Original article

Prediction of gene expression under drought stress in spring wheat using codon usage pattern

Meshal M. Almutairi *, Abdullah A. Alrajhi

National Center of Agricultural Technology, King Abdulaziz City for Science and Technology KACST, Box 6086, Riyadh 11442, Saudi Arabia

ABSTRACT

Spring wheat (*Triticum aestivum*) is a staple food providing sources of essential proteins for human. In fact, gene expressions of wheat play an important role in growth and productivity that are affected by drought stress. The objective of this work focused on analysis gene feature on spring wheat represented by nucleotide and gene expressions under drought stress. It was found that the higher codon adaptation index was in both wheat root and L-galactono-1, 4-lactone dehydrogenase. It was also found that guanine and cytosine content were high (55.56%) in wheat root. Whereas, guanine and cytosine content were low (41.28%) in L-galactono-1, 4-lactone dehydrogenase. Moreover, the higher relative synonymous codon usage value was observed in codon CAA (1.20), GAA (1.33), GAT (1.00), and ATG (1.00) in wheat root and thus about 62.95% of the total variation in relative synonymous codon was explained by principal component analysis. Additionally, high averages frequency number of codon were (above 15.76) in Met, Lys, Ala, Gly, Phe, Asp, Glu, His, and Tyr; whereas, low averages were in remaining amino acids and majority (90%) of modified relative codon bias values was between 0.40 and 0.90. Shortly, calculations and analysis of codon usage pattern under drought stress would help for genetic engineering, molecular evolution, and gene prediction in wheat studies for developing varieties that associate with drought tolerance.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Spring wheat that belongs to common wheat (*Triticum aestivum* L., $2n = 6x = 42$ = AABBDD genomes) is a staple food providing sources of proteins and dietary energy for human nutrition and is the main target for genome sequences. The AABBDD genomes is hexaploid and contains 21 pairs of chromosomes that are grouped into seven homologous and orthologous genes response to many characters of common wheat (Sleper and Poehlman, 2006). In addition, fusion protein can be developed by homologous recombination through DNA repair mechanism for introducing modified genes. However, wheat has a larger gene pool showing difficult to improve its characters. Genetic transfers from A and D genomes occur through direct hybridization and homologous recombination

during breeding program following by selection progress (Sleper and Poehlman, 2006). Spring wheat has shown to acclimate under drought stress through different mechanisms such as early maturity, deep root system, and ability to close stomata during period of drought stress. Some of important genes under drought stress have selected for purpose of this study. In fact, it is important to explain the functions of these selected genes in this study as following. MYB3R gene known as the cellular morphogenesis, secondary metabolism and reflecting important role in controlling cell process (Matsui et al., 2008). The novel allele of GLDH gene is essential for the ascorbic acid AsA and accumulation of respiratory complex in plant (Zhang et al., 2016). TaDREB gene is extraction from shoot and root during 10 days of growing seedlings under drought stress. Transcriptional and post-transcription factor such as nuclear factor (NFYA5) gene is importance for dehydration stresses that confer drought tolerance in oxidative stress (Jones-Rhoades and Bartel, 2004). Moreover, wheat under drought stress has different mechanisms such as osmotic adjustment (proline, betaine, and glycine) and protective metabolites. These mechanisms play important roles for maintaining normal photosynthesis under drought stress (Ma et al., 2016).

* Corresponding author.
E-mail address: malmetairy@kacst.edu.sa (M.M. Almutairi).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

Amino acids are important assessments of potential growth because of building the protein structures and functions (Tian et al., 2020). Thus, amino acids are manifest the cornerstone template for genetic information. Moreover, most amino acids are encoded by more than one codon that could be revealed by a synonymous codon (Hershberg, 2016). The triple codon (nucleotide) is the basic unit of mRNA encoding that reflects vital roles during a generation of organism. This basic unit encodes amino acids that reflect important genetic and evolution system. Furthermore, codon adaptation index (CAI) is one of the most algorithm to demonstrate a difference in the frequency of the synonymous codon for the mRNA sequence. This CAI helps to understand the gene expression, new genes discovery, and molecular mechanism for genes evolution contribution (Hershberg, 2016). Thus, the theoretical basis for genetic studied such as genetic engineering, gene prediction, and molecular studies in wheat is critical concept by analysis the codon usage pattern (Yu et al., 2015). This CAI has influence by several factors such as nucleotide composition, tRNA abundance, and structure of protein, translation processes, and more (Carbone et al., 2003; Lithwick and Margalit, 2005). At the same amino acid, genetic code associates with sibling codons that may occur more frequently resulting from several factors such as GC content (Grantham et al., 1980; Wada et al., 1990).

The objective of this paper was to analyze gene feature in spring wheat under drought stress represented by nucleotide, used codon adaptation index, relative synonymous codon usage, synonymous codon and amino acid usage, modified relative codon bias, and Principal component analysis.

2. Material and methods

2.1. Retrieval of sequences:

Five mRNA sequences of Chinese Spring wheat (*Triticum aestivum*) were retrieved from the National Center for Biotechnology NCBI (www.ncbi.nlm.nih.gov) Gen-Bank and the accession number are BU672415, ES466900, HV510735, KU695148, and HQ236494 (Table 1). All these sequences were studied under drought stress at different levels during root wheat, MYB gene, enzyme named L-galactono-1,4-lactone dehydrogenase located in mitochondrion, and transcription factor for drought resistance.

2.2. Measures of codon usage patterns

A variety of analyses tool were used to measure the codon usage patterns such as codon adaptation index (CAI) given by sharp and Li. (1987), relative synonymous codon usage (RSCU) given by Yu, et al. (2015). GC content, relative codon bias strength (RCBS), and modified relative codon bias (MRCBS) applied in this study (Shoo et al., 2019). Some of codons were excluded form calculation such as AUG (methionine), UGG (tryptophan), and three stop codons such as UAA, UAG, and UGA. Therefore, these codons for all mRNA sequences were normalized (Zhou et al., 2005).

2.3. Codon adaptation index

CAI is represented by the formula: CAI = $\left(\prod_1^N w_i\right)\frac{1}{n}$. Where N is the number of codons in the gene and relativeness w_i is described as $w_i = \frac{f_i}{f_{aa,max}}$ where f_i is the frequency of the ith codon and $f_{aa,max}$ is maximum frequency of the codons most often used for encoding amino acid aa.

2.4. Relative synonymous codon usage

The RSCU can be calculated by way of following formula: RSCU = $\frac{x_{ij}}{\sum_j^{ni} x_{ij}}$ where x_{ij} shows the frequency of codon j coding for the ith amino acid and n_i represented the number of synonymous codon encoding the ith amino acids. The interpretation for the RSCU is that RSCU greater than one indicates more frequency while RSCU small than 1 indicates less frequency in a gene for the particular codon. Whereas RSCU > 1.6 indicates over represented codon for the corresponding amino acid (Wong et al., 2010).

2.5. Modified relative codon bias

MRCSB measures the expression level of a gene (Shoo et al., 2019). MRCSBs were computed through this formula: MRCBS = $\prod_{i=1}^N (MRCBS_{xyz})$ where $MRCBS_{xyz} = \frac{RCBS_{(xyz)}}{RCBS_{aa,max}}$ where $RCBS_{(xyz)} = \frac{f_{xyz}}{f1(x)f2(y)f3(z)}$ whereas f_{xyz} is the normalized codon frequency of a codon *xyz* and $f_n(m)$ is the normalized frequency of base m at codon position in a gene. The $RCBS_{aa,max}$ reflects the maximum value of RCBS of codon encoding the same amino acid aa. The score of MRCBS ranges from zero to one.

2.6. Statistical analysis

Calculations were done through Microsoft Excel and codonW software (Peden, 2000). Moreover, two ways analyses of variances were computed using PROC GLM procedure at significate level of *p*-value (0.001) in SAS computer packages version 9.2 for windows (SAS Institute Inc.). Analysis of principal component helps to reset data into a few data set and was computed using PROC PRINCOMP in SAS in order to analysis total variations.

3. Results

3.1. The CAI and GC content

At early stages of development in spring wheat, randomly selected transcriptional factors are known to various functions. The overall result of the codon adaptation index was 0.722 and the highest GC content was 50.03 in the wheat root while the lowest GC content was 41.28 in the GLDH gene (Table 1). The ANOVA analysis reflecting significant differences for the variables under study such as genes, amino acids, and codons with p-value < 0.0003, <0.0001, and <0.0001 respectively (Table 2).

Table 1
Codon adaptation index (CAI) and GC Content under drought stress for root and genes.

Accession	Root and genes	Length bp	CAI	GC%	Authors
BU672415	Wheat root	2370	0.74	55.56	Anderson et al. (2011)
ES466900	TaDREB gene	645	0.73	51.47	Sadeghi et al. (2007)
KU695148	GLDH gene	5526	0.74	41.28	Zhang et al. (2016)
HQ236494	MYB3R gene	2370	0.73	46.29	Cai et al. (2011)
HV510735	NFYA5 gene	729	0.67	55.56	Li and Zhu. (2011)

Table 2
The ANOVA analysis for spring wheat at different growth level, amino acids, and codons.

Source	DF	SS	MS	F-Value	Pr > F
Root and genes	4	1.43	0.36	5.53	0.0003
Amino acids	19	5.97	0.31	4.62	<0.0001
Codons	63	14.40	0.23	3.54	<0.0001

Codon adaptation index (CAI) helps to measure the synonymous codon usage bias toward codons that highly expressed genes. The high CAI shows in both the wheat root and GLDH gene. We found that GC content in the wheat root, TaDREB gene, and NFYA5 gene had the higher CAI value excluded the NFYA5 gene. The relative adaptiveness of a codon in the CAI calculation is computed as the ratio between the frequency in the reference and the largest frequency among synonymous. The Length of the mRNA sequence did not influence the GC contact and CAI.

The overall frequency of codon used was demonstrated in Fig. 1. In this figure, the black line represents the average (15.76) across all amino acid. This average divides amino acids into two groups. The First group was above average that were lysine, alanine, glycine, phenylalanine, aspartic acid, asparagine, glutamate, histidine, and tyrosine. The second group was below the average that was serine, proline, isoleucine, valine, leucine, threonine, cysteine, and arginine. This result showed the importance of amino acid metabolism under drought stress that is robust gene expression. Moreover, amino acids are more disturbed by drought stress during the transcriptional and metabolic processes.

3.2. The RSCU and MRCBS

More computational tools such as RSCU and MRCBS were used to explain more about the codon usage patterns and determine gene expression as presenting in Fig. 2. RSCU stands to the ratio between the frequency of synonymous and the expected frequency of codon. The majority of genes about 90% that had MRCBS values between 0.40 and 0.90 with mean, median, and standard deviation values were 0.623, 0.627, and 0.339 respectively. RSCU greater than or equal to 1 implies high frequency whereas RSCU small than 1 means less frequency usage. When RSCU equal to 0 implies no bias codon usage existing. The higher RSCU value was observed from codon CAA (1.20), GAA (1.33), GAT (1.00), and ATG (1.00) in
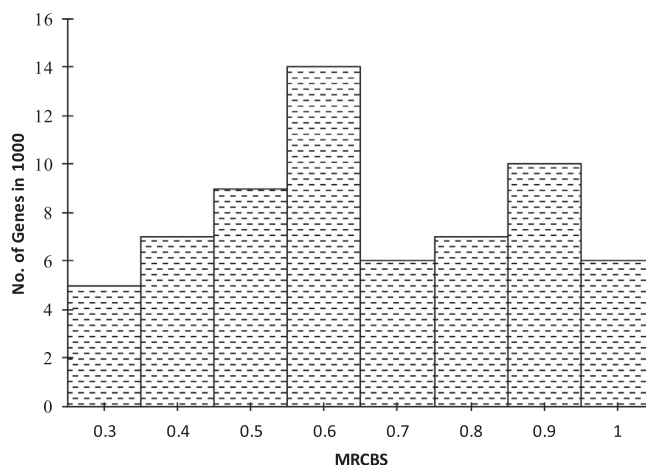

Fig. 2. Frequency of MRCBS for all codons in all traits for Wheat.

the wheat root, MYB3R, and TaDREB gene respectively. Besides, Pearson correlation coefficients analysis showed that the high correlation is between MRCBS and RSCU 0.64. While the second correlation is between MRCBS and RCBS 0.41 with a p-value < 0.0001.

3.3. The PCA

The principal component analysis (PCA) was able to distinguish the genetic variance based on RSCU, RCBS, and MRCBS (see Fig. 3). PCA illustrated the most important factor that explained 62.95% of the total variation. The most effective method for codon usage bias in spring wheat genes in the first component was RSCU with eigenvalue 1.88 and proportion 0.62 while the second component was MRCBS with eigenvalue 0.77 and proportion 0.25.
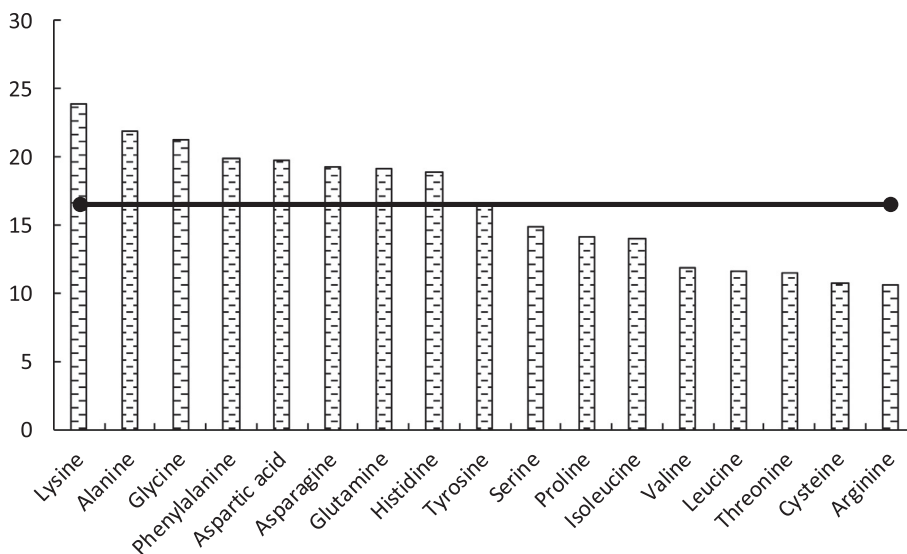

Fig. 1. Overall frequency of number of codon used obtained from genes in this study under drought stress.
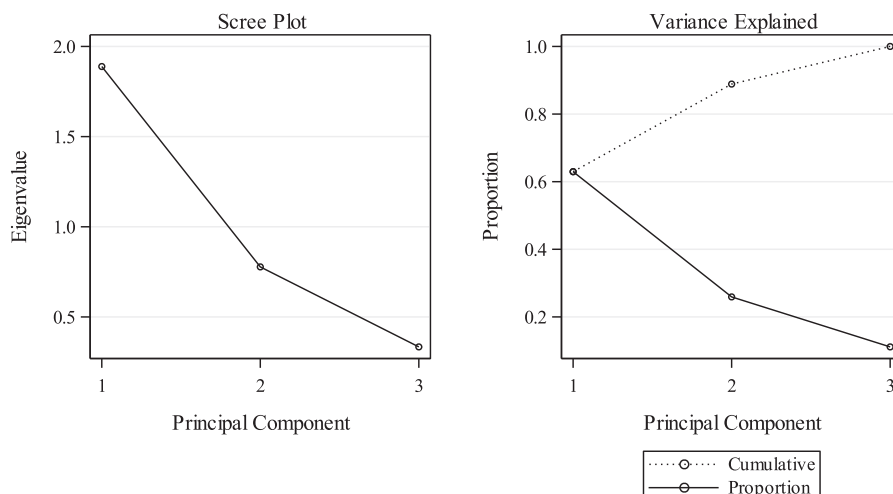
Fig. 3. The principal component analysis (PCA) using relative synonymous codon usage (RSCU), relative codon bias strength (RCBS), and modified relative codon bias (MRCBS).

4. Discussion

Spring wheat (*T. aestivum*) is a staple food providing sources of essential proteins but its production is drastically impacted by the drought stress. Thus, development cultivars that drought tolerance is a major objective for wheat breeding programs. Analyzed codon patterns are significant aspect because of connotation with the genome evolution concerning nucleotide and gene expression. These nucleotide and gene expression could benefit the translation efficiency of gene population (Das et al., 2006). This analysis is also importance to increase our understanding of wheat response and its mechanism under drought stress (Sanders et al., 2009). To enhance drought tolerance, it is importance to analysis codon to ensure high expression levels with removing repeated and unnecessary digestion sites during synthesized. Furthermore, The GC contents were used for estimated codon usage in a gene and these contents help breeders to predict the conserved amino acid patterns (Araus et al., 2008). However, this GC index is more sensitive for both rare amino acids and short CDS and thus it is poorly informative that could not provide details for codon usage. Unless analysis small gene that is less 100 codons could be used in GC index to be more informative. Several studies have argued that GC contents may reflect the primary influence on codon usage patterns that change the expression profile. These results imply variation in GC contents that might be due to genetic composition and mating behavior (Mazumdar et al., 2017). It was found that Arginine and proline metabolism, alanine, aspartate, glutamate were significantly disturbed by drought stress (You et al., 2019). The content of proline could be consider as main key in plant consistency with drought stress. In Arabidopsis, gene named Atlpk2α encoding IPK2 has several functions in plant development such as root growth and pollen germination association with drought stress and involving in abscisic acid (ABA) (Wang et al., 2017; Yang et al., 2018). This gene, Atlpk2α, could be benefit for enhancing drought tolerance of wheat through strategy of genetic engineering by a synthetic gene (Shu-juan et al., 2020). Furthermore, it is a useful purpose of wheat breeding program to determine gene feature and its mechanisms associated with drought stress providing ideas for genetic manipulation (Salehe Naderia et al., 2020).

5. Conclusion

In this study, the relative synonymous codon usage, relative codon bias strength, and modified relative codon bias may be useful tools predicting amino acids frequency and gene expression during growth stages of spring wheat under drought stress. More study on codon usage pattern for spring wheat is necessarily to identification of expressed genes. Thus, this study was importance to demonstrate significant heterogeneity in codon usage among genes in spring wheat. The overall result of codon adaptation index was 0.722, which indicates highest genes expressed represented bias of these genes. Thus, result of codon adaption index can help for more confidence of predictions. In addition, the overall of GC content was 50.03 implies that all codons are used uniformly, which is valuable for gene expression in wheat. The idea of support our method and finding is based on largely responsible of codon usage pattern for regulation gene translations that occur during protein translation. In addition, the relative synonymous codon usage is the ratio of observed frequency of codon to expected frequency at all synonymous codons. This relative synonymous codon usage was normalized by removing methionine (ATG), tryptophan (TGG), and three stop codons. The higher RSCU value was observed in following codons: CAA, GAA, GAT, and ATG in wheat root, MYB3R, and TaDREB gene respectively indicating high frequency in these codons. Further investigation in spring wheat genes would enhance our understanding of metabolism. You et al., (2019) Found that these amino acids: arginine and proline metabolism, alanine, aspartate, and glutamate were significantly disturbed by drought stress. Lastly, PCA is used to illustrate the most important factor because PCA provides the predication accuracy by eliminating random error and redundancy. PCA explained 62.95% of the total variation and the most effect method for codon usage bias was RSCU with eigenvalue 1.88 and proportion 0.62 in spring wheat genes (Ghosh and Barman, 2016).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Anderson, O., Chao, S., Close, T., Crossman, C., Fenton, R., Lazo, G., Nguyen H., Pham, J., Rausch, C., Turuspekov, Y., Wilson, C., Woo, J. and Zhang, D. 2011. The structure and function of the expressed portion of the wheat genomes - Chinese Spring drought stressed root cDNA library. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/BU672415.1.

Araus, J., Slafer, G., Royo, C., Serret, M., 2008. Breeding for yield potential and stress adaptation in cereals. Crit. Rev. Plant Sci. 27, 377–441.

Cai, H., Tian, S., Liu, C., Dong, H., 2011. Identification of a MYB3R gene involved in drought, salt and cold stress in wheat (Triticum aestivum L.). J. Gene 485 (2), 146–152.

Carbone, A., Zinovyev, A., Képès, F., 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19, 2005–2015. https://doi.org/10.1093/bioinformatics/btg272.

Das, S., Paul, S., Dutta, C., 2006. Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res. 117 (2), 227–236.

Ghosh, A., Barman, B., 2016. Application of Euclidean distance measurement and principal component analysis for gene identification. Gene 583, 112–120.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., Pave, A., 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 8, r49–r62.

Hershberg, R., 2016. Codon usage and translational selection. Encyclopedia Evolutionary Biol. 293–298. https://doi.org/10.1016/B978-0-12-800049-6.00178-5.

Jones-Rhoades, M., Bartel, D., 2004. Computational identification of plant microRNAs and their targets, including a stress induced miRNA. Mol. Cell 14, 787–799.

Li, W., Zhu, J., 2011. Transcriptional and post-transcription regulation of transcription factor for drought resistance. Patent: JP 2011519562-A 61 14-JUL-2011. University of California. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/HV510735.1/.

Lithwick, G., Margalit, H., 2005. Relative predicted protein levels of functionally associated proteins are conserved across organisms. Nucleic Acids Res. 33, 1051–1057. https://doi.org/10.1093/nar/gki261.

Ma, X., Xia, H., Liu, Y., Wei, H., Zheng, X., Song, C., Chen, L., Liu, H., Luo, L., 2016. Transcriptomic and metabolomic studies disclose key metabolism pathways contributing to well-maintained photosynthesis under the drought and the consequent drought-tolerance in rice. Front. Plant Sci. 7, 1886. https://doi.org/10.3389/fpls.2016.01886.

Matsui, R., Umemura, Y., Ohme, Takagi, M. 2008. AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in Arabidopsis. Plant J. 55, 954–967.

Mazumdar, P., Binti Othman, R., Mebus, K., Ramakrishnan, N., Ann Harikrishna, J., 2017. Codon usage and codon pair patterns in non-grass monocot genomes. Ann. Bot. 120, 893–909.

Peden, J., 2000. Analysis of codon usage. 90 (1), dissertation from university of Nottingham, 73–74.

Sadeghi, S., Mohsenzadeh, S., Mohabatkar, H., 2007. Expression of TaDREB gene related to drought stress in wheat. Unpublished. Retrieved from: https://www.ncbi.nlm.nih.gov/nuccore/ES466900.

Salehe Naderia, S., Fakheria, B., Maali-Amirib, R., Mahdinezhada, N., 2020. Tolerance responses in wheat landrace Bolani are related to enhanced metabolic adjustments under drought stress. Plant Physiol. Biochem. 150, 244–253.

Sanders, A., Collier, R., Trethewy, A., Gould, G., Sieker, R., Tegeder, M., 2009. AAP1 regulates import of amino acids into developing Arabidopsis embryos. Plant J. 59 (4), 540–552. https://doi.org/10.1111/j.1365-313X.2009.03890.x.

Sharp, P.M., Li, W.H., 1987. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15, 1281–1295.

Shoo, S., Das, S., Rakshit, R., 2019. Codon usage pattern and predicted gene expression in Arabidopsis thaliana. Gene X2, (2) 100012.

Shu-juan, Z., Yu-lian, L., Guo-qi, S., Jie, G., Rong-zhi, Z., Wei, L., Ming-li, C., Gen-yin, L., 2020. Heterologous expression of the ThIPK2 gene enhances drought resistance of common wheat. J. Integr. Agric. 19 (4), 941–952.

Sleper, D., Poehlman, J., 2006. Breeding Field crop. Blackwell, Iowa, pp. 221–225.

Tian, R., Yang, Y., Chen, Y., 2020. Genome-wide survey of the amino acid transporter gene family in wheat (Triticum aestivum L.): identification, expression analysis and response to abiotic stress. Int. J. Biol. Macromol. 162 (1), 1372–1387.

Wada, K., Aota, R., Tsuchiya, F., Ishibashi, T., Gojobori, T., Ikemura, T., 1990. Codon usage tabulated from GenBank genetic sequence data. Nucleic Acids Res. 18 (Suppl), 2367–2411.

Wang, P., Yang, Q., Sang, S., Chen, Y., Zhong, Y., Wei, Z., 2017. Arabidopsis inositol polyphosphate kinase AtIpk2beta is phosphorylated by CPK4 and positively modulates ABA signaling. Biochem. Biophys. Res. Commun. 490, 441–446.

Wong, E., Smith, D., Rabadan, R., Peiris, M., Poon, L., 2010. Codon usage bias and the evolution of influenza a viruses. Codon usage biases of influenza virus. BMC Evol. Biol. 10 (253). https://doi.org/10.1186/1471-2148-10-253.

Yang, Q., Sang, S., Chen, Y., Wei, Z., Wang, P., 2018. The role of Arabidopsis inositol polyphosphate kinase AtIPK2 beta in glucose suppression of seed germination and seedling development. Plant Cell Physiol. 59, 343–354.

You, J., Zhang, Y., donghua, A., Wang, X., dossa, K., Zhou, R., Yu, J., Zhang, Y., Wang, L., and Zhang, X. 2019. Transcriptomic and metabolic profiling of drought-tolerant and susceptible sesame genotypes in response to drought stress. MBC Plant Biol, 19, 267. Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/KU695148.1.

Yu, C., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M., Liu, Y., 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Mol. Cell 59 (5), 744–754. https://doi.org/10.1016/j.molcel.2015.07.018.

Zhang, J., Li, B., Yang, Y., Mu, P., Qian, W., Dong, L., Zhang, K., Liu, X., Qin, H., Ling, H., Wang, D., 2016. A novel allele of L-galactono-1,4-lactone dehydrogenase is associated with enhanced drought tolerance through affecting stomatal aperture in common wheat. J. Sci. Rep. 6, 30177.

Zhou, H., Wang, H., Huang, L., Naylor, M., Clifford, P., 2005. Heterogeneity in codon usages of sobemovirus genes. Arch. Virol. 150, 1591–1605.