Structural and functional similarities and differences in nucleolar Pumilio RNA-binding proteins between Arabidopsis and Charophyte Chara corallina

Su Hyun Park
Myongji University - Natural Science Campus
Hyung-Sae im
Myongji University - Natural Science Campus
Prakash Jyoti Kalita
Myongji University
Sang-Bong Choi (choisb@mju.ac.kr)
Myongji University https://orcid.org/0000-0001-6040-2472

Research article

Keywords: Arabidopsis thaliana, Chara corallina, Charophyte, ITS2, Puf, RNA-binding proteins, rRNA

Posted Date: November 27th, 2019

DOI: https://doi.org/10.21203/rs.2.17798/v1

Version of Record: A version of this preprint was published at BMC Plant Biology on May 24th, 2020. See the published version at https://doi.org/10.1186/s12870-020-02444-x.

Abstract

Background: Pumilio RNA-binding proteins are evolutionarily conserved throughout all eukaryotes and are involved in RNA decay, transport, and translation repression in the cytoplasm. Even though a majority of Pumilio proteins function in the cytoplasm, two nucleolar forms have been reported to have a function in rRNA processing in Arabidopsis. Chara is known to be most closely related to land plants as it shares several characteristics with modern embryophytes.

Results: In this study, we identified two putative nucleolar Pumilio protein genes, ChPUM2 and ChPUM3 , from the transcriptome of Chara corallina . Of the two ChPUM proteins, ChPUM2 was the most similar to Arabidopsis APUM23 in the amino acid sequence (27% identity and 45% homology) and the predicted protein structure, while ChPUM3 was similar to APUM24 (35% identity and 54% homology). The transient expression of 35S:ChPUM2-RFP and 35S:ChPUM3-RFP showed nucleolar localization of fusion proteins in tobacco leaf cells, similar to the expression of APUM23-GFP and APUM24-GFP. Moreover, 35S:ChPUM2 complemented the morphological defects of the apum23 phenotypes but not those of apum24, while 35S:ChPUM3 could not complement the apum23 and apum24 mutants. Similarly, the 35S:ChPUM2/apum23 plants rescued the pre-rRNA processing defect of apum23 , but 35S:ChPUM3/apum24 +/- plants did not rescue that of apum24 . Consistent with these complementation results, a known target RNA-binding sequence at the end of the 18S rRNA (5'-GAAUUGACGG) for APUM23 was conserved in Arabidopsis and Chara, whereas a target region of ITS2 pre-rRNA for APUM24 was 156 nt longer in Chara than in Arabidopsis. Moreover, ChPUM2 and APUM23 were predicted to have nearly identical structures, but ChPUM3 and APUM24 have different structures in the 5th C-terminal Puf RNA-binding domain as ChPUM3 has a long random coil in this domain.

Conclusions: Our results indicate that ChPUM2 has evolutionarily maintained functions in Arabidopsis, while ChPUM3 is not functional in Arabidopsis, most likely due to the long target ITS2 sequence of Arabidopsis and the distinct tertiary structure from Arabidopsis homologue APUM24.

Background

Pumilio proteins are a family of RNA-binding proteins that are evolutionarily conserved in eukaryotes [1]. Typical Pumilio proteins have tandem repeats of 8 Puf domains that recognize 8 RNA bases, and each Puf domain contains 35 ~ 39 amino acids that comprise three α-helical structures [2, 3]. The basis of RNA recognition by these proteins is the crescent-shaped structure, in which the conserved aromatic and basic amino acids on the concave side interact with RNA, whereas the amino acids on the convex side interact with partner proteins [4, 5]. Although Pumilio proteins have a variety of biological roles, their molecular functions are mRNA decay, localization, translational repression [6, 7], and rRNA processing [8]. Most of the Pumilio proteins are localized in the cytoplasm and are involved in the posttranscriptional regulation of mRNA. In contrast, only a small subset of Pumilio proteins is localized in the nucleolus and participates in rRNA processing. Nucleolar Nop9 of yeast [9] and TbPUF7 of trypanosome [8] are involved in 18S rRNA biosynthesis and ribosome maturation through proper pre-rRNA processing. In plants, two

nucleolar Pumilio proteins are implicated in rRNA processing [10–15] and include APUM23, a homologue of Nop9, and APUM24, a homologue of human Puf-A and yeast Puf6. APUM23 is not only required for normal growth patterning including leaf development and organ polarity [10, 16] but also involved in ABA signaling [17]. APUM24 is essential for plant development, as its homozygous mutant displays embryo lethality [15]. APUM23 participates in the processing of 18S and 5.8S rRNAs, while APUM24 is implicated in the maturation of 5.8S and 25S rRNAs.

pre-rRNA is a long single-stranded RNA transcribed from the rDNA repeat in the nucleolus. This transcript is subsequently cleaved to three mature rRNAs (5.8S, 18S, and 25S) by endoribonucleolytic activities [18, 19]. Misprocessed rRNA byproducts that are produced during rRNA processing are degraded by 5' to 3' and 3' to 5' exoribonucleolytic activities. These two pre-rRNA processing activities require additional accessory proteins, such as RNA exosome components, Pumilio, and many RNA-binding proteins. It has been reported that Arabidopsis and rice show similar pre-rRNA processing pathways, probably due to the similar flanking sequences around the endocleavage sites of A2 and A3 in the ITS1 [20], suggesting that RNA-binding specificity is important for the selection of cleavage sites. Recently, APUM23 was found to bind 11 nt in the 18S rRNA at positions 1141–1151 [12], and APUM24 interacts with rRNA segments encompassing the 5.8S and ITS2 region [15]. Therefore, it is very likely that APUM23 and APUM24 play crucial roles in the recruitment of target RNA sequences and interacting proteins for the maturation of 18S and 5.8S rRNAs in Arabidopsis.

Approximately 450 million years ago, land plants evolved from Charophytes living in freshwater and adapted to the terrestrial environment [21–24]. Charophyta and Embryophyta share molecular and physiological similarities that are not found in other green algae [25, 26]. Compared to other green algae, Charophyta and Embryophyta have higher similarities with land plants in plastidal atpB and rbcL, mitochondrial nad5, and nuclear-encoded small subunit rRNA genes [27].

In this study, all viridiplantae species that were examined have two nucleolar Pumilio proteins that are homologous to Arabidopsis APUM23 and APUM24. We found two nucleolar Pumilio genes in C. corallina, ChPUM2 and ChPUM3, that are homologous to Arabidopsis APUM23 and APUM24, respectively. We thought that these two Pumilio proteins are evolutionarily conserved as their pre-rRNA processing roles are crucial in the ribosome biogenesis required for proper protein synthesis. Therefore, we examined the nucleolar localization of ChPUM2-RFP and ChPUM3-RFP and performed complementation assays of Arabidopsis apum23 and apum24 mutants using 35S:ChPUM2 and 35S:ChPUM3. Our results showed that ChPUM2 and ChPUM3 are located on the nucleolus and that ChPUM2 rescued the rRNA processing defect of apum23, but ChPUM3 did not complement apum23 and apum24 mutants.

Results

Phylogeny of nucleolar Pumilio proteins

Pumilio proteins are a family of evolutionarily conserved RNA-binding proteins found in all eukaryotes. Among the organisms whose whole genome sequences are available, higher plants have a higher number

of Pumilio proteins than photosynthetic single-cell organisms and nonplant organisms; for example, 25 Pumilio proteins are found in Arabidopsis thaliana, 20 in Oryza sativa, 14 in Physcomitrella patens, 5 in Chlamydomonas reinhardtii, 11 in Caenorhabditis elegans, 7 in Saccharomyces cerevisiae, and 2 in humans [10]. Based on the similarity search using Arabidopsis nucleolar Pumilio proteins (APUM23 and APUM24) as queries and the existence of nucleolar localization signal(s) (NoLS) [28], viridiplantae species, including green algae, were examined and found to have two putative nucleolar Pumilio proteins. Using our transcriptome data obtained from PacBio iso-seq, we also identified two putative nucleolar Pumilio proteins, ChPUM2 and ChPUM3, in Chara corallina.

To gain insight into the evolutionary relationship of putative nucleolar Pumilio proteins, we constructed a phylogenic tree from the homologous proteins of representative viridiplantae species (Fig. 1a). All land plants and green algae analyzed in this study have a single protein homologous to APUM23 and APUM24. Chara ChPUM2 was closer to APUM23 than APUM24, whereas ChPUM3 was categorized as being in the APUM24 clade. Our phylogenetic tree showed that Chara ChPUM2 and ChPUM3 are more homologous to the Bryophyta counterparts than to the potential nucleolar Pumilios of other land plants.

We then examined the structural similarity and numbers of Puf domains in ChPUM2 and ChPUM3 and Arabidopsis nucleolar Pumilios using their primary sequences in a SMART web program (http://smart.embl-heidelberg.de) [29]. APUM23 and APUM24 have six and five Puf domains, respectively; ChPUM2 and ChPUM3 have one and five Puf domains, respectively. As the number of Puf domains may determine the RNA base numbers, this observation raises the possibility that ChPUM2 has a distinct structure-function relationship from APUM23 and that ChPUM3 may be an APUM24 homologue (Fig. 1b).

Structure of ChPUM2 and ChPUM3

Classical structural analysis of Pumilio proteins shows the typical motif of 8 Puf domains in the C-terminal end of the protein [30, 31]; however, a more detailed analysis using human Puf-A as a representative domain showed the existence of 3 additional Puf domains at the N-terminus [14]. Therefore, we analyzed the domain homology according to Qiu et al. (2014), which enabled us to identify cryptic Puf domains that had a low homology with typical amino acid sequences in the flanking regions of the Puf core domain. It was previously reported that APUM23 and its budding yeast homologue Nop1 have 10 Puf domains when analyzed using this tool [12, 32]. Consistently, all 14 plant proteins were categorized into the ChPUM2 and APUM23 families (Fig. 1a) and contained 10 Puf domains that are distributed at irregular intervals (Fig. 2a and Additional file 1a). Each domain of ChPUM2 showed an average 34.3% identity with the corresponding domains of APUM23 with a 12.9% gap. Despite low homology between the entire sequences of each domain of ChPUM2 and APUM23, a high degree of homology was found in the 1st, 2nd, and 5th residues of the 2nd α-helix, and these residues are pivotal for recognizing RNA (red box in Fig. 2a and Additional file 1a). Overall, the Puf domains among ChPUM2 relatives are well conserved throughout the viridiplantae species examined in this study when considering the RNA recognition residues.

Using the same methodology, all the proteins that belong to the ChPUM3 and APUM24 clades (Fig. 1a) possessed 11 Puf domains (N-R1 ~ N-R3 in the N-terminus and C-R1 ~ C-R8 in the C-terminus), as reported in human Puf-A and APUM24 [14]. ChPUM3 displayed an average of 47.9% identity with APUM24 in the N-terminal Puf domains but 34.4% identity and 14.7% gap in the C-terminus. Puf domains in the N-terminus of ChPUM3 showed high homology to APUM24 in the 1st, 2nd, and 5th residues of the 2nd α-helix, but those on the C-terminus of ChPUM3 had low homology to APUM24 (red box in Fig. 2b and Additional file 1b).

Although the comparison of amino acid sequences among these Puf domains showed that ChPUM2 and ChPUM3 are homologous nucleolar proteins to APUM23 and APUM24 (Fig. 2a-b), it cannot be ruled out that ChPUM2 has a distinct RNA-binding activity from APUM23 as ChPUM2, unlike APUM23, has only a single Puf domain in the classical domain analysis (Fig. 1b). Therefore, we used the SWISS-MODEL web server (https://swissmodel.expasy.org/) to predict the tertiary structure of these proteins. It is known that APUM23 has a long chain between the 2nd and 3rd α-helix of the 3rd Puf domain and has a twisted C-shaped structure [12]. Our prediction showed a high coverage of the ChPUM2 tertiary structure with the APUM23 reference and a twisted C-shaped structure that is similar to APUM23. Although ChPUM2 has many unfolded chains in the 3rd Puf domain compared with APUM23, it has nearly identical α-helix structures to APUM23 that contribute to RNA binding (Fig. 3a). ChPUM3 and APUM24 had an L-shape structure that was similar to that found in the human Puf-A reference [14] (Fig. 3b). However, ChPUM3 had a prominently longer random coil in the C-R5 domain than APUM24 (Fig. 2b and Fig. 3b) and contained negative (E265) and uncharged (Q305) amino acids in the N-R2 and N-R3 domains of patch 1B (Fig. 2b), unlike Puf-A and APUM24. Overall, ChPUM2 had a similar structure to APUM23 except for the unfolded structures positioned outside the core sequence of the 2th, 4th, and 8th Puf domains, and ChPUM3 had a longer side chain in the C-R5 domain than APUM24 but also contained distinct amino acids in the N-terminal Puf domains that were distinct from those in APUM24.

Subcellular localization of ChPUM2 and ChPUM3

Next, we examined the subcellular localization of ChPUM2-RFP and ChPUM3-RFP. Previously, the GFP fusions of APUM23 and APUM24 were known to preferentially localize in the nucleoli of the Arabidopsis root and tobacco leaf cells [10, 15]. We performed Agrobacterium-mediated coinfiltration into N. benthamiana leaf cells using 35S:APUM23-GFP and 35S:ChPUM2-RFP, and using 35S:APUM24-GFP and 35S:ChPUM3-RFP. All the GFP- and RFP-tagged Pumilio proteins were found in the nucleolus and weakly in the nucleoplasm, probably due to high constitutive expression under the 35S promoter (Fig. 4), indicating that ChPUM2 and ChPUM3 have nucleolar functions, such as pre-rRNA recognition and processing. Our results obtained from the structural prediction and localization assays suggested a strong possibility that ChPUM2 and ChPUM3 are Chara orthologues of Arabidopsis APUM23 and APUM24, respectively.

Expression of 35S:ChPUM2 restored apum23 mutant phenotype to normal

As APUM23 and APUM24 are closely related to ChPUM2 and ChPUM3, respectively, it is likely that apum23 and apum24 mutants could be complemented by the heterologous expression of ChPUM2 and ChPUM3. To address this possibility, we generated 35S:ChPUM2 and 35S:ChPUM3 plants in the apum23-2 mutant background that shows the phenotypes of delayed germination, light green leaves, short stem and roots, and streptomycin resistance. The 35S:ChPUM2/apum23-2 plants exhibited normal phenotypes, whereas the 35S:ChPUM3/apum23-2 plants showed an apum23-2 phenotype (Fig. 5a-d). The 35S:ChPUM2/apum23-2 plants recovered seed germination efficiency, root length, plant height, and color and shape of rosette leaves. Moreover, complemented plants were sensitive to streptomycin, making them similar to control Col-0 plants and suggesting that defects in ribosome biogenesis in the apum23-2 mutant might be recovered (Fig. 5D). The failure to restore the apum23-2 phenotype to the wild-type phenotype by 35S:ChPUM2 excluded the possibility of orthologues of ChPUM3 and APUM23.

In addition to the restoration of morphological phenotypes, 35S:ChPUM2/apum23-2 complemented the defects observed in the apum23-2 mutant that accumulates polyadenylated [poly(A)] 18S-ITS1 and 5.8S-ITS2 pre-rRNAs (Fig. 5e). We synthesized the oligo (dT)-primed cDNAs from 35S:ChPUM/apum23-2 and control Col-0 plants to measure the amount of poly(A) pre-rRNA as PCR templates. For the detection of unprocessed poly(A) rRNA, RT-PCR was performed using three primer sets for 5'ETS-18S, 18S-ITS1, and 5.8S-ITS2. In the apum23-2 mutant, all three quantitative RT-PCR products (5'ETS-18S, 18S-ITS1, and 5.8S-ITS2) were more accumulated than those in the Col-0 control. However, in 35S:ChPUM2/apum23-2 plants, the amounts of poly(A) 18S-ITS1 and 5.8S-ITS2 pre-rRNAs were greatly reduced compared with those of apum23-2, although that of poly(A) 5.8S-ITS2 was slightly decreased. In contrast to 35S:ChPUM2/apum23-2, 35S:ChPUM3/apum23-2 showed a nearly identical amount of poly(A) pre-rRNAs to the mutant (Fig. 5e, right panel). Hence, qRT-PCR results using poly(A) pre-rRNA indicated that ChPUM2 is involved in pre-rRNA processing that is similar to that in APUM23, although it was not fully functional for removing the 5.8S pre-rRNA byproducts.

Chara nucleolar Pumilio protein genes did not rescue the Arabidopsis apum24 mutant

APUM24, a homologue of ChPUM3, is involved in 5.8S rRNA processing and is essential for cell division and pattern formation in early embryogenesis [13, 15, 33]. To examine if nucleolar ChPUMs recovered the apum24 mutant to wild-type, we performed a complementation experiment using a heterozygous apum24-1$^{+/-}$ mutant line. Homozygous 35S:ChPUM2 and 35S:ChPUM3 transgenics on the apum24-1$^{+/-}$ background set the normal and aborted seeds at the same rate as the apum24$^{+/-}$ mutant (Fig. 6a-c and Table 1). Consistent with this result, 35S:ChPUM2$^{+/+}$/apum24-1$^{+/-}$ and 35S:ChPUM3$^{+/+}$/apum24-1$^{+/-}$ plants accumulated poly(A) 5.8S pre-rRNA similar to that in the apum24-1$^{+/-}$ plants (Fig. 6d), indicating that Chara ChPUM3 is not a functional orthologue of Arabidopsis APUM23 and APUM24.

Table 1
The ratio of normal seeds to aborted seeds.

Genotype	Total seeds (n*)	Normal seeds (n)	Aborted seeds + undeveloped ovules (n)	Ratio (N : A&U**)
Control (pB2GW7)	3016	2869	147	19.5 : 1
apum24-1$^{+/-}$	3043	2108	935	2.25 : 1
35S:ChPUM2 / apum24-1$^{+/-}$	3035	2050	985	2.08 : 1
35S:ChPUM3 / apum24-1$^{+/-}$	2988	1997	991	2.01 : 1
*n, numbers; **A&U, normal seeds : aborted seeds & undeveloped ovules.				

ChPUM2 restored the salt- and glucose-hypersensitive phenotypes of apum23-2, but ChPUM3 did not recover apum23-2 and apum24-1

The apum23 and apum24 have changes in the expression levels of their ribosomal biosynthetic genes, thereby resulting in a hypersensitivity to high concentrations of salt or glucose [13, 17] (Fig. 7). The 35S:ChPUM2/apum23-2 seedlings exhibited a similar degree of resistance to 150 mM NaCl and 200 mM glucose to that found in wild-type seedlings, while the 35S:ChPUM3/apum23-2 seedlings showed NaCl- and glucose-susceptibility that was similar to the apum23-2 seedlings (Fig. 7a). However, 35S:ChPUM2/apum24-1$^{+/-}$ and 35S:ChPUM3/apum24-1$^{+/-}$ failed to recover the salt sensitivity of apum24-1$^{+/-}$. Unexpectedly, when 35S:ChPUM2 or 35S:ChPUM3 was overexpressed in apum24-1$^{+/-}$, their transgenic seedlings showed a similar glucose resistance to that found in wild-type seedlings (Fig. 7a). It was previously reported that APUM24 gene expression was greatly increased in wild-type plants by exogenously supplying glucose [13, 15]. Our data showed similar expression levels of the APUM24 transcript in the control, apum24-1$^{+/-}$, 35S:ChPUM2/apum24-1$^{+/-}$, and 35S:ChPUM3/apum24-1$^{+/-}$ (Fig. 7b), indicating a haploid sufficiency of the APUM24 gene for the expression of the glucose-induced phenotype. Consistently, apum24-1$^{+/-}$ and transgenic plants reduced the amount of unprocessed 5.8S rRNA to normal control levels under high concentrations of glucose (Fig. 7c). Our results on the salt and glucose treatments support that ChPUM2 complements apum23-2 and ChPUM3 does not complement apum23-2 and apum24-1.

Discussion

From our transcriptome data, public database, and previous results [10, 15, 34], we found that green plant species have two nucleolar Pumilio proteins that are structurally well conserved throughout viridiplantae. The phylogeny was constructed on the basis of primary amino acid sequences and showed a closer relationship of nucleolar Pumilio proteins of land plants with C. corallina than with single cell and

multicellular Chlorophytes. Although it is not completely ruled out that there is a possibility that putative nucleolar Pumilios proteins other than the Chara and Arabidopsis counterparts do not localize in the nucleolus despite having nucleolar localization signals, land plant species appear to have evolved to have two Pumilio proteins for the scavenging of aberrant pre-rRNAs tagged with poly(A) tails. This is consistent with a previous report that C. corallina is more related to land plants than other green algae in other genes (atpB, rbcL and nad5) and small subunit rRNA [27].

We demonstrated that ChPUM2 functionally resembled APUM23 in Arabidopsis cells as evident from the complementation of the defective pre-rRNA processing and morphology of apum23. Nucleolar Pumilio proteins play roles in recognizing the target sequences on pre-rRNA and recruiting proteins, including ribonuclease, in Arabidopsis [10, 12, 13, 15]. Although the RNA sequence bound to ChPUM2 was not examined in our study, a known target sequence (5'-GAAUUGACGG-3') of APUM23 is located in the 18S rRNA of C. corallina at positions 1141–1151 [12] (Additional file 2). Therefore, ChPUM2 is very likely to bind this site that Chara and Arabidopsis have in common. This conclusion is further supported by the primary structures of these proteins, in which five residues in the 2nd α-helix of each Puf domain that are known to interact with target RNA bases are very similar between APUM 23 and ChPUM2 (Fig. 2a and Additional file 1).

However, given that ChPUM2 has an identical target sequence in Chara and Arabidopsis, it may not completely remove the polyadenylated 18S and 5.8S pre-rRNAs that are accumulated in the apum23 mutant (Fig. 5e). This might be due to a weak interaction of ChPUM2 with other unknown protein components compared with that of intrinsic APUM23 and/or due to the 156 nt shorter ITS2 in Arabidopsis than in Chara that is not fully recognized by ChPUM2 for the production of mature 5.8S rRNA. Indeed, the predicted structure of ChPUM2 has more unfolded chains than APUM23 (Fig. 3a), which thereby interferes with the interaction of ChPUM2 with other protein components that are required for efficient processing. To verify this possibility in planta, it is worthwhile to identify the components interacting with CPUM2 in Chara and APUM23 in Arabidopsis.

At first, ChPUM3 appeared structurally similar to APUM24 (Fig. 3b); however, ChPUM3 did not functionally replace APUM24 in Arabidopsis. We assume that the cause of this result is from the detailed structural differences between ChPUM3 and APUM24. Typical Pumilio proteins bind to a specific RNA base with the second α-helix of the Puf domain, but APUM24 and its homologues are not capable of binding to a specific RNA base through this α-helix domain [14, 15]. The reasons why ChPUM3 does not complement the apum24 mutant might be due to: (1) a very long random coil in the C-R5 domain, (2) the negative and uncharged amino acids in the N-terminal domains, and (3) a long ITS2 sequence in Chara. First, a long random chain at the C-R5 domains of ChPUM2 would interrupt the interaction of other C-terminal domains with RNA bases in the 5.8S-ITS junction of Arabidopsis pre-rRNA. It was reported that human Puf-A and its homologous APUM24 have a long random coil in C-R5 that prevents the C-terminal Puf domains from binding to RNA (Qiu et al., 2014). ChPUM3 has an 80 aa longer random coil than APUM24 (Fig. 2b); thus, ChPUM3 is not able to recognize Arabidopsis pre-rRNA. Second, in ChPUM3, the N-R2 and N-R3 of patch 1B include negative (E210) and uncharged (Q249) amino acids, unlike positive amino

acids (K) at both positions, which may result in differential binding characteristics from APUM24 towards the 5.8S-ITS2 region. Interestingly, the side chain of the C-R5 domain is very long in the green algae C. corallina, C. reinhardtii, and V. carteri, but not O. lucimarinus, when compared to land plants (Fig. 2b). Third, ChPUM3 might be adjusted to recognize the long ITS2 sequence for proper processing of the 3'-extended 5.8S pre-rRNA. The length of ITS2 is 156 nt longer in Chara than in Arabidopsis (Additional file 3). It may be possible that the long side chain of ChPUM3 may prevent ChPUM3 from binding to the short ITS2 sequence of Arabidopsis. Indeed, ITS2 evolved rapidly and has been used to evaluate the genetic divergence [35, 36].

Conclusions

In this study, we identified two nucleolar Pumilio proteins (ChPUM2 and ChPUM3) from Chara and showed that they are phylogenetically and structurally related to the Arabidopsis nucleolar Pumilios proteins, APUM23 and APUM24, respectively. Complementation analyses using *35S:ChPUM2* and *35S:ChPUM3* demonstrated that *ChPUM2* and *APUM23* are functional orthologues, but *ChPUM3* did not restore the *ampm24* mutant. Consistent with these complementation results, ChPUM2 showed a similar structure to APUM23 in 3D protein modeling, but ChPUM3 has a long random coil in the C-R5 domain and has distinct amino acids from those in APUM24 in the N-terminal domains. The ITS2 sequence of Arabidopsis is 156 nt shorter than that of Chara and therefore might not be recognized by ChPUM3 for proper 5.8S pre-rRNA processing. Further studies on pre-rRNA processing in viridiplantae species and Pumilio-interacting proteins will help us analyze the patterns of rRNA evolution from green algae to land plants.

Methods

Plant materials and growth conditions

Chara corallina was collected in South Korea (38°20’N, 128°30’E) and grown at room temperature in a small aquarium. Genomic DNA and voucher specimen of *C. corallina* were identified and deposited in National Institute of Biological Resources (https://www.nibr.go.kr/) under the number, NIBRGR0000609814. The *apum23* [10] and *apum24* [15] mutants, obtained from Arabidopsis Biological Resources Center and reported previously, were used for complementation analyses. The *35S:ChPUM2* and *35S:ChPUM3* transgenic Arabidopsis plants with *apum23* and *apum24* backgrounds were produced by transformation using *Agrobacterium tumefaciens* GV3101 with the floral dipping method [37]. *A. thaliana* wild-type Col-0 and control (Col-0 transformed with pB2GW7), *apum23-2*$^{-/-}$, *apum24-1*$^{+/-}$, *35S:ChPUM2*, and *35S:ChPUM3* overexpression lines were grown on MS medium or in soil in a conditioned room at 22°C under 16 h light (120 μmol photons $m^{-2}s^{-2}$) and 8 h dark cycles. For testing antibiotics, salt, and glucose resistance, seeds were grown on 1/2 MS plates supplemented with 50 μg mL^{-1} streptomycin, 150 mM NaCl, and 200 mM glucose, respectively, in a growth room for 10 to 12 days. All seeds were stratified at 4°C for 3 days before sowing.

Identification of *ChPUM2* and *ChPUM3* transcripts

The sequences for Chara Pumilio proteins (ChPUMs) were obtained by searching our PacBio iso-seq transcriptome data that were generated from entire *C. corallina* plants, including rhizoids, globules, and nucules. Expression of ChPUMs was verified using reverse transcription polymerase chain reaction (RT-PCR). We then identified ChPUMs that are homologous to Arabidopsis nucleolar Pumilio proteins (APUM23 and APUM24) and analyzed to have nucleolar localization signal (NoLS) [28].

Phylogenic analysis of ChPUM2 and ChPUM3 homologues

To perform the phylogenetic analysis, full-length amino acid sequences of the homologous proteins of ChPUM2, ChPUM3, APUM23, and APUM24 were obtained from representative species of viridiplantae in Phytozome (v 12.1) [34], and *Klebsormidium flaccidum* in the Klebsormidium genome database (http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome _project/klebsormidium/), and two red algae species in the Ensembl Plant database (http://plants. ensembl.org) were used. A phylogenetic tree was constructed by using the maximum likelihood method based on the LG+G+F model with MEGA7 software [38, 39]. Amino acid sequence alignments were performed using ClustalW and edited using BioEdit software.

Plasmid construction

For the construction of *35S:ChPUMs* plasmids, coding sequences (CDSs) for *ChPUM2* and *ChPUM3* were amplified by RT-PCR using the primers ChPUM2-F and ChPUM2-R1 and ChPUM3-F and ChPUM3R1, respectively (see Additional file 4 for the sequences of these primers). The PCR products were inserted into the pENTR-D-TOPO vector (Invitrogen) and then transferred to pB2GW7 or pK2GW7 (Vlaams Instituut voor Biotechnologie, Ghent University) by Gateway™ LR Clonase II (Invitrogen). For the construction of *35S:ChPUM-RFP* and *35S:APUM-GFP* plasmids, CDSs of *ChPUM2*, *ChPUM3*, *APUM*23, and *APUM24* were amplified by RT-PCR using the primer combinations of ChPUM2-F/ChPUM2-R2, ChPUM3-F/ChPUM3-R2, APUM23-F/APUM23-R, and APUM24-F/APUM24-R, respectively. PCR products were inserted into the pENTR-D-TOPO vector and then transferred to the pB7RWG2 or pK7FWG2 vector [40] by Gateway™ LR Clonase II.

Colocalization assay of ChPUM and APUM fusion proteins

Transient expression of C-terminal RFP fusions of ChPUM2 and ChPUM3 and C-terminal GFP fusions of APUM23 and APUM24 was performed using agroinfiltration into *Nicotiana benthamiana* leaves [41].

Cultures of Agrobacterium carrying fusion constructs were harvested at the stationary phase and resuspended in MMA buffer (10 mM MES, 10 mM $MgCl_2$, and 150 µM acetosyringone) to OD_{600} = 0.8. For coexpression of ChPUM-RFP and APUM-GFP, equal volumes of two Agrobacterium cultures that have either the *35S:ChPUM-RFP* or *35S:APUM-GFP* vector were mixed before infiltration. Infiltration was performed on the abaxial side of tobacco leaves using a needless syringe. Plants were kept in the dark at 22°C under high humidity for 30-34 h, and the infiltrated leaves were observed under a fluorescence microscope.

Quantitative RT-PCR (qRT-PCR) for analyzing unprocessed rRNA

Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, cat. # 74904) and treated with DNase I (Ambion, cat. # AM2238) at 37°C for 50 min. First-strand cDNA was synthesized from 5 µg total RNA using the oligo $(dT)_{18}$ primer in a 20 µL reaction and diluted 3-fold. Then, 1 µL of cDNA was mixed with 0.6 µL of 10 mM primers and 10 µL of 2 x SYBR® Green Supermix (Bio-Rad, cat. # 172-5261) in a 20 µL reaction and subjected to PCR according to the manufacturer's instructions. Tubulin (*Tub4,* At5g44340) cDNA was used as an internal control. For qPCR measurements, two technical and three biological replicates were used. Data were calculated using the $2^{-\Delta\Delta CT}$ method [42].

Declarations

Abbreviations

C-R: C-terminal repeat; ITS: Internal Transcribed Sequence; MS: Murashige and Skoog; N-R: N-terminal repeat; pre-rRNA: pre-ribosomal RNA**.**

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Sequence data for the cDNAs of *ChPUM2* and *ChPUM3* described in this study can be found in the GenBank database (https://www.ncbi.nlm.nih.gov/) under the accession numbers, …. and …., respectively. The transcriptome dataset (SRX….) analyzed during the identification of ChPUM2 and ChPUM3 transcripts are available in GenBank database.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financially supported by grants from the Next-Generation BioGreen21 Program, Rural Development Administration (No. PJ01366), South Korea, to SBC. The funding agencies supported this research projects, but played no role in the design of the study, data analysis, interpretation, and writing the manuscript. These were the sole responsibilities of the authors.

Authors' contributions

SHP and SBC designed the experiments and wrote the article. SHP generated the constructs and performed the complementation assays. HSK assisted in the subcellular localization study. PJK helped with the complementation of the heterozygous mutants. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Jungho Lee for the advice on the culture conditions.

References

1. Spassov DS, Jurecic R. The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? IUBMB Life. 2003;55:359-66.
2. Zhang B, Gallegos M, Puoti A, Durkin E, Fields S, Kimble J, Wickens MP. A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. Nature. 1997;390:477-

84.

3. Miller MT, Higgin JJ, Hall TM. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat Struct Mol Biol. 2008;15:397-402.
4. Edwards TA, Pyle SE, Wharton RP, Aggarwal AK. Structure of Pumilio reveals similarity between RNA and peptide binding motifs. Cell 2001;105:281-9.
5. Quenault T, Lithgow T, Traven A. PUF proteins: repression, activation and mRNA localization. Trends Cell Biol. 2011;21:104-12.
6. Murata Y, Wharton RP. Binding of pumilio to maternal *hunchback* mRNA is required for posterior patterning in Drosophila embryos. Cell. 1995;80:747-56.
7. Sonoda J, Wharton RP. Recruitment of Nanos to *hunchback* mRNA by Pumilio. Genes Dev 1999;13:2704-12.
8. Droll D, Archer S, Fenn K, Delhi P, Matthews K, Clayton C. The trypanosome Pumilio-domain protein PUF7 associates with a nuclear cyclophilin and is involved in ribosomal RNA maturation. FEBS Lett. 2010;584:1156-62.
9. Thomson E, Rappsilber J, Tollervey D. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. RNA. 2007;13:2165-74.
10. Abbasi N, Kim HB, Park NI, Kim HS, Kim YK, Park YI, Choi SB. APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. Plant J. 2010;64:960-76.
11. Abbasi N, Park YI, Choi SB. Pumilio Puf domain RNA-binding proteins in Arabidopsis. Plant Signal Behav. 2011;6:364-8.
12. Bao H, Wang N, Wang C, Jiang Y, Liu J, Xu L, Wu J, Shi Y. Structural basis for the specific recognition of 18S rRNA by APUM23. Nucleic Acids Res. 2017;45:12005-14.
13. Maekawa S, Ishida T, Yanagisawa S. Reduced expression of *APUM24*, encoding a novel rRNA processing factor, induces sugar-dependent nucleolar stress and altered sugar responses in *Arabidopsis thaliana*. Plant Cell. 2018;30:209-27.
14. Qiu C, McCann KL, Wine RN, Baserga SJ, Hall TM. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. Proc Natl Acad Sci USA. 2014;111:18554-9.
15. Shanmugam T, Abbasi N, Kim HS, Kim HB, Park NI, Park GT, Oh SA, Park SK, Muench DG, Choi Y, et al. An Arabidopsis divergent pumilio protein, APUM24, is essential for embryogenesis and required for faithful pre-rRNA processing. Plant J. 2017;92:1092-105.
16. Huang T, Kerstetter RA, Irish VF. APUM23, a PUF family protein, functions in leaf development and organ polarity in *Arabidopsis*. J Exp Bot. 2014;65:1181-91.
17. Huang KC, Lin WC, Cheng WH. Salt hypersensitive mutant 9, a nucleolar APUM23 protein, is essential for salt sensitivity in association with the ABA signaling pathway in *Arabidopsis*. BMC Plant Biol. 2018;18:40.

18. Henras AK, Plisson-Chastang C, O'Donohue M-F, Chakraborty A, Gleizes P-E. An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip Rev RNA. 2015;6:225-42.
19. Weis BL, Kovacevic J, Missbach S, Schleiff E. Plant-specific features of ribosome biogenesis. Trends Plant Sci. 2015;20:729-40.
20. Hang R, Wang Z, Deng X, Liu C, Yan B, Yang C, Song X, Mo B, Cao X. Ribosomal RNA biogenesis and its response to chilling stress in *Oryza sativa*. Plant Physiol. 2018;177:381.
21. Domozych D, Ciancia M, Fangel J, Mikkelsen M, Ulvskov P, Willats W. The cell walls of green algae: A journey through evolution and diversity. Front Plant Sci. 2012;3:82.
22. Sørensen I, Domozych D, Willats WGT. How have plant cell walls evolved? Plant Physiol. 2010;153:366.
23. Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F, O'Neill MA, Fei Z, Rose JK, Domozych DS, Willats WG. The charophycean green algae provide insights into the early origins of plant cell walls. Plant J. 2011;68:201-11.
24. Domozych D, Sørensen I, Popper ZA. Editorial: Charophytes: Evolutionary ancestors of plants and emerging models for plant research. Frontiers in Plant Science. 2017;8:338.
25. Lewis LA, McCourt RM. Green algae and the origin of land plants. Am J Bot. 2004;91:1535-56.
26. Nishiyama T, Sakayama H, de Vries J, Buschmann H, Saint-Marcoux D, Ullrich KK, Haas FB, Vanderstraeten L, Becker D, Lang D, et al. The Chara genome: secondary complexity and implications for plant terrestrialization. Cell. 2018;174:448-64 e24.
27. Karol KG, McCourt RM, Cimino MT, Delwiche CF. The closest living relatives of land plants. Science. 2001;294:2351-3.
28. Scott MS, Troshin PV, Barton GJ. NoD: a Nucleolar localization sequence detector for eukaryotic and viral proteins. BMC Bioinformatics. 2011;12:317.
29. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 2000;28:231-4.
30. Miller MA, Olivas WM. Roles of Puf proteins in mRNA degradation and translation. Wiley Interdiscip Rev RNA. 2011;2:471-92.
31. Wang X, McLachlan J, Zamore PD, Hall TM. Modular recognition of RNA by a human pumilio-homology domain. Cell. 2002;110:501-12.
32. Wang B, Ye K. Nop9 binds the central pseudoknot region of 18S rRNA. Nucleic Acids Res. 2017;45:3559-67.
33. Maekawa S, Yanagisawa S. Nucleolar stress and sugar response in plants. Plant Signal & Behav. 2018;13:e1442975.
34. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40:D1178-86.

35. Caisova L, Marin B, Melkonian M. A close-up view on ITS2 evolution and speciation - a case study in the Ulvophyceae (Chlorophyta, Viridiplantae). BMC Evol Biol. 2011;11:262.
36. Qin Y, Li M, Cao Y, Gao Y, Zhang W. Molecular thresholds of ITS2 and their implications for molecular evolution and species identification in seed plants. Sci Rep. 2017;7:17316.
37. Clough SJ, Bent AF. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 1998;16:735-43.
38. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Mol Biol Evol. 2008;25:1307-20.
39. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870-4.
40. Karimi M, Inzé D, Depicker A. GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. Trends Plant Sci. 2002;7:193-5.
41. Goodin MM, Dietzgen RG, Schichnes D, Ruzin S, Jackson AO. pGD vectors: versatile tools for the expression of green and red fluorescent protein fusions in agroinfiltrated plant leaves. Plant J. 2002;31:375-83.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25:402-8.

Figures

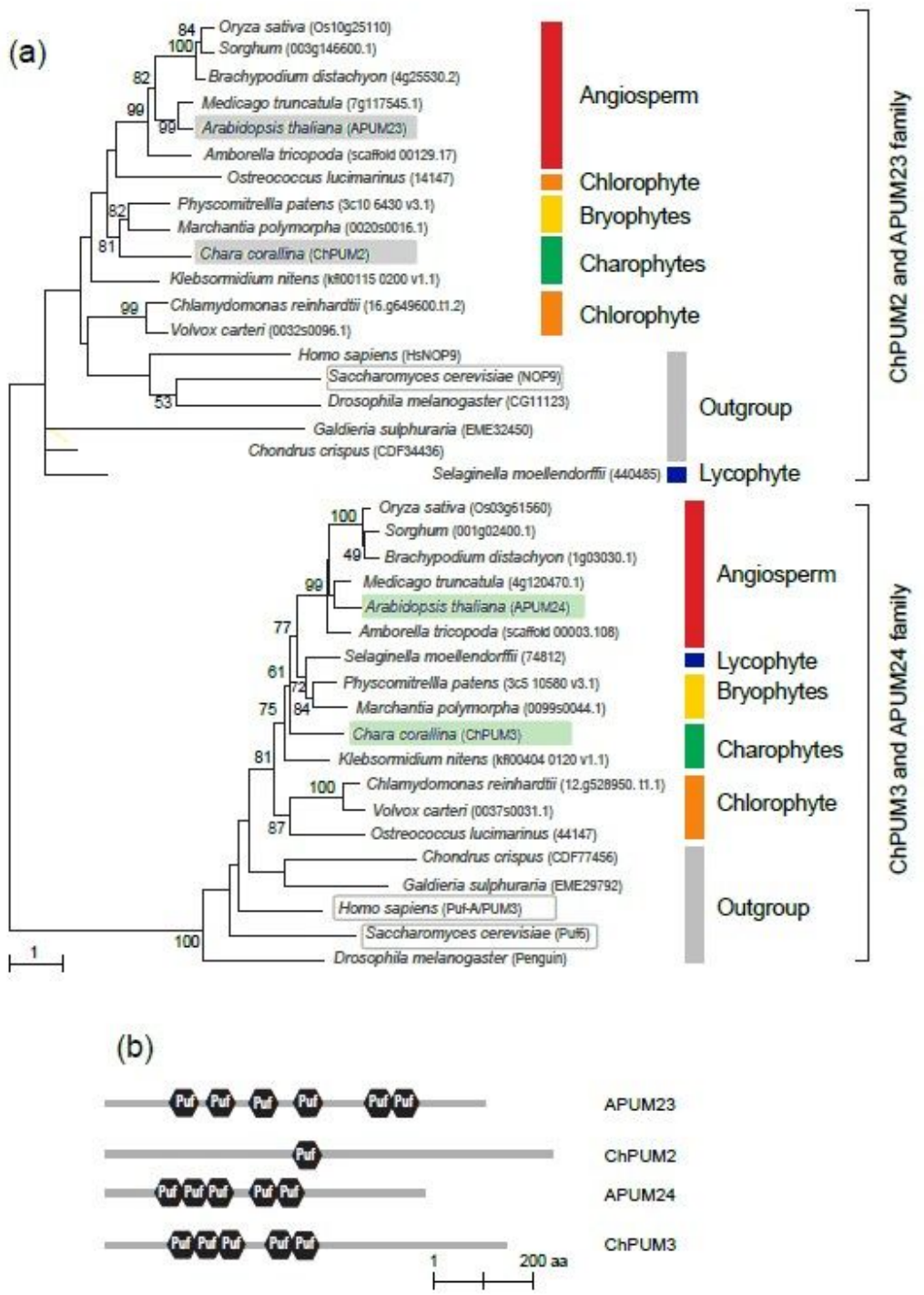

Figure 1

Phylogenetic tree of the putative nucleolar Pumilio proteins in representative viridiplantae species (a) and the protein structures of nucleolar APUMs and ChPUMs (b). a The phylogenetic relationship among the putative nucleolar Pumilio proteins in representative viridiplantae species. The phylogenetic tree was constructed using the maximum likelihood LG+G+F model with MEGA7, and the bootstrap value was 1000. Putative nucleolar Pumilio proteins of red algae (Chondrus crispus and Galdieria sulphuraria),

Drosophila melanogaster, Homo sapiens, and Saccharomyces cerevisiae were used as outgroups. b Primary protein structures of APUM23 and APUM24 from Arabidopsis thaliana and ChPUM2 and ChPUM3 from Chara corallina. Black hexagons indicate Puf RNA-binding domains.

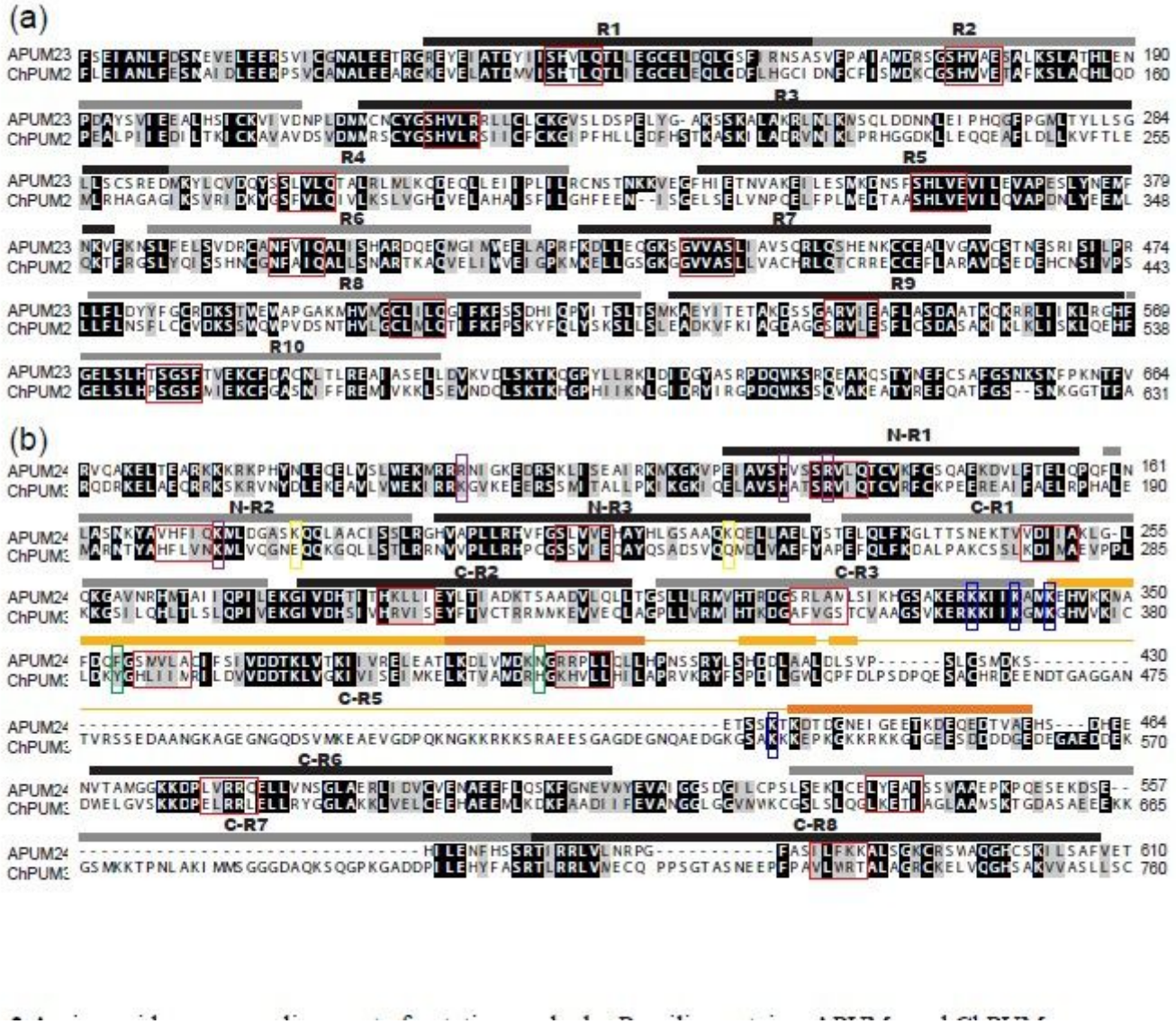

Figure 2

Amino acid sequence alignment of putative nucleolar Pumilio proteins, APUMs and ChPUMs. a Amino acid alignment of APUM23 and ChPUM2. b Amino acid alignment of APUM24 and ChPUM3. The Puf domains are indicated with black and gray sticks above the amino acid sequences, and the five residues in the 2nd α-helix of each Puf domain that potentially interact with RNA bases are indicated with red boxes. Basic amino acids in patches 1A and 1B are boxed with purple and yellow, respectively. Conserved aromatic amino acids are boxed with green, and basic amino acids in the C-terminal region are boxed with blue.

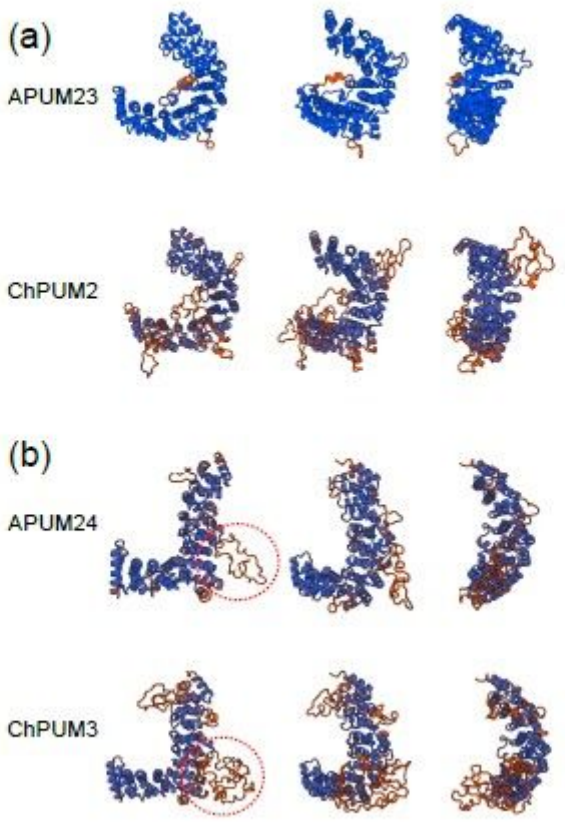

Figure 3

Predicted 3D structures of putative nucleolar APUMs and ChPUMs. a Tertiary structures of APUM23 and ChPUM2. b Tertiary structures of APUM24 and ChPUM3. Unfolded side chains in the C-R5 domain are marked with red circles.

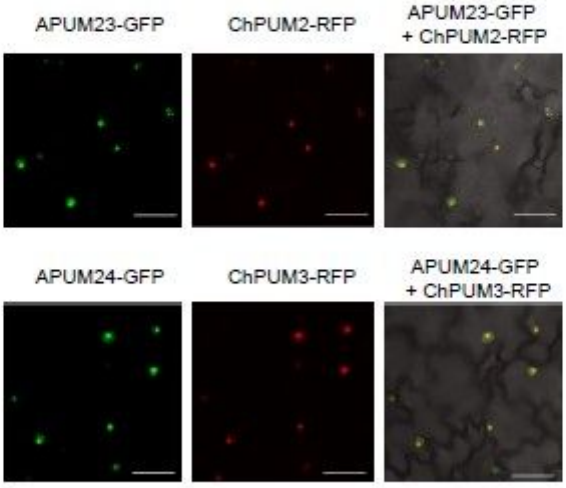

Figure 4

Nucleolar colocalization of APUM23 and ChPUM2 (upper panel) 1 and APUM24 and 2 ChPUM3 (lower panel) in N. benthamiana leaf cells. Scale bars = 50 μm.

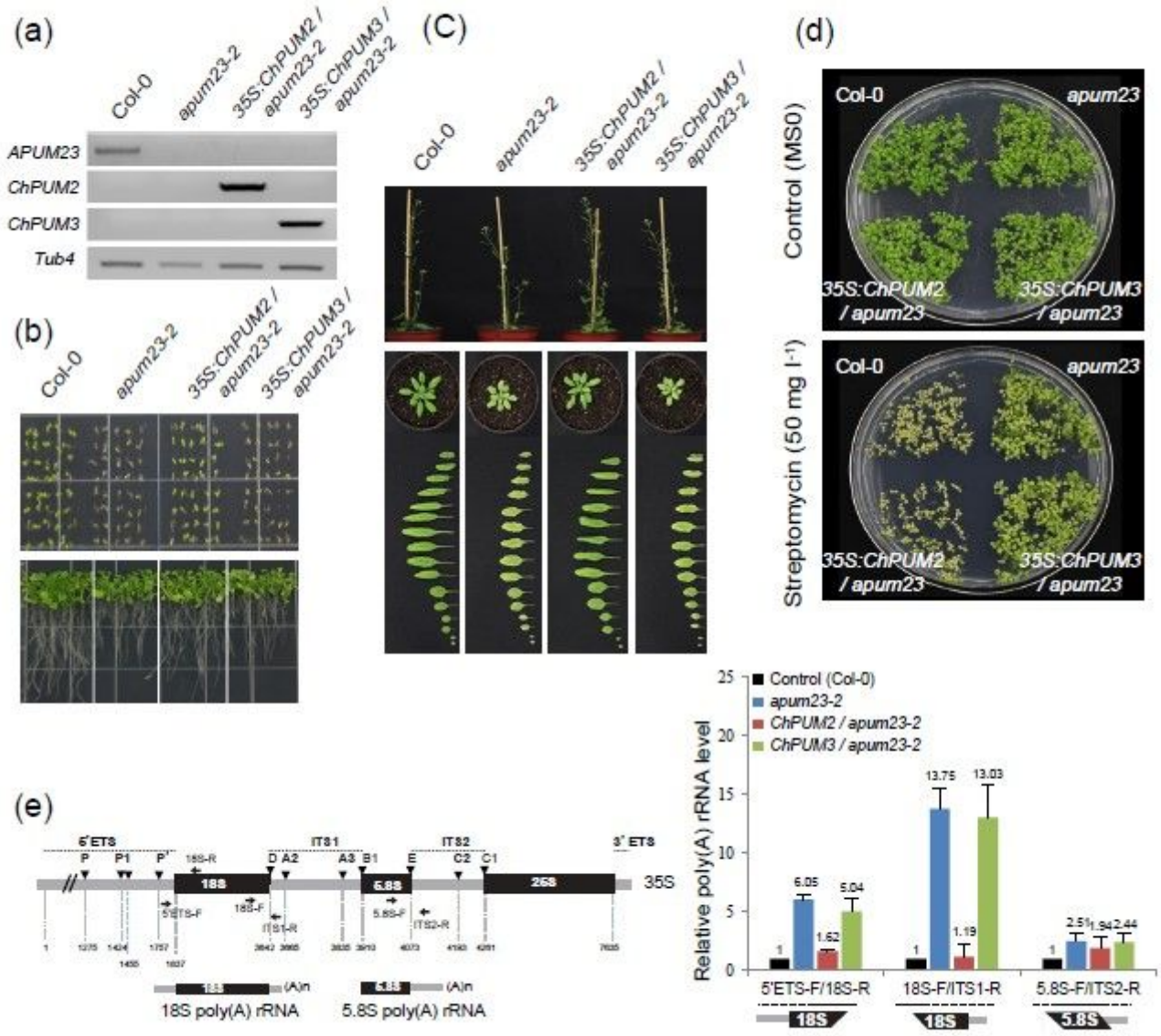

Figure 5

Complementation assays of 35S:ChPUM2 and 35S:ChPUM3 transgenic plants with the apum23 mutant background. a Confirmation of the expression of ChPUM2 and ChPUM3 transgenes in the apum23-2 mutant using RT-PCR. b Normal phenotypes of germination and root growth of the apum23-2 complemented with 35S:ChPUM2. Note the delayed germination in apum23-2 and 35S:ChPUM3/apum23-2 seeds. c Plant heights and rosette leaves in mature plants. Leaves were collected from 2-week-old plants. d Recovery of streptomycin susceptibility in the apum23-2 complemented 35S:ChPUM2. e qRT-PCR analysis for unprocessed rRNAs in wild-type Col-0, apum23-2, 35S:ChPUM2/apum23-2, and 35S:ChPUM3/apum23-2. Two technical and three biological replicates were performed for PCR measurements. Asterisks indicate the results of Student's t-test between apum23-2 and transgenic plants (**; $p<0.01$). Values represent means ± standard deviation, SD (n = 3).

The left panel shows a schematic diagram of the poly(A) prerRNA byproducts and primers. Unprocessed poly(A) 18S (~ 2.6 knt) and 5.8S (~ 300 nt) pre-rRNAs are shown below the 35S pre-rRNA.

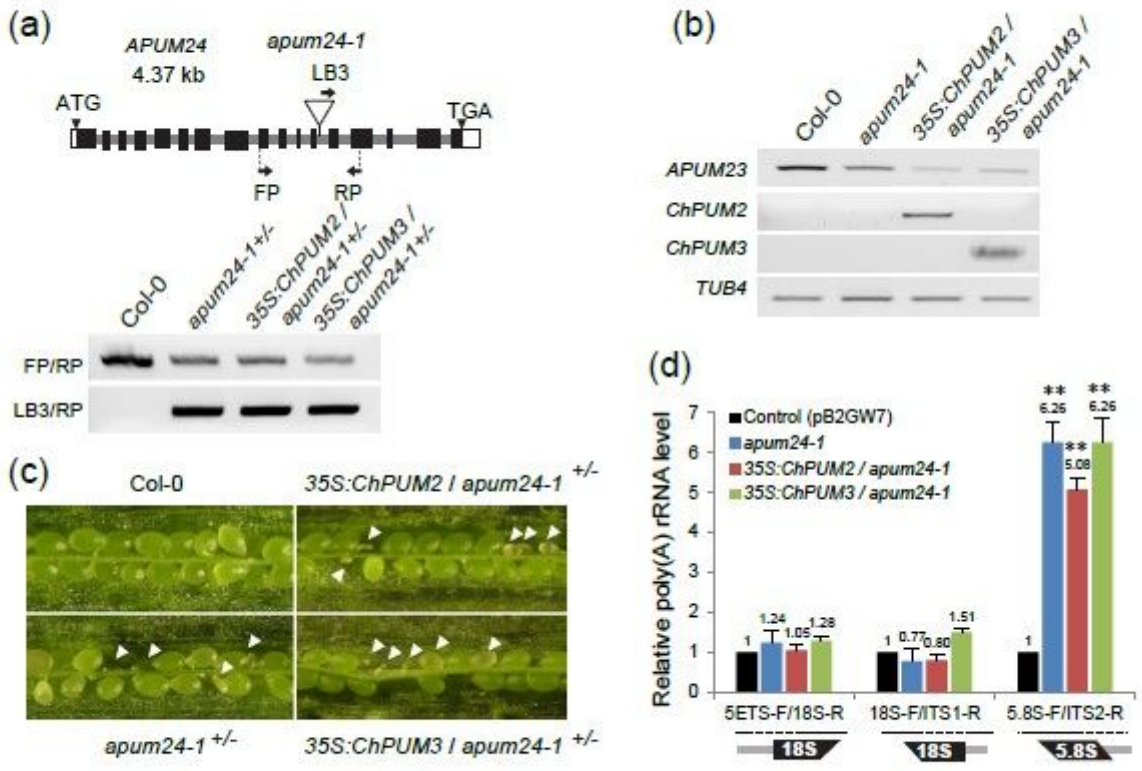

Figure 6

Complementation assays of 35S:ChPUM2 and 35S:ChPUM3 transgenic plants in the apum24+/- mutant background. a T-DNA insertion site of apum24-1 mutant alleles and genotyping. Primers used for genotyping are indicated with arrows. b Confirmation of the expression of ChPUM2 and ChPUM3 transgenes in the apum24-1+/- mutant. c Seed morphology of Control (pB2GW7), apum24-1+/-, and transgenic apum24-1+/- expressing 35S:ChPUM2 or 35S:ChPUM3. Note that none of the transgenics complemented the aborted seeds to normal. d qRT-PCR for analyzing relative unprocessed rRNA levels in Control (pB2GW7), apum24-1+/-, and 35:ChPUMN/apum24-1+/-, using the same primers were used as in Fig. 5. Two technical and three biological replicates were performed for PCR measurements. Values represent means ± SD (n = 3) (**; $p<0.01$).

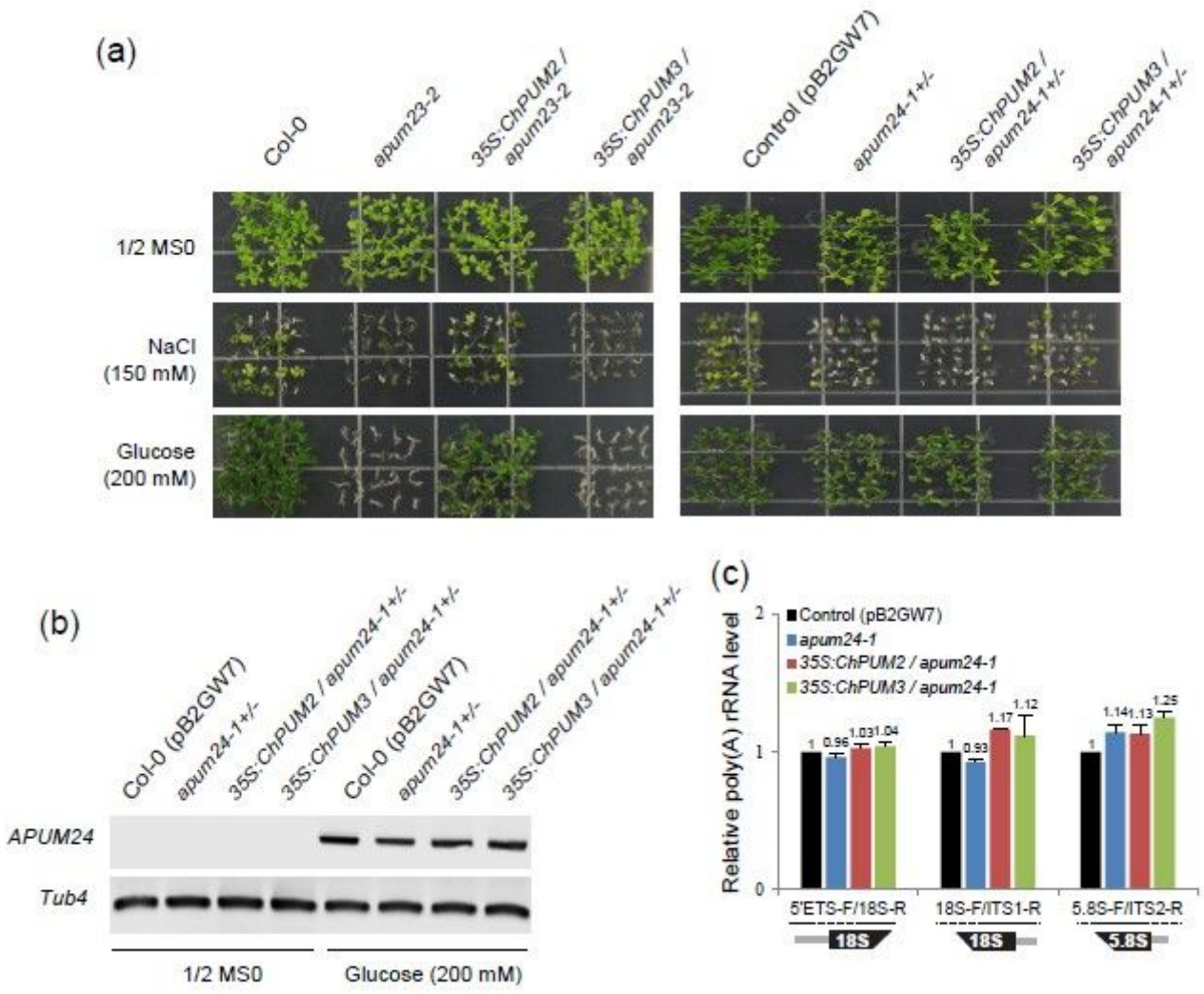

Figure 7

ChPUM2, but not ChPUM3, recovers the salt- and sugar-sensitive apum23-2 phenotype. a Complemented apum23-2 and apum24-1+/- seeds with 35S:ChPUM2 and 35S:ChPUM2. Transgenic apum23-2 seeds expressing 35S:ChPUM2 germinate normally in 1/2 MS medium supplemented with NaCl and glucose (left panel). Transgenic apum24-1+/- seeds expressing ChPUM3 show delayed germination in medium containing NaCl but display normal germination in medium supplemented with glucose (right panel). In the right panel, Control (pB2GW7), apum24-1+/-, and 35S:ChPUM2/apum24-1+/-, and 35S:ChPUM3/apum24-1+/- plants were grown on 1/2 MS medium, which includes NaCl or glucose in the presence of 10 µg l-1 Basta. b Expression levels of APUM24 in Control (pB2GW7), apum24-1+/-, and transgenic apum24-1+/- plants expressing 35S:ChPUM2 and 35S:ChPUM3 in the absence and presence of 200 mM glucose. Note similar expression levels of APUM24 in the presence of 200 mM. c qRT-PCR for analyzing the relative unprocessed rRNA levels in Control (pB2GW7), apum24-1+/-, and transgenic apum24-1+/- plants expressing 35S:ChPUM2 and 35S:ChPUM3 in the presence of 200 mM glucose. The same primers were used as in Fig. 5. Two technical and three biological replicates were performed for

PCR measurements. Values represent means ± SD (n = 3) (**; p<0.01).ChPUM2, but not ChPUM3, recovers the salt- and sugar-sensitive apum23-2 phenotype. a Complemented apum23-2 and apum24-1+/- seeds with 35S:ChPUM2 and 35S:ChPUM2. Transgenic apum23-2 seeds expressing 35S:ChPUM2 germinate normally in 1/2 MS medium supplemented with NaCl and glucose (left panel). Transgenic apum24-1+/- seeds expressing ChPUM3 show delayed germination in medium containing NaCl but display normal germination in medium supplemented with glucose (right panel). In the right panel, Control (pB2GW7), apum24-1+/-, and 35S:ChPUM2/apum24-1+/-, and 35S:ChPUM3/apum24-1+/- plants were grown on 1/2 MS medium, which includes NaCl or glucose in the presence of 10 µg l-1 Basta. b Expression levels of APUM24 in Control (pB2GW7), apum24-1+/-, and transgenic apum24-1+/- plants expressing 35S:ChPUM2 and 35S:ChPUM3 in the absence and presence of 200 mM glucose. Note similar expression levels of APUM24 in the presence of 200 mM. c qRT-PCR for analyzing the relative unprocessed rRNA levels in Control (pB2GW7), apum24-1+/-, and transgenic apum24-1+/- plants expressing 35S:ChPUM2 and 35S:ChPUM3 in the presence of 200 mM glucose. The same primers were used as in Fig. 5. Two technical and three biological replicates were performed for PCR measurements. Values represent means ± SD (n = 3) (**; p<0.01).