

RESEARCH ARTICLE

Correlation between gene expression levels under drought stress and synonymous codon usage in rice plant by in-silico study

Fatemeh Chamani Mohasses¹, Mahmood Solouki¹, Behzad Ghareyazie^{2*}, Leila Fahmideh¹, Motahhareh Mohsenpour^{2*}

1 Department of Plant Breeding and Biotechnology (PBB), Faculty of Agriculture, University of Zabol, Zabol, Iran, **2** Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research Education and Extension Organization (AREEO), Karaj, Iran

* ghareyazie@yahoo.com (BG); mthrh@yahoo.com (MM)



Abstract

We studied the correlation of synonymous codon usage (SCU) on gene expression levels under drought stress in rice. Sixty genes related to drought stress (with high, intermediate and low expression) were selected from rice meta-analysis data and various codon usage indices such as the effective number of codon usage (ENC), codon adaptation index (CAI) and relative synonymous codon usage (RSCU) were calculated. We found that in genes highly expressing under drought 1) GC content was higher, 2) ENC value was lower, 3) the preferred codons of some amino acids changed and 4) the RSCU ratio of GC-end codons relative to AT-end codons for 18 amino acids increased significantly compared with those in other genes. We introduce ARSCU as the Average ratio of RSCUs of GC-end codons to AT-end codons in each gene that could significantly separate high-expression genes under drought from low-expression genes. ARSCU is calculated using the program ARSCU-Calculator developed by our group to help predicting expression level of rice genes under drought. An index above ARSCU threshold is expected to indicate that the gene under study may belong to the “high expression group under drought”. This information may be applied for codon optimization of genes for rice genetic engineering. To validate these findings, we further used 60 other genes (randomly selected subset of 43233 genes studied for their response to drought stress). ARSCU value was able to predict the level of expression at 88.33% of the cases. Using third set of 60 genes selected amongst high expressing genes not related to drought, only 31.65% of the genes showed ARSCU value of higher than the set threshold. This indicates that the phenomenon we described in this report may be unique for drought related genes. To justify the observed correlation between CUB and high expressing genes under drought, possible role of tRNA post transcriptional modification and tRFs was hypothesized as possible underlying biological mechanism.

OPEN ACCESS

Citation: Chamani Mohasses F, Solouki M, Ghareyazie B, Fahmideh L, Mohsenpour M (2020) Correlation between gene expression levels under drought stress and synonymous codon usage in rice plant by in-silico study. PLoS ONE 15(8): e0237334. <https://doi.org/10.1371/journal.pone.0237334>

Editor: Niranjana Baisakh, Louisiana State University, UNITED STATES

Received: February 7, 2020

Accepted: July 23, 2020

Published: August 10, 2020

Copyright: © 2020 Chamani Mohasses et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its supporting information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Abiotic stresses such as drought, salinity, low or high temperatures and other environmental extremes negatively affect crop growth and reduce crop yield on a global scale [1]. Among

these stresses, drought stress is the most important factor-limiting yield in agricultural systems in arid and semi-arid regions [2]. Rice (*Oryza sativa*), a well-known cost-effective cereal in the world, is very sensitive to drought stress because of its limited adaptation to water-deficit conditions. Although rice germplasm has a large amount of functional genetic diversity for drought tolerance-related traits/mechanisms [3], breeding for drought tolerance is still a challenge for breeders because of its complex genetic nature with higher environmental plasticity and multiple metabolic pathways involved. There are hopes that transgenic technology can improve stress tolerance by introducing novel exogenous genes or altering the expression levels of endogenous genes [4]. Researchers have endeavored during recent decades to generate transgenic crops with improved tolerance against abiotic stresses. The first step of transgenic research is the identification of genes serving as key regulators of different metabolic pathways, including osmolyte synthesis, ion homeostasis through selective ion uptake, antioxidant defense system and other frontline defense pathways [5]. Rice is currently considered a model crop for genetic engineering with well-developed tissue culture and gene transfer protocols. There are numerous reports on the production of genetically modified (GM) rice for such traits as enhanced resistance against insect pests and/or herbicide tolerance [6–10]. In some cases, approvals have even been issued for such GM rice plants in Japan, Canada, Colombia, Mexico, Honduras, New Zealand, Philippines and the United States [11]. However, unlike in the case of drought tolerant maize (DroughtGard) [12], to date there is no verifiable report of production and/or approval of any GM drought-tolerant rice. Despite the ever-increasing discovery and introduction of genes affecting different traits, the challenge remains to discover and explore genes effectively rendering abiotic stress tolerance [13]. Plant species are reported to indicate wide diversity in terms of their gene expression, physiology and stress response under different environmental conditions [14]. A key to progress towards breeding better crops for stressed conditions is to understand the alterations in cellular, biochemical and molecular machinery that occur in response to stress [15]. A particular stress changes the expression level of specific genes in a species-dependent fashion. It causes differences in the efficiency of signal perception and subsequent transcriptional changes leading to elicitation of a specific response and adaptation and finally enhanced stress tolerance [16]. In addition, gene expression levels correlate with several features of the underlying genes and encoded proteins, including synonymous codon (codons encoding the same amino acid) usage, amino acid composition, rates of protein evolution, and the length of coding sequence [17,18]. Synonymous codons are not equally present in all genes and/or species. This phenomenon, called codon usage bias (CUB), is ubiquitous in all organisms. Mutational pressure and natural selection are considered the two major factors contributing to CUB [19]. The correlation of CUB with gene expression levels has been reported in organisms from all domains of life [20–29]. Codon usage presents the specific pattern of gene expression and it has been noted that the use of the same codon is preferred in gene expression under the same physiological state [30]. It has been shown that the overall preference of amino acid usages and codon usage preferences in the proteins of a given organism were significantly affected by major environmental factors [31]. This study notes the presence of a correlation between cellular function and codon usage profiles of the genes in the studied pairs. In addition, some previous studies showed the existence of correlation between gene expression levels and CUB [18,32,33]. Furthermore, several published studies support the idea that translation of specific codon-biased transcripts in stress response genes can be regulated by tRNA modifications [34–36]. Wobble-base tRNA modification levels have the potential to work in concert with codon usage patterns in specific transcripts to regulate translation of response proteins. Dedon et al. [37] designated these as modification tunable transcripts (MoTTs). Modified tRNA directly participate in translation. However, unmodified mature tRNAs or pre-tRNAs are processed into various types of transfer

RNA-derived fragments known as tRFs [38–41]. Production of tRFs is promoted by abiotic stress conditions to eventually control gene expression by transcriptional, post-transcriptional and translational regulation [42].

Hence, knowledge of the codon usage and codon-pair context patterns of plants and underlying evolutionary forces will be useful to understand the molecular mechanism of environmental adaptation and biological diversity of each species [43]. In addition, this information can be applied for codon optimizations of diverse sets of genes to be used in plant transformation programs. Here, we use genome sequence and genome-wide expression data of drought responsive genes to investigate the relationship between gene expression levels and CUB in rice.

Material and methods

Selection of genes with different expression levels under drought stress

Sixty genes associated with drought stress in rice were identified through a structured literature survey from meta-analysis data [44–46]. Then the genes were divided based on their expression levels under drought condition into three categories: high-, intermediate- and low-expression genes. Complete coding DNA sequences of these genes were accessed from the NCBI nucleotide database [47]. In the present study, high-expression genes included NAC transcription factors, zinc finger proteins, AP2/ERF transcription factors, LEA proteins, phosphatidylethanolamine binding proteins, peptide transporters, dioxygenases, actin-binding proteins, protein phosphatases, dehydrins and membrane proteins family. These genes have been reported to show high expression under drought stress in different experiments (Table 1). In addition to the high expression, some of these genes showed enhanced drought tolerance in reported genetic engineering experiments [48–57] (Table 2).

Analysis of the nucleotide composition

The program CAIcal [58] was used to calculate the AT content and GC content at first, second and the third nucleotide positions, (AT1, AT2, AT3) and (GC1, GC2, GC3), respectively.

Relative Synonymous Codon Usage (RSCU) analysis

RSCU was calculated following Hastings and Emerson [59], as the ratio of the observed frequency of a codon to the expected frequency of the same codon within a synonymous codon group (with no bias) in the entire coding sequence of the gene concerned. RSCU can be equal to one when there is no bias, more than one when there is preference to use that specific codon and less than one when the codon is underused [60]. RSCU values can range from zero where the specific codon is not used at all to 2, 3, 4 and 6 when only one codon is used for encoding amino acids with 2, 3, 4 and 6 synonymous codons, respectively. An RSCU value >1 for each codon shows that this codon is preferred. Higher RSCU value indicates the presence of higher CUB. The RSCU of 60 genes associated with drought stress in rice were calculated using ACUA software [61] and the program CAIcal [58], excluding the stop codons and the two amino acids which are encoded by a single codon (Trp and Met).

ENC analysis

Effective number of codon usage (ENC) refers to the number of unique codons found in a gene. This value can be in a range of 20 (where there is an extreme bias towards the use of only one codon for each amino acid) to 61 (representing the use of all synonymous codons). If

Table 1. Genes, their level of expression under drought condition (high, intermediate and low), their encoded proteins, log₂ of the fold change (gene expression level), the nucleotide composition, ENC value and CAI value. (Additional information about the genes used in genetic transformation programs (shown with asterisk) is provided in Table 2).

Expression Level	Gene Name	Nucleotide Length	Protein/ Function	Log ₂ of the Fold Change ¹	%A +T	%G +C	%G +C(1)	%G +C(2)	%G +C(3)	ENC	CAI
High	LOC_Os02g04780	1080	Uncharacterized protein/function	6.71	31.2	68.8	71.4	42.8	92.2	33.8	0.908
	LOC_Os05g39250	504	Phosphatidylethanolamine-binding protein	8.38	29.2	70.8	71.4	43.5	97.6	31.1	0.922
	LOC_Os01g50910	648	LEA protein	7.53	30.1	69.9	64.4	54.6	90.7	27.4	0.919
	LOC_Os05g33960	1815	Peptide transporter PTR2	5.03	31.4	68.7	63.5	48.1	94.4	32.2	0.918
	LOC_Os10g39920	387	Uncharacterized protein	6.01	26.6	73.4	67.4	57.4	95.3	28.1	0.926
	LOC_Os07g05940	1749	9-cis-Epoxycarotenoid dioxygenase 1, chloroplast	7.88	28.2	71.8	72	48	95.2	31.1	0.925
	LOC_Os03g60580	453	actin-Depolymerizing factor	5.89	30.9	69.1	63.6	46.4	97.4	30.3	0.946
	LOC_Os09g15670	1077	Protein phosphatase 2C	8.77	26.4	73.6	70.2	56.8	93.9	34.5	0.9
	LOC_Os11g26780	495	Dehydrin family	9.49	31.1	68.9	64.8	49.7	92.1	33.4	0.919
	LOC_Os07g24000	522	AWPM-19-like membrane family protein	6.41	31.8	68.2	64.4	46	94.3	31.2	0.934
	LOC_Os01g66120*	912	LEA protein	5.46	30.2	69.8	63.5	49.3	96.7	30.6	0.929
	LOC_Os03g60080*	951	LEA protein	6.96	32.8	67.2	61.2	48.6	91.8	32.8	0.908
	LOC_Os05g10670*	1395	NAC transcription factor	2.28	28.1	71.9	65.6	60.4	89.7	34.4	0.894
	LOC_Os03g60560*	513	NAC transcription factor	2.47	27.1	72.9	65.5	56.7	96.5	31.5	0.922
	LOC_Os12g39400*	753	Zinc finger gene family	2.45	25.2	74.8	72.9	59	92.4	30	0.895
	LOC_Os08g31580*	843	Zinc finger gene family	3.16	27.8	72.2	67.3	55.9	93.6	30.6	0.913
	LOC_Os03g20680*	1035	Zinc finger gene family	4.28	27.2	72.8	66.7	52.8	98.8	29.7	0.943
	LOC_Os05g46480*	603	AP2/ERF transcription factor	2.55	30.7	69.3	61.2	52.2	94.5	31.1	0.931
	LOC_Os03g57240*	906	C2H2 zinc finger protein	2.8	27.3	72.7	65.2	63.6	89.4	36.3	0.859
	LOC_Os05g34830*	969	NAC transcription factor	6.54	28.8	71.2	66.3	53.6	93.8	32	0.919
Intermediate	LOC_Os01g04840	1347	Vacuolar protein sorting-associated protein 4B	0	48.9	51.1	53.7	41.6	57.9	55.6	0.805
	LOC_Os01g04930	885	MYB family transcription factor	0	38.8	61.2	56.3	47.8	79.7	46.8	0.867
	LOC_Os01g05020	2016	Uncharacterized protein	0	43.5	56.5	54.6	36.8	78.3	46.8	0.859
	LOC_Os01g05090	657	Uncharacterized protein	0	54.9	45.1	47	44.7	43.4	50	0.750
	LOC_Os01g05100	1014	Uncharacterized protein	0	51.1	48.9	53.6	42.3	50.9	52.1	0.760
	LOC_Os01g05180	426	Uncharacterized protein	0	35.4	64.6	59.9	57.7	76.1	47.2	0.852
	LOC_Os01g05289	339	Uncharacterized protein	0	33.9	66.1	68.1	60.2	69.9	51.7	0.799
	LOC_Os01g05420	1602	Uncharacterized protein	0	47.0	53	61.8	40.8	56.4	55.9	0.789
	LOC_Os01g05440	1236	XH domain-containing protein	0	43.2	56.8	59.2	42.5	68.7	49.9	0.821
	LOC_Os01g05470	2103	Transcription factor X1	0	47.9	52.1	59.6	37.7	58.9	56	0.797
	LOC_Os01g05480	1665	Uncharacterized protein	0	44.0	56	55.3	42.7	69.9	49.4	0.832
	LOC_Os01g05490	762	Triose phosphate isomerase, cytosolic	0	50.7	49.3	57.9	41.7	48.4	50.4	0.772
	LOC_Os01g05520	321	Uncharacterized protein	0	39.6	60.4	71	52.3	57.9	48.9	0.744
	LOC_Os01g05540	1200	Uncharacterized protein	0	43.0	57	57	44.5	69.5	49.4	0.844
	LOC_Os01g05620	2970	NBS-LRR disease-resistant protein	0	57.8	42.2	50.3	33	43.1	51.1	0.733
	LOC_Os01g05744	678	Uncharacterized protein	0	44.9	46.2	40.7	38.5	59.3	53.7	0.792
	LOC_Os01g05940	4233	Receptor kinase	0	52.1	47.9	49.8	41.7	52.2	55.9	0.766
	LOC_Os01g06050	1377	Uncharacterized protein	0	49.3	50.7	55.6	48.6	47.9	56.1	0.776
	LOC_Os01g06140	924	Uncharacterized protein	0	46.4	53.6	51.6	45.5	63.6	52.6	0.809
	LOC_Os01g06320	921	MYB family transcription factor	0	44.3	55.7	53.1	44.3	69.7	50	0.837

(Continued)

Table 1. (Continued)

Expression Level	Gene Name	Nucleotide Length	Protein/ Function	Log ₂ of the Fold Change ¹	%A +T	%G +C	%G +C(1)	%G +C(2)	%G +C(3)	ENC	CAI
Low	LOC_Os04g11060	177	Uncharacterized protein	-7.59	53.1	46.9	54.2	28.8	57.6	47	0.777
	LOC_Os11g24140	375	Plastocyanin-like domain-containing protein	-10.6	39.7	60.3	57.6	48	75.2	51.3	0.831
	LOC_Os09g19820	1173	Aminopeptidase	-5.31	58.4	41.6	49.4	35.8	39.6	51.7	0.735
	LOC_Os02g54180	399	Uncharacterized protein	-6.39	46.9	53.1	52.6	45.9	60.9	53.8	0.802
	LOC_Os07g23430	1410	Fatty acid desaturase	-4.67	37.0	63	59.6	45.5	84	40.3	0.870
	LOC_Os07g23410	1173	Fatty acid desaturase	-6.44	37.2	62.8	59.3	41.7	87.5	38.6	0.894
	LOC_Os04g11040	270	Uncharacterized protein	-6.05	44.1	55.9	57.8	50	60	42.7	0.794
	LOC_Os12g38290	231	Metallothionein	-7.68	40.3	59.7	46.8	61	71.4	47.6	0.849
	LOC_Os11g20160	1107	O-methyltransferase	-9.68	38.2	61.8	64.8	41.2	79.4	44.4	0.869
	LOC_Os11g14910	1080	NADP-dependent oxidoreductase	-8.59	45.7	54.3	59.4	38.9	64.4	53.8	0.803
	LOC_Os11g04350	1071	Cell death-associated protein	-7.52	33.3	66.7	70.6	46.5	82.9	39.3	0.876
	LOC_Os06g18960	2571	embryogenesis Transmembrane protein	-5.9	41.8	58.2	58.8	43.3	72.5	48.6	0.830
	LOC_Os01g72140	696	Glutathione S-transferase	-5.76	39.5	60.5	64.7	37.1	79.7	41.8	0.864
	LOC_Os07g04960	456	app1	-5.75	29.2	70.8	70.4	58.6	83.6	41.6	0.884
	LOC_Os11g37700	4368	Pleiotropic drug-resistant protein	-5.74	53.7	46.3	49.5	41	48.4	56.5	0.753
	LOC_Os08g37250	858	Patatin	-5.69	46.0	54	51.4	35.7	74.8	50.6	0.861
	LOC_Os03g46150	732	LTPL7—Protease Inhibitor/seed storage/LTP family protein precursor	-5.67	30.3	69.7	64.8	64.8	79.5	40.2	0.846
	LOC_Os10g02070	981	Peroxidase precursor	-5.63	46.4	53.6	54.7	40.4	65.7	50.7	0.820
	LOC_Os05g16930	2679	Protein kinase	-5.48	56.1	43.9	48.2	41.9	41.7	54.9	0.729
	LOC_Os03g05530	1125	Nodulin	-5.23	44.2	55.8	48.5	40.5	78.4	44.1	0.842

¹ Relative expression of rice genes under drought stress. Values are shown in log₂ of the fold changes (FC) between expression values of each gene under control and drought conditions. Genes with FC above 1 were counted as up regulated, and, below -1, down regulated.

<https://doi.org/10.1371/journal.pone.0237334.t001>

there are fewer than 60 synonymous codons used in a particular gene, then CUB may be present (considering the fact that Trp has only one codon). We used CAIcal to calculate ENC.

Codon Adaptation Index (CAI) analysis

CAI is another widely used method for evaluation of CUB. It measures the similarity between the frequency of synonymous codons used by a gene and that of a reference set [58]. The range of the CAI value is between 0 and 1. The rice codon usage from the Kazusa database [62] was used as a reference set. The program CAIcal [58] was used for CAI calculation.

Searching for tRFs data under drought stress

Data on tRFs under drought stress in rice were obtained from the PtRFdb (<http://223.31.159.8/PtRFdb/plant.php?txt=Oryza%20sativa&bs=1&type=all>) [63] and their frequency was compared in the control and drought stress condition in different stages of plant growth. Codons related to each tRF were also compared with the rice preferred codons under drought stress.

Results

Base composition analysis

In the present study, GC content was higher than AT content in all three gene-expression categories (high, intermediate and low-expression genes), as is expected for the rice genome. The

Table 2. High expression genes that have been used in genetic engineering programs for enhanced stress tolerance.

Name	Gene	Protein	Observation	Reference
LOC_Os03g20680	<i>Os LEA3-2</i>	LEA protein	The transgenic rice showed significantly stronger growth performance than control under salinity or osmotic stress conditions and were able to recover after 20 days of drought stress.	[43]
LOC_Os05g46480	<i>Os LEA3-1</i>	LEA protein	Overexpression of an LEA gene in rice improved drought resistance under the field conditions without yield penalty.	[44]
LOC_Os01g66120	<i>SNC2</i>	NAC transcription factor	Overexpression of <i>SNAC2</i> improved the tolerance to PEG treatment. The transgenic plants had higher cell membrane stability and higher germination and growth rate than wild type during the cold and salt stresses respectively.	[45]
LOC_Os03g60080	<i>OsNAC9</i>	NAC transcription factor	Overexpression of <i>OsNAC9</i> gene increased rice drought tolerance.	[46]
LOC_Os12g39400	<i>ZFP252</i>	Zinc finger	Overexpression of <i>ZFP252</i> in rice increased the amount of free proline and soluble sugars enhanced rice tolerance to salt and drought stresses.	[47]
LOC_Os03g60560	<i>ZFP182</i>	Zinc finger	Overexpression of <i>ZFP182</i> significantly enhanced salt, cold and drought tolerance in transgenic rice.	[48]
LOC_Os05g10670	<i>OsTZF1</i>	Zinc finger gene family	Overexpression of <i>OsTZF1</i> in rice showed improved tolerance to high-salt and drought stresses.	[49]
LOC_Os08g31580	<i>OsDRAP1</i>	AP2/ERF transcription factor	Overexpressing <i>OsDRAP1</i> transgenic plants exhibited significantly improved drought tolerance.	[50]
LOC_Os03g57240	<i>DST</i>	C2H2 zinc finger protein	A zinc finger protein, <i>DST</i> , regulated drought and salt tolerance in rice via stomatal aperture control.	[51]
LOC_Os05g34830	<i>OsNAC52</i>	NAC transcription factor	Overexpression of <i>OsNAC52</i> activated the expression of downstream genes in transgenic <i>Arabidopsis</i> , resulting in enhanced tolerance to drought stresses but not growth retardation.	[52]

<https://doi.org/10.1371/journal.pone.0237334.t002>

means of GC content in categories 1, 2 and 3 were 70.9%, 56.95% and 54.2%, respectively. The results showed that the GC content at the three-codon positions was noticeably different. G or C at position three of each codon (GC3) was higher than that at position one (GC1), and it was lowest at position two (GC2) among most of the studied genes. The means of GC3 content in categories 1, 2 and 3 were 94%, 69.3% and 56.2%, respectively (Table 1, Fig 1).

ENC analysis

The CUB for each gene was also calculated using ENC. The ENC value of the genes of the first group was between 27 and 34, the second group was between 38 and 56 and the third group was between 46 and 56 (Table 1).

Analysis of CAI

The range of the CAI value in the first category was between 0.859 and 0.946; in the second and third categories were between 0.733 and 0.867 and between 0.729 and 0.894, respectively (Table 1). High expression genes showed higher CAI value as compared to the other categories.

RSCU analysis

The comparative study among the RSCU values of the studied genes in three expression categories under drought stress with that in general pattern of rice genes reveals the existence of CUB. In other words, when genes with different expression levels under drought in rice are compared with the standard rice codon usage table, a change in preferred codon of some amino acids of high expression genes and low expression genes under stress is observed. For example, alanine and serine are encoded by the most preferred codon GCG and TCG instead

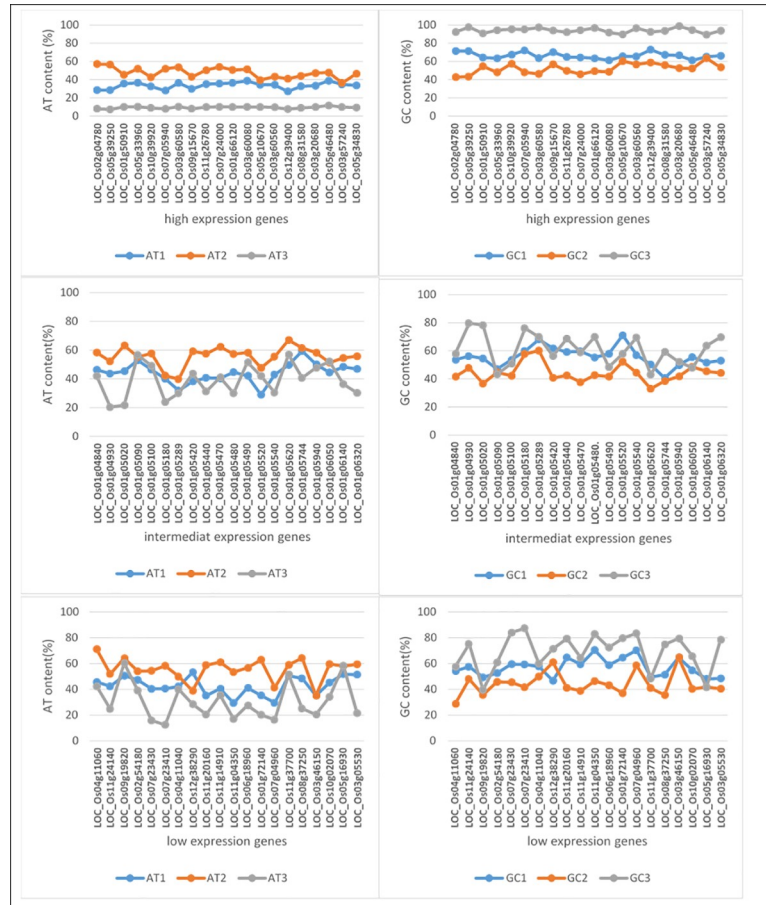


Fig 1. The nucleotide composition of three gene expression categories under drought stress (high, intermediate and low expression) in 60 selected rice genes.

<https://doi.org/10.1371/journal.pone.0237334.g001>

of rice’s most widely used codons GCC and TCC, respectively, in the high expression genes category. The third codon position changed to the nucleotide G in this category. Aspartic acid and proline are encoded by the most preferred codon GAT and CCA instead of rice standard codons for these two amino acids: GAC and CCG, respectively, in the intermediate expression genes category. The third codon position changed to the nucleotides A and T in this category (Fig 2).

The heat-plot of different gene categories based on the RSCU index is shown in Fig 3. This plot separates distinctly the high expression genes under drought conditions from other categories so that the colors were divided into two distinct groups (Fig 3). These data may indicate that genes with higher expression under drought stress have acquired their own preferred codons in the course of evolution. Also, the heat-plots showed that G or C end codons are highly preferred in this set of genes for most amino acid codons, but the TTG codon was not preferred for leucine in high expression genes under drought stress.

Average RSCU (GC/AT): A new index for estimation of CUB

After calculating the RSCU index for each codon per gene, it was observed that the ratio of RSCU of codons with GC end to codons with AT end in each amino acid was significantly higher in high expression genes than that in the other categories. Even in some high expression



Fig 2. RSCU values of each codon/amino acid across 20 genes in each high, intermediate and low gene expression categories in rice.

<https://doi.org/10.1371/journal.pone.0237334.g002>

genes, the RSCU of codons with AT end was zero. Hence, an RSCU-based index was defined and called ARSCU. It measures the average ratio of RSCUs with GC- to AT-end codons for all amino acids in one gene, providing only one value for each gene. Our findings indicate that ARSCU could significantly separate highly expressed genes under drought from the others. In this study, the ARSCU (GC/AT) index ranged between 9.5 to 18.5, 0.7 to 8.5, and 0.7 to 4.5 for high, intermediate and low expression genes, respectively (Fig 4A).

ARSCU was calculated as:

$$\text{ARSCU} = \left(\frac{\sum_{aa:1}^{aa:18} a/b}{18} \right)$$

Where aa is amino acid, a is RSCU of GC end codons and b is RSCU of AT end codons (any a and b with a value of zero is arbitrarily assigned a value of 0.1).

ARSCU- calculator

A program was developed in an Excel file for calculating the ARSCU index (S1 Table). Input data for this program is the RSCU index of all amino acid codons in each gene. This program calculates only one ARSCU (GC/AT) for each gene (whilst every gene has 59 RSCU values). Since RSCU may be zero for some codons, any RSCU with a value of zero is arbitrarily assigned a value of 0.1.

Based on our findings, ARSCU values can be used for discrimination of genes with different expressions under drought conditions, as follows:

- Genes with ARSCU above 13 (arbitrary threshold) are predicted to be high expression genes under drought conditions;
- Genes with ARSCU ranging from 9 to 13 are predicted to be genes with high or intermediate expression;
- Genes with ARSCU numbers less than 9 are predicted to be genes with low or intermediate expression.

Validation of discriminating power of ARSCU

For validation of discriminating power and accuracy of the ARSCU, 60 other rice genes were randomly selected from 43233 genes studied for their response to drought stress [44] and their ARSCU were calculated by the program (S2 Table). The results were correlated with the expression data from the published meta-analysis [44]. ARSCU calculator was able to predict genes with different expression level under drought stress in rice with an accuracy of 88.3% (Fig 4B).

In order to compare the codon usage pattern of different highly expressing genes unrelated to drought with that of drought responsive high expression genes, third set of 60 other highly expressed genes not related to drought [64,65] were used. The ARSCU value for these genes was also calculated by the program (S3 Table). Only 31.65% of the genes (18.33% plus 13.33% were correctly categorized as high and intermediate expressing genes, respectively) similar to those obtained for drought responsive genes (Fig 4C). ARSCU as an index worked for

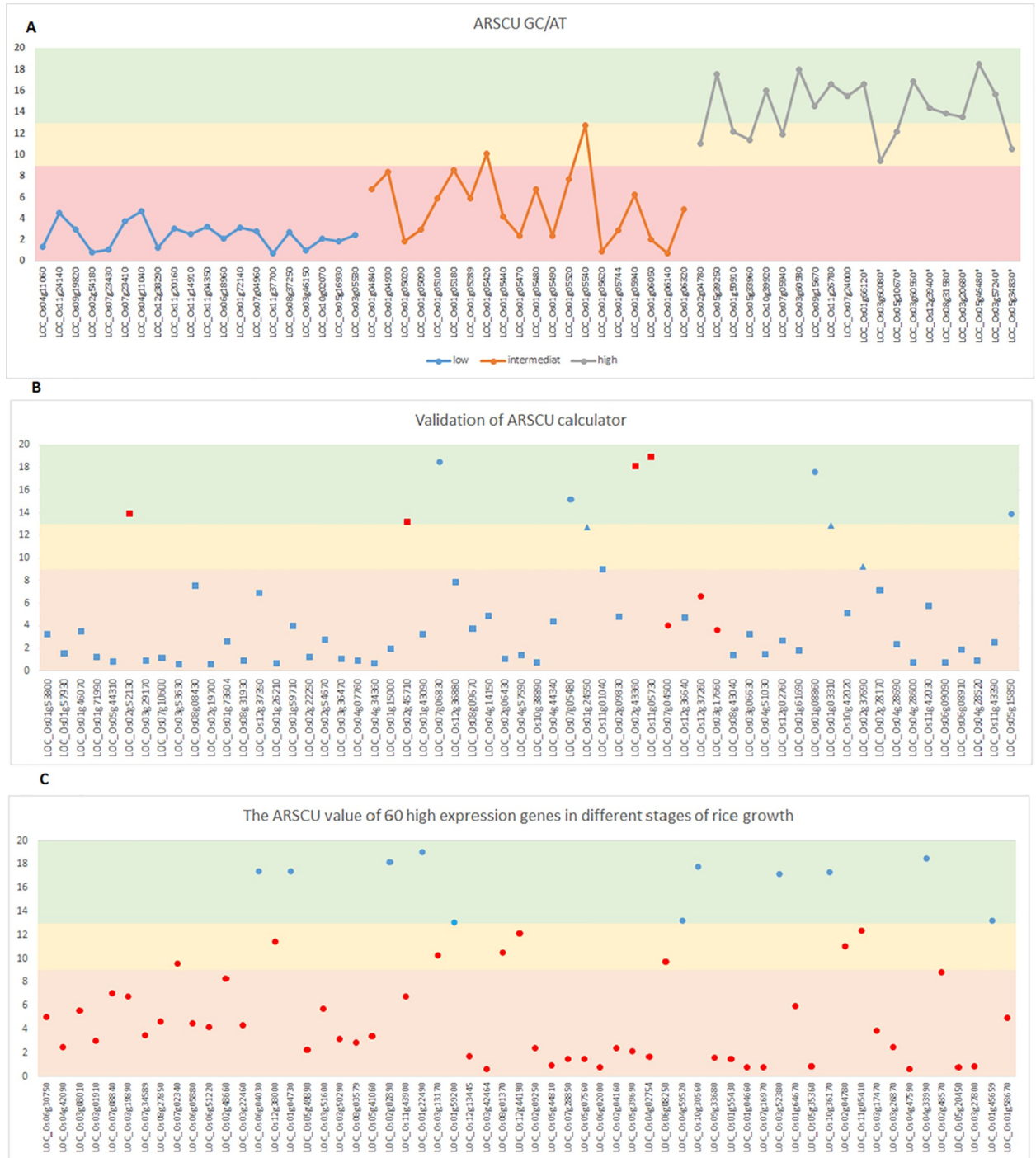


Fig 4. (A) The ARSCU GC/AT value of high, intermediate and low expression gene categories in rice; (B) Validation of ARSCU-Calculator using 60 rice drought stress related genes, randomly selected and calculated by the program. (C) The ARSCU value of 60 high expressing non drought related genes. Genes with ARSCU above 13 are estimated as high expression genes under drought conditions) green area. (ARSCU between 9 to 13 are estimated as genes with high or intermediate expression) (yellow area). ARSCU less than 9 are estimated as genes with low expression or intermediate expression (pink area). Blue and red shapes: correct and wrong prediction, respectively. High, intermediate and low-expression genes are indicated with solid circle, triangle and square, respectively.

<https://doi.org/10.1371/journal.pone.0237334.g004>

Table 3. Frequency of tRFs under control and drought stress conditions in rice.

Amino Acid	Anticodon*	frequency of tRFs under control condition	frequency of tRFs under drought stress	preferred codon**	tRF for preferred codon under drought
Cyc	ACA	9944	8089	TGC	No
Glu	CTC	506	0	GAG	Yes, but decreased under drought
Gly	CCC	33	11	GGC	No
His	GTG	33	17	CAC	Yes, but decreased under drought
Leu	AAG	396	655	CTC	No
Met	CAG	340	178	None	No preferred codon-
Asn	GTT	97	69	AAC	Yes, but decreased under drought
Arg	ACG	0	17	CGC	No
Ser	GCT	234	244	TCG	No
Val	GAC	206	220	GTG	No
Tyr	GTA	27	0	TAC	Yes, but decreased under drought
Ala	CGC	1377	1947	GCG	Yes
Asp, Phe, Ile, lys, Pro, Gln, Thr, Trp	-	tRF not reported	tRF not reported	-	No

*Anticodon of the type of tRNA that breaks into tRF

**Preferred codons of high expression genes under drought stress

<https://doi.org/10.1371/journal.pone.0237334.t003>

identification of drought related genes (with 88% accuracy) compared to only 32% accuracy on non-drought related genes.

tRFs under drought stress

Results of our search for tRFs under drought stress and control condition in rice, showed that:

1. In most cases, no tRF has been reported for the preferred codon under drought stress, which means the tRNA remains for longer time and participates in translation.
2. In few cases, tRF has been reported for tRNA related to preferred codon under drought stress. However, even in these cases the frequency of this tRF has decreased under drought stress compared with its frequency under normal condition.
3. Only in one case (Ala), tRF related to preferred codon has increased under drought stress. tRNA expression data is required in order to explain this observation. One mechanism may be over expression of the preferred tRNA in such amounts that even its break down to its tRF does not affect the over-expression of the gene in concern (Table 3).

Discussion

Drought tolerance in rice is governed by many genes with huge environmental interaction, with low heritability, and thus are difficult to study [66]. Development of GM plants with enhanced tolerance to drought is an important challenge in rice biotechnology research. For this purpose, further investigation of the mechanisms, pathways and genes involved in response to abiotic stress is essential. Expression of genes belonging to diverse functional and regulatory groups, such as transcription factors, protein kinases, and phosphatases, are influenced by abiotic stress conditions [67]. In addition to environmental conditions, gene

expression levels correlate with multiple aspects of gene sequence and structure including SCU [18]. Here, we report the existence of strong correlation between SCU and gene expression levels under drought stress in rice which has not previously been reported.

Correlation between GC content and higher expression of rice genes under drought

Our analysis of codon usage patterns of studied genes in different expression categories under drought stress shows that most of the genes in the three categories have more GC content as expected from rice genes. In this study, GC percent analysis was able to separate 60 selected genes in rice into two major classes: 1) the high expression genes that were also high in GC and 2) the other genes with lower GC content. Higher GC content in rice genome has been reported with no particular attention paid for presence or absence of differences among genes with different levels of expression under drought [68]. Wang and Hickey [69] showed that the means of GC content of 14005 rice-coding sequences was 57.8%. Also they separated the 14,005 rice genes into two classes: High (67.4%) and Low GC genes (50.1%). The means of the high GC rice genes and low GC rice genes' GC3 were 80.4% and 52.7%, respectively in their study. Also, Yi et al. [33] showed higher expression is associated with higher GC3 in three closely-related species of the genus *Misgurnus*. Our results further confirm the previous reports in the field. However, the novelty of our findings may be the existence of a correlation between genes with different levels of expression under drought condition and their GC content. We showed that genes with high expression under drought conditions had overall 70.9% GC content. Furthermore, we showed that on average 94% of the high expressing genes contained a G or C at third position of their codons. Therefore, we have demonstrated that high expression genes under drought are more dependent on high GC percentages, especially in the third nucleotide position.

Correlation between SCU and higher expression of rice genes under drought

As we expected, the range of the CAI value was higher in the high expression genes compared with other categories. In addition, the ENC value for the high expression genes was less than 36, which is an indication of the presence of a relatively strong CUB. This value for the second and third categories of genes was above 38 and 46, respectively. Sharp et al. [60] reported that, highly expressed genes generally show a tendency of preferentially using a limited number of codons. Also, Liu et al. [70] showed a positive correlation between ENC value and gene expression level in the rice plant. In their report, genes with $ENC \leq 30$ and $ENC \geq 55$ were correlated with high and low expression genes, respectively.

Alanine and serine in the high expression genes under drought stress preferred the GCG and TCG codons, respectively, whereas according to the common rice codon usage table, the preferred codons for alanine and serine were GCC and TCC, respectively. Our results about those amino acids encoded by six codons (leucine, arginine and serine) showed that among the C- or G-ending codons, using G or C in the first and second positions is also more preferred. For example, for the case of leucine, the use of the TTG codon, although ending with a G, is not preferred and the CTC and CTG codons are more preferred. Therefore, using TTG is not recommended for codon optimization for enhanced expression of the concerned gene under drought. Intermediate and low expression genes under drought stress end with A or T codons in some cases. This may further indicate the effect of more efficient codons on higher expression of genes under drought stress. For example, proline and aspartic acid preference the GAT and CCA codons in the intermediate expression genes under drought stress, whereas

the preferred rice codon for these two amino acids is GAC and CCG that end with a G or C. Therefore, the use of A and T end codons for these amino acids has reduced the expression of these genes under drought stress. A number of studies of eukaryotes and prokaryotes have showed that groups or families of genes involved in stress responses systematically overuse and underuse specific “non-optimal” synonymous codons. Chionh et al. [71] showed the 48 genes in the DosR regulon in *Mycobacterium bovis* BCG, which are essential for survival under hypoxic stress, are enriched in G- and C-ending codons. Others showed that hypoxic stress increased the translation of proteins from genes enriched with non-optimal codons ACG (Thr), CTA (Leu), GCG (Ala) and GGA (Gly), whereas their synonymous partners ACC (Thr), CTT (Leu), GCT (Ala) and GGT (Gly) were all overrepresented in downregulated proteins in hypoxia [72]. High-expression genes under drought in rice use G/C-end codons more frequently. Therefore, we created a different codon usage table for these genes compared to the standard rice codon usage table (S4 Table).

Correlation of CUB with gene expression levels for different species including: *Caenorhabditis*, *Drosophila*, *Arabidopsis* [20], *Tribolium castaneum* [18], arthropods [32,72], *Paeonia lactiflora* [73] and *Populus tremula* [74] has been shown by several authors. Selection favors specific codons promoting the efficient and accurate translation of genes that are expressed at high levels.

ARSCU as a single value for prediction of rice genes expression under drought

Previous reports have used the RSCU number for different cases, but this number is calculated for the codons of each amino acid in the gene. In this report, we introduced the ARSCU GC/AT as a single value/index for each gene with the potential to predict gene expression under drought stress. In addition, we developed a simple program in an Excel file for calculating the ARSCU index (ARSCU calculator). It is expected that using this program, we will be able to predict high expression genes in high throughput data with relatively high confidence. This program was developed based on calculations made on a set of genes expressed under drought stress conditions in the rice genome. We validated 1) the discriminating power of ARSCU and 2) its specificity for higher expressing genes under drought. This index was not able to predict highly expressed non-drought related genes. These findings may reflect the different codon usage pattern of the genes under drought stress compared with other highly expressed genes in rice. In other words, the observed phenomenon (strong correlation between ARSCU and high expressing genes) is restricted to higher expressing genes under drought stress and not every high expressing genes in rice.

It remains unclear if ARSCU will also be useful for prediction of gene expression in other plant species under drought or even other stressed conditions. We are currently attempting validation of this program for prediction of gene expression under drought for other cereals. It may require a different threshold definition than the one reported by this index (ARSCU) in this article. There is room for improvement of this program to be suitable for use with the codon usage features of other plant species. The prediction made by this program potentially indicates the extent to which such a codon composition can be expressed under drought. It is, however, clear that in addition to ORF, regulatory elements in particular the promoter and transcription factors affect gene expression and that these variables were not considered in this program. It is therefore expected that even in rice, genes with high ARSCUs may exhibit low expression under drought. The high ARSCUs in such genes only indicate their codon potential for high expression under stressed conditions. In the case of such genes, alteration of regulatory elements may materialize their expression potential.

Possible role of tRNA post transcriptional modification and tRFs justifying the correlation between CUB and high expressing genes under drought

We observed different codon usage pattern in high expression genes under drought stress in our research. This observation termed Modification Tunable Transcripts (MoTTs) was also reported as a model combining genome-wide CUB analytics and gene expression studies. This model implies that modification of tRNA drives the translational regulation of critical response proteins whose transcripts display a distinct codon bias [75].

Average transcripts not exhibiting CUB, do not need specific tRNA modifications, as they are efficiently translated under all conditions. In contrast, the translation of MoTTs in cells lacking tRNA specific modifications is severely reduced [37].

The number of genes that encode tRNA modification enzymes are far more abundant than the tRNA coding genes, which further highlights the importance of tRNA modification [76,77]. It has been long postulated that modified nucleosides on tRNA molecules may function as “bio-sensor” for environmental and physiological changes [78,79] as a fast module to regulate gene expression at translational level. In agreement with this hypothesis, the abundance of tRNA modified nucleosides do change in response to various stresses [34]. Stress-induced degradation of tRNA has been reported by several investigators [80,81].

We speculate that those transcripts that are more efficiently expressed due to the presence of CUB during the course of stress, can be more efficiently used by translational machinery to express proteins due to the presence of specific modifications. However, this require more direct evidence to draw a final conclusion.

Methylation sensitive degradation of tRNA as another potential underlying mechanism for these observations has also the merit to look at. Wang et al. [82] investigated the mechanism of changes in the composition and abundance of tRNA-modified nucleosides in response to drought, salt and cold stress, as well as in different tissues during the whole growth season in two model plants—*O. sativa* and *Arabidopsis thaliana*. They identified 22 and 20 candidate genes for methyl-transferases in rice and Arabidopsis, respectively. Based on bioinformatics analysis, nucleoside abundance assessments and gene expression profiling, they found four methylated nucleosides (Am, Cm, m¹A and m⁷G) that are critical for stress response in rice and Arabidopsis.

However, the relationship of these modifications with CUB is not yet known. On the other hand, several articles have discussed tRNA modification of the animal species and its association with codon usage. Chan et al [74] showed a model for translational control mechanism for survival under oxidative stress in yeast. Exposure to H₂O₂ leads to an elevation in the level of m⁵C at the wobble position of the leucine tRNA for translating the codon UUG on mRNA. This modification in tRNA enhanced the translation of the UUG-enriched RPL22A mRNA relative to its paralog RPL22B and leads to changes in ribosome composition in their study. This reprogramming of tRNA and ribosomes ultimately causes selective translation of proteins from genes enriched with the codon TTG. More comprehensive studies are needed to better correlate the expression of genes with specific codon preference under drought stress in rice with tRNA modifications.

The fact that in most cases, no tRF has been reported for the preferred codon under drought stress, may support our hypothesis that during the course of evolution plants with preferred codons for which the tRNA does not break down into tRF gained selective advantage under drought condition. This may be the result of preservation of already existing such preferred codon or even occurrence of the mutation creating the preferred codon. This may be considered for further studies with the objective of identifying the underlying biological mechanism of CUB observed for high expression genes under drought. Based on these observations we propose an illustrated model based on codon characteristics, tRNA modification and tRFs under drought stress in rice (Fig 5).

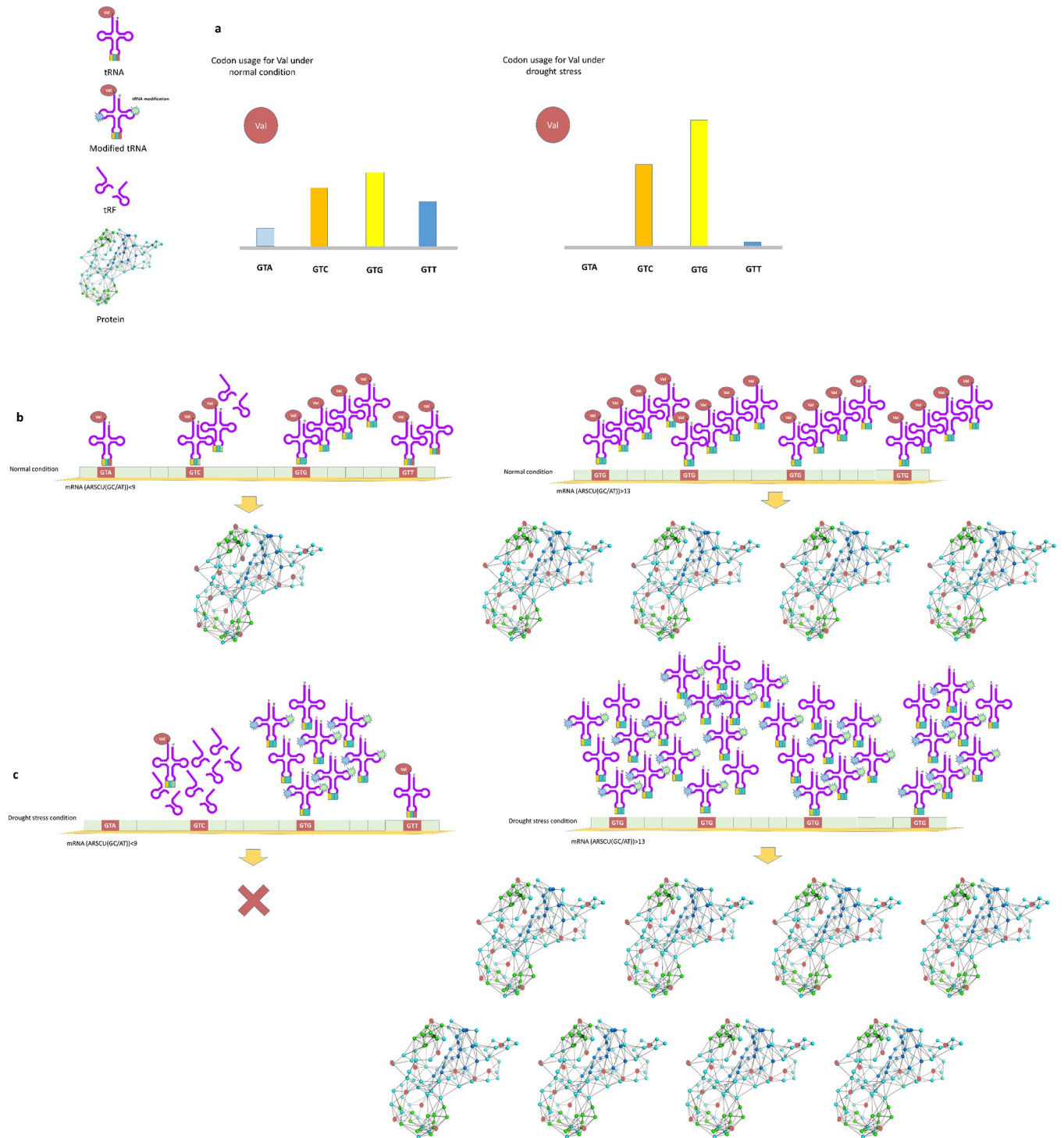


Fig 5. Proposed model justifying codon optimization in genetic engineering programs based of the observed CUB, codon characteristics, tRNA modification and tRFs under drought stress in rice. a) The codon usage difference under normal and drought condition for amino acid Valine. b) A segment of mRNA is shown under normal condition, before (left) and after codon optimization (right) symbolizing a gene. Under normal conditions, both genes (before and after codon optimization) are expressed. However, the codon-optimized gene is more expressed [83]. c) Native genes are minimally expressed or no expressed under stress (left) compared to the optimized genes (right). Under stress conditions (C), tRNA modifying enzymes mark a series of tRNAs that carry preferred codons leading to increased translation of codon optimized gene (right). In the case of amino acid Valine, the tRNA associated with GTC codon is broken under stress in the form of tRFs. It makes limited access to the codon for translation.

<https://doi.org/10.1371/journal.pone.0237334.g005>

In summary, the results of this paper suggests that codon optimization may play more significant role for improved gene expression under drought. Therefore, it is suggested to be considered in genetic engineering programs. It is expected that drought tolerant rice will be available from our ongoing projects on engineering rice for enhanced drought tolerance using several different strategies. The genes used in our projects have been optimized based on the findings reported in the current study.

Supporting information

S1 Table. ARSCU- calculator.

(XLSX)

S2 Table. Validation of ARSCU-calculator using 60 rice drought stress related genes.

(XLSX)

S3 Table. The ARSCU value of 60 high expression genes in different stages of rice growth.

(XLSX)

S4 Table. Rice codon usage table compare with rice high expression genes under drought stress.

(XLSX)

Author Contributions

Conceptualization: Behzad Ghareyazie, Motahhreh Mohsenpour.

Data curation: Fatemeh Chamani Mohasses.

Formal analysis: Fatemeh Chamani Mohasses, Motahhreh Mohsenpour.

Investigation: Fatemeh Chamani Mohasses.

Methodology: Fatemeh Chamani Mohasses, Motahhreh Mohsenpour.

Project administration: Behzad Ghareyazie, Motahhreh Mohsenpour.

Software: Fatemeh Chamani Mohasses.

Supervision: Behzad Ghareyazie, Motahhreh Mohsenpour.

Validation: Fatemeh Chamani Mohasses, Behzad Ghareyazie, Motahhreh Mohsenpour.

Visualization: Fatemeh Chamani Mohasses, Motahhreh Mohsenpour.

Writing – original draft: Fatemeh Chamani Mohasses.

Writing – review & editing: Mahmood Solouki, Behzad Ghareyazie, Leila Fahmideh, Motahhreh Mohsenpour.

References

1. Acquaah G. Principles of plant genetics and breeding. John Wiley & Sons; 2009.
2. Mollasadeghi V, Valizadeh M, Shahryari R, Imani AA. Evaluation of end drought tolerance of 12 wheat genotypes by stress indices. *World Appl Sci J. IDOSI Publications*; 2011; 13: 545–551.
3. Nachimuthu V V, Sabariappan R, Muthurajan R, Kumar A. Breeding rice varieties for abiotic stress tolerance: Challenges and opportunities. *Abiotic stress management for resilient agriculture*. Springer; 2017. pp. 339–361.
4. Yamaguchi T, Blumwald E. Developing salt-tolerant crop plants: challenges and opportunities. *Trends Plant Sci. Elsevier*; 2005; 10: 615–620. <https://doi.org/10.1016/j.tplants.2005.10.002> PMID: 16280254

5. Ahmad P, Ashraf M, Younis M, Hu X, Kumar A, Akram NA, et al. Role of transgenic plants in agriculture and biopharming. *Biotechnol Adv. Elsevier*; 2012; 30: 524–540.
6. Bennett J, Cohen MB, Katiyar SK, Ghareyazie B, Khush GS. Enhancing insect resistance in rice through biotechnology. *Adv insect Control role transgenic plants. Taylor and Francis, London*; 1997; 75–93.
7. Ghareyazie B, Alinia F, Menguito CA, Rubia LG, de Palma JM, Liwanag EA, et al. Enhanced resistance to two stem borers in an aromatic rice containing a synthetic *cryIA (b)* gene. *Mol Breed. Springer*; 1997; 3: 401–414.
8. Ye R, Huang H, Yang Z, Chen T, Liu L, Li X, et al. Development of insect-resistant transgenic rice with Cry1C*-free endosperm. *Pest Manag Sci Former Pestic Sci. Wiley Online Library*; 2009; 65: 1015–1020.
9. Goto S, Sasakura-Shimoda F, Suetsugu M, Selvaraj MG, Hayashi N, Yamazaki M, et al. Development of disease-resistant rice by optimized expression of WRKY45. *Plant Biotechnol J. Wiley Online Library*; 2015; 13: 753–765. <https://doi.org/10.1111/pbi.12303> PMID: 25487714
10. Fartyal D, Agarwal A, James D, Borphukan B, Ram B, Sheri V, et al. Developing dual herbicide tolerant transgenic rice plants for sustainable weed management. *Sci Rep. Nature Publishing Group*; 2018; 8: 1–12.
11. Eggers B, Mackenzie R. The Cartagena protocol on biosafety. *J Int Econ Law. Oxford University Press*; 2000; 3: 525–543.
12. James C. Global Status of Commercialized Biotech/GM Crops in 2017: Biotech crop adoption surges as economic benefits accumulate in 22 years. ISAAA Br No 53. 2017; ISAAA.
13. Wani SH, Sah SK. Biotechnology and abiotic stress tolerance in rice. *J Rice Res. 2014*; 2: e105.
14. De La Torre AR, Lin Y-C, Van de Peer Y, Ingvarsson PK. Genome-wide analysis reveals diverged patterns of codon bias, gene expression, and rates of sequence evolution in picea gene families. *Genome Biol Evol. Oxford University Press*; 2015; 7: 1002–1015. <https://doi.org/10.1093/gbe/evv044> PMID: 25747252
15. Bhatnagar-Mathur P, Vadez V, Sharma KK. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. *Plant Cell Rep. Springer*; 2008; 27: 411–424. <https://doi.org/10.1007/s00299-007-0474-9> PMID: 18026957
16. Ahanger MA, Akram NA, Ashraf M, Alyemeni MN, Wijaya L, Ahmad P. Plant responses to environmental stresses—from gene to biotechnology. *AoB Plants. Oxford University Press*; 2017; 9.
17. Gout J-F, Kahn D, Duret L, Consortium PP-G. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. *PLoS Genet. Public Library of Science*; 2010; 6.
18. Williford A, Demuth JP. Gene expression levels are correlated with synonymous codon usage, amino acid composition, and gene architecture in the red flour beetle, *Tribolium castaneum*. *Mol Biol Evol. Oxford University Press*; 2012; 29: 3755–3766. <https://doi.org/10.1093/molbev/mss184> PMID: 22826459
19. Hershberg R, Petrov DA. Selection on codon bias. *Annu Rev Genet. Annual Reviews*; 2008; 42: 287–299. <https://doi.org/10.1146/annurev.genet.42.110807.091442> PMID: 18983258
20. Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci. National Acad Sciences*; 1999; 96: 4482–4487. <https://doi.org/10.1073/pnas.96.8.4482> PMID: 10200288
21. Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast. Wiley Online Library*; 2000; 16: 1131–1145. [https://doi.org/10.1002/1097-0061\(20000915\)16:12<1131::AID-YEA609>3.0.CO;2-F](https://doi.org/10.1002/1097-0061(20000915)16:12<1131::AID-YEA609>3.0.CO;2-F) PMID: 10953085
22. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. *Nature. Nature Publishing Group*; 2003; 425: 737–741. <https://doi.org/10.1038/nature02046> PMID: 14562106
23. Urrutia AO, Hurst LD. The signature of selection mediated by expression on human genes. *Genome Res. Cold Spring Harbor Lab*; 2003; 13: 2260–2264. <https://doi.org/10.1101/gr.641103> PMID: 12975314
24. Comeron JM. Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics. Genetics Soc America*; 2004; 167: 1293–1304. <https://doi.org/10.1534/genetics.104.026351> PMID: 15280243
25. Cutter AD, Wasmuth JD, Blaxter ML. The evolution of biased codon and amino acid usage in nematode genomes. *Mol Biol Evol. Oxford University Press*; 2006; 23: 2303–2315. <https://doi.org/10.1093/molbev/msl097> PMID: 16936139

26. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. Elsevier; 2008; 134: 341–352. <https://doi.org/10.1016/j.cell.2008.05.042> PMID: 18662548
27. Ingvarsson PK. Molecular evolution of synonymous codon usage in *Populus*. *BMC Evol Biol*. Springer; 2008; 8: 307. <https://doi.org/10.1186/1471-2148-8-307> PMID: 18983655
28. Qin F, Shinozaki K, Yamaguchi-Shinozaki K. Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol*. Oxford University Press; 2011; 52: 1569–1582. <https://doi.org/10.1093/pcp/pcr106> PMID: 21828105
29. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*. 1985; 2: 13–34. <https://doi.org/10.1093/oxfordjournals.molbev.a040335> PMID: 3916708
30. Chiapello H, Lisacek F, Caboche M, Hénaut A. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene*. Elsevier; 1998; 209: GC1–GC38.
31. Goodarzi H, Torabi N, Najafabadi HS, Archetti M. Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. *Gene*. Elsevier; 2008; 407: 30–41. <https://doi.org/10.1016/j.gene.2007.09.020> PMID: 17977670
32. Whittle CA, Extavour CG. Codon and amino acid usage are shaped by selection across divergent model organisms of the Pancrustacea. *G3 Genes, Genomes, Genet*. G3: Genes, Genomes, Genetics; 2015; 5: 2307–2321.
33. Yi S, Li Y, Wang W. Selection shapes the patterns of codon usage in three closely related species of genus *Misgurnus*. *Genomics*. Elsevier; 2018; 110: 134–142. <https://doi.org/10.1016/j.ygeno.2017.09.004> PMID: 28911975
34. Chan CTY, Pang YLJ, Deng W, Babu IR, Dyavaiah M, Begley TJ, et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. *Nat Commun*. Nature Publishing Group; 2012; 3: 1–9.
35. Begley U, Dyavaiah M, Patil A, Rooney JP, DiRenzo D, Young CM, et al. Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol Cell*. Elsevier; 2007; 28: 860–870. <https://doi.org/10.1016/j.molcel.2007.09.021> PMID: 18082610
36. Hatfield DL, Gladyshev VN. How selenium has altered our understanding of the genetic code. *Mol Cell Biol*. *Am Soc Microbiol*; 2002; 22: 3565–3576. <https://doi.org/10.1128/mcb.22.11.3565-3576.2002> PMID: 11997494
37. Dedon PC, Begley TJ. A system of RNA modifications and biased codon use controls cellular stress response at the level of translation. *Chem Res Toxicol*. ACS Publications; 2014; 27: 330–337. <https://doi.org/10.1021/tx400438d> PMID: 24422464
38. Thompson DM, Parker R. Stressing out over tRNA cleavage. *Cell*. Elsevier; 2009; 138: 215–219. <https://doi.org/10.1016/j.cell.2009.07.001> PMID: 19632169
39. Kumar P, Kuscu C, Dutta A. Biogenesis and function of transfer RNA-related fragments (tRFs). *Trends Biochem Sci*. Elsevier; 2016; 41: 679–689. <https://doi.org/10.1016/j.tibs.2016.05.004> PMID: 27263052
40. Schaefer M, Pollex T, Hanna K, Tuorto F, Meusburger M, Helm M, et al. RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev*. Cold Spring Harbor Lab; 2010; 24: 1590–1595. <https://doi.org/10.1101/gad.586710> PMID: 20679393
41. Blanco S, Dietmann S, Flores J V, Hussain S, Kutter C, Humphreys P, et al. Aberrant methylation of tRNA s links cellular stress to neuro-developmental disorders. *EMBO J*. 2014; 33: 2020–2039. <https://doi.org/10.15252/embj.201489282> PMID: 25063673
42. Park EJ, Kim T-H. Fine-tuning of gene expression by tRNA-derived fragments during abiotic stress signal transduction. *Int J Mol Sci*. Multidisciplinary Digital Publishing Institute; 2018; 19: 518.
43. Mazumdar P, Binti Othman R, Mebus K, Ramakrishnan N, Ann Hari Krishna J. Codon usage and codon pair patterns in non-grass monocot genomes. *Ann Bot*. Oxford University Press US; 2017; 120: 893–909. <https://doi.org/10.1093/aob/mcx112> PMID: 29155926
44. de Abreu Neto JB, Frei M. Microarray meta-analysis focused on the response of genes involved in redox homeostasis to diverse abiotic stresses in rice. *Front Plant Sci*. Frontiers; 2016; 6: 1260. <https://doi.org/10.3389/fpls.2015.01260> PMID: 26793229
45. Frei M, Wang Y, Ismail AM, Wissuwa M. Biochemical factors conferring shoot tolerance to oxidative stress in rice grown in low zinc soil. *Funct Plant Biol*. CSIRO; 2010; 37: 74–84.
46. Shaik R, Ramakrishna W. Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant Physiol*. *Am Soc Plant Biol*; 2014; 164: 481–495.
47. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res*. Oxford University Press; 2000; 28: 10–14. <https://doi.org/10.1093/nar/28.1.10> PMID: 10592169

48. Duan J, Cai W. *OsLEA3-2*, an abiotic stress induced gene of rice plays a key role in salt and drought tolerance. PLoS One. Public Library of Science; 2012; 7.
49. Xiao B, Huang Y, Tang N, Xiong L. Over-expression of a *LEA* gene in rice improves drought resistance under the field conditions. Theor Appl Genet. Springer; 2007; 115: 35–46. <https://doi.org/10.1007/s00122-007-0538-9> PMID: 17426956
50. Hu H, You J, Fang Y, Zhu X, Qi Z, Xiong L. Characterization of transcription factor gene *SNAC2* conferring cold and salt tolerance in rice. Plant Mol Biol. Springer; 2008; 67: 169–181. <https://doi.org/10.1007/s11103-008-9309-5> PMID: 18273684
51. Redillas MCFR, Jeong JS, Kim YS, Jung H, Bang SW, Choi YD, et al. The overexpression of *OsNAC9* alters the root architecture of rice plants enhancing drought resistance and grain yield under field conditions. Plant Biotechnol J. Wiley Online Library; 2012; 10: 792–805. <https://doi.org/10.1111/j.1467-7652.2012.00697.x> PMID: 22551450
52. Huang J, Sun S, Xu D, Lan H, Sun H, Wang Z, et al. A TFIIIA-type zinc finger protein confers multiple abiotic stress tolerances in transgenic rice (*Oryza sativa* L.). Plant Mol Biol. Springer; 2012; 80: 337–350. <https://doi.org/10.1007/s11103-012-9955-5> PMID: 22930448
53. Jan A, Maruyama K, Todaka D, Kidokoro S, Abo M, Yoshimura E, et al. OsTZF1, a CCH-tandem zinc finger protein, confers delayed senescence and stress tolerance in rice by regulating stress-related genes. Plant Physiol. Am Soc Plant Biol; 2013; 161: 1202–1216.
54. Huang L, Wang Y, Wang W, Zhao X, Qin Q, Sun F, et al. Characterization of transcription factor gene *OsDRAP1* conferring drought tolerance in rice. Front Plant Sci. Frontiers; 2018; 9: 94. <https://doi.org/10.3389/fpls.2018.00094> PMID: 29449862
55. Huang X-Y, Chao D-Y, Gao J-P, Zhu M-Z, Shi M, Lin H-X. A previously unknown zinc finger protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. Genes Dev. Cold Spring Harbor Lab; 2009; 23: 1805–1817. <https://doi.org/10.1101/gad.1812409> PMID: 19651988
56. Gao F, Xiong A, Peng R, Jin X, Xu J, Zhu B, et al. OsNAC52, a rice NAC transcription factor, potentially responds to ABA and confers drought tolerance in transgenic plants. Plant Cell, Tissue Organ Cult. Springer; 2010; 100: 255–262.
57. Xu D-Q, Huang J, Guo S-Q, Yang X, Bao Y-M, Tang H-J, et al. Overexpression of a TFIIIA-type zinc finger protein gene ZFP252 enhances drought and salt tolerance in rice (*Oryza sativa* L.). FEBS Lett. Wiley Online Library; 2008; 582: 1037–1043. <https://doi.org/10.1016/j.febslet.2008.02.052> PMID: 18325341
58. Puigbò P, Bravo IG, Garcia-Vallve S. CAIcal: a combined set of tools to assess codon usage adaptation. Biol Direct. BioMed Central; 2008; 3: 38.
59. Hastings KEM, Emerson CP. Codon usage in muscle genes and liver genes. J Mol Evol. Springer; 1983; 19: 214–218. <https://doi.org/10.1007/BF02099968> PMID: 6887263
60. Sharp PM, Li W-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. Oxford University Press; 1987; 15: 1281–1295. <https://doi.org/10.1093/nar/15.3.1281> PMID: 3547335
61. Vetrivel U, Arunkumar V, Dorairaj S. ACUA: a software tool for automated codon usage analysis. Bioinformatics. Citeseer; 2007; 2: 62.
62. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. Oxford University Press; 2000; 28: 292. <https://doi.org/10.1093/nar/28.1.292> PMID: 10592250
63. Gupta N, Singh A, Zahra S, Kumar S. PtRFdb: a database for plant transfer RNA-derived fragments. Database. Narnia; 2018; 2018.
64. Jung K-H, Kim S-R, Giong H-K, Nguyen MX, Koh H-J, An G. Genome-wide identification and functional analysis of genes expressed ubiquitously in rice. Mol Plant. Elsevier; 2015; 8: 276–289. <https://doi.org/10.1016/j.molp.2014.10.013> PMID: 25624149
65. Xu H, Gao Y, Wang J. Transcriptomic analysis of rice (*Oryza sativa*) developing embryos using the RNA-Seq technique. PLoS One. Public Library of Science; 2012; 7.
66. Verulkar SB, Verma SK. Screening protocols in breeding for drought tolerance in rice. Agric Res. Springer; 2014; 3: 32–40.
67. Bashir K, Matsui A, Rasheed S, Seki M. Recent advances in the characterization of plant transcriptomes in response to drought, salinity, heat, and cold stress. F1000Research. Faculty of 1000 Ltd; 2019; 8. <https://doi.org/10.12688/f1000research.17047.1>
68. Guo X, Bao J, Fan L. Evidence of selectively driven codon usage in rice: implications for GC content evolution of Gramineae genes. FEBS Lett. Elsevier; 2007; 581: 1015–1021. <https://doi.org/10.1016/j.febslet.2007.01.088> PMID: 17306258

69. Wang H-C, Hickey DA. Rapid divergence of codon usage patterns within the rice genome. *BMC Evol Biol*. BioMed Central; 2007; 7: S6.
70. Liu Q, Dou S, Ji Z, Xue Q. Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems*. Elsevier; 2005; 80: 123–131. <https://doi.org/10.1016/j.biosystems.2004.10.008> PMID: 15823411
71. Chionh YH, McBee M, Babu IR, Hia F, Lin W, Zhao W, et al. tRNA-mediated codon-biased translation in mycobacterial hypoxic persistence. *Nat Commun*. Nature Publishing Group; 2016; 7: 1–12.
72. Whittle CA, Extavour CG. Expression-linked patterns of codon usage, amino acid frequency, and protein length in the basally branching arthropod *Parasteatoda tepidariorum*. *Genome Biol Evol*. Oxford University Press; 2016; 8: 2722–2736. <https://doi.org/10.1093/gbe/evw068> PMID: 27017527
73. Wu Y, Zhao D, Tao J. Analysis of codon usage patterns in herbaceous peony (*Paeonia lactiflora* Pall.) based on transcriptome data. *Genes (Basel)*. Multidisciplinary Digital Publishing Institute; 2015; 6: 1125–1139. <https://doi.org/10.3390/genes6041125> PMID: 26506393
74. Ingvarsson PK. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*. *Mol Biol Evol*. Oxford University Press; 2007; 24: 836–844. <https://doi.org/10.1093/molbev/msl212> PMID: 17204548
75. Endres L, Dedon PC, Begley TJ. Codon-biased translation can be regulated by wobble-base tRNA modification systems during cellular stress responses. *RNA Biol*. Taylor & Francis; 2015; 12: 603–614. <https://doi.org/10.1080/15476286.2015.1031947> PMID: 25892531
76. Björk GR, Durand JMB, Hagervall TG, Leipunien R, Lundgren HK, Nilsson K, et al. Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett*. Wiley Online Library; 1999; 452: 47–51. [https://doi.org/10.1016/s0014-5793\(99\)00528-1](https://doi.org/10.1016/s0014-5793(99)00528-1) PMID: 10376676
77. El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet*. Annual Reviews; 2012; 46: 69–95. <https://doi.org/10.1146/annurev-genet-110711-155641> PMID: 22905870
78. Björk GR, Ericson JU, Gustafsson CED, Hagervall TG, Jönsson YH, Wikström PM. Transfer RNA modification. *Annu Rev Biochem*. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303–0139, USA; 1987; 56: 263–285. <https://doi.org/10.1146/annurev.bi.56.070187.001403> PMID: 3304135
79. Gustilo EM, Vendeix FAP, Agris PF. tRNA's modifications bring order to gene expression. *Curr Opin Microbiol*. Elsevier; 2008; 11: 134–140. <https://doi.org/10.1016/j.mib.2008.02.003> PMID: 18378185
80. Motorin Y, Helm M. tRNA stabilization by modified nucleotides. *Biochemistry*. ACS Publications; 2010; 49: 4934–4944. <https://doi.org/10.1021/bi100408z> PMID: 20459084
81. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, et al. Rapid tRNA decay can result from lack of nonessential modifications. *Mol Cell*. Elsevier; 2006; 21: 87–96.
82. Wang Y, Pang C, Li X, Hu Z, Lv Z, Zheng B, et al. Identification of tRNA nucleoside modification genes critical for stress response and development in rice and Arabidopsis. *BMC Plant Biol*. BioMed Central; 2017; 17: 261. <https://doi.org/10.1186/s12870-017-1206-0> PMID: 29268705
83. Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol cell Biol*. Nature Publishing Group; 2018; 19: 20–30. <https://doi.org/10.1038/nrm.2017.91> PMID: 29018283