Scientific Research Publishing

American Journal of Plant Sciences, 2019, 10, 1090-1101
http://www.scirp.org/journal/ajps
ISSN Online: 2158-2750
ISSN Print: 2158-2742

Comparative Analysis of the 5' Flanking Region in *Arabidopsis* and *O. sative* RP Genes Revealed Conserved and Divergent Regulating Mechanisms

Donggen Zhou*, Jie Luo

Ningbo International Travel Healthcare Center, Ningbo, China
Email: *dongdong1004@163.com

How to cite this paper: Zhou, D.G. and Luo, J. (2019) Comparative Analysis of the 5' Flanking Region in *Arabidopsis* and *O. sative* RP Genes Revealed Conserved and Divergent Regulating Mechanisms. *American Journal of Plant Sciences*, **10**, 1090-1101.
https://doi.org/10.4236/ajps.2019.106079

Received: October 8, 2018
Accepted: June 25, 2019
Published: June 28, 2019

Abstract

Ribosome is one of the most abundant organelles in all living cells and plays a crucial role in cell growth. Synthesis of ribosomal components is tightly related with the change of growth conditions. We have comparatively analyzed the 5' flanking region of ribosomal protein (RP) genes in *Arabidopsis and O. sativa*. In both *Arabidopsis* and *O. sativa*, there are two putative transcriptional factor binding motifs (telo box and site II elements) overrepresented in the proximal promoter region with a strong positional bias in most of the RP genes, which suggests the conserved mechanism of transcription-level control in RP genes of these two organisms. Tri-nucleotide repeats motif CTT and CCG were also common in 5' flanking region of RP genes in *Arabidopsis* and *O. sativa*. However, we only found CCG repeat motif was enriched in *O. sativa* RP genes and most of them were clustered in the 5' UTR region. This finding reveals molecular mechanism for divergent regulation of RP genes in *Arabidopsis* and *O. sativa*, and gives the possible clue to the mechanism of controlling *O. sativa* RP genes expression at the translation level.

Keywords

Ribosome, Motif, Genes Expression, Mechanism

1. Introduction

The ribosome is a large ribonucleoprotein complex with high conservation. Eucaryotic ribosome is composed of four RNA molecules (rRNAs) and about 80 different ribosomal proteins (RPs). It is commonly known as the translational machinery for protein biosynthesis. Researchers also found its importance for its distinct role in consuming large cellular energy in the process of ribosome bio-

genesis [1]. Because of high resource-consuming in ribosome biogenesis process, the synthesis of ribosomal components in equimolar amounts is very important for cell growth.

The coordinated regulation of RP genes expression may accomplish at different level in different organisms. Rap1p binds to the site of 250 to 500 bp upstream from the translational start codon of 90% RP genes promoters [2], fascinating other factors such as Ifh1, FHL1, Crf1 and Hmo1 to bind to RP gene promoters and permit their efficient regulation at transcriptional level [3]. Unlike *S. cerevisiae*, studies in human RP genes suggest that the coordinated regulation of RP genes is mainly at translational level [4]. An oligopyrimidine tract located exact at the 5' end of human RP mRNA with 5 - 25 bases in length was found to be an essential cis-regulatory factor for their translational expression control [5]. The *C. elegans* RP genes have very short 5' UTR element and a few bases resided in this region, but study had found a TTGTT element located at 3' UTR in most RP genes, implicating a important translation-level control in *C. elegans* RP genes [6].

In contrast to the relatively well-defined regulatory mechanisms for yeast and animal RP genes, little is known about the regulation of expression of plant RP genes. Recent studies in *Arabidopsis* have identified two cis elements presented in most of the RP genes promoters, suggesting that it could be involved in the coordinated expression of this class of genes at transcriptional level [7]. One of these cis elements is the proliferating cell nuclear antigen (PCNA) site II motif (5'-GCCCR-3') [8]. It is presented in most of the RP genes promoters and can be recognized by TCP-domain proteins [9]. Another cis element is telo box motif (consensus 5'-AAACCCTA-3'). It is similar in sequence to the *Arabidopsis* telomere repeat sequence [10]. This motif was first observed within the promoter of the *Arabidopsis* genes encoding the translation elongation factor *EF*1*α* promoters, and has been identified in the 5' flanking regions of 80% *Arabidopsis* r-protein genes [11].

Microarray analysis of differential gene expression between quiescent and germinated maize embryo stages had found the expression of mRNAs encoding ribosomal proteins to remain mostly unchanged throughout the germination process, suggesting that the transcriptional control is not so important for these genes during this developmental period. However, further analysis of these two stages revealed that RP mRNAs largely accumulate in polysomes of the growth-stimulated tissues compared to quiescent tissues, indicating a translational control mechanism to account for the rapid ribosomal protein synthesis in this organism [12] [13]. All these indicate that like other organisms, the expression of plant ribosomal protein genes can also be regulated at different expression levels. These characters prompt us to analyze the gene structures around transcriptional start site of *Arabidopsis* and *O. sativa* RP genes to compare the sequence features in these two organisms and search for the possible mechanisms for co-regulation of RP genes in plant organisms.

2. Methodology

The sequences of 250 *Arabidopsis* RP genes were downloaded from TAIR10 website (www.*arabidopsis*.org). 233 *O. sativa* RP genes were downloaded from Rice Genome Annotation Project website (http://rice.plantbiology.msu.edu/). MEME [14] is used to characterize conserved features of the 5' flanking region at RP genes. We restricted our analysis to the region [−800, 200] bp relative to the transcriptional start sites. We run MEME with a fixed minimum motif length of 6 and a maximum of 15, requested 10 motifs using ZOOP model. The corresponding sequence logos for each motif were created using Weblogo [15].

The MEME motifs were converted to the Position-specific scoring matrices (PSSM). The PSSM was used as input for Patser [16] program to scan two groups of [−1000, 500] bp 5' flanking regions from RP genes and background genome for each PSSM from MEME and a p value cutoff of 0.00001 was select to calculate the target match promoters for each MEME pattern.

Then we used the hypergeometric distribution to measure the probability that the observed number of motif matches in the RP genes would be found if the sequences had been selected at random from the genome [17]. A group specificity score was calculated for each MEME motif with the following formula:

$$S = \sum_{i=r}^{\min(R,g)} \frac{\binom{R}{i}\binom{G-R}{g-i}}{\binom{G}{g}}$$

where R is the number of promoters in RP genes and G is the total number of sequences in the genome. The quantities r and g represent the promoter subset of B and G that match the motif. A cutoff value of e−10 was select to define the highly specific motifs for RP genes when compared to the genome background.

We next used χ^2-test to access the motifs with statistical significance of the local positional bias. It was hypothesized that the positions of the sampled motif were generated from a model where all positions were equally probable. We divided the 5' flanking regions of 1500 bases pairs into 15 windows. The distribution profile with $\chi^2 > 36.12$ corresponds to the probability of 0.001.

We removed RP genes and performed a functional association analysis for other genes in the genome containing motifs found by MEME. Only functional groups with ≥10 genes were selected. We compared these groups of genes with Gene Ontology categories [biological process (BP), cellular compartment (CC) and molecular function (MF)] using the online tool agriGO [18]. The agriGO program calculated appropriate *P*-values and Yekutieli correction for multiple tests. A multiple corrected FDR $P < 0.05$ was considered significant.

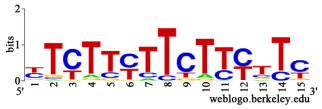
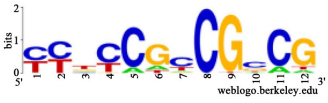
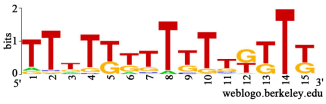
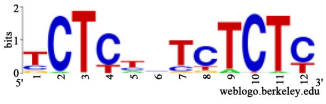
3. Results and Discussion

3.1. Identification of Putative DNA-Binding Motifs in the 5' Flanking Regions of *Arabidopsis* RP Genes

According to materials and methods, we identified 10 motifs in the 5' flanking

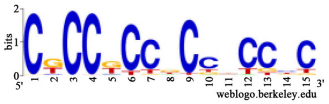
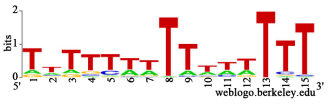
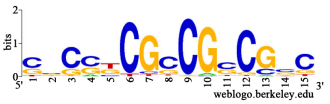
region of *Arabidopsis* RP genes. The results are listed in Table 1 and Table 2.

Table 1. Identified motifs in the 5' flanking region of ribosomal protein genes in *Arabidopsis*.

Motif	Weblogo	E-value	Hits	χ^2	SpecScore
Motif1		1.4e−142	243	71.55	0.57
Motif2		2.9e−134	185	777	3.2E−60
Motif3		8.4e−120	154	973.4	3.6E−37
Motif4		2.0e−102	231	25.33	0.63
Motif5		5.9e−072	183	724.68	5.3E−59
Motif6		7.0e−055	134	800.3	2.7E−30
Motif7		2.5e−035	102	143.82	1.2E−5
Motif8		8.0e−017	86	83.42	0.005
Motif9		3.3e−027	163	26	0.095
Motif10		3.1e−019	116	111.36	0.04

Motif stands for motifs found by MEME. Weblogo shows the motif logos output by weblogo program based on MEME results. The E-value is calculated using MEME ZOOP model. Hits refer to the highest number of ribosomal protein genes which contain one particular motif identified by MEME or Patser program. χ^2 score refers to the positional bias score for motifs found in the RP gene promoters. The higher this score is, the greater the chance is in finding a particular motif at certain position in the 5' flanking regionsof RP gene. The SpecScore refers to the probability that a motif is found with equal or higher likelihood in the 5' flanking regionsof RP gene when compared to the whole genome background using hypergeometric distribution method.

Table 2. Identified motifs in the 5' flanking region of ribosomal protein genes in *O. sativa*.

Motif	Weblogo	E-value	Hits	χ^2	SpecScore
Motif1		2.2e−592	214	907.36	2.7E−19
Motif2		9.0e−226	217	935.27	6.1E−28
Motif3		1.2e−144	178	20.4	1.4E−7
Motif4		1.9e−132	206	77.76	0.0028
Motif5		3.2e−104	214	806.78	1.5E−28
Motif6		1.5e−085	145	526.35	1.1E−65
Motif7		1.5e−084	181	378.12	7.2E−40
Motif8		1.5e−067	131	594.74	1.5E−55
Motif9		4.8e−065	59	54.28	0.007
Motif10		7.0e−064	69	111.09	3.6E−16

For the 10 motifs, there are two highly over-represented motifs exit in most of the *Arabidopsis* RP genes, Motif2 and Motif3 (Table 1). Motif2 is related to the telo box with the core consensus of TAGGGTTT, which was previously reported as the binding site for a MYB-related telomeric DNA-binding protein conserved in *S. cerevisiae*, plants, and animals [6]. This motif is found in 185 *Arabidopsis* RP promoters by MEME and is highly specific to RP genes with a group speci-

ficity score of 3.2e−60. Interestingly, it shows an extremely strong positional preference in the [−100, 100] bp region with the highest chi square score of 777 (Figure 1).

Motif3 is known as site II motif, whose core consensus is GCCCA. It is found in 154 RP genes. This motif is highly specific to RP genes with a group specificity score of 3.6e−37. It also exhibits extremely strong positional bias in the [−200, 0] bp promoter regions of RP genes with a chi square value of 973.4. In addition to its general existence in RP genes, Motif3 is also found to be associated with dark-induced genes and over-represented in genes under circadian regulation in *Arabidopsis* [8]. Motif5 and motif6 are structurally related to motif2 and motif3 and can be considered as derivatives of these two motifs.

When concentrated on the other 6 motifs, we found none of them are over-represented in RP genes. Motif1 is a poly-pyrimidine tract resembling the (GAA/TTC)n microsatellite [9]. This poly-pyrimidine patch was identified at 5' flanking region of most *Arabidopsis* RP genes. Detailed characterization of transcriptional start site had found that few RP genes have this motif as start site exactly from their 5' terminal site (data not shown), which is a common feature in mammals [11]. The role of this element in *Arabidopsis* RP genes need to be deep analyzed.

3.2. Both Telo Box and Site II Elements Are Conserved in *Arabidopsis* and *O. sativa* RP Genes

Once we have identified those two significant cis-elements telo box and site II in *Arabidopsis*, we next hoped to find commonalities of these two cis-regulatory motifs of RP genes in other plant organisms. Our analysis had found that these two cis element are also significantly enriched in *O. sativa* RP genes (Table 2) and the residential positions were highly conserved compared to *Arabidopsis* (Figure 2).

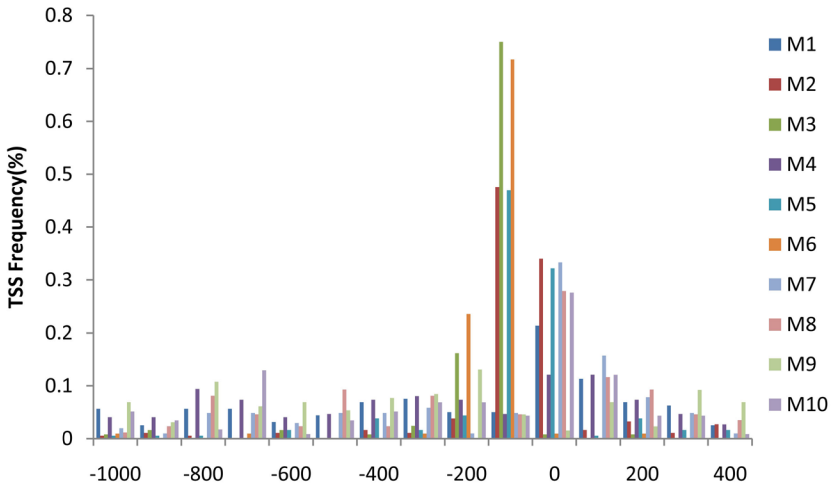


Figure 1. Positional distributions of the occurrence of the 10 motifs relative to the transcription start site, as determined by MEME in *Arabidopsis*. The positions of base 1 of the motifs were binned in 100-bp intervals.

Motif6 and Motif8 in *O. sativa* RP genes correspond to the telo box element. Motif8 is a reverse-complement to Motif6 and we only take Motif6 as an example. About 60% of rice RP genes have this cis-element in their 5' flanking regions. It is highly specific to RP genes with a group specificity score of 1.1e−65 after normalized to the genome background and biased at the [−100, 100] bp region around the 5' start site with a high Chi score of 526.35. The conservation of mechanism for telo box regulating RP genes in plant can trace back to early green plant. In two recently sequenced land plant species, *S. moellendorffii* and *P. patens*, telo box is also enriched in the promoter regions of their RP genes [12].

Motif7 is the same as the site II element in sequence. We found 80% RP genes have this motif in their core promoter regions and is enriched in RP genes with a group specificity of 7.2E−40. Further analysis also found the position of this motif in RP genes is very conserved, most of their binding sites are clustered in the [−200, 0] bp region, which was previous showed at the counterpart position in *Arabidopsis* (**Figure 2**).

3.3. *O. sativa* RP Genes Have Special Sequence Features Compared to *Arabidopsis*

Interestingly, our results indicated that the RP genes in *O. sativa* had many unique sequence features compared to *Arabidopsis*. Aside from telo box and site II cis element, there are a lot of other motifs overrepresented in *O. sativa* RP genes, which had not been found in *Arabidopsis*.

Among these motifs, Motif1 is CCG trinucleotide repeat motif and represents the $(CCG/CGG)^n$ microsatellite. The results reported in **Table 2** reveal that it was significantly enriched in most of the rice RP genes. We found Motif1 is clustered in the 5' flanking region downstream from the transcription start site

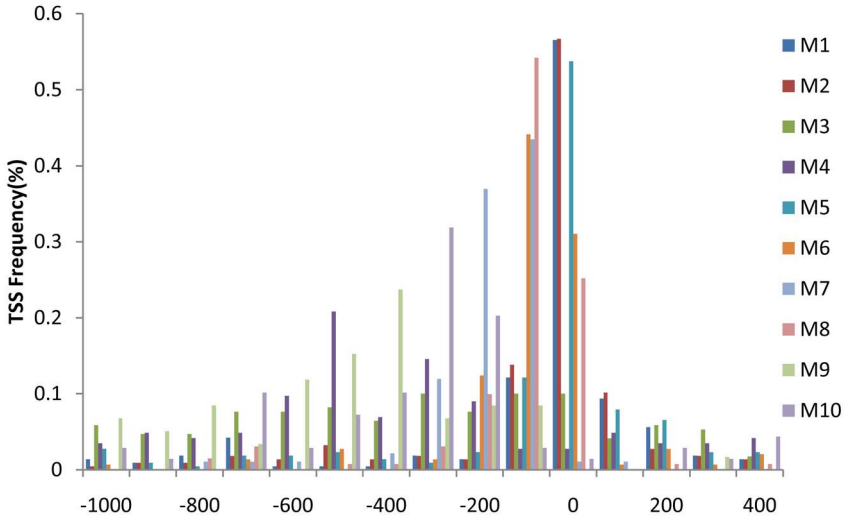


Figure 2. Positional distributions of the occurrence of the 10 motifs relative to the transcription start site in *O. sativa*.

in RP genes. About 60% of this motif located in the [0, 100] bp region downstream from the transcription start site with a significantly strong positional bias score of 907.36.

For other three over-represented motifs uniquely found in *O. sativa* RP genes, Motif2 and Motif5 are also commonly exit in most RP genes. Similar with Motif1, both these two motifs had strong positional bias score and more than half of sites were also clustered in the [0, 100] bp region downstream from the transcriptional start site. However, another motif Motif10 is only occurred in a small set of RP genes and had a moderated strong local positional bias score and most of these site were biased in the [−200, −600] bp region upstream from transcriptional start site.

3.4. Comparative Distribution of Trinucleotide Repeat Motif in *Arabidopsis* and *O. sativa* RP Genes

Our results indicated that the TTC and CCG trinucleotide repeat motifs are common in *Arabidopsis* and *O. sativa* RP genes separately. The specific distribution of simple sequence repeats may suggest its potential roles in regulating gene expression. Previous findings had implications in the common features of the over-represented microsatellites for gene regulation in plant-specific pathways [5] [7]. To make a more detailed studies of any possible functions of these two trinucleotide repeat motif found in *Arabidopsis* and *O. sativa* RP genes, we comparatively analyzed the distribution of these two motifs in the 5' flanking region and different genomic functional regions using the genome background as a reference.

Both the two motifs are clustered in the [0, 100] bp region downstream from the transcriptional start site in RP genes and genome background (**Figure 3**).

However, there are a lot of differences between them. A lot of TTC repeat motif sites is distributed outside [0, 100] bp region downstream from the transcriptional start site and the frequency of appearance of this motif in the 5' flanking region of *Arabidopsis* RP genes has the similar profile compared to the genome background, especially in the [0, 100] bp region. This is not the same as the CCG repeat motif in *O. sativa* RP genes. Few sites have been found to be outsite [0, 100] bp region and the distribution profile is significant different in the [0, 100] bp between RP genes and the genome background. Nearly 60% sites of this motif in RP genes are clustered in this region, but only 20% in the genome background.

Previous report showed that simple sequence repeats in different genic regions have different function and could regulate gene expression by affecting translation in 5'-UTR [13]. So we searched these two motifs in 5' UTR, CDS, intron and the first intron in the 5' terminal sites in their own genomes respectively. **Figure 4** summarizes the frequencies of these two motifs in these three functional elements.

As previous studies showed, the frequency of TTC repeat motif in *Arabidopsis*

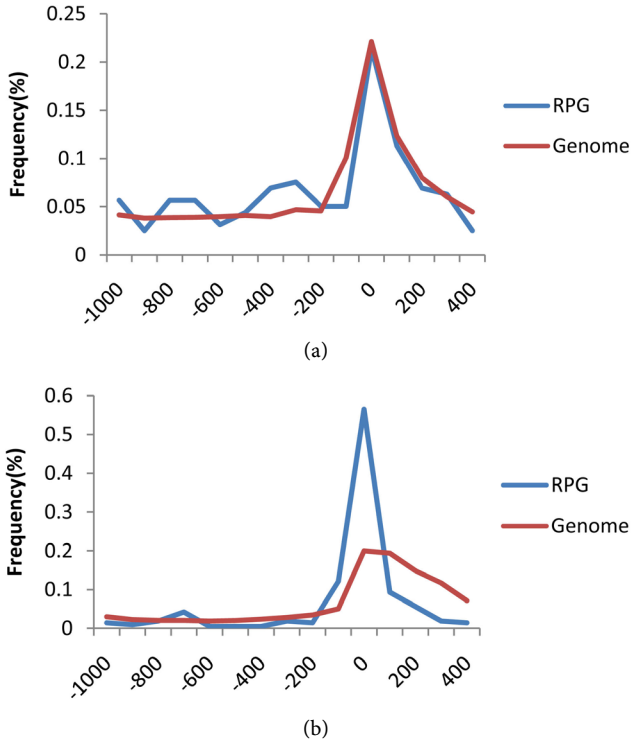


Figure 3. Distributions of trinucleotide repeat motif in the 5' flanking regions of plant RP genes and the genome background. (a) TTC repeat motif in *Arabidopsis*; (b) CCG repeat motif in *O. sativa*.

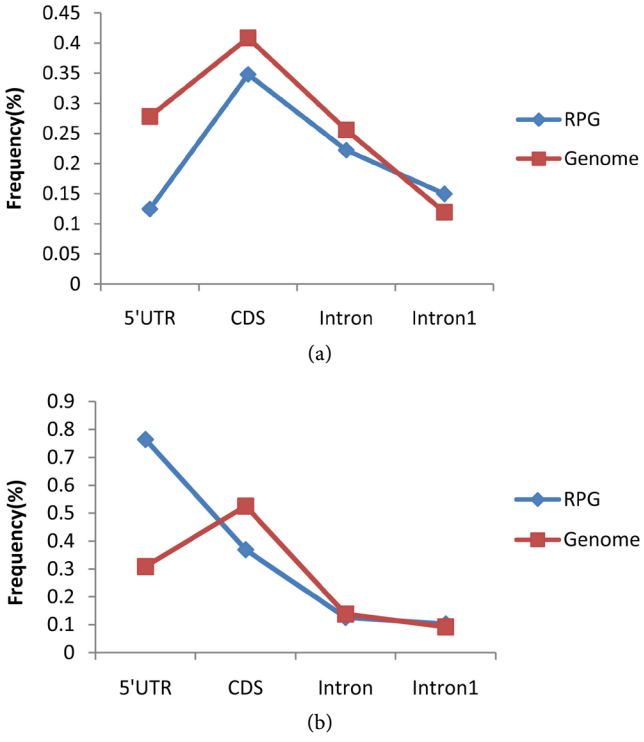


Figure 4. Distributions of trinucleotide repeat motif in the functional genomic regions of plant RP genes and the genome background. (a) TTC repeat motif in *Arabidopsis*; (b) CCG repeat motif in *O. sativa.*

is higher in CDS than that in 5' UTR and intron. This may be as a result of codon usage bias [8]. However, compared to the genome background, there is no enrichment of this motif in the genic regions in RP genes, especially in the 5' UTR region, and only 10% RP genes has this motif in their 5' UTR regions and is significant lower than the genome background. Like *Arabidopsis*, The CCG repeat motif in *O. sativa* is also higher in CDS than other genomic region in the genome background genes. But we found that about 80% RP genes of rice had this motif in their 5' UTR regions, which is significantly higher than the genome background. This result indicates CCG repeat motif is high specific to 5' UTR in rice RP genes and may take a special function in regulating translation.

3.5. CCG Repeat Element Is Associated with Ribosome Biogenesis Genes in *O. sativa*

We first analyze the functional categories of genes containing telo box or site II motifs in both *Arabidopsis* and *O. sativa*, but not including RP genes. Previous studies had showed that these two motifs are associated with genes overrepresented as ribosomal protein gene or ribosome biogenesis gene [6]. **Table 3** showed that when removed RP genes, the telo box was still over-represented in genes functioned as ribosome biogenesis both in *Arabidopsis* and *O. sativa*. There was some difference for site II motif. This motif is not enriched for the functional set of ribosome biogenesis in *O. sativa* when compared to *Arabidopsis*. We next examined the possible function of *O. sativa* genes that have CCG repeat element in the 5' flanking region. The enrichment genes associated with this motif was also included into the category of ribosome biogenesis. It is not a common feature for tri-nucleotide sequence repeats in plant. Analysis of TTC repeat motif in *Arabidopsis* indicated that this motif showed no enrichment for functional category of ribosome biogenesis.

4. Conclusions

In this study, we comparatively analyzed the 5' flanking region of RP genes

Table 3. Functional categorize of the enriched motif in plant RP genes.

Organism	Motif	GO term	Description	P value	FDR
Arabidopsis	Motif1	GO0042254	ribosome biogenesis	/	/
Arabidopsis	Motif2	GO0042254	ribosome biogenesis	2.0E−15	7.70E−13
Arabidopsis	Motif3	GO0042254	ribosome biogenesis	2.00E−07	4.60E−05
O. sativa	Motif1	GO0042254	ribosome biogenesis	0.00034	0.021
O. sativa	Motif2	GO0042254	ribosome biogenesis	/	/
O. sativa	Motif5	GO0042254	ribosome biogenesis	/	/
O. sativa	Motif6	GO0042254	ribosome biogenesis	0.0001	0.0063
O. sativa	Motif7	GO0042254	ribosome biogenesis	/	/

"/" in P value and FDR refer to the FDR p value for this motif is higher than 0.05.

between *Arabidopsis* and *O. sativa*. The results of our study demonstrated both conserved and divergent features in RP genes between these two organisms. For our analysis, we found *Arabidopsis* and *O. sativa* RP genes have conserved 5' flanking structures, both of them have two common, strong local positional biased and group specialized telo box and site II sites motifs when compared to the genome background.

Trinucleotide repeats motifs are also common in *Arabidopsis* and *O. sativa* RP genes. However, we only found CCG trinucleotide repeats motif in *O. sativa* was enriched in RP genes, but not for TTC trinucleotide repeats motif in *Arabidopsis*.

We found the over-representation of telo box and site II motif in 5' flanking region of *Arabidopsis* and *O. sativa* RP genes indicated that both of these two motifs are conserved in plant RP genes and contributed to the co-regulation of RP genes transcription. Both these two elements were close to each other and may as a module to coordinate RP gene expression in plant [13].

The difference of the enrichment of trinucleotide repeat motif between the two genomes revealed a divergent regulating mechanism in RP genes expression control, and may be the combined results of ancient species divergences and the individual evolution of these plants.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Yu, X., Willmann, M.R., Anderson, S.J. and Gregory, B.D. (2016) Genome-Wide Mapping of Uncapped and Cleaved Transcripts Reveals a Role for the Nuclear mRNA Cap-Binding Complex in Cotranslational RNA Decay in *Arabidopsis*. *Plant Cell*, **28**, 2385-2397. https://doi.org/10.1105/tpc.16.00456

[2] Polito, L., Bortolotti, M., Maiello, S., Battelli, M.G. and Bolognesi, A. (2016) Plants Producing Ribosome-Inactivating Proteins in Traditional Medicine. *Molecules*, **21**, 1560. https://doi.org/10.3390/molecules21111560

[3] Yabuki, Y., Katayama, M., Kodama, Y., *et al.* (2017) Arp2/3 Complex and Mps3 Are Required for Regulation of Ribosome Biosynthesis in the Secretary Stress Response. *Yeast*, **34**, 155-163. https://doi.org/10.1002/yea.3221

[4] Neben, C.L., Lay, F.D., Mao, X., Tuzon, C.T. and Merrill, A.E. (2017) Ribosome Biogenesis Is Dynamically Regulated during Osteoblast Differentiation. *Gene*, **612**, 29-35. https://doi.org/10.1016/j.gene.2016.11.010

[5] Englund, E., Liang, F. and Lindberg, P. (2016) Evaluation of Promoters and Ribosome Binding Sites for Biotechnological Applications in the Unicellular Cyanobacterium *Synechocystis* sp. PCC 6803. *Scientific Reports*, **6**, Article No. 36640. https://doi.org/10.1038/srep36640

[6] Zhao, D., Hamilton, J.P., Hardigan, M., *et al.* (2017) Analysis of Ribosome-Associated mRNAs in Rice Reveals the Importance of Transcript Size and GC Content in Translation. *Genetics*, **7**, 203-219. https://doi.org/10.1534/g3.116.036020

[7] Espinar-Marchena, F.J., Fernández-Fernández, J., Rodríguez-Galán, O., *et al.* (2016)

Role of the Yeast Ribosomal Protein L16 in Ribosome Biogenesis. *The FEBS Journal*, **283**, 2968-2985. https://doi.org/10.1111/febs.13797

[8] Cao, D., Liu, Y., Ma, L., *et al*. (2018) Transcriptome Analysis of Differentially Expressed Genes Involved in Selenium Accumulation in Tea Plant (*Camellia sinensis*). *PLoS ONE*, **13**, e0197506. https://doi.org/10.1371/journal.pone.0197506

[9] Liu, X., Yang, J., Qian, B., *et al*. (2018) MoYvh1 Subverts Rice Defense through Functions of Ribosomal Protein MoMrt4 in *Magnaporthe oryzae*. *PLoS Pathogens*, **14**, e1007016. https://doi.org/10.1371/journal.ppat.1007016

[10] Doamekpor, S.K., Lee, J.-W., Hepowit, N.L., *et al*. (2016) Structure and Function of the Yeast Listerin (Ltn1) Conserved N-Terminal Domain in Binding to Stalled 60S Ribosomal Subunits. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, E4151-E4160. https://doi.org/10.1073/pnas.1605951113

[11] von Walden, F., Liu, C., Aurigemma, N. and Nader, G.A. (2016) The Role of mTOR Signaling Regulates Myotube Hypertrophy by Modulating Protein Synthesis, rDNA Transcription, and Chromatin Remodeling. *American Journal of Physiology-Cell Physiology*, **311**, C663-C672.

[12] Ogasawara, R., Fujita, S., Hornberger, T.A., *et al*. (2016) The Role of mTOR Signalling in the Regulation of Skeletal Muscle Mass in a Rodent Model of Resistance Exercise. *Scientific Reports*, **6**, Article No. 31142. https://doi.org/10.1038/srep31142

[13] Jossé, L., Xie, J., Proud, C.G. and Smales, C.M. (2016) mTORC1 Signalling and eIF4E/4E-BP1 Translation Initiation Factor Stoichiometry Influence Recombinant Protein Productivity from GS-CHOK1 Cells. *Biochemical Journal*, **473**, 4651-4664. https://doi.org/10.1042/BCJ20160845

[14] Bailey, T.L., Boden, M., Buske, F.A., *et al*. (2009) MEME SUITE: Tools for Motif discovery and Searching. *Nucleic Acids Research*, **37**, W202-W208. https://doi.org/10.1093/nar/gkp335

[15] Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Research*, **14**, 1188-1190. https://doi.org/10.1101/gr.849004

[16] Hertzand, G.Z. and Stormo, G.D. (1999) Identifying DNA and Protein Patterns with Statistically Significant Alignments of Multiple Sequences. *Bioinformatics*, **15**, 563-577. https://doi.org/10.1093/bioinformatics/15.7.563

[17] Friberg, M., von Rohr, P. and Gonnet, G. (2005) Scoring Functions for Transcription Factor Binding Site Prediction. *BMC Bioinformatics*, **6**, 84. https://doi.org/10.1186/1471-2105-6-84

[18] Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: A GO Analysis Toolkit for the Agricultural Community. *Nucleic Acids Research*, **38**, W64-W70. https://doi.org/10.1093/nar/gkq310