



Original article

# Analysis of chromosomes and nucleotides in rice to predict gene expression through codon usage pattern

Meshal M. Almutairi

National Center of Agricultural and Technology, King Abdulaziz City for Science and Technology (KACST), P.O. Box 6086, Riyadh 11442, Saudi Arabia



## ARTICLE INFO

## Article history:

Received 7 October 2020

Revised 7 April 2021

Accepted 21 April 2021

Available online 30 April 2021

## Keywords:

Amino acids

Codon adaptation index (CAI)

Relative synonymous codon usage (RSCU)

Modified relative codon bias (MRCB),

guanine and cytosine content (GC %)

## ABSTRACT

Amino acids are essential measurements for the potential growth stage because of connecting to protein structures and functions. The objective of this paper was to analyze chromosomes feature at plastid region of rice represented by nucleotide, synonymous codon, and amino acid usage to predict gene expression through codon usage pattern. The results showed that the values of the codon adaption index ranged from 0.733 in chromosome 9 to 0.631 in chromosome 8 with full length of these two chromosomes were 3738 and 1635 respectively. The higher value of guanine and cytosine content was 60% in chromosomes 9 while the lower values was 37% in chromosomes 11. Eight chromosomes (ch1, ch2, ch3, ch5, ch7, ch8, ch10, and ch12) were greater value of modified relative codon bias than threshold (threshold: 0.66) especially in cysteine for ch1, ch2, ch5, ch10, and ch12. While other remaining chromosomes were less than the threshold. Relative synonymous codon usage found that the over-represented of amino acids were asparagine, aspartate, cysteine, glutamate, and phenylalanine across all 12 chromosomes. These results would establish a platform for more and further projects concerning rice breeding and genetics and codon optimization in the amino acids for developing varieties. These results also will help breeders to select desirable genes through the genome for improve target traits.

© 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

Rice (*Oryza sativa* L.) is one of the most food sources among cereal crops and is stable food for all world's population. This gene, *Oryza*, has basic 12 chromosomes with six genome groups (A, B, C, D, E, and F;  $2n = 2X = 24$ ). These genomes and the efficiency of genetic transformation are considered excellent models for biological studies (Goff, 1999). The rice quality trait is important to enhance the nutritional values. Rice grains however, are deficient in amino acids especially lysine and methionine. To undertake the quality trait of rice, analysis of the amino acid is an effective approach and main goal for breeders (Ufaz and Galili, 2008).

One of the important rice traits is plastid that differentiates into several forms one of them is the chloroplast. The function of plastid is important due to its manufacture and storage of food used in photosynthesis. The chlorophyll carries photosynthesis that known

as chloroplasts to producing energy. The genome of chloroplast consists of several copies of homogeneous circular with its size of DNA reaches about 120 kb and contains six tRNA, a pair large of invert repeats (IR), and seven genes (Palmer 1991; Sugita and Sugiura, 1996; Guat, 1998). Development transformation of this chloroplast helps to address and modify functional aspects for the plastid genome (Drescher et al., 2000).

Amino acids are the essential measurements of the potential growth stage because of connecting to protein structures and functions and involving many metabolic processes (Tian et al., 2020). Thus, these amino acids establish the basis of genetic information. In addition, several of these amino acids are encoded to more than one codon that can be revealed as synonymous codon (Hershberg, 2016). The bias in the synonymous codon may arise from a consequence of natural selection or mutation according to some of studies showed that selection prefers specific codons (Hershberg and Petrov, 2008). Nucleotides are the basic units of mRNA reflecting fundamental roles throughout the generation of organism. Furthermore, codon adaption index (CAI) is one of the most methods to calculate the frequency of the synonymous codon for the mRNA sequence to gain more understanding of gene expression, new genes discovery, and molecular mechanism contributed to gene evolution. Thus, the theoretical basis of analyzing codon usage is

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

E-mail address: [malmetairy@kacst.edu.sa](mailto:malmetairy@kacst.edu.sa)<https://doi.org/10.1016/j.sjbs.2021.04.059>

1319-562X/© 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/>).

an important concept for genetic engineering, gene prediction, and molecular studies in rice (Yu et al., 2015). This CAI has influenced by several factors such as nucleotide composition, tRNA abundance, the structure of a protein, and translation processes (Lithwick and Margalit, 2005). CAI values close to one suggest increasing intense selection that would help to reach efficient translation (Carbone et al., 2003).

Prediction of gene expression from nucleotide sequence is considered a key knowledge for modern bioinformatics. This gene expression is controlled by several factors such as protein biosynthesis, mutation, and natural selection. This gene expression feature reflects the importance of codon usage patterns that vary from one gene to another. Several measurements for gene expression are available such as codon adaptation index (CAI), relative synonymous codon usage (RSCU), guanine and cytosine content (GC content), relative codon bias strength (RCBS), and modified relative codon bias (MRCBS) (Das, 2017; Shoo and Das, 2014; (Sharp and Li, 1987; Zhou et al., 2013). The objective of this work was to analyze chromosomes feature in rice represented by nucleotide, synonymous codon, and amino acid usage to predict gene expression through codon usage pattern.

## 2. Materials and methods

### 2.1. Retrieval of sequences

The sequences of 12-rice chromosome (*Oryza sativa* ssp japonica cv. Nipponbare) was retrieved from rice genome annotation project (Kawahara et al., 2013; Ouyang et al., 2007) through <http://rice.plantbiology.msu.edu/>. The locations of all chromosomes are at plastid provided with Landmark IDs (Table 1).

### 2.2. Measures of codon usage patterns

A variety of analyses tool were used to measure the codon usage patterns such (CAI), (RSCU), GC content, (RCBS), and (MRCBS) that have applied in this study (Das, 2017; Shoo and Das, 2014; Sharp and Li, 1987; Zhou et al., 2013). While other calculations were done through Microsoft Excel and codonW software via bioinformatics.org.

### 2.3. Codon adaptation index and GC%

CAI is calculated from the formula =  $\left(\prod_{i=1}^N w_i\right)^{\frac{1}{n}}$  where N is the number of codons in the gene and relative adaption  $w_i$  is described as  $w_i = \frac{f_i}{f_{aa,max}}$  where  $f_i$  is the frequency of the  $i^{\text{th}}$  codon and  $f_{aa,max}$  is maximum frequency of the codon most often used for encoding amino acid aa.

**Table 1**

The basic 12 chromosome attributes for rice plastids.

Chromosomes	Landmarks	Average	CAI	GC%	Length (bp)
1	LOC_Os01g02170	0.71	0.67	44.50	5299
2	LOC_Os02g26014	0.72	0.64	43.73	19,596
3	LOC_Os03g09830	0.70	0.67	43.36	2765
4	LOC_Os04g52100	0.62	0.68	40.46	3574
5	LOC_Os05g15320	0.70	0.66	46.43	3151
6	LOC_Os06g27760	0.58	0.66	43.73	245
7	LOC_Os07g39590	0.71	0.64	54.60	3024
8	LOC_Os08g26800	0.65	0.63	51.13	1635
9	LOC_Os09g36470	0.60	0.73	60.06	3738
10	LOC_Os10g42500	0.68	0.66	46.93	2936
11	LOC_Os11g14950	0.57	0.65	37.83	5218
12	LOC_Os12g37530	0.65	0.67	41.74	4880

Determination of guanine and cytosine (GC) content (%) was according to this formula:

$$GCcontent = \frac{TotalnumberofGandC}{TotalnumberofA,T,C,andG} \times 100$$

### 2.4. Relative synonymous codon usage

RSCU can be estimated as following formula =  $\frac{x_{ij}}{\sum_j^n x_{ij}}$  where  $x_{ij}$  shows the frequency of codon j for the  $i^{\text{th}}$  amino acid and  $n_i$  represents the number of synonymous codons encoding the  $i^{\text{th}}$  amino acids. The interpretation for the RSCU is that the value of RSCU greater than or equal to 0.60 suggests over-represented in a gene for the particular codon. Whereas RSCU small than 0.6 indicates low-represented codon for the corresponding amino acid (Wong et al., 2010).

### 2.5. Modified relative codon bias

MRCBS measures the expression level of a gene and is estimated by these formulas =  $\prod_{i=1}^N (MRCBS_{xyz})$  then this  $MRCBS_{xyz} = \frac{RCBS_{xyz}}{RCBS_{aa,max}}$ . Then  $Relativecodonbiasstrength(RCBS)_{(xyz)} = \frac{f_{xyz}}{f_1(x)f_2(y)f_3(z)}$  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 n 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

All above formulas were calculated through Microsoft Excel program and codonW software. Moreover, two ways analysis of variance was computed using PROC GLM procedure at significance level of p-value (0.05) in SAS computer packages version 9.2 for windows (SAS Institute Inc.). Z test was calculated using formula:  $\frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$  to describe the impact score. Some of codons such as methionine, tryptophan, and three stop codons were excluded from calculation in order to normalize these codons for all mRNA sequences (Zhou et al., 2005).

## 3. Results

### 3.1. Data CAI and GC%

The rice's chromosomes at the plastid region were analyzed to calculate codon adaption index (CAI) as shown in Table 1. The CAI values range from 0.73 in chromosome 9 to 0.63 in chromo-

some 8 and full length of these two chromosomes were 3738 and 1635 respectively. The high CAI values represents highly gene expression while the lower CAI values represent low gene expression. In addition, the higher value of the GC% showing was in chromosomes 9 (60%); while, the lower values was in chromosomes 11 (37%). The low GC% indicates that not all codons are used uniformly. It was found that chromosome 9 has higher in both CAI and GC%. However, the high length was in chromosome 2 (19596 bp) but its CAI value was 0.64 and GC% was 43.73. This chromosomes (2) reflects surprising results due to long length with moderate value in both CAI and GC%.

3.2. Data MRCBS, Euclidean distance, and RSCU

The MRCBS was calculated based on rice chromosome sequences to report gene expressions for several amino acids that varied from one to 0.09 (Table 2). The MRCB values that greater than the threshold were taken as standard for determining the high gene expression. Therefore, the threshold of amino acid is set to 0.66 due to the overall of CAI among chromosomes. Chromo-

somes that were bigger than the threshold were ch1, ch2, ch3, ch5, ch7, ch8, ch10, and ch12, while other chromosomes were less than the threshold. Among these chromosomes, the higher value of MRCBS was found in cysteine in these following chromosomes: ch1, ch2, ch5, ch10, and ch12; however, the lower MRCBS value was in serine among other amino acids (Fig. 1). Moreover, the distance tree shows that high similarity was found between chromosomes 11 and 4, between 12 and 10, and between 5 and 3. While the fewer similarity chromosomes were found between 11 and 7, between 12 and 8, and between 5 and 6 (Fig. 2). At codons, the higher MRCBS value (>0.90) was found in TTT (Phe), AAA (Lys), GAT (Asp), TAT (Tyr), TGT (Cys), and CAT (His) (Fig. 1). Using the Z test, we found some significant impact score in some codons that contribute to an expression such as TGT (Cys), GAT (Asp), GAA (Glu), TTT (Phe), GGA (Gly), AAA (Lys), TTA (Leu), CAA (Gln), ACA (Thr), TAT (Tyr), and TAA (End) < 0.05. These significant codons may influence translational efficiency. (Duret, 2002; Roymondal et al., 2009).

The RSCU values were calculated to estimate synonymous codon usage pattern for 12 rice chromosomes (Table 3). The

Table 2  
The MRCBS for Rice chromosomes in the Protein Location at Plastid excluded Methionine and Tryptophan.

Amino Acid	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12
Alanine	0.67	0.67	0.64	0.43	0.51	0.26	0.47	0.72	0.36	0.76	0.64	0.58
Arginine	0.46	0.81	0.73	0.46	0.55	0.65	0.64	0.32	0.40	0.56	0.35	0.48
Asparagine	0.77	0.76	0.75	0.84	0.91	0.78	0.77	0.78	0.62	0.76	0.72	0.88
Aspartic	0.75	0.85	0.62	0.82	0.57	0.75	0.93	0.85	0.73	0.61	0.66	0.98
Cysteine	0.94	0.82	0.70	0.99	0.88	0.85	0.76	0.62	0.68	0.90	0.80	0.84
Glutamic	0.75	0.74	0.93	0.75	0.97	0.85	0.81	0.76	0.69	0.66	0.63	0.73
Glutamic acid	0.92	0.72	0.58	0.66	0.72	0.52	0.68	0.58	0.98	0.60	0.69	0.63
Glycine	0.68	0.72	0.51	0.46	0.73	0.28	0.78	0.67	0.60	0.42	0.43	0.70
Histidine	0.75	0.81	0.76	0.67	0.75	0.83	0.80	0.82	0.80	0.87	0.60	0.84
Isoleucine	0.83	0.54	0.73	0.66	0.60	0.35	0.91	0.58	0.48	0.68	0.70	0.70
Leucine	0.71	0.55	0.58	0.65	0.56	0.33	0.71	0.36	0.72	0.57	0.63	0.73
Lysine	0.73	0.94	0.65	0.65	0.75	0.51	0.77	0.59	0.87	0.60	0.56	0.72
Phenylalanine	0.99	0.72	0.70	0.80	0.76	0.68	0.80	0.96	1.00	0.93	0.72	0.66
Proline	0.50	0.79	0.70	0.52	0.75	0.59	0.55	0.59	0.69	0.71	0.54	0.47
Serine	0.51	0.66	0.60	0.43	0.59	0.55	0.52	0.62	0.33	0.53	0.37	0.48
Threonine	0.66	0.79	0.78	0.59	0.92	0.62	0.77	0.82	0.36	0.68	0.45	0.48
Tyrosine	0.95	0.68	0.69	0.61	0.73	0.88	0.85	0.92	0.59	0.97	0.73	0.81
Valine	0.77	0.54	0.76	0.58	0.67	0.57	0.47	0.69	0.47	0.70	0.62	0.63

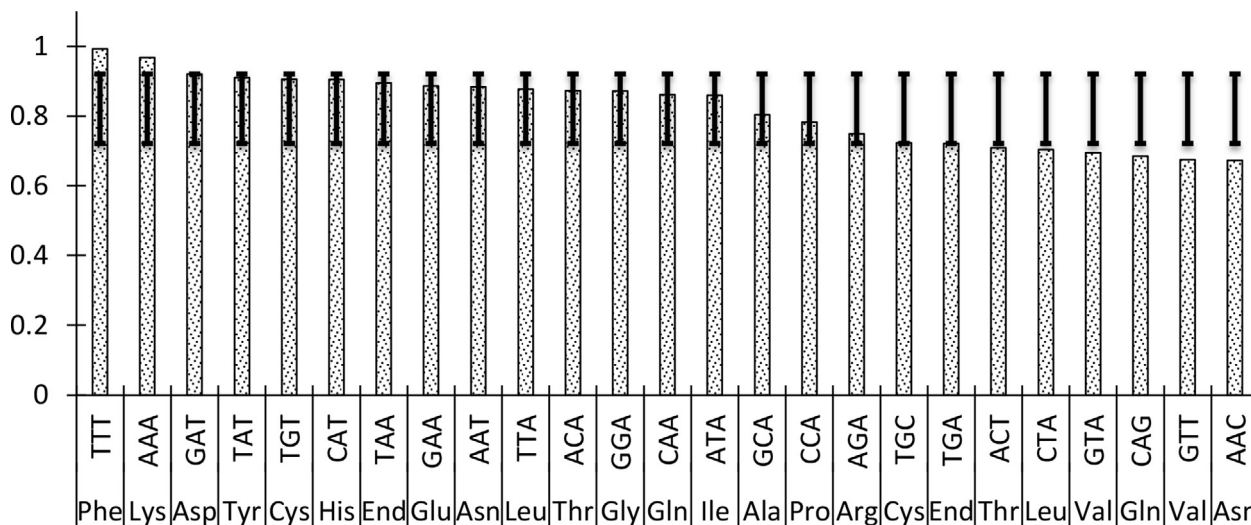


Fig. 1. The higher MRCBS values (greater than or equal 0.66) for amino acid codons with standard deviation excluded Methionine and Tryptophan.

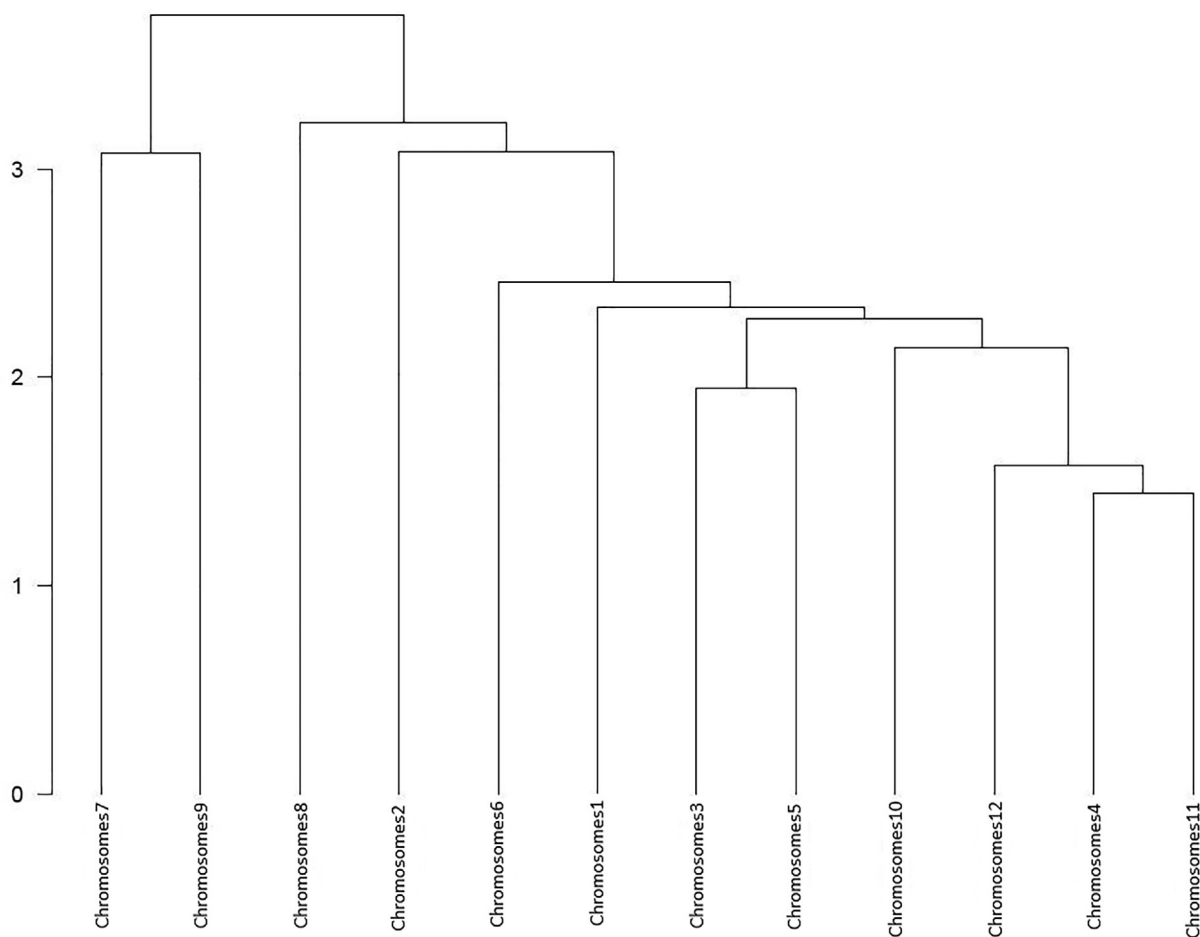


Fig. 2. The tree relationship among Rice chromosomes using Euclidean distance.

Table 3

The RSCU values for Rice chromosomes Location at Plastid, excluded Methionine and Tryptophan.

Amino Acids	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12
Alanine	0.29	0.33	0.36	0.28	0.29	0.29	0.31	0.31	0.48	0.35	0.30	0.28
Arginine	0.17	0.20	0.21	0.18	0.19	0.18	0.22	0.18	0.24	0.19	0.18	0.18
Asparagine	<u>0.62*</u>	<u>0.63</u>	<u>0.64</u>	<u>0.66</u>	<u>0.71</u>	<u>0.61</u>	<u>0.73</u>	<u>0.62</u>	<u>0.80</u>	<u>0.61</u>	<u>0.62</u>	<u>0.65</u>
Aspartic	<u>0.69</u>	<u>0.67</u>	<u>0.73</u>	<u>0.67</u>	<u>0.83</u>	<u>0.69</u>	<u>0.67</u>	<u>0.68</u>	<u>0.68</u>	<u>0.80</u>	<u>0.71</u>	<u>0.67</u>
Cysteine	<u>0.67</u>	<u>0.65</u>	<u>0.61</u>	<u>0.71</u>	<u>0.66</u>	<u>0.64</u>	<u>0.74</u>	<u>0.56</u>	<u>0.75</u>	<u>0.64</u>	<u>0.65</u>	<u>0.64</u>
Stop codon	0.43	0.47	0.45	0.42	0.47	0.47	0.47	0.40	0.40	0.44	0.48	0.47
Glutamic	<u>0.72</u>	<u>0.73</u>	<u>0.64</u>	<u>0.70</u>	<u>0.65</u>	<u>0.69</u>	<u>0.65</u>	<u>0.62</u>	<u>0.62</u>	<u>0.77</u>	<u>0.75</u>	<u>0.70</u>
Glutamic acid	<u>0.66</u>	<u>0.67</u>	<u>0.64</u>	<u>0.66</u>	<u>0.67</u>	<u>0.67</u>	<u>0.60</u>	<u>0.66</u>	<u>0.66</u>	<u>0.65</u>	<u>0.66</u>	<u>0.64</u>
Glycine	0.31	0.32	0.30	0.29	0.32	0.32	0.35	0.33	0.35	0.31	0.27	0.33
Histidine	<u>0.62</u>	<u>0.64</u>	<u>0.64</u>	<u>0.60</u>	<u>0.61</u>	<u>0.63</u>	<u>0.72</u>	<u>0.73</u>	<u>0.68</u>	<u>0.64</u>	0.56	<u>0.64</u>
Isoleucine	0.41	0.40	0.42	0.42	0.42	0.41	0.47	0.44	0.60	0.42	0.39	0.41
Leucine	0.19	0.19	0.19	0.19	0.20	0.19	0.21	0.21	0.21	0.20	0.18	0.19
Lysine	<u>0.73</u>	<u>0.76</u>	<u>0.70</u>	<u>0.71</u>	<u>0.66</u>	<u>0.73</u>	<u>0.67</u>	<u>0.71</u>	0.59	<u>0.74</u>	<u>0.86</u>	<u>0.71</u>
Phenylalanine	<u>0.67</u>	<u>0.66</u>	<u>0.66</u>	<u>0.67</u>	<u>0.66</u>	<u>0.64</u>	<u>0.66</u>	<u>0.67</u>	<u>0.66</u>	<u>0.67</u>	<u>0.66</u>	<u>0.64</u>
Proline	0.28	0.30	0.30	0.29	0.29	0.31	0.30	0.36	0.30	0.32	0.29	0.29
Serine	0.19	0.19	0.20	0.19	0.18	0.18	0.20	0.22	0.26	0.19	0.19	0.19
Threonine	0.30	0.31	0.31	0.31	0.32	0.30	0.32	0.30	0.45	0.32	0.29	0.30
Tyrosine	<u>0.70</u>	0.59	<u>0.61</u>	0.57	<u>0.61</u>	<u>0.65</u>	<u>0.63</u>	<u>0.69</u>	<u>0.84</u>	<u>0.66</u>	<u>0.62</u>	<u>0.63</u>
Valine	0.29	0.30	0.34	0.30	0.30	0.29	0.31	0.33	0.37	0.31	0.29	0.27

\* Underline RSCU values (RSCU greater than or equal to 0.66) regarded as an over-represented.

over-represented with amino acids were asparagine, aspartic acid, cysteine, glutamic acid, and phenylalanine across all 12 rice chromosomes. Moreover, the over-represented with codon were AAA, AAT, CAT, GAT, and TAT across all 12 chromosomes (data not shown).

#### 4. Discussion

Rice (*Oryza sativa* L.) is one of the most food sources among cereal crops and is stable food for all world's population. The basic 12 chromosomes features in the rice were analyzed through nucleotides and codon patterns because of connotation with the genome evolution in order to study the gene expressions. Overall, it was found that high CAI values represented high gene expression in chromosome 9 meanwhile the higher value of the GC% was in chromosome 9 (60%). This chromosome 9 could benefit the translation efficiency of the rice gene population (Das, 2006). Zhang et al. (2007) found that in wheat codon usages, the RSCU value was found to be 0.08 as preferred codons in the chloroplast, nuclear, and mitochondrion. They concluded that the intensity of natural selection might reflect the variation of codon usage pattern. Some studies have shown that GC contents are the main influence of codon usage patterns that effect ultimately gene expression resulted from mutation bias that could be used to determine the major trends in codon usage. 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. (Duret, 2002; (Almutairi and Alrajhi, 2020). While other studies showed different results. Thus, the difference among these studies may be due to the use of different methods for calculation expressions (Roymondal et al., 2009). This GC content, however, is known to be more sensitive for rare amino acids and thus poorly provides information about the codon usage unless a short gene sequence of fewer than 100 codons has analyzed (Mazumdar et al., 2017). Kamjijam et al. (2020) noted that a longer period for germination about 48 h for rice seed developed glutamic acid with a high concentration value of 0.89 and all amino acids reached their peaks after 96 h. During 24 h of germination water-uptake was observed and most amino acids were nourished with glutamic acid, aspartic acid, leucine, lysine, proline, and tryptophan. They also found that a higher concentration of amino acids was located inside the endosperm of rice. Thus, these results show dissimilar with other studies such as (Das, 2017) and (Yan-qing et al., 2018) who found that codon closed to U was preferred in more observations.

#### 5. Conclusion

In this study, we analyzed chromosomes feature in Rice to predict gene expression through codon usage index. This analysis open a new window for plant breeders for more developments growth habits. More study on codon usage pattern for rice is needed for detection of expressed genes. Thus, this study was importance to demonstrate significant heterogeneity in codon usage among genes in rice chromosomes feature of rice MRCBS shows higher values at some chromosomes while other chromosomes are less than the threshold. Furthermore, we found that high similarity was found between chromosomes 11 and 4 between 12 and 10, and between 5 and 3. Therefore, we concluded that some chromosomes such as 12, 10, 5, and 3 might play a critical role for rice breeding. Furthermore, some amino acids have both higher MRCBS and RSCU (over-represented) such as asparagine, phenylalanine, and aspartic. This relative synonymous codon usage was normalized by removing ATG (methionine), TGG (tryptophan),

and three stop codons. Similarity, this codon CAT, GAT, and TAT found to be more benefit and considered in rice breeding program. Analysis of amino acids have more benefits to advance understanding their metabolism. More combinations studies are required to understand the amino acid network metabolism such as biochemical, molecular, and genomics (Galili et al., 2008). These results would determine a good breeding program in the further projects concerning rice breeding and genetics and codon optimization in the amino acids for developing varieties. These results also will help breeders to select desirable genes through the genome for improve target traits.

#### Funding

This paper did not receive any funding resources or specific grant from agencies in the public, commercial, or any other sectors.

#### 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.

#### Acknowledgements

Authors would like to thank KACST for provided facilities and materials that help to setting and finished this research. We also thank the reviewers for their valuable comments and improvements.

#### References

- Almutairi, M., Alrajhi, A., 2020. Prediction of gene expression under drought stress in spring wheat using codon usage pattern. Saudi Journal of Biological Sciences 28 (7), 4000–4004. <https://doi.org/10.1016/j.sjbs.2021.04.015>.
- Carbone, A., Zinovyev, A., Képès, F., et al., 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 19 19, 2005–2015. <https://doi.org/10.1093/bioinformatics/btg272>.
- Das et al., Das, S., Paul, S., and Dutta, C., 2006. Synonymous codon usage in adenoviruses: Influence of mutation, selection and protein hydrophathy. Virus Res 117 (2), 227–236.
- Das, S., 2017. Analysis of Gene Expression using modified relative codon bias strength in Nanoarcheum Equitans. Biosci. Biotechnol. Res. Asia 14 (2), 793–799.
- Drescher, A., Ruf, S., Calsa Jr, T., Carrer, H., Bock, R., 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. Plant J. 22 (2), 97–104.
- Duret, L., 2002. Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Develop. 12 (6), 640–649.
- Galili, S., Amir, R., Galili, G., 2008. Genetic Engineering of Amino Acid Metabolism in Plants. Adv. Plant Biochem. Mol. Biol. 1, 49–80.
- Goff, 1999. Rice as model for cereal genomics. Curr. Opin. Plant Biol. 2 (4), 86–89.
- Guat, B., 1998. Molecular clocks and nucleotide substitution rates in higher plants. In: Hecht, M.K. (Ed), evolutionary Biology, vol. 30.
- Hershberg, R., 2016. Codon usage and translational selection. Encyclopedia Evolut. Biol. 293–298. <https://doi.org/10.1016/B978-0-12-800049-6.00178-5>.
- Hershberg, R., Petrov, D., 2008. Selection on codon bias. Annu. Rev. Genet. 78, 271–300.
- Kawahara, Y., Bastide, M., Hamilton, J., Kanamori, H., McCombie, W., Ouyang, S., Schwartz, D., Tanaka, T., Wu, J., Zhou, S., Childs, K., Davidson, R., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S., Kim, J., Numa, H., Itoh, T., Buell, C., Matsumoto, T., 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice 6, 4. <https://doi.org/10.1186/1939-8433-6-4>.
- Lithwick, G., Margalit, H., 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>.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissan, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, C., 2007. The TIGR Rice Genome Annotation Resource: Improvements and new features. Nucleic Acids Research 35 (Database issue):D883–D887; <http://doi.org/10.1093/nar/gkl976>.
- Palmer, J., 1991. Plastid chromosome: structure and evolution. In: Bogorad, L., Vasil, I. (Eds.), Cell Culture and somatic cell genetics of plants, vol. 7A. Academic Press, New York, pp. 5–53.

- Roymondal, U., Das, S., Sahoo, S., 2009. Predicting Gene Expression Level from Relative Codon Usage Bias: An Application to *Escherichia coli* Genome. *DNA Res.* 16 (1), 13–30.
- Sharp, P., Li, W., 1987. The codon adaptation index a measure of directional synonymous codon usage bias and its potential applications. *Nucl. Acids Res.* 15, 1281–1295.
- Shoo, S., Das, S., 2014. Analyzing gene expression and codon usage bias in diverse genomes using a variety of model. *Curr. Bioinforma.* 9, 102–112.
- Sugita, M., Sugiura, M., 1996. Regulation of gene expression in chloroplasts of higher plants. In: Filipowicz, W., Hohn, T. (Eds.), *Post-Transcriptional Control of Gene Expression in Plants*. Springer, Dordrecht. [https://doi.org/10.1007/978-94-009-0353-1\\_14](https://doi.org/10.1007/978-94-009-0353-1_14).
- Tian, R., Yang, Y., Chen, M., 2020. Genome-wide survey of the amino acid transporter gene family in wheat (*Triticum aestivum* L.): Identification, expression analysis and response to abiotic stress. *International J. of Biological. Macromolecules* 162, 1372–1387.
- Ufaz, S., Galili, G., 2008. Improving the content of essential amino acids in crop plants: goals and opportunities. *Plant Physiol.* 147, 954–961.
- 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>.
- Yan-qing, W., Zhi-yuan, L., Da-qiu, Z., Jun, T., 2018. Comparative analysis of flower-meristem-identity gene APETALA2 (AP2) codon in different plant species. *J. Integrative Agric.* 17 (4), 867–877.
- 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, Wen, Zhou, J., Li, Z., Wang, L., Gu, X., Zhong, Y., 2007. Comparative Analysis of Codon Usage Patterns Among Mitochondrion, Chloroplast and Nuclear Genes in *Triticum aestivum* L. *J. Integr. Plant Biol.* 49 (2), 246–254.
- Zhou, J., Zhang, J., Sun, D., Ma, Q., Chen, H., Ma, L., Ding, Y., Liu, Y., 2013. The distribution of synonymous codon choice in the translation initiation region of dengue virus. *PLoS ONE* e77239.
- 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.