

Identification of 5'-untranslated regions that function as effective translational enhancers in monocotyledonous plant cells using a novel method of genome-wide analysis

Shotaro Yamasaki, Atsunobu Suzuki, Yasuaki Yamano, Harunori Kawabe,
Daishin Ueno, Taku Demura, Ko Kato*

Division of Biological Science, Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

*E-mail: kou@bs.naist.jp Tel: +81-743-72-5464 Fax: +81-743-72-5469

Received June 21, 2018; accepted September 3, 2018 (Edited by M. Yamaguchi)

Abstract High expression of a transgene is often necessary to produce useful substances in plants. The efficiency of mRNA translation is an important determinant of the level of transgene expression. In dicotyledonous plants, the 5'UTR of certain mRNAs act as translational enhancers, dramatically improving transgene expression levels. On the other hand, translation enhancers derived from dicotyledonous plants are not so much effective in monocotyledonous plants, which are important as industrial crops and as hosts for production of useful substances. In this study, we evaluated the polysome association on a large scale with high resolution for each 5'UTR variant from multiple transcription start site in normal and heat-stressed *Oryza sativa* suspension cultures. Translational enhancer candidates were selected from the resultant large-scale data set, and their enhancer activities were evaluated by transient expression assay. In this manner, we obtained several translational enhancers with significantly higher activities than previously reported enhancers. Their activities were confirmed in a different monocotyledonous plant, *Secale cereale*, and using a different reporter gene. In addition, enhancer activities of tested 5'UTRs were different between monocotyledonous and dicotyledonous plants, suggesting that the enhancer activities were not compatible between them. Overall, we demonstrate these useful 5'UTRs as enhancer sequence for transgene expression in monocotyledonous plants.

Key words: 5'-untranslated region, cap analysis of gene expression, genome-wide analysis, transgene overexpression in monocotyledonous plant, translational enhancer.

Introduction

Currently, useful substances such as vaccines and antibodies are being produced in transgenic plants (Yao et al. 2015). As with dicotyledonous plants, monocotyledonous plants such as rice, wheat, and corn are attracting attention as the hosts for production of these substances. Monocotyledonous plants typified by grains are widely used as staple food, feed, and bioethanol raw materials, and research aimed at developing and utilizing their agricultural characteristics is actively ongoing. A famous example of a transgenic plant monocotyledonous is golden rice, which expresses the carotenoid gene to produce beta-carotene, and is expected to prevent vitamin A deficiency in many children around the world (Tang et al. 2009). Other examples include transgenic rice that accumulates cedar

pollen peptides via long-term preservation characteristics in endosperm (Takaiwa 2007).

A great deal of research has been performed to date, and the utilization of monocotyledonous plants to generate valuable substances is predicted to expand. However, there are some obstacles to further development, including the low expression levels of transgenes in monocotyledonous plants. Thus far, various approaches have been tested for achieving higher-level transgene expression. For example, in the dicotyledonous plants *Arabidopsis thaliana* and *Nicotiana benthamiana*, 5'UTR sequences functioning as translation enhancers can achieve 100-fold higher expression than existing expression systems (Satoh et al. 2004; Yamasaki et al. 2018). However, translational enhancers derived from dicotyledonous plants do not cause dramatic improvement in transgene expression

Abbreviations: 5'UTR, 5'-untranslated region; CAGE, cap analysis of gene expression; TSS, transcription start site; TPM, tag per million; PR, polysome ratio.

This article can be found at <http://www.jspcmb.jp/>

Published online November 29, 2018

in rice, a monocotyledonous plant (Sugio et al. 2008). On the other hand, the 5'UTR of *Oryza sativa* alcohol dehydrogenase (OsADH) and *Oryza sativa* MAP-kinase-activating protein (OsMAC1) act as translation enhancers in rice (Aoki et al. 2014; Sugio et al. 2008). Therefore, large-scale searches in monocotyledonous plants represent a promising strategy for identifying translation enhancers that function effectively in these species. To date, only a few genome-wide analyses of mRNA translation in monocotyledonous plants have been performed, and it is possible that more enhancers could be discovered by applying next-generation sequencing technology, which has advanced remarkably in recent years. Moreover, mRNA translation changes under abiotic stress and in developed tissues, with the majority of mRNA species being translationally repressed under these conditions (Juntawong et al. 2014; Kawaguchi et al. 2004; Matsuura et al. 2010; Yamasaki et al. 2015; Yángüez et al. 2013). In dicotyledonous plants, the use of an appropriate 5'UTR sequence can avoid this repression (Ueda et al. 2014; Yamasaki et al. 2018). In other words, to construct a better expression system, it is preferable to use a 5'UTR that not only has strong translation ability but is also resistant to repression.

In this study, we searched for 5'UTR sequences that could be used as translational enhancers in monocotyledonous plants. To this end, we applied a novel form of genome-wide analysis of mRNA translation, polysome/CAGE analysis, which combines polysome fraction analysis and cap analysis of gene expression (CAGE). This method can reveal the polysome association and 5'UTR sequence of each 5'UTR variant transcribed from multiple transcription start sites (TSS) within each gene. On the other hand, conventional methods such as microarray and RNA-seq with polysome fraction analysis and ribosome footprint profiling (King and Gerber 2016) only evaluate the translational efficiency at each gene level and it cannot be determine 5'UTR sequence. The polysome/CAGE analysis can be used to analyze mRNA translation genome-wide and at high resolution, capturing slight differences in translation as a function of variations in the 5'UTR sequence. In addition, we selected candidates for effective translational enhancer by considering translational repression in two data set, normal conditions and heat stress, in *O. sativa* suspension cultures. Reporter assays in transiently expressed cells revealed two 5'UTR sequences with very high translation enhancer activities. The activities of both sequences were confirmed in another monocotyledonous plant, *Secale cereale*, as well as with another reporter gene. These 5'UTRs should be powerful and universal translational enhancer sequences for achieving high transgene expression in monocotyledonous plants.

Materials and methods

Plant materials, culture conditions, and growth conditions

Oryza sativa L. cv. Nipponbare suspension cells were cultured in R2S medium with rotary shaking at 90 rpm at 30°C in the dark. For genome-wide analysis of polysome association, 3-day-old suspension cultures and samples additionally incubated at 41°C for 15 min were collected as control and heat-stressed samples, respectively. In addition, Oc suspension cells from roots of *Oryza sativa* L. accession C5924 (Baba et al. 1986), from which it is easy to isolate protoplasts, were cultured under the same conditions and used for transient expression assays. *Arabidopsis thaliana* T87 suspension cells (Axelos et al. 1992) were cultured in modified Murashige–Skoog medium (Matsui et al. 2015) with constant agitation at 80 rpm at 22°C with a 16 h light/8 h dark photoperiod. All suspension cultures were subcultured once per week. *Secale cereale* were grown on Petri dishes at 25°C with an 11 h light/13 h dark photoperiod.

Polysome fractionation assays and RNA isolation from sucrose gradients

Polysome fractionation analysis was performed as previously described (Yamasaki et al. 2015). Cell extracts were layered on a 26.25–71.25% sucrose density gradient and centrifuged. After centrifugation, gradients were separated into two fractions using a piston gradient fractionator (BioComp Instruments, Fredericton, NB, Canada). The bottom-half fraction (polysome fraction) and both fractions (total fraction) were individually collected and purified. We prepared RNA in two independent biological replicates.

Cap analysis of gene expression (CAGE) and data analysis

nAnT-iCAGE library preparation, sequencing, filtering, mapping, and gene annotation were performed as previously described (Yamasaki et al. 2018). In this analysis, to more accurately identify the TSS, additional quality control was performed to remove tags with mismatches within three bases of the 5' end. In addition, tag counts on each TSS were converted to tag per million (TPM) values as TPM_TSS. Finally, we calculated the polysome ratio (PR) for each TSS (PR_TSS) as an indicator of polysome association and averaged between the replicates. The PR_TSS, value in the n number data sets is expressed by Eq. 1, shown below. To obtain more reliable data, for calculation of PR_TSS we used a subset of TSSs that mapped to more than 50 tags in the total fraction data in both replicates. In addition, to create a ranking that could evaluate the translational activity among different conditions at same importance, PR_TSS was normalized and averaged. The normalized PR_TSS, value in the n number data sets is expressed by Eq. 2 below. CAGE library preparation and sequencing were performed by DNAFORM (Yokohama, Japan). Genome information from IRGSP-1.0 in The Rice Annotation Project Database (<http://rapdb.dna.affrc.go.jp>) was used as

a reference for rRNA tag removal, mapping, and annotation. All of the raw CAGE sequences data are available on the DDBJ Sequence Read Archive (DRA) database with accession number DRA006661 (<https://trace.ddbj.nig.ac.jp/DRAsearch/submission?acc=DRA006661>).

$$PR_TSS_i = \frac{TPM_TSS_i \text{ in polysome fraction}}{TPM_TSS_i \text{ in total fraction}} \quad (\text{Eq. 1})$$

$$\begin{aligned} & \text{normalized } PR_TSS_i \\ &= \frac{PR_TSS_i - \overline{PR_TSS}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (PR_TSS_i - \overline{PR_TSS})^2}} \quad (\text{Eq. 2}) \end{aligned}$$

$\overline{PR_TSS}$ represents the average of all PR_TSS in the n number data set from all TSSs of all genes.

Plasmid construction

According to previously described methods (Yamasaki et al. 2018), the OsADH 5'UTR (Sugio et al. 2008) was amplified by PCR using gene-specific primers (Supplementary Table S1), and the PCR fragment was introduced into plasmid pBlue-p35S-FL-HSPT (Yamasaki et al. 2018), yielding plasmid pBlue-p35S-OsADH_5'UTR-FL-HSPT. The test 5'UTRs were amplified by PCR using gene-specific primers that contained a portion of the modified 35S promoter or a portion of the firefly luciferase (F-Luc) coding region (Supplementary Table S2). Each PCR fragment was introduced into plasmid pBlue-p35S-OsADH_5'UTR-FL-HSPT using the In-Fusion cloning kit (Takara Bio, Kusatsu, Japan) to generate plasmids pBlue-p35S-test_5'UTR-FL-HSPT. These plasmids were designed to initiate transcription immediately upstream of the test 5'UTR under the control of the modified 35S promoter (...TTCATTATC GAT GGA GAG AA::test 5'UTR). Similarly, plasmid pRI-p35S-test_5'UTR-RL-HSPT was generated from plasmid pBlue-p35S-RL-HSPT (Yamasaki et al. 2018) using the Renilla luciferase (R-Luc) gene-specific primer with a region of homology to each test 5'UTR (Supplementary Table S3), and then using the In-Fusion cloning kit. Inserted fragments were verified by sequencing.

Transient expression assay in T87 suspension cells using the PEG method

Protoplasts were generated from 3-day-old *A. thaliana* T87 suspension cells using a slight modification of a previously reported method (Kawabe et al. 2018). Some solutions were altered according to the method of Yamasaki et al. (2018): enzyme solution (1% Cellulase RS, 0.05% Pectolyase, 0.4 M mannitol, pH 5.6), W5 solution (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, 2 mM MES-KOH, pH 5.6), and MaMg solution (0.4 M mannitol, 15 mM MgCl₂, 4 mM MES-KOH, pH 5.7). In addition, the resuspension volume in W5 solution was changed to 20 ml, and the final protoplast density was adjusted to 1×10^6 protoplasts ml⁻¹ in MaMg solution. After protoplast preparation, a mixture of 10 μl plasmid DNAs containing 0.8 μg

of pBlue-p35S-test_5'UTR-FL-HSPT plasmids and 0.08 μg of plasmid pBlue-p35S-RL-HSPT (internal control) was added to 190 μl of protoplast solution. Immediately thereafter, an equal volume of polyethylene glycol-CMS (PEG-CMS) solution [0.2 M mannitol, 0.1 M Ca(NO₃)₂, 40% PEG 4000] was added to each sample and gently mixed. The protoplast mixture was incubated at room temperature for 20 min, and then mixed with 400 μl protoplast medium (Satoh et al. 2004). The transiently transfected protoplasts were incubated at 22°C for 16 h, lysed in Passive Lysis Buffer (Promega, Madison, WI, USA), and assayed for R-Luc and F-Luc activities using the Dual-Luciferase Reporter Assay system (Promega) on a plate reader (TriStar LB 941: Berthold Technologies, Bad Wildbad, Germany).

Transient expression assay in *Oc* suspension cells using the PEG method

Three-day-old *O. sativa* *Oc* suspension cells were collected and gently shaken in *Os* protoplast isolation enzyme solution (4% Cellulase RS, 1% macerozyme R10, 0.1% CaCl₂·6H₂O, 0.1% MES, 0.4 M mannitol, pH 5.6) at 30°C for 3 h. Isolation solution containing crude protoplasts was filtered through a 40 μm nylon sieve and mixed with the same volume of W5 solution. After centrifugation for 4 min at 100×*g*, pelleted protoplasts were collected and washed once more in W5 solution, followed by another round of centrifugation. The washed, pelleted protoplasts were mixed with W5 solution and incubated on ice for 30 min. The final protoplast density was adjusted to 1×10^6 protoplasts ml⁻¹. Afterwards, to evaluate the translation activity of test 5'UTRs for the *F-Luc* gene, a mixture of 10 μl plasmid DNAs containing 1 μg of pBlue-p35S-test_5'UTR-FL-HSPT and 0.04 μg of plasmid pBlue-p35S-RL-HSPT (internal control) was added to 190 μl of protoplast solution. Similarly, to evaluate the translation activity of the test 5'UTR for the *R-Luc* gene, a mixture of 10 μl plasmid DNAs containing 1 μg of pRI-p35S-test_5'UTR-RL-HSPT and 0.04 μg of plasmid pBlue-p35S-OsADH_5'UTR-FL-HSPT (internal control) was added to 190 μl of protoplast solution. An equal volume of PEG-CMS solution was immediately added to each sample and gently mixed. The protoplast mixture was incubated at room temperature for 20 min, and then mixed with 1 ml of *Os* protoplast medium (R2S supplemented 400 mM mannitol). The transiently transfected protoplasts were incubated at 30°C for 16 h, then in Passive Lysis Buffer, and assayed for R-Luc and F-Luc activities using the Dual-Luciferase Reporter Assay system on a plate reader.

Transient expression assay of *Secale cereale* leaves by the PEG method

Protoplasts from leaves in 4-day-old *Secale cereale* seedlings were isolated using a slight modification of a previously reported method (Matsui et al. 2015): all centrifugation steps were performed at 60×*g* instead of 40×*g*. Isolated protoplasts were resuspended in 5 ml of W5 solution and incubated on ice for 30 min. The final protoplast density was adjusted to 5×10^5

protoplasts ml^{-1} . Afterwards, a mixture of $5 \mu\text{l}$ plasmid DNAs containing $1 \mu\text{g}$ of pBlue-p35S-test_5'UTR-FL-HSPT plasmids and $0.04 \mu\text{g}$ of plasmid pBlue-p35S-RL-HSPT (internal control) was added to $50 \mu\text{l}$ of protoplast solution. Immediately, an equal volume of PEG-CMS solution was added to each sample and gently mixed. The protoplast mixture was incubated at room temperature for 20 min, and then mixed with 1 ml of *Os* protoplast medium. Incubation and assay of transiently transfected protoplasts were performed as described above for transiently transfected *Oc* suspension protoplasts.

Results and discussion

Different behavior of translation enhancers between dicotyledonous and monocotyledonous plants

In a previous study, *AtADH* 5'UTR, which acts as a strong translational enhancer in dicotyledonous plants, was not as effective in the monocotyledonous plant *O. sativa*, although the *OsADH* 5'UTR did exhibit translational enhancer activity (Satoh et al. 2004; Sugio et al. 2008). Here, we first confirmed the difference between the activities of translation enhancers in dicotyledonous and monocotyledonous plants, using the *AtADH* and *OsADH* 5'UTRs along with the *AtCOR47* 5'UTR, a powerful translation enhancer in dicotyledonous plants that was discovered from ranking of genome-wide analysis of polysome association under various conditions (Yamasaki et al. 2018). To evaluate their translational activities, we constructed plasmids in which the test 5'UTRs were fused to the 5'ends of *F-Luc*, used as a reporter gene, under the control of the constitutive CaMV 35S promoter. The translational efficiencies of these reporter genes were examined by transient expression assays in protoplasts from *A. thaliana* T87 and *O. sativa* *Oc* suspension cells, using plasmid DNA with an *R-Luc* internal control (Figure 1). In the T87 suspension cells, the known translational enhancers *AtADH* and *AtCOR47* 5'UTRs exhibited significantly higher translational enhancer ability than that of the *OsADH* 5'UTR (Figure 1A). By contrast, in the *Oc* suspension cells, both 5'UTRs had significantly lower activities (Figure 1B). In other words, the trend of enhancer activity of 5'UTRs was completely opposite between the dicotyledonous and monocotyledonous species (Figure 1C), suggesting that sequence features important for efficient translation differ between the two types of plants. Therefore, to develop a high transgene expression system in monocotyledonous plants, it will be necessary to identify translational enhancers from these species.

Novel genome-wide analysis of ribosome association

To search for translation enhancers from large detailed data sets, we performed a novel method of genome-

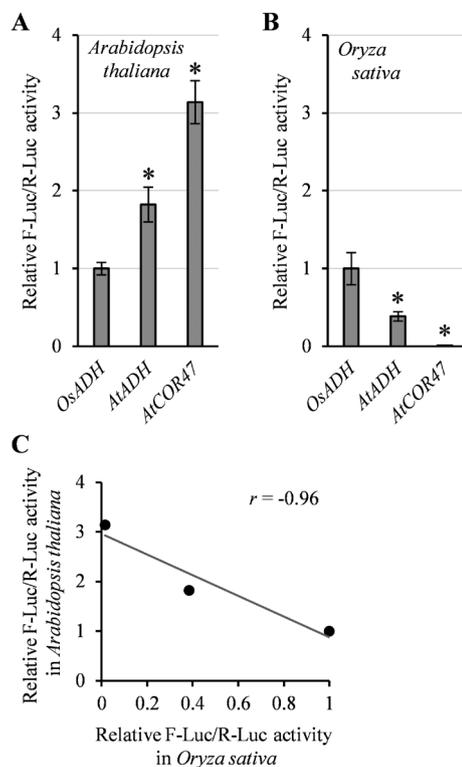


Figure 1. Evaluation by transient expression assay of the translational enhancer activity of known monocotyledonous and dicotyledonous 5'UTRs in protoplasts from *A. thaliana* T87 (A) and *O. sativa* *Oc* suspension cells (B). Effects of the 5'UTRs on expression of a reporter gene were evaluated as F-Luc activity, normalized against a co-introduced internal control, R-Luc. The data for each sample are means \pm SD of three measurements (* $p < 0.05$; Welch's *t*-test). (C) Scatter plots of relative F-Luc/R-Luc activity of the 5'UTRs between *A. thaliana* and *O. sativa* (shown in A and B, respectively). Pearson correlation coefficients and regression line (solid line) for each pair of data sets are shown.

wide analysis, polysome/CAGE analysis, that combines polysome fraction analysis and cap analysis of gene expression (CAGE) (Figure 2A). Polysome fractionation can separate mRNAs based on the number of bound ribosomes, whereas CAGE can determine the TSS, and evaluate the expression level of each 5'UTR variant transcribed from a given TSS as TPM_TSS (see Materials and methods for definitions). Using these methods, we evaluated the ratio of polysomal RNA at each 5'UTR variant levels as PR_TSS values in control (30°C) and heat-stressed (41°C , 15 min) *O. sativa* L. cv. Nipponbare suspension cells. CAGE sequencing identified more than 10 million tags, even after contaminating rRNA was filtered out, and more than 85% of the tags mapped to 5'UTR regions; thus, data of sufficient quality and quantity were obtained. All analyses were carried out in two biologically independent replicate samples, and the log-transformed TPM_TSS values were highly correlated ($r = 0.85\text{--}0.89$) (Supplementary Figure S1). Finally, we calculated the PR_TSS values for 19,917 5'UTR variant transcripts derived from multiple TSSs in 7,048 genes

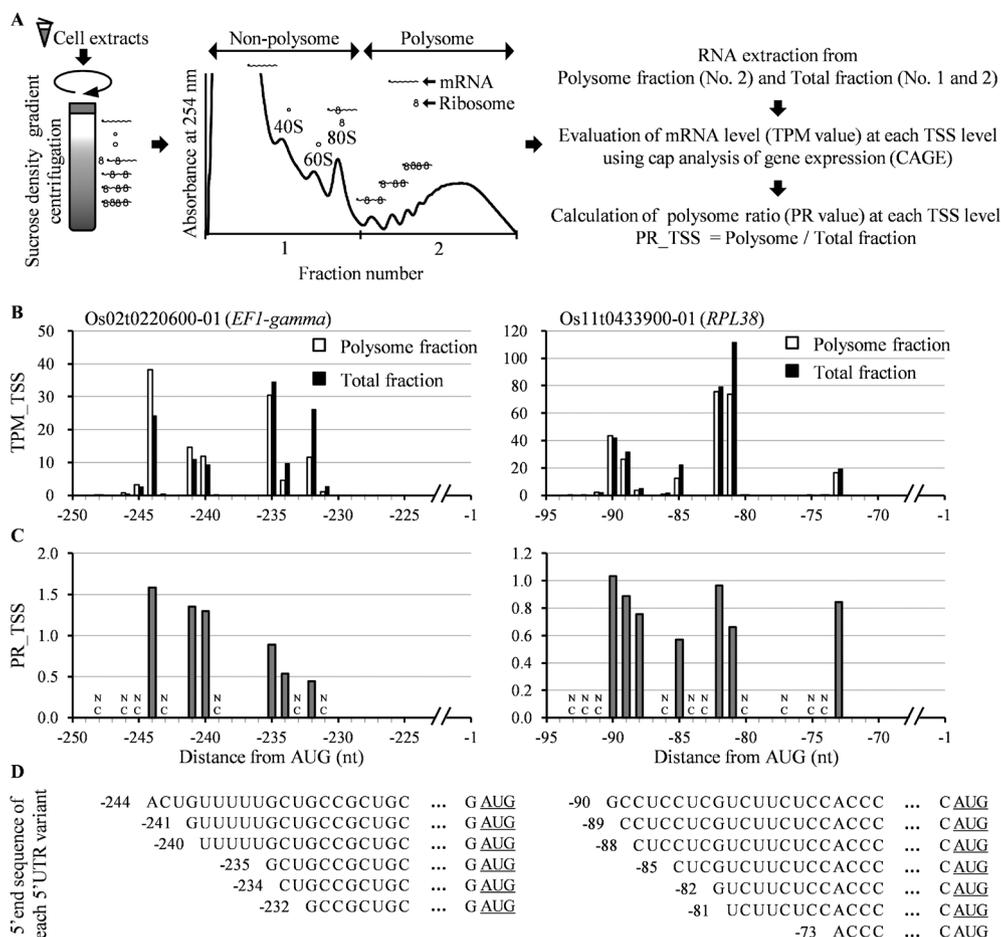


Figure 2. Experimental strategy for polysome/CAGE. (A) Cell extracts from samples were subjected to sucrose density gradient centrifugation, and all mRNAs in the extracts were separated based on the number of bound ribosomes. Subsequently, absorbance was measured at 254 nm to generate polysome profiles; the corresponding polysome and non-polysome fractions, as well as the 40S, 60S, and 80S peaks, are shown. RNA was extracted from the polysomes (fraction 2) and the whole gradient (fractions 1 and 2), and accumulated mRNA level was calculated as a tag per million (TPM) value at each TSS level (TPM_{TSS}). By comparing mRNA levels between the polysome and total fraction, the PR_{TSS} value, an indicator of polysome association at each TSS levels, was calculated. Examples of TPM_{TSS} (B), PR_{TSS} (C) and 5' end sequence (D) of each 5'UTR variant transcripts that transcribed from multiple TSSs of polysome/CAGE results in control suspension cells are shown. Minor TSS (i.e., those with frequencies <5% of the sum of TPM_{TSS}) outside the x-axis range in the figure are omitted. Some TSSs with low expression levels, for which the PR_{TSS} value was not calculated (NC), are described.

under both conditions. This data set provides a high-resolution view of the relationship between the PR_{TSS} value, a reflection of translational efficiency, and the 5'UTR sequence. For example, Os02t0220600-01 has multiple TSSs, and the major 5'UTR variant mRNA was transcribed from the TSS at position -235 relative to the AUG (Figure 2B, black bars). On the other hand, the most enriched transcript in the polysome fraction was the 5'UTR variant transcribed from the -244 position (Figure 2B, white bars). Consequently, the PR_{TSS} value varied among the TSSs (Figure 2C), suggesting that translational activity differs depending on the 5'UTR sequence even within the same gene. For example, the PR_{TSS} value in 5'UTR variant mRNA transcribed from the -244 position is 1.8-fold higher than the transcribed from the -235 position (Figure 2C), indicating the mRNA transcribed from the -244 position has high

possibility of forming polysomes. Similarly, PR_{TSS} values in Os11t0433900-01 also differed among 5'UTR variants, and in this case translational activity was affected by a single-base difference in the 5'UTR sequence such as among -90 to -88 or between -82 and -81 (Figure 2D).

Selection of candidate 5'UTRs for translational enhancers

We prepared a large data set of PR_{TSS} values (a reflection of polysome association) and the 5'UTR sequence in transcripts from each TSS. Next, to select candidate 5'UTR sequences with high translational activity under both normal and heat-stressed conditions, the transcripts were ranked based on the mean normalized PR_{TSS} values that data of both conditions are included as equal importance. 5'UTR sequences

harboring an uAUG or intron were removed from the ranking, because it was possible that these target genes were not correctly translated. In addition, genes with a CDS length longer than the median (864 nt) were excluded from the ranking because polysome association is positively correlated with CDS length (Kawaguchi and Bailey-Serres 2005); consequently, high PR values in some genes arise due to long CDSs, but this effect is independent of the translational activity of the 5'UTR. Ultimately, the 5'UTR sequences of the top 15 transcripts were selected as candidate translational enhancers (Table 1, Supplementary Table S4). On the other hand, these transcripts showed no tendency of high TPM_TSS value (accumulated mRNA level). Moreover, in the whole data set, the PR_TSS value did not correlate with the TPM_TSS value (Supplementary Figure S2). Thus, the translational step seems to be generally independent of the transcriptional step. In addition, low correlation between translational step and transcriptional step was also observed in *A. thaliana* (Juntawong et al. 2014; Yamasaki et al. 2015). Therefore it is important to use enhancer sequences to consider the translational step as with the transcriptional step.

Transient expression assay to evaluate translational efficiency of candidate 5'UTR sequences

To evaluate the translational activity of the candidate 5'UTRs, we performed a transient expression assay using plasmids in which the test 5'UTRs were fused to the 5'ends of a reporter gene expressed under the control of the 35S promoter. First, we tested 15 candidate 5'UTRs and the known translational enhancer *OsADH* 5'UTR with the *F-Luc* reporter and *R-Luc* internal control,

using protoplasts from *O. sativa* Oc suspension cells (Figure 3). Preliminarily, we selected rank 8, 12, 13, and 14 5'UTRs, which had greater than 1.5-fold *F-Luc/R-Luc* activity relative to *OsADH*. Some candidate 5'UTRs exhibited significantly lower activity. One reason for this is that the ranking was based on the translation activity of the whole sequences of the transcripts, including not only the 5'UTR but also the CDS and 3'UTR. Therefore, there is possibility that the high PR_TSS value of those transcripts depended on regions other than the 5'UTR such as codon usage and U-rich motif in 3'UTR (Kim et al. 2017; Miyazaki et al. 2015). Note that the sequences with rank 12 and 15 are 5'UTR variants (5'UTR length=129 and 116, respectively) of the same gene, but had significantly different translational activities. These differences were caused by differences of only 13 bases. The importance of 5'end sequence of 5'UTR in regulation of translational efficiency is reported in *A. thaliana* (Matsuura et al. 2013; Yamasaki et al. 2018). These studies revealed only several bases at the 5'end of 5'UTR sequence changed translation efficiency. These results suggest that 5'end sequence is important for the translation initiation such as recruitment of ribosomes. Therefore, slight differences in the 5'UTR sequence greatly affect translational activity, demonstrating the usefulness of polysome/CAGE analysis. Moreover, this polysome/CAGE data set provides important information for further challenges, that is detailed data analysis of the sequence characteristics involved in determination of translational efficiency in *O. sativa*.

Next, the translation enhancer activities of the preliminary 5'UTR candidates of rank 8, 12, 13, and 14 were evaluated by transient expression assay with the *F-Luc* reporter gene and internal control *R-Luc* gene,

Table 1. Polysome/CAGE data of candidate 5'UTRs.

Rank	5'UTR ID ^a	Gene name	Mean of normalized PR_TSS	PR_TSS CS	PR_TSS HS	TPM_TSS CS	TPM_TSS HS
1	Os10t0530900-01_00076	Glutathione transferase U50	2.81	1.67	2.11	7.78	9.03
2	Os04t0556300-01_00118	Glutathione peroxidase 1	2.62	1.64	2.02	8.53	9.21
3	Os03t0180400-01_00112	20S proteasome alpha1 subunit	2.61	1.67	1.97	15.21	17.34
4	Os11t0592100-01_00063	Pathogenesis-related protein 4b	2.61	1.66	1.98	11.33	10.23
5	Os05t0119200-01_00064	Glycine-rich cell-wall structural protein 1	2.51	1.7	1.85	7.43	9.75
6	Os01t0303800-01_00106	UspA domain containing protein	2.47	1.53	2.07	10.69	7.61
7	Os06t0147400-01_00074	Conserved hypothetical protein	2.45	1.59	1.97	9.76	8.47
8	Os02t0684500-00_00094	Similar to Histone H4	2.44	1.49	2.11	16.10	9.69
9	Os08t0117400-01_00160	Calmodulin-like protein 7	2.42	1.7	1.77	10.50	9.91
10	Os01t0559000-01_00115	DUF1000 family protein	2.38	1.74	1.68	12.29	13.61
11	Os07t0230700-01_00066	DUF567 family protein	2.36	1.67	1.78	6.87	8.55
12	Os08t0136800-01_00129	DUF26 domain containing protein	2.34	1.31	2.31	11.35	7.22
13	Os04t0583600-01_00068	Similar to Histone H4	2.3	1.43	2.08	7.01	5.45
14	Os06t0342200-01_00113	Translation initiation factor 1A-2	2.29	1.45	2.05	7.09	6.32
15	Os08t0136800-01_00116	DUF26 domain containing protein	2.27	1.55	1.89	27.34	21.97
Mean value ^b			0.00	1.02	0.94	30.27	30.12

Abbreviations: CS, control suspension cells; HS, heat-stressed suspension cells (42°C). ^a Candidate 5'UTR ID consists of Gene ID from IRGSP-1.0 and TSS ID, defined based on TSS position. ^b Mean value was calculated from data set of 19,917 5'UTR variant transcripts derived from multiple TSSs in 7,048 genes.

using protoplasts from *Oc* suspension cells, *S. cereale* leaves, and T87 suspension cells. Relative to the known translational enhancer *OsADH* 5'UTR, these 5'UTRs exhibited high F-Luc/R-Luc activities in both types of monocotyledonous plants (Figure 4A, B), particularly the sequences with rank 12 and 13. It is possible that these 5'UTRs would exhibit strong translational enhancer

activity even in other monocotyledonous plants. By contrast, all preliminarily selected 5'UTR sequences exhibited lower translational activity than *OsADH* in the dicotyledonous species *A. thaliana* (Figure 4C). Although the F-Luc/R-Luc activities were slightly different between the *O. sativa* and *S. cereale*, these activities were positively correlated between them (Figure 4D).

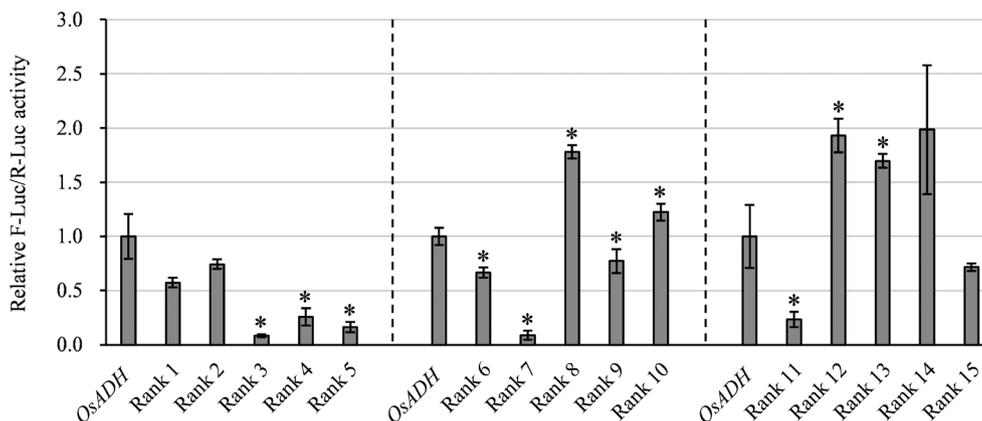


Figure 3. Transient expression assay to evaluate translational enhancer activity in 15 candidate 5'UTRs in *Oc* protoplasts. Effects of candidate 5'UTRs on expression of a reporter gene were evaluated as F-Luc activity normalized against the activity of the co-introduced internal control, R-Luc. These experiments were carried out in three independent groups. Data for each sample are means \pm SD of three measurements (* p < 0.05; Welch's t -test).

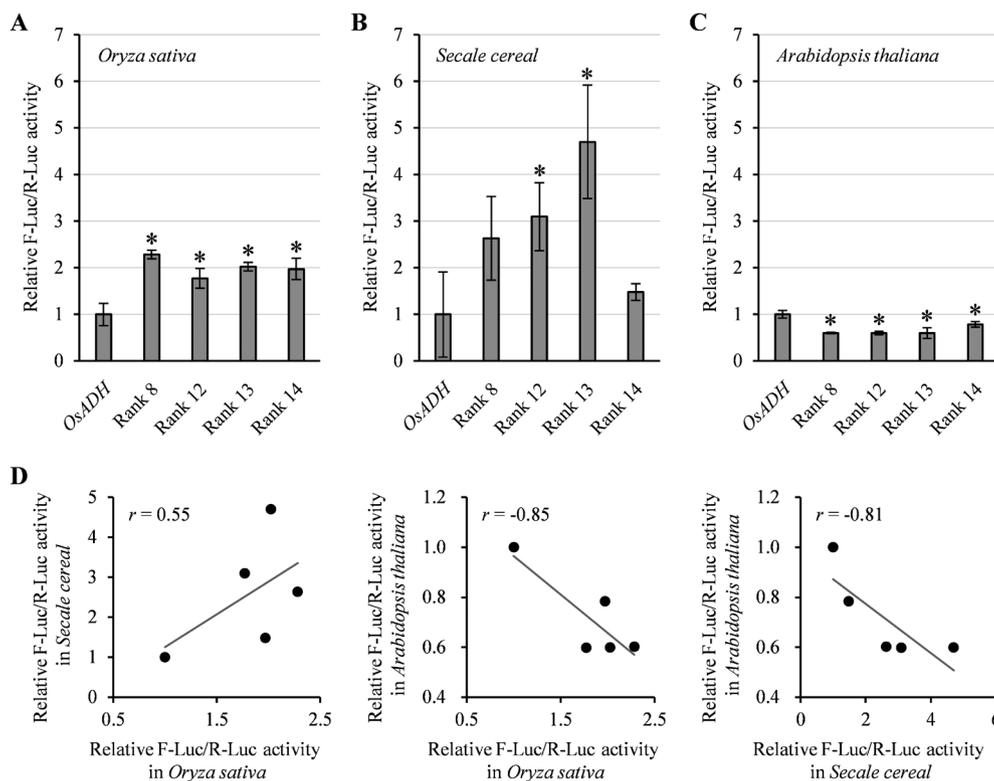


Figure 4. Evaluation by transient expression assay of the translational enhancer activity of candidate 5'UTRs in protoplasts from *O. sativa* *Oc* suspension cells (A), *S. cereale* leaves (B), and *A. thaliana* T87 suspension cells (C). Effects of the 5'UTRs on expression of a reporter gene were evaluated as F-Luc activity normalized against the activity of the co-introduced internal control, R-Luc. Data for each sample are means \pm SD of three measurements (* p < 0.05; Welch's t -test). (D) Scatter plots of relative F-Luc/R-Luc activity of candidate 5'UTRs between *O. sativa*, *S. cereale* and *A. thaliana* (shown in A, B and C, respectively). Pearson correlation coefficients and regression line (solid line) for each pair of data sets are shown.

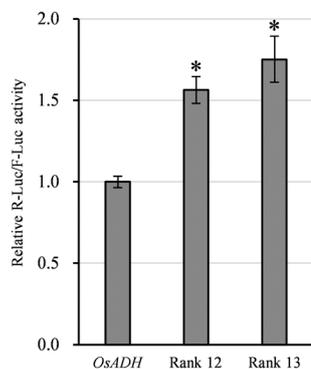


Figure 5. Evaluation of the translational enhancer activities of candidate 5'UTRs to assess universality (i.e., independence of the target gene). Protoplasts from *Oc* suspension cells were subjected to transient expression assay. Effects of the 5'UTRs on expression of a reporter gene were evaluated as R-Luc activity normalized against the activity of the co-introduced internal control, F-Luc. Data for each sample are means \pm SD of three measurements (* $p < 0.05$; Welch's t -test).

However, these activities between monocotyledonous plants and *A. thaliana* showed strong negative correlation (Figure 4D). This results indicate that the enhancer activities in *A. thaliana* were extremely different in the monocotyledonous plants. In addition, similar results were observed in previous study (Matsui et al. 2015). They tested the translational enhancer activity of *NtADH* 5'UTR sequence, which is known as dicotyledonous-type, in 13 species of monocotyledonous plants and 29 species of dicotyledonous plants. The activities varied among plant species, especially in monocotyledonous plants. Together, these results show that the sequence of effective translational enhancers in monocotyledonous plants differ from those in dicotyledonous plants (Figures 1C, 4D), and suggest that simple diversion of high-expression vectors from dicotyledonous to monocotyledonous plants has a significant risk of failure.

Finally, we confirmed that the translational enhancer activity was not target gene-specific by replacing the *F-Luc* reporter gene with *R-Luc* (Figure 5). Activity of the reporter R-Luc, normalized against the activity the F-Luc internal control was significantly higher than that of *OsADH*, indicating that these translational enhancer activities are independent of the CDS. We did observe slight differences between the monocotyledonous plants *O. sativa* and *S. cereale*, as well as between the *F-Luc* and *R-Luc* reporter genes, presumably due to differences in the structure of a *trans* factor involved in translation among monocotyledonous plants or differences in the interaction between the 5'UTR and each CDS sequence. Regardless of the cause, however, these differences were small. Nevertheless, if any effort to high expression is required, one approach would be to test several 5'UTR sequences, such as rank 8, 12, 13, and 14. In addition, it is possible that these sequences contribute to active translation under environmental stress. Thus, the

sequences discovered in this study provide powerful enhancers as part of vector for achieving high transgene expression in monocotyledonous plants, and their usefulness will be further demonstrated by testing under conditions of environmental stress.

Acknowledgements

We thank Dr. Kawasaki's laboratory at the Kindai University for helpful technical advice regarding *O. sativa* suspension cultures. This work was supported by Grants-in-Aid from the "Development of Fundamental Technologies for the Production of High-Value Materials Using Transgenic Plants" project from the Ministry of Economy, Trade and Industry of Japan (METI) (to K.K.), and the "Development of Production Techniques for Highly Functional Biomaterials Using Smart Cells of Plants and Other Organisms" project from the New Energy and Industrial Technology Development Organization (NEDO).

References

- Aoki H, Teramura H, Schepetilnikov M, Ryabova LA, Kusano H, Shimada H (2014) Enhanced translation of the downstream ORF attributed to a long 5'untranslated region in the *OsMac1* gene family members, *OsMac2* and *OsMac3*. *Plant Biotechnol* 31: 221–228
- Axelos M, Curie C, Mazzolini L, Bardet C, Lescure B (1992) A protocol for transient gene expression in *Arabidopsis thaliana* protoplasts isolated from cell suspension cultures. *Plant Physiol Biochem* 30: 123–128
- Baba A, Hasezawa S, Syono K (1986) Cultivation of rice protoplasts and their transformation mediated by *Agrobacterium* spheroplasts. *Plant Cell Physiol* 27: 463–471
- Juntawong P, Girke T, Bazin J, Bailey-Serres J (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in *Arabidopsis*. *Proc Natl Acad Sci USA* 111: 203–212
- Kawabe H, Ohtani M, Kurata T, Sakamoto T, Demura T (2018) Protein S-nitrosylation regulates xylem vessel cell differentiation in *Arabidopsis*. *Plant Cell Physiol* 59: 17–29
- Kawaguchi R, Bailey-Serres J (2005) mRNA sequence features that contribute to translational regulation in *Arabidopsis*. *Nucleic Acids Res* 33: 955–965
- Kawaguchi R, Girke T, Bray EA, Bailey-Serres J (2004) Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in *Arabidopsis thaliana*. *Plant J* 38: 823–839
- Kim MY, Kim TG, Yang MS (2017) Production and immunogenicity of *Actinobacillus pleuropneumoniae* ApxIIA protein in transgenic rice callus. *Protein Expr Purif* 132: 116–123
- King HA, Gerber AP (2016) Translatome profiling: Methods for genome-scale analysis of mRNA translation. *Brief Funct Genomics* 15: 22–31
- Matsui T, Sawada K, Takita E, Kato K (2015) Compatibility of translational enhancers with various plant species. *Plant Biotechnol* 32: 309–316
- Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K (2010) Genomewide analyses of early translational responses to elevated temperature and high salinity in *Arabidopsis thaliana*. *Plant Cell Physiol* 51: 448–462
- Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, Hirata K, Demura T, Kanaya S, Kato K (2013) A computational

- and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis. *Plant Cell Physiol* 54: 474–483
- Miyazaki S, Sato Y, Asano T, Nagamura Y, Nonomura K (2015) Rice MEL2, the RNA recognition motif (RRM) protein, binds in vitro to meiosis-expressed genes containing U-rich RNA consensus sequences in the 3'-UTR. *Plant Mol Biol* 89: 293–307
- Satoh J, Kato K, Shinmyo A (2004) The 5'-untranslated region of the tobacco *alcohol dehydrogenase* gene functions as an effective translational enhancer in plant. *J Biosci Bioeng* 98: 1–8
- Sugio T, Satoh J, Matsuura H, Shinmyo A, Kato K (2008) The 5'-untranslated region of the *Oryza sativa* alcohol dehydrogenase gene functions as a translational enhancer in monocotyledonous plant cells. *J Biosci Bioeng* 105: 300–302
- Takaiwa F (2007) A rice-based edible vaccine expressing multiple T-cell epitopes to induce oral tolerance and inhibit allergy. *Immunol Allergy Clin North Am* 27: 129–139
- Tang G, Qin G, Dolnikowski G, Russell R, Grusa KM (2009) Golden Rice is an effective source of vitamin A. *Am J Clin Nutr* 89: 1776–1783
- Ueda K, Okawara R, Yamasaki S, Sanada Y, Kinoshita E, Yoneda A, Demura T, Kato K (2014) Efficient transgene expression by alleviation of translational repression in plant cells. *J Biosci Bioeng* 118: 434–440
- Yamasaki S, Matsuura H, Demura T, Kato K (2015) Changes in polysome association of mRNA throughout growth and development in *Arabidopsis thaliana*. *Plant Cell Physiol* 56: 2169–2180
- Yamasaki S, Sanada Y, Imase R, Matsuura H, Ueno D, Demura T, Kato K (2018) *Arabidopsis thaliana* cold-regulated 47 gene 5'-untranslated region enables stable high-level expression of transgenes. *J Biosci Bioeng* 125: 124–130
- Yángüez E, Castro-Sanz AB, Fernández-Bautista N, Oliveros JC, Castellano MM (2013) Analysis of genome-wide changes in the transcriptome of Arabidopsis seedlings subjected to heat stress. *PLoS One* 8: e71425
- Yao J, Weng Y, Dickey A, Wang KY (2015) Plants as factories for human pharmaceuticals: Applications and challenges. *Int J Mol Sci* 16: 28549–28565