BMC Plant Biology

RESEARCH ARTICLE   **Open Access**

GhYGL1d, a pentatricopeptide repeat protein, is required for chloroplast development in cotton

Peng He[1], Shuyin Wu[1], Yanli Jiang[2], Lihua Zhang[1], Meiju Tang[1], Guanghui Xiao[3*] and Jianing Yu[1*]

Abstract

Background: The pentatricopeptide repeat (PPR) gene family, which contains multiple 35-amino acid repeats, constitutes one of the largest gene families in plants. PPR proteins function in organelles to target specific transcripts and are involved in plant development and growth. However, the function of PPR proteins in cotton is still unknown.

Results: In this study, we characterized a PPR gene YELLOW-GREEN LEAF (*GhYGL1d*) that is required for cotton plastid development. The *GhYGL1d* gene has a DYW domain in C-terminal and is highly express in leaves, localized to the chloroplast fractions. GhYGL1d share high amino acid-sequence homology with AtECB2. In *atecb2* mutant, overexpression of *GhYGL1d* rescued the seedling lethal phenotype and restored the editing of *accD* and *ndhF* transcripts. Silencing of *GhYGL1d* led to the reduction of chlorophyll and phenotypically yellow-green leaves in cotton. Compared with wild type, *GhYGL1d*-silenced cotton showed significant deformations of thylakoid structures. Furthermore, the transcription levels of plastid-encoded polymerase (PEP) and nuclear-encoded polymerase (NEP) dependent genes were decreased in *GhYGL1d*-silenced cotton.

Conclusions: Our data indicate that *GhYGL1d* not only contributes to the editing of *accD* and *ndhF* genes, but also affects the expression of NEP- and PEP-dependent genes to regulate the development of thylakoids, and therefore regulates leaf variegation in cotton.

Keywords: Chloroplast, PPR, Leaf variegation, Cotton

Background

Life on earth depends on the process of photosynthesis, which converts light energy into chemical energy, in chloroplasts to provide us with food and oxygen [1]. Chloroplasts are specialized organelles in plants and an abundance of chloroplasts are found in the leaf, which is the primary location of photosynthesis and sugar manufacturing. The development of a leaf directly affects photosynthetic efficiency of an individual plant, which further determines the yield of plants [2]. Dozens of studies have demonstrated that coordinated expression and regulation between both nuclear and chloroplast genes is very important for biogenesis of chlorophyll and the development of chloroplasts [3]. Numerous nuclear-encoded genes were found to be involved in RNA processing events such as editing, splicing and degradation of chloroplast genes [4–8]. As nuclear factors, dozens of PPR proteins have been proved to be involved in the expression of chloroplast genes [9, 10].

The PPR protein family was one of the largest protein families in plants, which contained several repeating motifs consisting of 35 amino acids [11, 12]. In land plants, the PPR family has more than 400 protein members [13]. Proteins in the PPR family in plants are classified into two major subfamilies according to their different motifs [14]. The P subfamily of proteins contain only the P motif, while the PLS subfamily consists of degenerated P, L, and S motifs, where the S motif has 31 amino acids and L motif has 35 or 36 amino acids [15]. PLS

He *et al. BMC Plant Biology*   (2019) 19:350

Page 2 of 12

subfamily proteins are plant-specific and always possess C-terminal domains named E, E+ and DYW, which reportedly are involved in organelle RNA editing in plants [16]. Most PPR proteins are anchored into either the mitochondria or the chloroplast. There are 466 PPR proteins found in the *Arabidopsis thaliana* genome. Among them, 88 members belong to PPRs with the DYW domain [17].

Pentatricopeptide repeat proteins are involved in many post-transcriptional processes in chloroplasts and mitochondria, including RNA cleavage [18, 19], alternative splicing [20–22], transcriptional regulation [23, 24], RNA editing [25, 26], mRNA stabilization [27, 28], and translation [29, 30]. The PPR protein EMP9 with 16 PPR motifs is required for the editing of mitochondrial *ccmB-43* and *mrps4-335*, which affect seed development in maize [31]. The MEF13 protein, which consists of 21 PPR motifs, is required for RNA editing at eight sites in mitochondrial mRNAs in *Arabidopsis* [32]. CHLOROR-ESPIRATORY REDUCTION 4 (CRR4), the first RNA-editing factors with 11 PPR repeats identified in chloroplasts, is responsible for RNA editing of the initiation codon of *ndhD* [33]. Growing slowly 1 (GRS1), a PLS-type PPR protein, is involved in RNA editing at four specific sites and affects plant development [34]. *Arabidopsis* PDM1/SEL1 encode a PLS-type PPR protein, the *pdm1* loss-of-function mutant exhibits pigment-deficient phenotype with a defect in RNA splicing of *trnK* and *ndhA* [35]. PDM2, a plastid-localized PPR, plays an important role in the accumulation of plastid-encoded transcripts and plastid RNA editing [36]. PDM3 is a chloroplast protein with 12 PPR repeats domains and responsible for the RNA splicing of *trnA*, *ndhB*, and *clpP1* transcript [37]. In rice, a plastid-localized PPR protein, OsPPR6, reportedly mediates both RNA editing and splicing [38].

The DYW domain is named for the frequent presence of an Asp-Tyr-Trp tripeptide at the C terminal, and it contributes to the discrimination of target and non-target editing sites [39]. The *AtECB2* gene, encoding a PPR protein with a C-terminal DYW domain, is required for editing of *accD* genes and chloroplast biogenesis [40]. In *Arabidopsis*, DYW proteins CRR22 and CRR28 play important roles in RNA editing and RNA cleavage [13]. The *YS1* gene encoding a DYW protein is essential for editing of *rpoB* transcripts and the development of leaves [41]. The DYW domain of the OTP85 protein catalyzes site-specific cleavage and editing of target RNA [42]. In moss, *PPR_71* gene, encoding a PPR protein with a DYW domain, is required for RNA editing of the *ccmFc* transcript [43]. The *PPR_43* gene, which encodes a mitochondrial-localized PPR protein with a C-terminal DYW domain, is responsible for the splicing of *cox1* pre-mRNA at the second intron [44]. In rice, OGR1, a PPR-DYW protein, is essential for RNA editing in mitochondria and is normal growth and development [45].
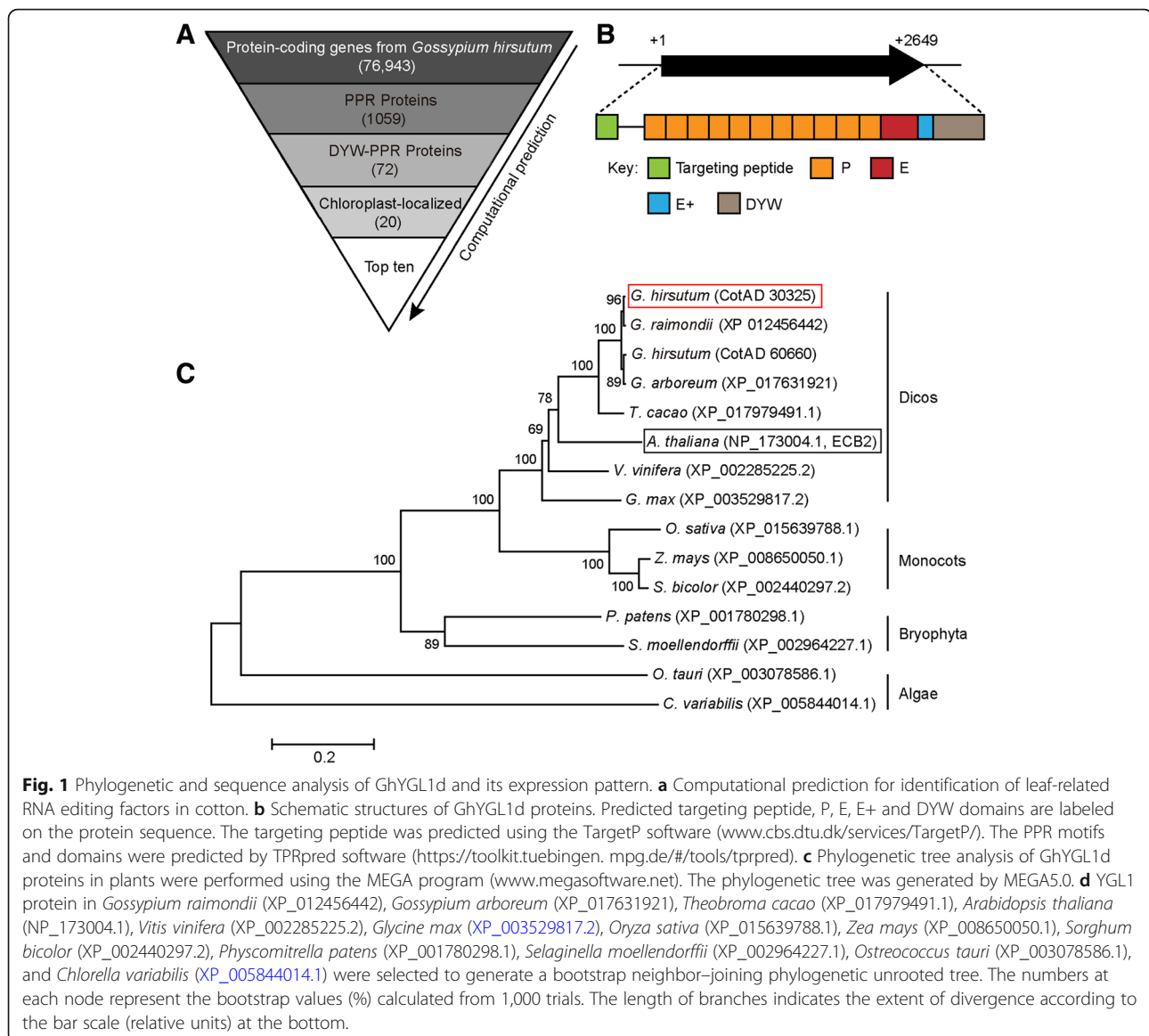
Cotton is the most important textile fiber and also a significant oilseed crop. The first report about PPR proteins in cotton was the identification of five PPR proteins in upland cotton [46]. Furthermore, eight PPR family genes were cloned and their expression patterns investigated [47]. Recently, a genome-wide identification of PPR genes with DYW domains was described [48]. However, the biological function of PPR proteins in cotton is largely unknown.

In this study, using computational prediction followed by verification with a virus-induced gene silencing (VIGS) experiment, we identified a PPR-DYW family gene, *GhYGL1d*, which was essential for cotton leaf development. Overexpression of *GhYGL1d* in *Arabidopsis* mutant, *atecb2*, resulted in a similar phenotype as the wild type plants. Reduction of *GhYGL1d* transcripts decreased the content of chlorophyll, led to the disruption of thylakoid structure and produced albino leaves in cotton. In addition, we also found that silencing the *GhYGL1d* gene caused dramatic reductions in transcription levels of PEP-dependent genes. Our findings will not only characterize mechanisms of cotton leaf variegation regulated by *GhYGL1d*, but also provide the strategy to study the function of other PPR-DYW genes in the future.

Results

Screening and identification of the GhYGL1d gene

PPR-DYW proteins play important roles in many biological processes, especially in RNA editing events [49]. To identify all of the PPR-DYW proteins in G. hirsutum, we screened for all PPR proteins in the cotton genome database (https://www.cottongen.org) and a total of 1, 059 PPR proteins were identified (Fig. 1a). Among them, 72 PPR-DYW proteins with 20 members localized in chloroplast were found (Additional file 1 Table S1). We selected ten PPR-DYW proteins with cTP values more than 0.6 (predicted by ChloroP or TargetP program) to investigate their biological functions using the VIGS experiment. We amplified nine gene fragments with specific primers (Additional file 1 Table S2) to silence their expression. In contrast to the silenced cotton with an empty vector, the *CotA_30325*-silenced plant displayed variegated leaf phenotypes (Additional file 1 Figure S1). An analysis of the CotA_30325 protein in the Pfam database showed that it contained a tandem repeat of 11 PPR motifs and belonged to a member of the DYW subfamily (Fig. 1b). A BLAST search of the NCBI database showed that *CotAD_30325* is a double-copy gene in the cotton genome which shared 98.98% similarity with *CotAD_60660* in their cDNA sequence over the 2649 bp length. To better present the evolutionary relationship

Fig. 1 Phylogenetic and sequence analysis of GhYGL1d and its expression pattern. **a** Computational prediction for identification of leaf-related RNA editing factors in cotton. **b** Schematic structures of GhYGL1d proteins. Predicted targeting peptide, P, E, E+ and DYW domains are labeled on the protein sequence. The targeting peptide was predicted using the TargetP software (www.cbs.dtu.dk/services/TargetP/). The PPR motifs and domains were predicted by TPRpred software (https://toolkit.tuebingen. mpg.de/#/tools/tprpred). **c** Phylogenetic tree analysis of GhYGL1d proteins in plants were performed using the MEGA program (www.megasoftware.net). The phylogenetic tree was generated by MEGA5.0. **d** YGL1 protein in *Gossypium raimondii* (XP_012456442), *Gossypium arboreum* (XP_017631921), *Theobroma cacao* (XP_017979491.1), *Arabidopsis thaliana* (NP_173004.1), *Vitis vinifera* (XP_002285225.2), *Glycine max* (XP_003529817.2), *Oryza sativa* (XP_015639788.1), *Zea mays* (XP_008650050.1), *Sorghum bicolor* (XP_002440297.2), *Physcomitrella patens* (XP_001780298.1), *Selaginella moellendorffii* (XP_002964227.1), *Ostreococcus tauri* (XP_003078586.1), and *Chlorella variabilis* (XP_005844014.1) were selected to generate a bootstrap neighbor–joining phylogenetic unrooted tree. The numbers at each node represent the bootstrap values (%) calculated from 1,000 trials. The length of branches indicates the extent of divergence according to the bar scale (relative units) at the bottom.

among CotAD_30325 proteins, we constructed a phylogenetic tree and found that the CotA_30325 protein shared high sequence homology with AtECB2 in *Arabidopsis* (Fig. 1c). According to the phylogenetic relationships, the *CotAD_30325* gene was from the D-subgenome named as *GhYGL1d*. We also performed an amino acid sequence alignment of GhYGL1d and AtECB2 and the result showed that GhYGL1d shared 11 conserved PPR motifs with AtECB2 (Additional file 1 Figure S2).

Expression pattern and subcellular localization of GhYGL1d

In order to understand the potential physiological functions of *GhYGL1d* gene in cotton, the expression patterns of the GhYGL1d gene were investigated in various tissues. Results show that GhYGL1d was highly expressed in leaf tissues (Fig. 2a), suggesting that GhYGL1d may play an important role in the leaf, which is consistent with the result of the leaf phenotype in the GhYGL1d-silenced plant.

GhYGL1d is predicted to be targeted to the chloroplast by Target P (http://www.cbs.dtu. dk/services/TargetP/). To confirm this prediction, a construct with the full-length *coding sequence* of GhYGL1d fused to green fluorescent protein (GFP) driven by the CaMV 35S promoter was transiently expressed in Arabidopsis protoplasts. Using confocal laser scanning microscopy, we *observed that* GhYGL1d-*GFP fusion protein signals* highly overlapped with the red auto-fluorescent signals of chlorophyll *in the protoplasts* (Fig. 2b). To further elucidate the *precise* location of GhYGL1d in chloroplasts,

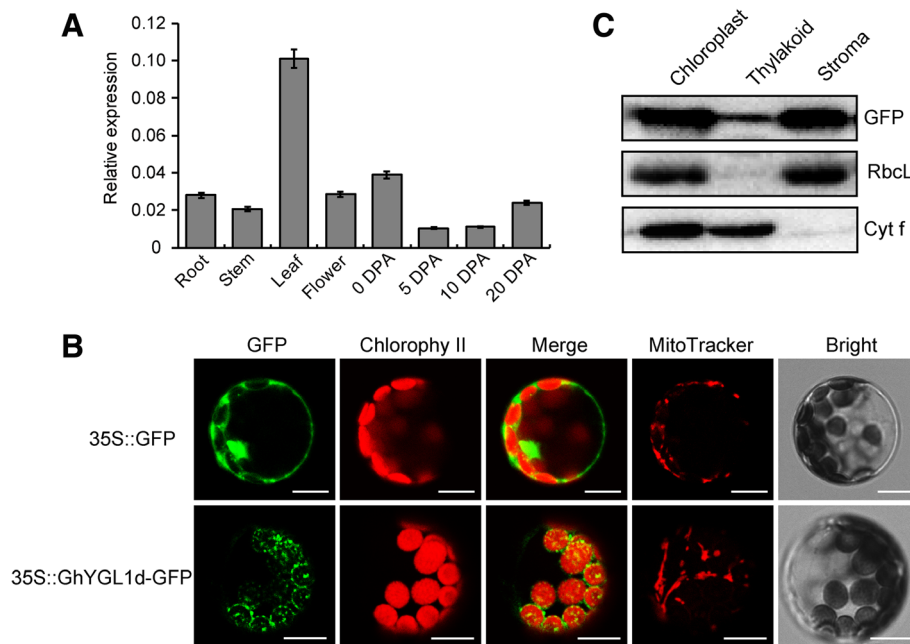He *et al. BMC Plant Biology*   (2019) 19:350

Page 4 of 12

Fig. 2 Expression and subcellular localization of GhYGL1d. **a** qRT-PCR analysis of *GhYGL1d* expression in different tissues. Total RNA was isolated from roots, stems, leaves, flowers and fibers. The *GhUBQ7* gene was used as a reference gene for qRT-PCR. DPA, day post-anthesis. The values shown are means ± SE of three replicates. **b** GhYGL1d localizes to the chloroplast. Protoplasts from wild-type *Arabidopsis* were transiently transformed with a control GFP vector (designated 35S::GFP) or with a GhYGL1d-GFP vector. Fluorescence was observed by confocal microscopy of single protoplasts; green fluorescence = GFP; red = chlorophyll autofluorescence. Bars = 10 μm. **c** GhYGL1d localizes to the stroma and thylakoid fractions of chloroplasts. Total proteins extracted from the 35S::GhYGL1d-GFP transgenic line were used to confirm the specificity of the anti-GFP antibody. Intact chloroplasts were isolated from 35S::GhYGL1d-GFP transgenic seedlings and separated into the thylakoid and stroma fractions. Chloroplasts protein Cyt f and RbcL were used as a marker of thylakoid membrane and stroma fractions, respectively. The GFP antibody was used to detect the GhYGL1d-GFP fusion protein.

we generated transgenic plants expressing the full-length sequence of *GhYGL1d* fused with GFP. We extracted chloroplasts from transgenic plants and separated the chloroplast into thylakoid membrane and stroma fractions. Strong immunoblotting signals were observed in the stroma fraction, which suggests that GhYGL1d highly accumulated in the stroma fraction of plastids (Fig. 2c).

Expressing GhYGL1d in an atecb2 mutant restores the defect

In A. thaliana, the atecb2 mutant displayed albino cotyledons and unorganized chloroplast structure. Moreover, the RNA editing at accD-794 and ndhF-C290 sites was abolished. *T*o test the function of GhYGL1d, we expressed GhYGL1d, driven by the constitutive 35S promoter, in the atecb2 mutant. After confirming by the PCR method, we successfully obtained the transgenic plant. As shown in Fig. 3, complement 35S::GhYGL1d mutant plants restored chlorophyll production, avoiding the production of albino cotyledons (Fig. 3a), and also seedlings were rescued from lethality (Fig. 3b). At the same time, the atecb2 mutant expressing GhYGL1d produced almost identical amounts of chlorophyll as compared with the wild-type plants (Fig. 3c). In the atecb2

mutant, the editing of accD and ndhF transcripts was deficient, which resulted in an alteration of the amino acid serine to leucine in AccD and NdhF proteins, respectively [50]. Given that GhYGL1d could rescue the phenotype of the atecb2 mutant, it is likely that GhYGL1d restored the editing of accD and ndhF transcripts. To test this, we examined the editing sites of accD and ndhF transcripts in wild-type, atecb2 mutant and GhYGL1d complemented plants. Indeed, the editing sites of accD and ndhF in the transgenic plants were identical to that in wild-type plants (Fig. 3d). These results indicate that GhYGL1d can restore the function of AtECB2 in Arabidopsis.

GhYGL1d is required for chloroplast development

To further elucidate the possible function of GhYGL1d in cotton chloroplast development, A. tumefaciens containing the *CLCrVA*-GhYGL1d construct was infiltrated into wild-type cotton cotyledons. Four weeks later, the cotton plants infiltrated with *CLCrVA*-GhYGL1d showed variegated leaves unlike the normal leaf phenotype of plants infiltrated with the empty vector (Fig. 4a). We detected the transcript level of GhYGL1d and the RT-qPCR experiments showed that the transcripts of
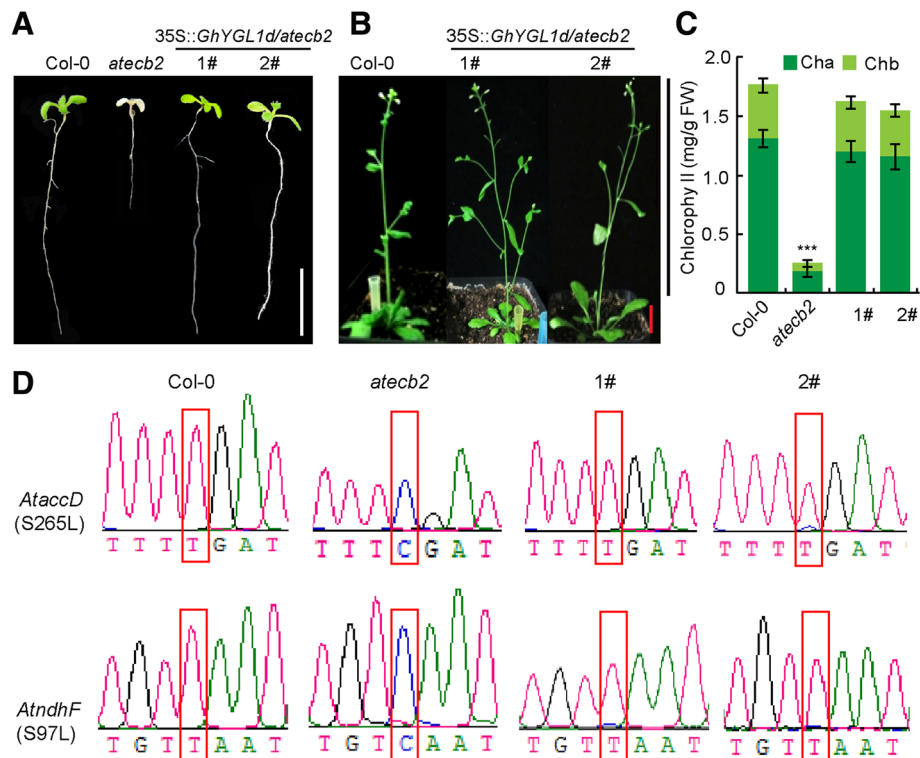
Fig. 3 *GhYGL1d* partially restored the function of *AtECB2* in *Arabidopsis*. (A, B) Phenotypes of a wild-type (Col-0) plant, an *atecb2* mutant and GhYGL1d complemented plants at two (**a**), and six (**b**) weeks after sowing. Scale bars are 2 cm. (C) Chlorophyll contents in a wild-type (Col-0) plant, an *atecb2* mutant and GhYGL1d complemented plants. FW, fresh weight. Chl a, chlorophyll *a*. Chl b, chlorophyll *b*. Error bars indicate SD for three biological replicates. Asterisks indicate significant differences (*P* < 0.001) from Col-0 plants. (D) RNA editing of *accD*, *ndhF* transcripts in wild-type (Col-0) plants, *atecb2* mutants and GhYGL1d complemented plants. 1# and 2# indicate GhYGL1d complemented lines.

GhYGL1d significantly decreased in CLCrVA-GhYGL1d RNAi lines (Fig. 4b). In addition, the contents of both chlorophyll a and b in CLCrVA-GhYGL1d RNAi plants were reduced in contrast to that in control plants (Fig. 4c). And the chlorophyll fluorescence of protoplasts isolated from the CLCrVA-GhYGL1d RNAi plants also showed a decreased intensity (Fig. 5a-h). We also compared the ultrastructure of chloroplasts by TEM. The chloroplasts of control plants were crescent-shaped and contained obvious starch grains and well-formed thylakoid structures, while in GhYGL1d RNAi plants, the chloroplasts *showed an unorganized chloroplast structure* (Fig. 5i-l). Taken together, our data suggest that GhYGL1d is required for leaf chloroplast development.

GhYGL1d effects on chloroplast-encoded gene transcripts and photosynthetic protein accumulation

The expression of plastid-encoded genes is closely linked with chloroplast developmental status. To investigate whether defective chloroplast development was linked to the changes in gene expression, we performed RNA sequencing with samples prepared from control plants and three independent GhYGL1d RNAi lines. The genes expression of ATP synthase and NADH dehydrogenase

and photosystems I and II were significantly reduced in GhYGL1d RNAi plants (Fig. 6a, Additional file 2 Table S3). In order to confirm the RNA-seq data, we examined transcription levels of some of these genes by RNA-qPCR. Photosystem I (PSI) subunit-*encoding* genes, psaA and psaB, and photosystem II (PSII) subunit-*encoding* genes, psaA, psaB and psaE, were drastically decreased in GhYGL1d RNAi plants compared with those in the WT. We noticed that the expression of ATP synthase genes (atpA and atpB) were significantly down-regulated in the RNAi plants, while the chloroplast caseinolytic protease gene ClpP and RNA polymerase subunit gene rpoA, exhibited a minimal reduction of expression in GhYGL1d RNAi plants (Fig. 6b). These results suggest that GhYGL1d may regulate the expression of plastid-encoded genes for chloroplast biogenesis. *To confirm whether the photosynthetic proteins are impaired in GhYGL1d RNAi plants, we explored the accumulation of photosynthetic proteins by immunoblot analyses. These results showed that* PSI, PSII and ATPase complex subunits *were clearly reduced in* GhYGL1d RNAi plants *(*Fig. 7a). Furthermore, we investigated the photosynthetic complexes of the thylakoid membrane in both *GhYGL1d RNAi* and WT *plants* by BN-PAGE. We
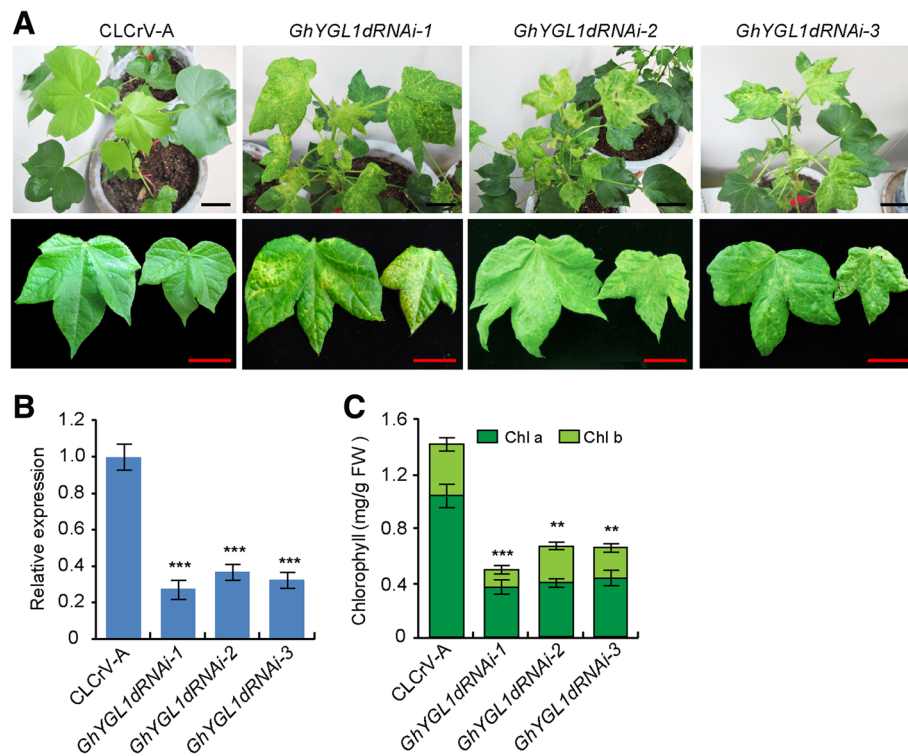
Fig. 4 Silencing of *GhYGL1d* showed variegated leaves. **a** Cotton plants infiltrated with CLCrVA:*GhYGL1d* (*GhYGL1d-RNAi*) showed variegated leaves. The photographs were taken at approximately five weeks after infiltration. Wildtype plants transformed by an empty CLCrVA vector was used as the control. Scale bars are 2 cm. **b** qRT-PCR analysis of GhYGL1d transcripts in RNAi plants. The *GhUBQ7* gene was used as a reference. The values shown are means ± SE of three biological replicates. Significant differences between RNAi plants and CLCrVA control plants were calculated using Student's t-test: ***, $P < 0.001$; (C) Chlorophyll contents in gene-silenced plants. FW, fresh weight. Chl a, chlorophyll *a*. Chl b, chlorophyll *b*. Values are means ± SD of three replicates. Student's t-test: ***, $P < 0.001$;

found that the accumulation of PSII and PSI complexes was significantly reduced in *GhYGL1d RNAi plants compared with that in the WT* (Figure 7B).

Discussion

Chloroplast development requires the balance between cytosolic and plastid co-expression and both nuclear- and chloroplast-encoding genes are coordinated to regulate chloroplast development [51]. In this study, we screened and identified a gene termed GhYGL1d, which regulates chloroplast development in cotton. GhYGL1d is a member of the PPR-DYW gene family and localized to the stroma fraction of chloroplasts. This gene might be involved in the C-to-U editing of accD and ndhF transcripts in cotton chloroplast. Down-regulation of GhYGL1d led to variegated leaves in cotton (Fig. 4a). These results suggest that GhYGL1d plays an important role in plastid development of cotton.

Loss of function of GhYGL1d causes defective chloroplast development

The nucleus-encoded PPR proteins are important to plastid development. The loss-of-function of targeted

plastid PPR proteins generally result in defects in chloroplast biogenesis. GhYGL1d is a typical PPR protein and *shared 11 conserved PPR motifs with AtECB2* of Arabidopsis. AtECB2 is required for early chloroplast biogenesis and the loss of its function resulted in the lack of thylakoid membranes, albino cotyledons and lethality of seedlings [52]. We demonstrated that *expressing GhYGL1d in the atecb2 mutant rescued cotyledons from albinism and the lethal* phenotype to be more similar to the wild type phenotype (Fig. 3a and b). We also observed that the leaves of GhYGL1d-RNAi lines became a yellow-green variegated phenotype and had unorganized chloroplast structure. All these RNAi plants exhibited a significant reduction in *gene* transcription levels of PSI and PSII core subunits (psaA and psaB, psbA, psbB and psbE), and a concomitantly reduced protein expression levels of PSII supercomplexes. Photosystem II is a large protein complex in the membranes of chloroplasts, which is essential to plant growth. Some PSII proteins are encoded by nuclear genes and are regulated by post-transcriptional processes [53]. The psbA gene encoded the reaction center protein D1, which is important to the biogenesis and functional maintenance of PSII in
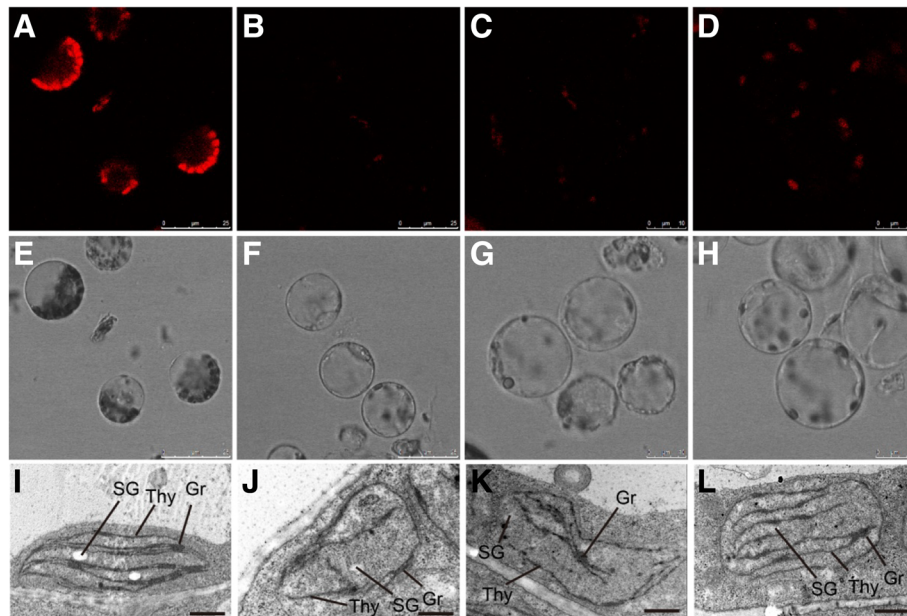
Fig. 5 Loss of GhYGL1d expression affects chloroplast development. **a-h** Chlorophyll fluorescence and morphology of the protoplasts isolated from the CLCrVA (**a**, **e**), GhYGL1d-RNAi-1(**b**, **f**), GhYGL1d-RNAi-2 (**c**, **g**) and GhYGL1d-RNAi-3 (**d**, **h**) plant leaves. (**i-l**) Transmission electron micrographs of plastid ultrastructures in the CLCrVA (I), GhYGL1d-RNAi-1(J), GhYGL1d-RNAi-2 (K) and GhYGL1d-RNAi-3 (L) plants. Plastids were from leaves of six-week-old plants. Three biological replicates were performed, and similar results were obtained. SG, starch grain. Thy, thylakoid. Gr, granum. Bars = 1 μm.
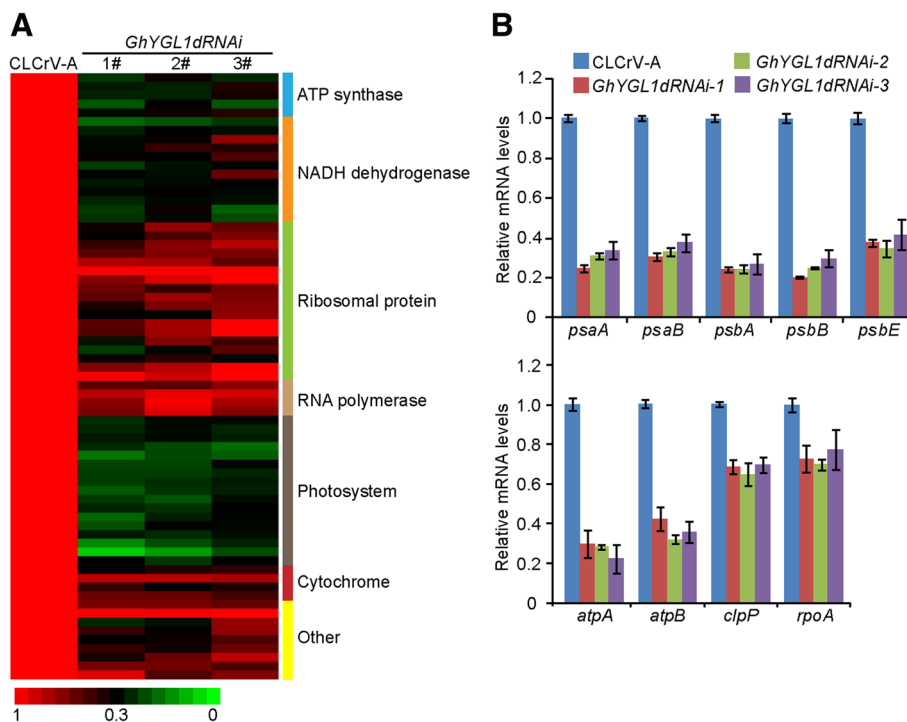
Fig. 6 GhYGL1d regulates plastid function-related genes. **a** Plastid transcriptomic comparison of CLCrVA and GhYGL1d-RNAi variegated leaves. **b** Expression analysis of plastid-encoded genes in CLCrVA and GhYGL1d-RNAi plants. Transcription levels were measured via quantitative real-time RT-PCR, and GhUBQ7 was used as a reference. Mean and SD values were obtained from three replicates.
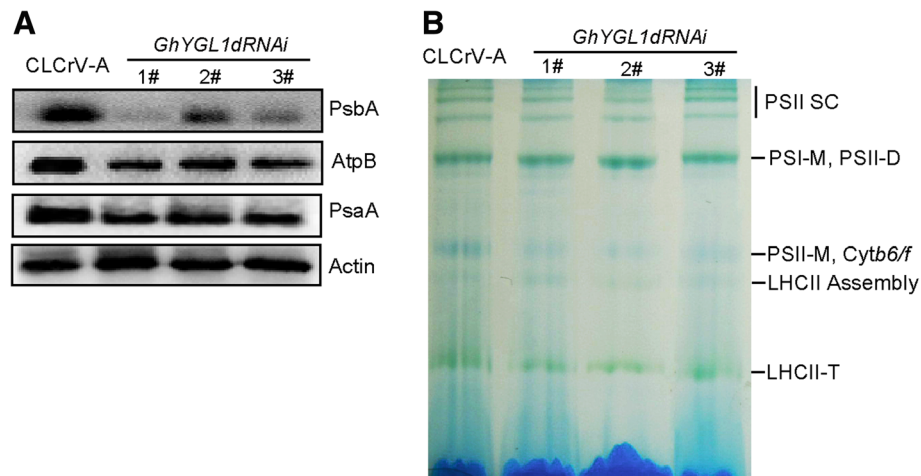
He *et al. BMC Plant Biology* (2019) 19:350

Page 8 of 12

Fig. 7 Analysis of photosynthetic complexes in RNAi plants and CLCrVA control plants. **a** Immunoblot analysis of photosynthetic proteins accumulated in CLCrVA and GhYGL1d-RNAi variegated leaves. Actin was used to check the difference in sample loading. **b** BN-PAGE analysis of photosynthetic complexes in CLCrVA and GhYGL1d-RNAi variegated leaves. Each lane was loaded with equal amounts of thylakoid membrane. PSI SC, PSI supercomplexes; PSI-M, PSI monomers; PSII-D, PSII dimers; PSII-M, PSII monomers; Cytb6/f, cytochrome f; and LHCII-T: PSII LHC trimers.

higher plants. Multiple factors, PPR protein, LPE1, regulated D1 translation, and the down-expression of D1, resulted in a noticeable retardation of photoautotrophic growth [46]. Thus, the reduced growth and photosynthetic activity of GhYGL1d-RNAi plants probably resulted from comprehensive effects including inhibited assembly of PSI and PSII supercomplexes and other indirect targets. Additionally, we found that the transcription levels of both PEP dependent genes (PEPs) and NEP dependent genes (NEPs) were decreased in *GhYGL1*-silenced cotton (Fig. 6b), which is similar to the result that the transcription of PEPs (*psaA, psaB, psbA,* and *rbcL*) and NEPs (*rpoA, rpoB, rps7* and *accD*) were reduced in rice al1 mutant [54]. There are two possible reasons, one is that the decreased transcription level of NEPs is the second effects of PEP reduction. The other is that chloroplast-to-nucleus retrograde signaling may affect the expression of nuclear encoded genes [55].

GhYGL1d may be involved in RNA editing of chloroplast genes

Pentatricopeptide repeat proteins with a DYW domain in the C terminus were reported to be involved in RNA editing [47, 12]. For example, CRR22 and CRR28 genes played important roles in the editing of ndhB7, ndhD5 and rpoB3 [17]. In Arabidopsis, AtECB2 was involved in chloroplast transcript-RNA editing in accD and ndhF [50]. The gene accD encoded the b-subunit of the acetyl-CoA carboxylase complex, which is important for fatty acid synthesis. The gene ndhF encoded subunit F of NADH dehydrogenase. Both genes that lost their editing sites in transcripts corresponded to the albino phenotype

in Arabidopsis [40]. In our research, RNA editing sites of accD and ndhF transcripts in the atecb2 mutant were rescued and more clearly resembled that of wild-type plants when the GhYGL1d gene was introduced into the mutant (Fig. 3d). This result suggests that GhYGL1d is involved in RNA editing of accD and ndhF transcripts. Both GhYGL1d and AtECB2 have similar sequences and might share the same biological functions in plastid development. We detected the editing events of plastid-genes transcripts and found that the editing efficiencies of accD-812 and ndhF-290 were reduced in CLCrVA-GhYGL1d RNAi plants compared with that in the *CLCrVA* plants (sss1 Fsigure S3). The results may be attributed to the cotton material used in the experiment: an RNAi plant in which the knock-down of GhYGL1d may have been insufficient to completely abolish RNA editing. As a result, the GhYGL1d RNAi cotton leaves exhibited yellow-green variegation instead of albinism. In the future, we plan to knock out GhYGL1d in cotton using CRISPR-Cas9 to further uncover its effects on RNA editing and assembly of PSI and PSII. In addition, determining which editing factors interact with GhYGL1d will help resolve the mechanisms of how GhYGL1d is involved in RNA editing.

Conclusionss

The *GhYGL1d* gene has a DYW domain in C-terminal and is highly express in leaves, localized to the chloroplast fractions. Silencing of *GhYGL1d* led to the reduction of chlorophyll and phenotypically yellow-green leaves in cotton. Compared with wild type, *GhYGL1d*-silenced cotton showed significant deformations of thylakoid structures. Furthermore, the transcription levels of

plastid-encoded polymerase- (PEP) dependent genes were significantly decreased in *GhYGL1d*-silenced cotton. Our data showed that *GhYGL1d* not only contributes to the editing of *accD* and *ndhF* genes, but also affects the expression of NEP- and PEP-dependent genes to regulate the development of thylakoids, and therefore regulates leaf variegation in cotton.

Materials and Methods
Plant Materials and Growth Conditions
Gossypium hirsutum (Xuzhou 142) seeds were planted into pots containing soil and grown in a climate chamber with 16:8 day:night light cycle at 30°C, as previously reported [56]. Two-week-old cotton plants were used for *Agrobacterium*-mediated VIGS assay. After inoculation, plants were transferred into a growth chamber set at 23°C. A total of nine seedlings were used for each treatment and three biological triplicates were performed per assay.

Arabidopsis T-DNA insertion lines of *atecb2* (SALK_112251, AT1G15510) in the Col-0 background were obtained from the ABRC (Arabidopsis Biological Resource Center) and identified by PCR with gene-specific primers. The plants were grown in standard soil in a growth chamber under a 12 h light/12 h dark cycle, 60–70% humidity, and a constant temperature of 22°C.

Chlorophyll Content Measurement
The total chlorophyll content (ChI a+b) of leaves was measured as previous described [28]. Approximately 100 mg FW of leaf samples were ground in a mortar, then the homogenate was transferred to a tube with 5 ml of acetone 80% (v/v) and kept in the dark overnight. After that, the mixture were centrifuged at 2600 g for 15 min 4°C. The supernatant was transferred to a clean tube and used for chlorophyll content measurement. The chlorophyll content was determined by absorption measurements at 645 nm and 663 nm and calculated according to the following equation: C_{a+b} (mg/g) = $[20.29A_{645}+8.04A_{663}] \times V/1000 \times W$, where V represents the volume of the extraction buffer (ml) and W is the weight of the fresh leaves (g).

Virus-induced gene silencing (VIGS)
To knock down the expression of *GhYGL1d*, a 325-bp fragment of *GhYGL1d* cDNA was PCR-amplified using TransStart FastPfu DNA Polymerase (Transgen, Beijing) and gene specific primers. The PCR product was cloned into the pCLCrVA vector to produce a VIGS vector named pCLCrVA-*GhYGL1d* and the empty pCLCrVA vector was used as control. The vectors pCLCrVB and pCLCrVA-*GhYGL1d* were introduced into the *Agrobacterium* strain LBA4404. For the VIGS array, the co-transformed *Agrobacterium* colonies containing pCLCrVB and pCLCrVA-*GhYGL1d* were grown for 24 h at 28°C on medium containing the proper antibiotic to select for transformants. *Agrobacterium* cells were collected and resuspended in infiltration medium (10 mM $MgCl_2$, 10 mM MES, and 200 mM acetosyringone) and adjusted to an OD_{600} of 1.2. *Agrobacterium* cells containing pCLCrVB and pCLCrVA-*GhYGL1d* were mixed at a ratio of 1:1. The resuspended cells were injected into cotton cotyledons (approximately 10 days after germination) via a syringe. Thereafter, the plants were grown at 22°C in a growth chamber under a 16-h light, 8-h dark cycle.

RNA isolation and quantitative RT-PCR analysis
Total RNA was isolated from cotton leaves using an RNA isolation kit (RNAqueous™ Total RNA Isolation Kit, Invitrogen™). RNA concentration was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific). The RNA was digested with RNase-Free DNase (Invitrogen) and checked for integrity by capillary gel electrophoresis. After that, total RNA was used to synthesize cDNA by One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen). Next, qRT-PCR was performed using SYBR Green PCR Master Mix (Takara). *GhUBQ7* was selected as the reference gene and each experiment had three biological repeats, each with three technical replicates.

Subcellular localization of GhYGL1d
To investigate the subcellular localization of GhYGL1d, the open reading frame of *GhYGL1d* without a stop codon was then cloned into the pTF486 expression vector to generate a GhYGL1d-GFP fusion protein using a pEASY-Uni Seamless Cloning and Assembly Kit (Transgen). The fusion constructs GhYGL1d-GFP and empty vector 35S-GFP were introduced into *Arabidopsis* leaf protoplasts via polyethylene glycol-mediated transformation as described previously [57]. The transformed protoplasts were incubated at 23°C for 18 h. The fluorescence was visualized using a Leica TCS SP8 fluorescence confocal microscope. The GFP was visualized with excitation at 488 nm and emission at 505-530 nm. Chlorophyll fluorescence was visualized with excitation at 488 nm and emission at 650-710 nm, whereas the mitochondria was stained with MitoTracker® Orange (M7510, Invitrogen) and visualized with excitation at 543 nm and emission at 560-600 nm.

Chloroplast fractionation
Chloroplast fraction was isolated as described previously with modifications [36]. Briefly, 20-day-old plants were homogenized with ice cold extraction buffer (0.33 M sorbitol, 5 mM $MgCl_2$, 5 mM EGTA, 5 mM EDTA, 50 mM HEPES-KOH pH 8.0 and 10 mM $NaHCO_3$), the

He *et al. BMC Plant Biology*   (2019) 19:350

Page 10 of 12

suspension was filtered through Miracloth and centrifuged at 2600 *g* at 4°C for 5 min. The pellets were resuspended in a extraction buffer and loaded onto Percoll step gradients (40%/70% Percoll, 0.33 M sorbitol, 2 mM EDTA, 50 mM HEPES, pH 8.0), and the collected chloroplast fractions were washed twice with wash buffer (0.33 M sorbitol and 50 mM HEPES-KOH, pH 8.0). Intact chloroplasts were further fractionated into stromal and thylakoid membrane fractions as previously reported [37].

Transmission electron microscope (TEM) analysis

To assess the changes of plastid structure, we conducted a TEM analysis. The young leaves were fixed with 4% glutaraldehyde for over 48 h at 4°C and washed three times with PBS buffer. After that the sample was fixed with 1% osmic acid for 4 h and washed three times with PBS buffer. The sample was dehydrated using different concentration ethanol (30%, 50%, 70%, 90% and 100%). Next, it was embed and aggregated with ethoxyline resin and cut into slices. After stained with uranyl acetate and alkaline lead citrate for 15 min, the sample was observed using a Model H-7650 transmission electronmicroscope (HITACHI).

Immunoblot analysis

Fresh leaves were ground into a fine powder in liquid nitrogen and used for protein extraction. The extracted protein samples were resolved on 12% SDS polyacrylamide gel electrophoresis (PAGE) gels, and then transferred to polyvinylidene difluoride membranes (0.45 μm, Millipore), followed by incubation with specific antibodies. The polyclonal antibodies of GFP, Cyt f, and RbcL were obtained from Agrisera.

Thylakoid isolation and blue-native polyacrylamide gel electrophoresis

The Thylakoid was isolated as described previously with modification [38]. Thylakoids were isolated from fresh leaves with ice cold extraction buffer. The suspension was filtered through Miracloth and centrifuged at 2600 *g* at 4°C for 5 min. The pellet was resuspended in a wash buffer followed by centrifugation at 2400 *g* at 4°C for 5 min. Finally, the thylakoid pellet was suspended in storage buffer and treated with n-Dodecyl-beta-D-maltoside for 15 min. The insoluble material was removed by centrifugation at 16,000 *g* at 4°C for 10 min. For Blue Native-Polyacrylamide Gel Electrophoresis (BN-PAGE), 15 μg of solubilized thylakoid protein was mixed with a sample buffer and loaded onto a 4–15% native-PAGE gradient gel. Electrophoresis was performed at 4°C at a constant 120 V for 2.5 h.

Additional file

Additional file 1: Figure S1. A virus-induced gene silencing (VIGS) assay for nine PPR-DYW genes. **Figure S2.** Alignment of amino acid sequences of the highest identity with the GhYGL1d protein. **Figure S3.** The editing efficiency of *accD*-812 and *ndhF*-290 in CLCrVA and GhYGL1d-RNAi plants. **Table S1.** Subcellular localization prediction of 72 PPR-DYW proteins in cotton. **Table S2.** Primer sequences used in this study (PDF 1207 kb)

Additional file 2: Table S3. RNA-seq dataset of CLCrVA and GhYGL1d-RNAi plants. (XLSX 17526 kb)

Abbreviations

PPR: pentatricopeptide repeat; VIGS: virus-induced gene silencing; GFP: green fluorescent protein; TEM: transmission electron microscope; PEP: plastid-encoded polymerase; NEP: nuclear-encoded polymerase; BN-PAGE: blue native polyacrylamide gel electrophoresis; cTP: chloroplast transit peptide

Author details
[1]College of Life Sciences, Shaanxi Normal University, Xi'an 710119, China. [2]Shanxi Academy of Agricultural Sciences, Cotton Research Institute, Yucheng 044000, China. [3]Key Laboratory of the Ministry of Education for Medicinal Plant Resources and Natural Chemistry, National Engineering Laboratory for Resource Development of Endangered Crude Drugs in the Northwest of China, College of Life Sciences, Shaanxi Normal University, Xi'an 710119, China.

References
1. Kobayashi K, Endo K, Wada H. Roles of Lipids in Photosynthesis. Subcell Biochem. 2016;86:21–49.
2. Wang K, Froehlich JE, Zienkiewicz A, Hersh HL, Benning C. A plastid phosphatidylglycerol lipase contributes to the export of acyl groups from plastids for seed oil biosynthesis. Plant Cell. 2017;29(7):1678–96.
3. Nelson N, Yocum CF. Structure and function of photosystems I and II. Annu. Rev. Plant Biol. 2006;57:521–65.

4.   Marín-Navarro J, Manuell AL, Wu J, P Mayfield S. Chloroplast translation regulation. Photosynth. Res. 2007;94(2-3):359–74.

5.   Ruwe H, Kupsch C, Teubner M, Schmitz-Linneweber C. (2011) The RNA-recognition motif in chloroplasts. J. Plant Physiol. 2011;168(12):1361–71.

6.   Zoschke R, Kupsch C, Schmitz-Linneweber C. RNA-binding proteins required for chloroplast RNA processing. In: *Plant Mitochondria*. New York, NY: Springer; 2011.

7.   Shi X, Bentolila S, Hanson MR. Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in *Arabidopsis*. Plant Signal Behav. 2016;11(5):e1167299.

8.   Wang S, Bai G, Wang S, Yang L, Yang F, Wang Y, Zhu JK, Hua J. (2016) Chloroplast RNA-binding protein RBD1 promotes chilling tolerance through 23S rRNA processing in *Arabidopsis*. Plos Genet . 2016;12(5):e1006027.

9.   Börner T, Zhelyazkova P, Legen J, Schmitz-Linneweber C. Chloroplast gene expression: RNA synthesis and processing. Plastid Biology. 2014;5:3–47.

10. Sun T, Bentolila S, Hanson MR. The unexpected diversity of plant organelle RNA editosomes. Trends Plant Sci. 2016;21(11):962–73.

11. Barkan A, Small I. (2014) Pentatricopeptide repeat proteins in plants. Annu. Rev. Plant Biol. 2014;65:415–42.

12. Fujii S, and Small I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol. 2011;191(1):37-47

13. Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. Plant Cell. 2009; 21(1):146–56.

14. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère 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. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell. 2004;16(8):2089–103.

15. Rivals E, Bruyère C, Toffano-Nioche C, Lecharny A. Formation of the *Arabidopsis* pentatricopeptide repeat family. Plant Physiol. 2006;141(3):825–39.

16. 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. Plant J. 2013;74(6):935–45.

17. Okuda K, Shikanai T. A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids. Nucleic Acids Res. 2013;40(11):5052–64.

18. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. HCF152, an *Arabidopsis* RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. Plant Cell. 2003; 15(6):1480–95.

19. Hattori M, Miyake H, Sugita M. A pentatricopeptide repeat protein is required for RNA processing of clpP pre-mRNA in moss chloroplasts. J Biol Chem. 2017;282(14):10773-10782.

20. Dai D, Luan S, Chen X, Wang Q, Feng Y, Zhu C, Qi W, Song R. Maize Dek37 encodes a P-type PPR protein that affects cis-splicing of mitochondrial nad2 intron 1 and seed development. Genetics. 2018; 208(3):1069-1082.

21. Ren X, Pan Z, Zhao H, Zhao J, Cai M, Li J, Zhang Z, Qiu F. (2017) EMPTY PERICARP11 serves as a factor for splicing of mitochondrial nad1 intron and is required to ensure proper seed development in maize. J. Exp. Bot. 2017; 68(16):4571–81.

22. Sun Y, Huang J, Zhong S, Gu H, He S, Qu LJ. Novel DYW-type pentatricopeptide repeat (PPR) protein BLX controls mitochondrial RNA editing and splicing essential for early seed development of Arabidopsis. J Genet Genomics. 2018;45(3):155-168.

23. McCormick S. Binding sites for pentatricopeptide repeat proteins differentially activate chloroplast transgenes. Plant J. 2018;94(1):6-7.

24. Ito A, Sugita C, Ichinose M, Kato Y, Yamamoto H, Shikanai T, Sugita M. An evolutionarily conserved P-subfamily pentatricopeptide repeat protein is required to splice the plastid *ndhA* transcript in the moss *Physcomitrella patens* and *Arabidopsis thaliana*. Plant J. 2018;94(4):638–48.

25. Arenas-M A, González-Durán E, Gómez I, Burger M, Brennicke A, Takenaka M, Jordana X. The Pentatricopeptide repeat protein MEF31 is required for editing at site 581 of the mitochondrial tatC transcript and indirectly influences editing at site 586 of the same transcript. Plant Cell Physiol. 2018; 59(2):355-365.

26. Xiao H, Xu Y, Ni C, Zhang Q, Zhong F, Huang J, Liu W, Peng L, Zhu Y, Hu J. A rice dual-localized pentatricopeptide repeat protein is involved in organellar RNA editing together with OsMORFs. J. Exp. Bot. 2018;69(12):2923–36.

27. B Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Mol Cell Biol. 2008;28(17):5337-5347.

28. Zhang Z, Cui X, Wang Y, Wu J, Gu X, Lu T. The RNA editing factor WSP1 is essential for chloroplast development in rice. Mol. Plant. 2017;10(1):86–98.

29. Legen J, Ruf S, Kroop X, Wang G, Barkan A, Bock R, Schmitz-Linneweber C. Stabilization and translation of synthetic operon-derived mRNAs in chloroplasts by sequences representing PPR protein-binding sites. Plant J. 2018;94(1):8–21.

30. Williams PM, Barkan A. A chloroplast-localized PPR protein required for plastid ribosome accumulation. Plant J. 2013;36(5):675-86.

31. Yang YZ, Ding S, Wang HC, Sun F, Huang WL, Song S, Xu C, Tan BC. The pentatricopeptide repeat protein EMP9 is required for mitochondrial *ccmB* and *rps4* transcript editing, mitochondrial complex biogenesis and seed development in maize. New Phytol. 2017;214(2):782–95.

32. Glass F, Härtel B, Zehrmann A, Verbitskiy D, Takenaka M. MEF13 requires MORF3 and MORF8 for RNA editing at eight targets in mitochondrial mRNAs in *Arabidopsis thaliana*. Mol. Plant. 2015;8(10):1466–77.

33. Kotera E, Tasaka M, Shikanai T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature. 2005;433(7023):326–30.

34. Xie T, Chen D, Wu J, Huang X, Wang Y, Tang K, Li J, Sun M, Peng X. *Growing Slowly 1* locus encodes a PLS-type PPR protein required for RNA editing and plant development in *Arabidopsis*. J. Exp. Bot. 2016;67(19):5687–98.

35. Zhang HD, Cui YL, Huang C, Yin QQ, Qin XM, Xu T, He XF, Zhang Y, Li ZR, Yang ZN. PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid genes in Arabidopsis thaliana. Photosynth. Res. 2015;126(2-3):311–21.

36. Du L, Zhang J, Qu S, Zhao Y, Su B, Lv X, Li R, Wan Y, Xiao J. The pentatricopeptide repeat protein pigment-defective mutant 2 is involved in the regulation of chloroplast development and chloroplast gene expression in *Arabidopsis*. Plant Cell Physiol. 2017;58(4):747–59.

37. Zhang J, Xiao J, Li Y, Su B, Xu H, Shan X, Song C, Xie J, Li R. PDM3, a pentatricopeptide repeat-containing protein, affects chloroplast development. J. Exp. Bot. 2017;68(20):5615–27.

38. Tang J, Zhang W, Wen K, Chen G, Sun J, Tian Y, Tang W, Yu J, An H, Wu T, Kong F, Terzaghi W, Wang C, Wan J. OsPPR6, a pentatricopeptide repeat protein involved in editing and splicing chloroplast RNA, is required for chloroplast biogenesis in rice. Plant Mol. Biol. 2017;95(4-5):345–57.

39. Ichinose M, Sugita M. The DYW domains of pentatricopeptide repeat RNA editing factors contribute to discriminate target and non-target editing sites. Plant Cell Physiol. 2018;59(8):1652–9.

40. Huang C, Yu QB, Li ZR, Ye LS, Xu L, Yang ZN. Porphobilinogen deaminase HEMC interacts with the PPR-protein AtECB2 for chloroplast RNA editing. Plant J. 2017;92(4):546–56.

41. Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, Delannoy E, Hammani K, Small I, Huang J. The Arabidopsis gene *YS1* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. Plant J. 2009;58(1):82–96.

42. Nakamura T, Sugita M. A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. FEBS lett. 2008; 582(30):4163–8.

43. Tasaki E1, Hattori M, Sugita M. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. Plant J. 62:560–70.

44. Ichinose M, Tasaki E, Sugita C, Sugita M. (2012) A PPR-DYW protein is required for splicing of a group II intron of *cox1* pre-mRNA in *Physcomitrella patens*. Plant J. 2010;62(4):560–70.

45. Kim SR, Yang JI, Moon S, Ryu CH, An K, Kim KM, Yim J, An G. Rice *OGR1* encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. Plant J. 2009;59(5):738–49.

46. Yang L, Zhu H, Guo W, Zhang T. Molecular cloning and characterization of five genes encoding pentatricopeptide repeat proteins from Upland cotton (*Gossypium hirsutum* L.). Mol. Biol. Rep. 2010;37(2):801–8.

47. Yan J, Zhang Q, Yin P. RNA editing machinery in plant organelles. Sci. China Life Sci. 2018;61(2):162–9.

48. Zhang B, Liu G, Li X, Guo L, Zhang X, Qi T, Wang H, Tang H, Qiao X, Zhang J, Xing C, Wu J. A genome-wide identification and analysis of the DYW-deaminase genes in the pentatricopeptide repeat gene family in cotton (*Gossypium* spp.). PloS one. 2017;12(3):e0174201.

49. Bentolila S, Babina AM, Germain A, Hanson MR. Quantitative trait locus mapping identifies REME2, a PPR-DYW protein required for editing of specific C targets in *Arabidopsis* mitochondria. RNA Biol. 2013;10(9):1520–5.

50. Yu QB, Jiang Y, Chong K, Yang ZN. AtECB2, a pentatricopeptide repeat protein, is required for chloroplast transcript *accD* RNA editing and early chloroplast biogenesis in *Arabidopsis thaliana*. Plant J. 2009;59(6):1011–544.

51. Wang R, Zhao J, Jia M, Xu N, Liang S, Shao J, Qi Y, Liu X, An L, Yu F. (2018) Balance between cytosolic and chloroplast translation affects leaf variegation. Plant Physiol. 2018;176(1):804–18.

52. Cao ZL, Yu QB, Sun Y, Lu Y, Cui YL, Yang ZN. A point mutation in the pentatricopeptide repeat motif of the AtECB2 protein causes delayed chloroplast development. J. Integr. Plant Biol. 2011;53(4):258–69.

53. Cai W, Ji D, Peng L, Guo J, Ma J, Zou M, Lu C, Zhang L. LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. Plant Physiol. 2009; 150(3):1260–71.

54. Zhang Z, Tan J, Shi Z, Xie Q, Xing Y, Liu C, Chen Q, Zhu H, Wang J, Zhang J, Zhang G. *Albino Leaf1* that encodes the sole octotricopeptide repeat protein is responsible for chloroplast development. Plant Physiol. 2016; 171(2):1182–91.

55. Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J. Signals from chloroplasts converge to regulate nuclear gene expression. Science. 2007;316(5825):715–9.

56. He P, Zhao P, Wang L, Zhang Y, Wang X, Xiao H, Yu J, Xiao G. The *PIN* gene family in cotton (*Gossypium hirsutum*): genome-wide identification and gene expression analyses during root development and abiotic stress responses. BMC Genomics. 2017;18(1):507.

57. Sheen J. Signal transduction in maize and *Arabidopsis* mesophyll protoplasts. Plant Physiol. 2001;127(4):1466–75.

Publisher's Note