in Fig. S4, the *ZmPPR* gene density per chromosome is uneven. Most *ZmPPR* genes are localized toward the chromosome ends, and only four genes (*ZmPPR038*, *ZmPPR172*, *ZmPPR336*, and *ZmPPR432*) are found near centromeres. From Table 1, we found that the largest maize chromosome 1 has the most number of *PPR* genes and a higher ratio of *PPR* genes per Mb than do the other chromosomes. Chromosome 5 has

the highest density (0.31 genes per Mb) of *ZmPPR* genes with the number of 68, while chromosome 9 has the lowest density (0.18 genes per Mb). Compared with the *indica* and *japonica*, the average density of *PPR* genes in maize has a lower ratio per Mb. Moreover, each subclass of *PPR* genes was not randomly distributed over the chromosomes, such as the chromosome 8 and the short arm of chromosome 3 lacking PLS subclass genes, the short arm of chromosome 2 in the absence of E+ subclass, the short arm of chromosome 4 without PLS and E+ subclass, and the chromosome 6 lack of P class and PLS subclass on the short arm. Gene duplication, which usually occurs through segmental duplication, tandem duplication and retrotransposition, plays an extremely important role in gene family expansion and protein functional diversification. To investigate potential gene duplications in maize, the occurrences of tandem duplication and large-scale segmental duplication during the evolution of the *PPR* family was analyzed in this study. As a consequence, we identified 37 clusters of tandem repeat (including 79 *ZmPPR* genes) with the criteria that tandem repeat genes in one cluster are separated by no more than ten non-*PPR* genes. As depicted in Table S2, cluster 12, which was the largest cluster, consisted of four genes belonging to the P class. Three clusters (clusters 13, 25, and 26), each containing three genes, which were mapped on chromosome 2, 4, and 5, respectively, were members of the P class. We found that 25 of 37 clusters were from the same subclass; the rest (clusters 2, 5, 7, 16, 18, 20, 24, 27, 28, 30, 31, and 36) were from different classes/subclasses. For instance, the tandem repeat gene pair of *ZmPPR078/ZmPPR079* was grouped into PLS and E subclasses, respectively. It has been speculated that tandem gene duplication plays an indispensable role in the creation of major evolutionary novelty. In this study, a total of ten segmental duplication events were detected using the program CoGe SynMap. Within identified duplication events, nine of ten pairs were from the same class, while only one of the sister pair (*ZmPPR025/ZmPPR481*) belonged to different classes. The results revealed that both tandem and segmental duplication contributed to the expansion of the *PPR* gene family in the maize genome. However, the distribution of tandem duplication of *PPR* gene is not evident in *Arabidopsis* (Geddy and Brown 2007). The ω is widely used as an indicator of selective pressure acting on a protein-encoding gene (Wei et al. 2013).

The ratios of 48 tandem duplicated pairs varied from 0.23 to 1.26 with an average of 0.67 and those of ten segmental duplicated pairs ranged from 0.21 to 0.70 with a mean of 0.41. The Ka/Ks ratios for 9 segmental duplications and 11 tandem duplicated pairs were <0.5, suggesting strong purifying selection. However, six pairs of tandem duplications (including *ZmPPR005/ZmPPR006*, *ZmPPR067/ZmPPR068*, *ZmPPR107/ZmPPR108*, *ZmPPR110/ZmPPR111*, *ZmPPR149/ZmPPR151*, and *ZmPPR211/ZmPPR212*) are under positive selection, as their Ka/Ks ratios were >1. It is worth noting that the ratios of Ka and Ks for *ZmPPR104/ZmPPR105* and *ZmPPR136/ZmPPR138* were equal to zero, suggesting a higher degree of similarity within the duplicated pairs.

Analysis of 3D structure of ZmPPR479 protein

In plants, *PPR*, sequence-specific RNA-binding proteins, are involved in RNA editing, mRNA stabilization, and splicing of many genes in chloroplasts and mitochondria. The result of homology model showed that *ZmPPR479* share a high similarity with the synthetic consensus *PPR* protein (cPPR, PDB: 4WSL) at the resolution of 3.70 Å (Coquille et al. 2014). As can be seen from Table S1, *ZmPPR479* contains 26 *PPR* repeats, whereas only 8 *PPR* repeats reveals a high degree of similarity with the template. Therefore, only eight repeats of *ZmPPR479* could be built, and the repeats were clearly visible in Fig. S1. The 35 amino acids in each *PPR* motif form two antiparallel α -helices, each containing four helical turns and interacting to produce a helix-turn-helix motif (Yin et al. 2013). The two helices within each repeats, designated helix a and helix b, are connected by a short turn of two amino acids (Yin et al. 2013). As depicted in Fig. 1c, helices a and b of each repeat constitute the inner and outer layers of the superhelical assembly, respectively. The series of helix-turn-helix motifs are stacked together to form a superhelix with a RNA binding groove. Previous experiments indicated that amino acid residues of P motifs within cPPR at positions 2, 5, and 35 are responsible for sequence-specific recognition of RNA bases in the groove (Coquille et al. 2014). Therefore, we speculate that these amino acid residues within *ZmPPR479* have similar functions in organelles. To understand the roles of the protein in regulating mRNA stability and translation, more experimental evidences are needed.

Expression analysis in different organs and developmental stages

Genome-wide microarray data Zm37 provides 70,065 probe sets to profile transcription patterns across the 60 distinct tissues representing 11 major organ systems. However, 50 of 521 *ZmPPR* genes do not have corresponding probe sets in the dataset; the plausible explanations are that the version of B73 maize genome does not correspond to the microarray data or these genes are pseudogenes or are only expressed at specific developmental stages or under special conditions (Wei et al. 2012). The gene encoding E2 enzyme was selected as a reference in the expression analysis. Based on hierarchical clustering, the expression patterns of *ZmPPR* genes could be classified into four major lineages (I–IV). As presented in Fig. S5, 275 genes belong to lineage I, which account for 58.39% of the total. Lineage II and III contain 93 and 59 members, accounting for 19.75 and 12.53% of the total respectively. However, lineage IV including 44 members (excluding the E2 gene) has only 9.34% of the total. To explore the gene differential expression patterns in specific organs or developmental stages, we calculated the coefficient of variation (CV value; $CV = S/X$ mean, where S represents the standard deviation and X mean indicates the mean expression of a gene across all the tissues) of each gene in 60 tissues. In addition, a CV value of <15% indicates that the expression of those genes are in a low variability, whereas a value of >15% imply the existence of stage-specific gene expression during development. As listed in Table S3, the values of *ZmPPR* genes ranged from 1.01 to 37.44%, indicating that the levels of expression were highly variable among different developmental stages. Both lineage I and lineage II genes were expressed at a midlevel. The CV values of lineage I ranged from 3.75 to 37.44 and those of the lineage II from 2.66 to 17.71. In lineage I, the expression of *ZmPPR093* with the highest CV values were very specific to the leaf, primary root, apical meristem (SAM), anthers, and husk. Previous study has demonstrated that ATP4 corresponding to *ZmPPR017* is indispensable for the translation of chloroplast *atpB* open-reading frame (Zoschke et al. 2012). The *ZmPPR017* gene with a CV value of >15% was found to display transcript accumulation at several stages of leaf development (V1, V5, V7, V9, VT, and R2). *PPR10* (*ZmPPR341*) encodes a protein specifically binding to *ATPH* and *PSAJ* RNA oligonucleotides and also

showed leaf-specific expression (Yin et al. 2013). Arabidopsis OTP87 gene encoding an E-subclass PPR protein is required for mitochondria RNA editing of the *nad7-C24* and *atp1-C1178* sites, whose mutation will lead to small plants with growth and developmental delays (Hammani et al. 2011a). *ZmPPR392*, which is the homolog of *otp87*, showed high-abundance transcript levels mainly in whole seed, endosperm, and embryo. Two genes (*ZmPPR093* and *ZmPPR393*) belonging to P class were specifically expressed in leaf with CV values >30. Nevertheless, *ZmPPR264* in lineage I was highly expressed in anthers and *ZmPPR268* in lineage II also displayed high transcript levels in the tip of stage 2 leaf. Rice *MPR25* (LOC_Os04g51350) belonging to the E subclass is essential for RNA editing of *nad5* and preferentially expressed in leaves (Toda et al. 2012). The *ZmPPR091* genes grouped into E subclass in lineage II that shows homology to *MPR25* were typically expressed in leaves, apical meristem (SAM), and shoot tip and primary root. Besides, both *ZmPPR059* and *ZmPPR200*, which belong to the P class, showed high expression levels in primary root. Genes in lineage III exhibited a low abundance in all developmental stages with the mean of the log signal values of each gene ranging from 4.99 to 6.94 and the CV value ranging from 3.59 to 31%. There were only three genes (*ZmPPR479*, *ZmPPR095* and *ZmPPR472*) with CV values over 15% in lineage III. Arabidopsis proton gradient regulation3 (PGR3) is the homolog of *ZmPPR479* and performs great function in many aspects, such as stabilization of *photosynthetic electron transport L* (*petL*) operon RNA and translational activation of *petL* and *ndhA* (Fujii et al. 2013). Compared to lineage III, genes in lineage IV showed relatively high expression levels, with the mean of the log signal values of each gene changing from 11.59 to 14.08 and CV values ranging from 2.48 to 12%. Arabidopsis mitochondrial editing factor 8 (MEF8) and MEF8S were reported to differentially influence RNA editing in pollen and leaves (Verbitskiy et al. 2012). The *ZmPPR199* gene, which is homologous to *MEF8*, displayed high level of transcript accumulation throughout maize development. In addition, *ZmPPR420*, a homolog of Arabidopsis *ppr596* (Doniwa et al. 2010), was highly and stably expressed at nearly all of the maize organs and/or stages of development. Arabidopsis PPR596, which is the only example so far of a member of the P class involved in mitochondrial editing, plays an important role in the

maintenance of partial editing status both in leaf and flower tissues (Doniwa et al. 2010).

Differential gene expression in the developing leaves

Different transcription abundance patterns of 478 *ZmPPRs* and 80 organelle genes (36 chloroplastic and 44 mitochondrial genes) were identified from RNA-Seq dataset. E2 enzyme was used as an internal control. Obviously, the heat map (Figs. 2a and S6) can be divided into two lineages (lineage I and II), and genes in lineage I were highly expressed but relatively lowly expressed in lineage II. The CV value of each gene in the two lineages was calculated to examine the variability at transcriptional levels of expressed genes among 15 distinct sections. The higher CV values, the greater the fluctuation range in their expression levels (Table S4). *ZmPPR404* (*crp1*), whose encoding protein is required for the translation of the chloroplast *petA* and *petD* transcripts and for the processing of the *petD* mRNA and the translation of the mitochondrial *cox1* mRNA (Fisk et al. 1999), showed increased transcript abundance in *sec4*. *ZmPPR419*, which bears significant homology to Arabidopsis *ptac2* required for plastid gene expression through posttranscriptional processes (Pfalz et al. 2006), showed high expression in five sections (*sec1*, 2, 3, 4, and 6). Besides, chloroplast gene GRMZM5G834496 and mitochondrial *orf14-b* (GRMZM5G834666) were highly expressed in *sec9* and *sec1*, respectively. To further investigate the functions of *ZmPPR* genes, all of the homologs in Arabidopsis and rice studied previously were collected and listed in Table S5. Among them, the expression levels of 43 *ZmPPR* genes were reliably detected by RNA-seq data. Moreover, Pearson correlation coefficient (PCC) values were calculated between the *ZmPPRs* gene and the organelle genes to explore the *PPR* gene functions (Figs. 2b and S7). Genes pairs with PCC value of >0.7 have strong positive correlation and are marked with the same colors in Fig. 2a. Interestingly, the chloroplast gene GRMZM5G845244 exhibited strongly positive correlation with five nuclear genes (*ZmPPR193*, *ZmPPR244*, *ZmPPR270*, *ZmPPR369*, and *ZmPPR447*). A homolog of *ZmPPR244*, CLB19, is involved in editing of two distinct chloroplast transcripts (*rpoA* and *clpP*) in Arabidopsis (Chateigner-Boutin et al. 2008). DYW1 with homolog to *ZmPPR270* is required for RNA editing of the *ndhD-1* site in Arabidopsis chloroplasts

(Boussardon et al. 2014). AtECB2, which is the homolog of *ZmPPR369*, participates in the editing of *accD* RNA in chloroplast and early chloroplast biogenesis (Yu et al. 2009). BIR6, the homolog of *ZmPPR447*, is involved in efficient splicing of mitochondrial *nad7* transcripts in Arabidopsis (Koprivova et al. 2010). The *ZmPPR521-psbB* gene pair had a correlation higher than 0.8. Arabidopsis OTP82, a homolog of *ZmPPR521*, is required for RNA editing of the plastid genes (Okuda et al. 2010). However, only one pair of genes (*ZmPPR270/rbcL*) showed a significant negative correlation with PCC value of <-0.6. The relationship between *ZmPPR420* and 11 mitochondrial genes had a very strong positive correlation, suggesting that the *ZmPPR420* gene may be involved in the mitochondrial gene expression regulation. PNM1 is dual localized to mitochondria and nuclei in Arabidopsis and functions in mitochondria. In the nucleus of plants, PNM1 interacts with the nucleosome assembly protein NAP1 and binds with the transcription factor TCP8 in the promoter region, suggesting that the gene might play a role in self-regulating of gene expression in the nucleus and could thus be very essential for the coordination between mitochondria and the nucleus (Hammani et al. 2011b). There was a strong positive correlation between *ZmPPR130* that shared homology with *PNM1* and mitochondria *rps3*. PPR336 identified as part of a high-molecular-weight complex in Arabidopsis mitochondria is an unusual representative of the PPR family because of its relatively short tandem repeats and high expression level (Uyttewaal et al. 2008). *ZmPPR193*, which is homologous to *PPR336*, exhibited strong positive correlations with two mitochondria genes (*orf179* and *orf14-b*). *ZmPPR214*, whose homologous gene *otp71* is involved in the RNA editing of *ccmFN₂* in Arabidopsis mitochondria (Chateigner-Boutin et al. 2013), is strongly positively correlated with *nad7*. Five genes (*MEF18-MEF22*) are involved in Arabidopsis mitochondrial RNA editing at specific sites (*nad4-1355*, *ccb206-566*, *rps4-226*, *cox3-257*, and *nad3-149*, respectively) (Takenaka et al. 2010). Two mitochondria genes (GRMZM5G816772 and GRMZM5G897755) and *ZmPPR397* with homology to *MEF18* showed strong positive correlations. Especially, two gene pairs (*ZmPPR278/GRMZM5G875789* and *ZmPPR506/GRMZM5G841635*) had complete positive correlations with PCC values of 1, suggesting that the gene pairs have similar expression patterns in leaf development.

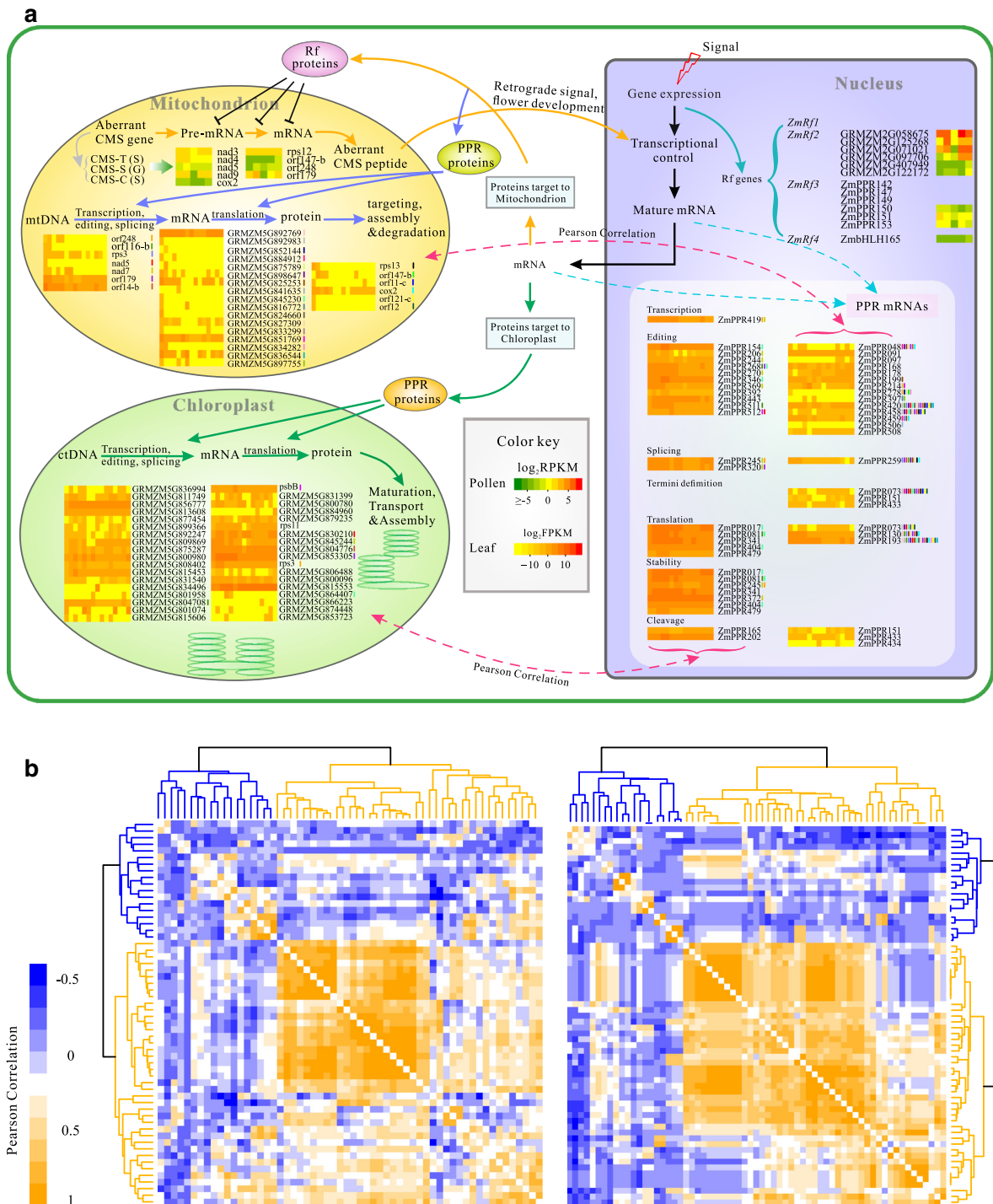


Fig. 2 Overview of ZmPPRs's major functions between the nucleus and intracellular organelles. **a** PPR proteins are involved in multiple aspects of RNA metabolism. CMS caused by mitochondrial encoded proteins can be restored by *Rf* genes. Expression profiles of nucleus and organelle genes at successive stages of leaf development and during the pollination

processes were visualized. PCC values higher than 0.7 are represented by *colored lines* besides the heat map. **b** Hierarchical clustering on pairwise correlation coefficients for *ZmPPRs* and intracellular organelle genes [chloroplast (*left*) and mitochondria (*right*)]

Differential gene expression during pollination

The critical step of sexual reproduction is pollination in flowering plants, during which pollen adhesion, germination, and tube growth require communication and coordination between male pollen and the female stigma (Xu et al. 2012). Transcript abundance profiles of *ZmPPR* genes of four tissues [immature silk (IMS), mature ovary (MO), mature pollen (MP), mature silk (MS)] in maize inbred line Zheng 58 compared to the expression levels in SL (6-day-old seedling) were used to elucidate the molecular mechanisms of pollen-stigma interactions. Because the mutation in the mitochondrial genome can lead to pollen abortion, we also analyzed maize mitochondrial genome and *Rf* gene expression profiles. Nine CMS (CMS-T, CMS-S, and CMS-C) mitochondrial genes in maize were obtained from the GenBank Data Libraries under accession nos. DQ490953, DQ490951, and DQ645536. As a result, 443 *ZmPPR* genes, 36 mitochondrial genes, and 10 *Rf* genes were expressed in at least 1 of the 5 tissues in maize (Figs. 2a and S8). A total of 219 *ZmPPR* genes expressed in MS were also detected in IMS, MO, and seedling (SL) but not in MP, suggesting that the transcriptome of MS is closer to IMS, MO, and SL than to MP. However, *ZmPPR267* expressed at high levels in MP, while 156 nuclear genes (149 *ZmPPR* genes and 7 *ZmRf* genes) and 11 mitochondrial genes showed significant differential expression from IMS to MS; we speculate that the genes might play essential roles in silk growth and development. *ZmPPR434*, which is homologous to *rpf2* required for 5' end processing of *nad9* and *cox3* mRNAs in mitochondria of Arabidopsis (Jonietz et al. 2010), was expressed at midlevel in MO whereas relatively low in the other tissues. The Arabidopsis gene *rpf3* is involved in the posttranscriptional generation of *ccmC* transcripts (Jonietz et al. 2011), and its homolog *ZmPPR151* showed a higher expression level in the IMS. *ZmPPR433* showed low expression in MP but was relatively higher in MO, and its homologue *rpf5* in Arabidopsis is necessary for the efficient 5' maturation of three different mitochondrial RNAs (*nad6*, *atp9* mRNA, and 26S *rRNA*) (Hauler et al. 2013). Obviously, two mitochondrial genes, GRMZM5G825253 and GRMZM5G834666, were highly expressed in MO but lowly in MP. The CMS gene *rps12* exhibited high and stable expression, whereas three genes (*cox2*, *orf147-b* and *orf248*) showed low expressions in five tissues. *ZmRf2* (GRMZM2G058675) was expressed at the

highest levels in MS, suggesting the genes may be related to the biological processes of cell wall function, such as lipid transport and carbohydrate and aminoglycan metabolism (Xu et al. 2012). *ZmPPR153*, whose homolog *opt1* plays an important role in 5'-end efficient processing of *nad4-228* mRNAs in Arabidopsis mitochondria (Holzle et al. 2011), showed low expression level in MP but was expressed at a midlevel in the rest of the tissues. However, *rf4* (*ZmbHLH165*) was weakly expressed in the five tissues. The result broadens our knowledge about the maize floral organogenesis and development and pollination process and provides information for the experimental validation of the gene functions involved in the pollen-silk interaction.

Expression profiles of *ZmPPR* genes under stresses

Expression of ZmPPR genes in response to fungal pathogen infection

Ustilago maydis (*U. maydis*) is a ubiquitous and pathogenic basidiomycete fungus that stunts maize growth and reduces yield. To get insights into the defense programs, we used microarrays to detail the global program of gene expression during the infection process. Samples from infected leaves were taken at 12 and 24 h postinfection, as well as 2, 4, and 8 days postinfection, and those from uninfected control plants were taken at the same time points. A total number of 64 probe sets (Table S6) for maize transcripts, which represent 64 *ZmPPR* genes, were used to detect whether these genes were responsive to fungal infection. As illustrated in Fig. S9, the heat map can be divided into three lineages. Obviously, *ZmPPRs* in lineage I maintained high gene transcription levels at 12 h after infection. Eleven members in the P class belonging to lineage I were obviously upregulated after 4.5 days of infection. Genes were expressed at a midlevel in lineage II, whereas lineage III genes had low expressions. Two genes (*ZmPPR117* and *ZmPPR462*) were downregulated at 4 days postinfection when compared with the control, and six genes (*ZmPPR263*, *ZmPPR454*, *ZmPPR202*, *ZmPPR393*, *ZmPPR123*, and *ZmPPR405*) were downregulated after 4.5 days of infection. However, only one gene *ZmPPR393* was downregulated after 8 days of infection. Our results showed that the differentially expressed *ZmPPR* genes under the fungal infection may be involved in maize immune response to disease.

Differential expression of *ZmPPR* genes under salinity stress

Salinity stress involves ionic and osmotic stress, resulting to membrane dysfunction, metabolic disorder, and oxidative stress. The maize root as a whole is sensitive to salt stress, but whether the primary root (PR), seminal root (SR), and crown root (CR) display differential growth response to the stress remains unknown. Expression patterns of 450 *ZmPPR* genes were detected and the result (Figs. 3a and S10 and Table S7) showed that three (*ZmPPR160*, *-207* and *-472*) and two genes (*ZmPPR095* and *ZmPPR503*) were significantly upregulated and downregulated in CR under salt stress, respectively. Besides, six *ZmPPR* genes (*ZmPPR059*, *ZmPPR216*, *ZmPPR222*, *ZmPPR352*, *ZmPPR404*, and *ZmPPR418*) significantly increased their expression levels in PR, whereas four genes (*ZmPPR158*, *ZmPPR217*, *ZmPPR382*, and

ZmPPR477) exhibited decreased levels of expression. However, only three *ZmPPR* genes were expressed significantly in SR. Among them, *ZmPPR202* has significant homology to Arabidopsis HCF152, which is involved in the processing chloroplast *psbB-psbT-psbH-petB-petD* transcripts (Meierhoff et al. 2003). Seven genes (*ZmPPR088*, *ZmPPR214*, *ZmPPR254*, *ZmPPR311*, *ZmPPR407*, *ZmPPR414*, and *ZmPPR444*) were downregulated in SR. The Arabidopsis homolog of *ZmPPR444*, mitochondrial RNA editing factor1 (MEF1), is required for RNA editing at multiple sites in mitochondria, such as *rps4-956*, *nad7-963*, and *nad2-1160* (Zehrmann et al. 2009). To confirm the expression patterns of *ZmPPRs* under salt stress, nine genes were selected for RT-PCR analyses; primer pairs used in this study are listed in Table S8. As shown in Fig. 3b, *ZmPPR186* and *ZmPPR207* were obviously upregulated in CR, while *ZmPPR095* and *ZmPPR515* displayed downregulation. *ZmPPR059* and

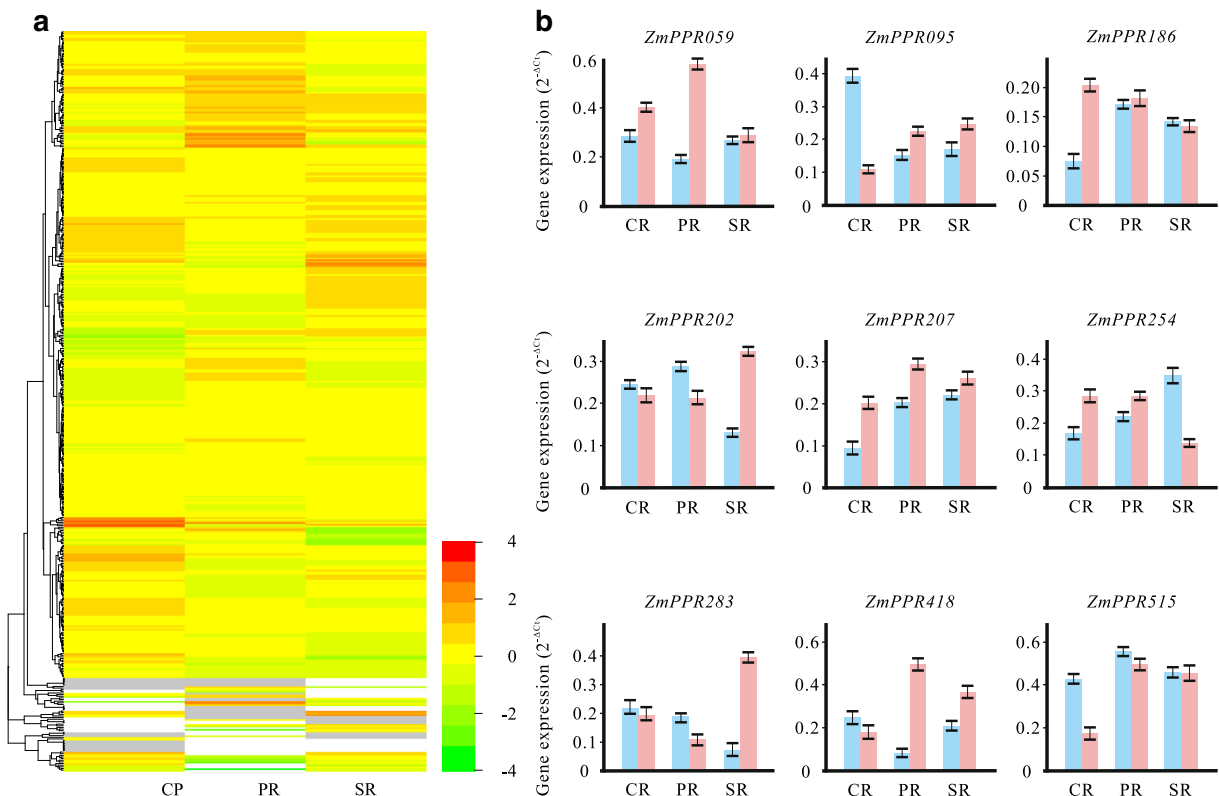


Fig. 3 Differential gene expression of *ZmPPR* genes under salt stress. **a** Expression patterns of *ZmPPR* genes in crown root (CR), primary root (PR), and seminal root (SR) after salt treatment. The relative expression values were \log_2 transformed. White box means that gene was only expressed in control or not expressed in both

control and salt stress treatment, while gray box indicates that gene was specifically expressed under salt treatment. **b** Gene expression analysis under salt treatment using quantitative real-time PCR. Blue, control; red, salt treatment

ZmPPR418 were specially expressed in PR. In addition, *ZmPPR202* and *ZmPPR283* were obviously upregulated in SR, whereas *ZmPPR254* was notably downregulated.

Expression of ZmPPR genes in drought stress response

In Arabidopsis, two *PPR* genes, *slg1* and *slo2*, were reported to respond to drought stress. Slow growth 1 (SLG1) is required for RNA editing of *NAD3* transcript of complex I (Yuan and Liu 2012), and SLO2 is involved in ABA signaling regulation (Zhu et al. 2014). However, so far, it is still unknown whether *ZmPPRs* improves plant tolerance to drought. In this study, microarray and RNA-seq data were used to explore differentially expressed *ZmPPRs* under drought stress. The gene expression profiling of 64 *ZmPPRs* was compared between the drought-tolerant line Han21 and the drought-sensitive line Ye478 using microarray-based RNA expression analysis (Fig. 4a). In general, *ZmPPR* genes in Ye478 respond to water deficit more sensitively than do those in Han21. The transcript levels of *ZmPPR401* were remarkably upregulated under moderate and severe drought conditions but decreased after rewatering both in Han21 and Ye478, which may be a response to the increased ABA level resulting from drought. Besides, the expression levels of two genes, *ZmPPR372* and *ZmPPR468*, were quickly recovered after rewatering in two maize inbred lines. Subsequently, the transcription abundance patterns of *ZmPPRs* in fertilized ovary and basal leaf meristem tissue under drought and well-watered conditions were analyzed using RNA-seq data. As shown in Table S9, the level of seven *ZmPPR* transcripts increased evidently and ten significantly decreased in leaf in response to drought treatment, and 25 and 145 were apparently upregulated and downregulated in cob, respectively. Among them, *ZmPPR075* and *ZmPPR200* showed the strongest upregulation in leaf and cob, respectively. However, the expression levels of *ZmPPR350* were obviously reduced in leaf and *ZmPPR187* in cob. Finally, real-time PCR analysis was used to determine the significantly different expression of the genes under drought stress and the primers are listed in Table S8. As illustrated in Fig. 4b, during drought treatment in leaf and cob, *ZmPPR075* upregulated obviously. *ZmPPR093* and *ZmPPR138* showed decreased transcript abundance in leaf, so did *ZmPPR078* and *ZmPPR120* in cob. However, *ZmPPR131* was

specifically expressed in leaf. Three genes (*ZmPPR195*, *ZmPPR200*, and *ZmPPR419*) had significantly increased mRNA levels in cob after drought treatment, while *ZmPPR093* was significantly downregulated. Taken together, the results suggested that differentially expressed genes may play specific roles in response to drought stress and quantitative real-time PCR analysis for these selected genes confirmed the outcome of RNA-seq analysis.

Analysis of regulatory motif and miRNA target

Transcription factors (TFs) and microRNAs (miRNAs) are prominent gene regulatory factors (Hobert 2008). *Cis*-regulatory (or *cis*-acting) elements function in the regulation of gene expression by controlling promoter efficiency. Statistical analysis showed that 99% of *ZmPPRs* contain core promoter elements TATA-box and CAAT-box. Skn-1_motif is distributed in more than 90% *ZmPPRs*, which might account for endosperm expression. The light responsiveness element G-box was predicted in 471 out of 521 *ZmPPRs*. Two *cis*-acting regulatory elements associated with MeJA-responsiveness, CGTCA-motif and TGACG-motif, were distributed in the promoter regions of more than 77% *ZmPPRs*. The promoters of 377 *ZmPPRs* contain MYB binding site (MBS) elements, suggesting that these genes may be involved in drought stress response and tolerance. The *cis*-acting element (ABRE) that participated in the abscisic acid responsiveness was observed in the upstream of 325 *ZmPPRs*. The circadian regulatory element was recorded in the 345 *ZmPPRs* promoter region which may be responsible for regulating biological circadian. Other elements were also predicted in the promoter regions of *ZmPPRs*, such as regulatory elements (AREs) that were essential for anaerobic induction, 5'-UTR Py-rich stretch (conferring high transcription levels), and several other light-responsive elements. *ZmPPRs* have significant discrepancies in both the number and distribution of *cis*-elements that might affect the gene expression, suggesting that these genes have generated functional divergence during evolution. The predicted *cis*-acting elements were displayed in Table S10. Besides, miRNAs regulate the mRNA stability and translation through the action of the RNA-induced silencing complex (RISC) (Bartel 2009). The Target-align program analysis showed that 93 *ZmPPRs* were targeted by 133 maize miRNAs (Table S11). It was reported that miR156,

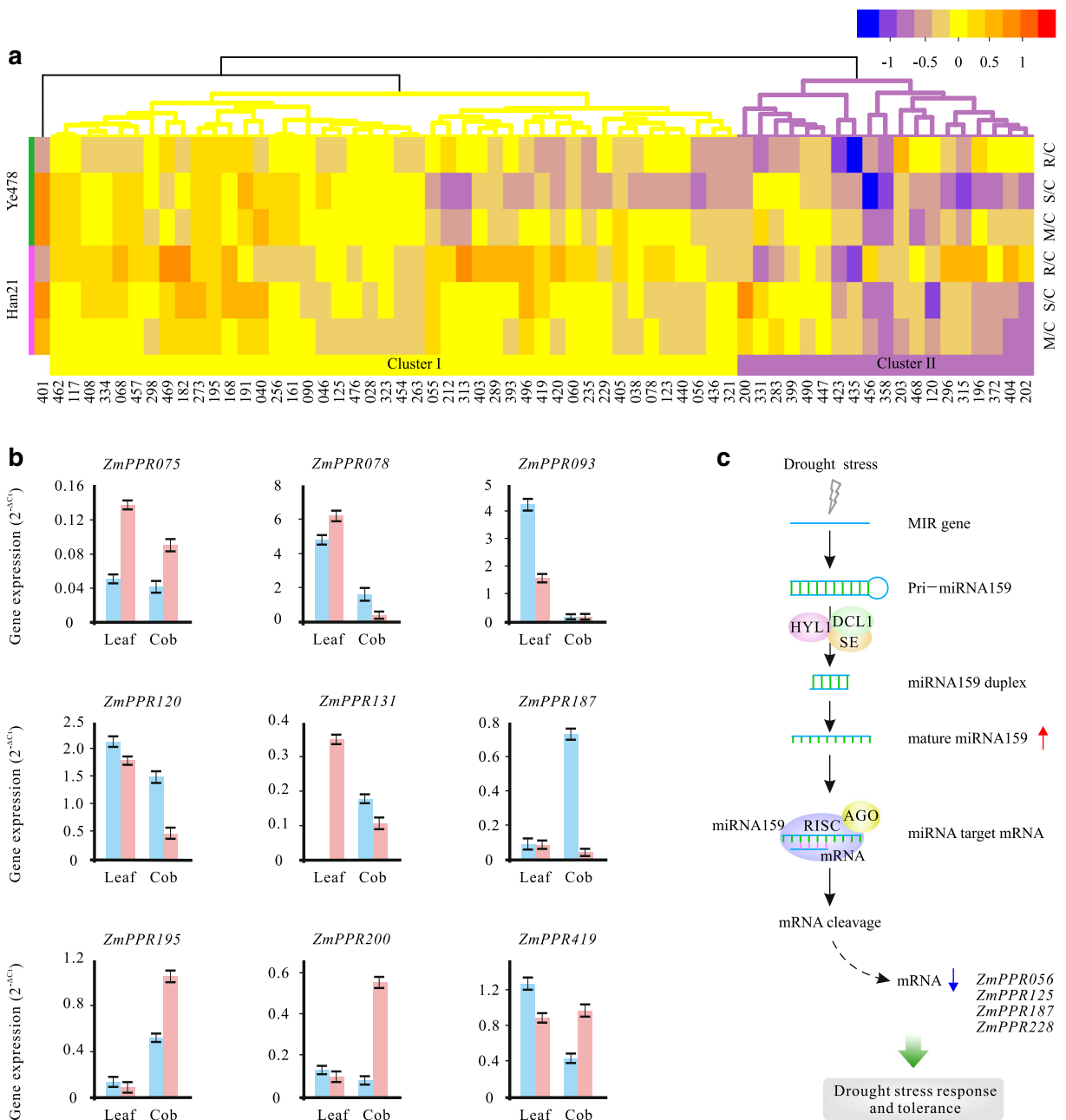


Fig. 4 Differential gene expression of *ZmPPRs* under drought stress. **a** Expression profiles of *ZmPPR* genes under moderate drought stress (M/C), severe drought stress (S/C), and rewatering (R/C) as compared to control seedlings in Han21 and Ye478,

respectively. Log₂-based fold changes were used to create the heat map. **b** Gene expression analysis during drought stresses by quantitative real-time PCR. Blue, control; red, salt treatment. **c** A putative miRNA regulation of *ZmPPR* genes under drought stress

miR164, miR166, miR168, miR171 and miR319 were involved in high salt and drought stresses (Kong et al. 2010). Salt-responsive gene *ZmPPR158* in cob was the predicted target of two miRNAs: miR164c-3p and miR164h-3p. miR529-3p and miR171b-3p might target *ZmPPR352* which showed a strong response to salt

stress in PR and drought condition in cob. *ZmPPR382* which exhibited decreased expression in PR under salt stress was predicted to be a target of several miRNAs (miR164c-5p, miR164b-5p, miR164a-5p, miR164g-5p, and miR164d-5p). *ZmPPR483*, which was a predicted target of two miRNAs (miR164c-3p and miR164h-3p),

was upregulated in seminal root (SR) under salt stress whereas downregulated in cob under drought treatment. *ZmPPR255* predicted as a target of miR164g-3p displayed significantly decreased expression level in leaf under drought stress. *ZmPPR329*, possibly targeted by miR164d-3p, was upregulated in leaf and downregulated in cob in response to drought stress. In addition, a total of eight *ZmPPRs* which showed significantly decreased expression in cob under drought stress were the targets of miR164a-3p. Previous study indicated that miR159 play critical roles in signaling pathways of maize response to drought stress (Wang et al. 2014). Figure 4c shows that the expression of four genes (*ZmPPR056*, *ZmPPR125*, *ZmPPR187*, and *ZmPPR228*) complementary to miR159 was apparently downregulated in cob under water deficit. In summary, miRNA-mRNA interactions would assist in understanding the regulatory mechanisms underlying maize gene expression in response to environmental stresses.

Conclusions

In the present study, we have performed an exhaustive, systematic bioinformatic analysis for maize *PPR* gene family. A total of 521 *PPR* members were identified in the maize genome, and further analyses for gene structure, duplicated event, and three-dimensional structure provided important features for the family. What is more, the expression profiles of some *ZmPPRs* across the different developmental stages showed tissue-specific expression. Besides, analyses of expression profiles via microarray and RNA-seq method revealed *ZmPPR*'s probable functions under abiotic stresses. There is no longer any doubt that the *PPR* protein family plays broad and essential roles in maize organelle gene expression and stress response. Our results provide a solid and unified platform on which future genetic and functional studies can be based.

Acknowledgments We are grateful to the providers who submitted microarray and RNA-seq data to the public expression databases, which can be freely applied.

Compliance with ethical standards

Funding The project was supported by the Science and Technology Cooperation Project of Fujian Province, China (Grant No. 2015I0006)

References

- Arnal N, Quadrado M, Simon M, Mireau H (2014) A restorer-of-fertility like pentatricopeptide repeat gene directs ribonucleolytic processing within the coding sequence of rps 3-rpl 16 and orf 240a mitochondrial transcripts in *Arabidopsis thaliana*. *Plant journal: for cell and. molecular biology* 78(1):134–145
- Bartel DP (2009) Micro RNAs: target recognition and regulatory functions. *Cell* 136(2):215–233
- Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc Natl Acad Sci U S A* 99(16):10887–10892
- Bentolila S, Heller WP, Sun T, Babina AM, Friso G, van Wijk KJ, Hanson MR (2012) RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc Natl Acad Sci U S A* 109(22):E1453–E1461
- Bosacchi M, Gurdon C, Maliga P (2015) Plastid genotyping reveals the uniformity of cytoplasmic male sterile-T maize cytoplasms. *Plant Physiol* 169(3):2129–2137
- Boussard C, Avon A, Kindgren P, Bond CS, Challenor M, Lurin C, Small I (2014) The cytidine deaminase signature HxE (x) n CxxC of DYW1 binds zinc and is necessary for RNA editing of ndh D-1. *New Phytol* 203(4):1090–1095
- Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I (2013) The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. *The Plant journal: for cell and. molecular biology* 74(6):935–945
- Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, de la Luz G-NM, Cantero A, Delannoy E, Jimenez LF, Lurin C, Small I, Leon P (2008) CLB19, a pentatricopeptide repeat protein required for editing of rpo a and clp P chloroplast transcripts. *The Plant journal: for cell and. molecular biology* 56(4):590–602
- Chen L, Liu YG (2014) Male sterility and fertility restoration in crops. *Annu Rev Plant Biol* 65:579–606
- Coquille S, Filipovska A, Chia T, Rajappa L, Lingford JP, Razif MF, Thore S, Rackham O (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat Commun* 5:5729
- Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, Rfo, in radish, as a member of the pentatricopeptide-repeat protein family. *EMBO Rep* 4(6):588–594
- Ding YH, Liu NY, Tang ZS, Liu J, Yang WC (2006) Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. *Plant Cell* 18(4):815–830
- Doniwa Y, Ueda M, Ueta M, Wada A, Kadowaki K, Tsutsumi N (2010) The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. *Gene* 454(1–2):39–46
- Duvick DN, Snyder RJ, Anderson EG (1961) The chromosomal location of Rfl, a restorer gene for cytoplasmic pollen sterile maize. *Genetics* 46:1245–1252

- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39(Web Server issue):W29–W37
- Fisk DG, Walker MB, Barkan A (1999) Molecular cloning of the maize gene *crp 1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J* 18(9):2621–2630
- Fujii S, Sato N, Shikanai T (2013) Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON gradient regulation 3. *Plant Cell* 25(8):3079–3088
- Geddy R, Brown GG (2007) Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8:130
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: micro RNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34(Database issue):D140–D144
- Hammani K, des Francs-Small CC, Takenaka M, Tanz SK, Okuda K, Shikanai T, Brennicke A, Small I (2011a) The pentatricopeptide repeat protein OTP87 is essential for RNA editing of *nad 7* and *atp 1* transcripts in Arabidopsis mitochondria. *J Biol Chem* 286(24):21361–21371
- Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giege P (2011b) An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. *Plant Cell* 23(2):730–740
- Hauler A, Jonietz C, Stoll B, Stoll K, Braun HP, Binder S (2013) RNA processing factor 5 is required for efficient 5' cleavage at a processing site conserved in RNAs of three different mitochondrial genes in *Arabidopsis thaliana*. *The Plant journal: for cell and molecular biology* 74(4):593–604
- Hobert O (2008) Gene regulation by transcription factors and micro RNAs. *Science* 319(5871):1785–1786
- Holzle A, Jonietz C, Torjek O, Altmann T, Binder S, Forner J (2011) A restorer of fertility-like PPR gene is required for 5'-end processing of the *nad 4* mRNA in mitochondria of *Arabidopsis thaliana*. *The Plant journal: for cell and molecular biology* 65(5):737–744
- Jiang SC, Mei C, Liang S, Yu YT, Lu K, Wu Z, Wang XF, Zhang DP (2015) Crucial roles of the pentatricopeptide repeat protein SOAR1 in Arabidopsis response to drought, salt and cold stresses. *Plant Mol Biol* 88(4–5):369–385
- Jimenez-Lopez JC, Gachomo EW, Seufferheld MJ, Kotchoni SO (2010) The maize ALDH protein superfamily: linking structural features to functional specificities. *BMC Struct Biol* 10:43
- Jonietz C, Forner J, Hildebrandt T, Binder S (2011) RNA processing factor3 is crucial for the accumulation of mature *ccm C* transcripts in mitochondria of Arabidopsis accession Columbia. *Plant Physiol* 157(3):1430–1439
- Jonietz C, Forner J, Holzle A, Thuss S, Binder S (2010) RNA processing factor2 is required for 5' end processing of *nad 9* and *cox 3* mRNAs in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 22(2):443–453
- Kakumanu A, Ambavaram MM, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A (2012) Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-seq. *Plant Physiol* 160(2):846–867
- Kong YQM, Elling AA, Chen BB, Deng XW (2010) Differential expression of micro RNAs in maize inbred and hybrid lines during salt and drought stress. *Am J Plant Sci* 1:69–76
- Kopriovova A, des Francs-Small CC, Calder G, Mugford ST, Tanz S, Lee BR, Zechmann B, Small I, Kopriva S (2010) Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial *nad 7* transcripts. *J Biol Chem* 285(42):32192–32199
- Kotera E, Tasaka M, Shikanai T (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433(7023):326–330
- Langewisch T (2012) Localization of the Rf3 restorer-of-fertility gene for maize S-type cytoplasmic male sterility.
- Laughnan JR, Gabay-Laughnan S (1983) Cytoplasmic male sterility in maize. *Annu Rev Genet* 17:27–48
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452
- Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16(8):2089–2103
- Lv HX, Huang C, Guo GQ, Yang ZN (2014) Roles of the nuclear-encoded chloroplast SMR domain-containing PPR protein SVR7 in photosynthesis and oxidative stress tolerance in Arabidopsis. *J Plant Biol* 57(5):291–301
- Manthey GM, McEwen JE (1995) The product of the nuclear gene PET309 is required for translation of mature mRNA and stability or production of intron-containing RNAs derived from the mitochondrial COX1 locus of *Saccharomyces cerevisiae*. *EMBO J* 14(16):4031–4043
- Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psb B-psb T-psb H-pet B-pet D* RNAs. *Plant Cell* 15(6):1480–1495
- Mootha VK, Lepage P, Miller K, Bunkenborg J, Reich M, Hjerrild M, Delmonte T, Villeneuve A, Sladek R, Xu F, Mitchell GA, Morin C, Mann M, Hudson TJ, Robinson B, Rioux JD, Lander ES (2003) Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc Natl Acad Sci U S A* 100(2):605–610
- Okuda K, Hammani K, Tanz SK, Peng L, Fukao Y, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndh B* and *ndh G* transcripts. *The Plant journal: for cell and molecular biology* 61(2):339–349
- Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmuller R (2006) pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. *Plant Cell* 18(1):176–197
- Saha D, Prasad AM, Srinivasan R (2007) Pentatricopeptide repeat proteins and their emerging roles in plants. *Plant physiology and biochemistry: PPB/Societe francaise de physiologie vegetale* 45(8):521–534
- Skibbe DS, Liu F, Wen TJ, Yandean MD, Cui X, Cao J, Simmons CR, Schnable PS (2002) Characterization of the aldehyde dehydrogenase gene families of *Zea mays* and Arabidopsis. *Plant Mol Biol* 48(5–6):751–764

- Small ID, Peeters N (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25(2):46–47
- Soderlund C, Bomhoff M, Nelson WM (2011) SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res* 39(10):e68
- Takenaka M (2010) MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the nad 7 transcript in mitochondria of *Arabidopsis*. *Plant Physiol* 152(2):939–947
- Takenaka M, Verbitskiy D, Zehrmann A, Brennicke A (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 285(35):27122–27129
- Tan J, Tan Z, Wu F, Sheng P, Heng Y, Wang X, Ren Y, Wang J, Guo X, Zhang X, Cheng Z, Jiang L, Liu X, Wang H, Wan J (2014) A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in rice. *Mol Plant* 7(8):1329–1349
- Toda T, Fujii S, Noguchi K, Kazama T, Toriyama K (2012) Rice MPR25 encodes a pentatricopeptide repeat protein and is essential for RNA editing of nad 5 transcripts in mitochondria. *The Plant journal: for cell and molecular biology* 72(3):450–460
- Uyttewaal M, Mireau H, Rurek M, Hammani K, Arnal N, Quadrado M, Giege P (2008) PPR336 is associated with polysomes in plant mitochondria. *J Mol Biol* 375(3):626–636
- Verbitskiy D, Zehrmann A, Hartel B, Brennicke A, Takenaka M (2012) Two related RNA-editing proteins target the same sites in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* 287(45):38064–38072
- Wang YG, An M, Zhou SF, She YH, Li WC, Fu FL (2014) Expression profile of maize micro RNAs corresponding to their target genes under drought stress. *Biochem Genet* 52(11–12):474–493
- Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG (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(3):676–687
- Wei K, Pan S (2014) Maize protein phosphatase gene family: identification and molecular characterization. *BMC Genomics* 15:773
- Wei KF, Chen J, Chen YF, Wu LJ, Xie DX (2012) Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA research: an international journal for rapid publication of reports on genes and genomes* 19(2):153–164
- Wei K, Chen Y, Xie D (2013) Genome-scale evolution and phylogenetics of H5N1 influenza virus in China during 1996–2012. *Vet Microbiol* 167(3–4):383–393
- Wei K, Wang Y, Zhong X, Pan S (2014) Protein kinase structure, expression and regulation in maize drought signaling. *Mol Breed* 34(2):583–602
- Wen LY, Chase CD (1999) Mitochondrial gene expression in developing male gametophytes of male-fertile and S male-sterile maize. *Sex Plant Reprod* 11(6):323–330
- Wise RP, Fliss AE, Pring DR, Gengenbach BG (1987) Urf 13-T of T cytoplasm maize mitochondria encodes a 13 kD polypeptide. *Plant Mol Biol* 9(2):121–126
- Xie F, Zhang B (2010) Target-align: a tool for plant micro RNA target identification. *Bioinformatics* 26(23):3002–3003
- Xu XH, Chen H, Sang YL, Wang F, Ma JP, Gao XQ, Zhang XS (2012) Identification of genes specifically or preferentially expressed in maize silk reveals similarity and diversity in transcript abundance of different dry stigmas. *BMC Genomics* 13:294
- Yang Z, Gu S, Wang X, Li W, Tang Z, Xu C (2008) Molecular evolution of the CPP-like gene family in plants: insights from comparative genomics of *Arabidopsis* and rice. *J Mol Evol* 67(3):266–277
- Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, Wang J, Zhu JK, Shi Y, Yan N (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504(7478):168–171
- Yu QB, Jiang Y, Chong K, Yang ZN (2009) AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript acc D RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. *The Plant journal: for cell and molecular biology* 59(6):1011–1023
- Yuan H, Liu D (2012) Functional disruption of the pentatricopeptide protein SLG1 affects mitochondrial RNA editing, plant development, and responses to abiotic stresses in *Arabidopsis*. *The Plant journal: for cell and molecular biology* 70(3):432–444
- Zabala G, Gabay-Laughnan S, Laughnan JR (1997) The nuclear gene Rf3 affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. *Genetics* 147(2):847–860
- Zehrmann A, van der Merwe J, Verbitskiy D, Hartel B, Brennicke A, Takenaka M (2012) The DYW-class PPR protein MEF7 is required for RNA editing at four sites in mitochondria of *Arabidopsis thaliana*. *RNA Biol* 9(2):155–161
- Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell* 21(2):558–567
- Zhang M, Kong X, Xu X, Li C, Tian H, Ding Z (2015) Comparative transcriptome profiling of the maize primary, crown and seminal root in response to salinity stress. *PLoS One* 10(3):e0121222
- Zhu Q, Dugardeyn J, Zhang C, Muhlenbock P, Eastmond PJ, Valcke R, De Coninck B, Oden S, Karampelias M, Cammue BP, Prinsen E, Van Der Straeten D (2014) The *Arabidopsis thaliana* RNA editing factor SLO2, which affects the mitochondrial electron transport chain, participates in multiple stress and hormone responses. *Mol Plant* 7(2):290–310
- Zoschke R, Kroeger T, Belcher S, Schottler MA, Barkan A, Schmitz-Linneweber C (2012) The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast atp B/E mRNA. *The Plant journal: for cell and molecular biology* 72(4):547–558