

Stress induced MAPK genes show distinct pattern of codon usage in *Arabidopsis thaliana*, *Glycine max* and *Oryza sativa*

H Surachandra Singha, Supriyo Chakraborty* & Himangshu Deka

Department of Biotechnology, Assam University, Silchar-788011, Assam, India; Supriyo Chakraborty - Email: supriyoch_2008@rediffmail.com; *Corresponding author

Received June 23, 2014; Accepted June 27, 2014; Published July 22, 2014

Abstract:

Mitogen activated protein kinase (MAPK) genes provide resistance to various biotic and abiotic stresses. Codon usage profiling of the genes reveals the characteristic features of the genes like nucleotide composition, gene expressivity, optimal codons etc. The present study is a comparative analysis of codon usage patterns for different MAPK genes in three organisms, viz. *Arabidopsis thaliana*, *Glycine max* (soybean) and *Oryza sativa* (rice). The study has revealed a high AT content in MAPK genes of *Arabidopsis* and soybean whereas in rice a balanced AT-GC content at the third synonymous position of codon. The genes show a low bias in codon usage profile as reflected in the higher values (50.83 to 56.55) of effective number of codons (N_c). The prediction of gene expression profile in the MAPK genes revealed that these genes might be under the selective pressure of translational optimization as reflected in the low codon adaptation index (CAI) values ranging from 0.147 to 0.208.

Keywords: MAPK, *Arabidopsis*, synonymous codons, codon usage bias, mutational pressure.

Background:

Many codons in the genetic code are functionally synonymous. A single amino acid is encoded by two to six codons, a phenomenon called synonymous codon usage. All the synonymous codons of an amino acid show a variation in the occurrence in a gene frequency. Some codons show a high frequency indicating that the usage of the particular codon is biased. The study of the codon usage bias may assist to design the DNA primers [1]. The synonymous codons show variation between as well as within the genome of an organism [2, 3]. This variation could be a consequence of natural and/or mutational pressure to determine the accuracy and the efficiency of the translation process of the organisms. Study of the synonymous codon usage patterns of a gene in various organisms could provide an insight into the evolution profile and the level of gene expression as well [4, 5]. In due course of evolution, plants have developed their own mechanisms to combat the stresses- both biotic as well as abiotic which they are continually subject to. Mitogen activated protein kinase

(MAPK) genes present in the plants actively respond to various stresses. Abiotic stresses mainly include drought, high salinity, heat, cold, freezing, limited nutrient availability, heavy metals etc. [6, 7]. The MAPK genes are generally classified into three distinct types-MAPKKK, MAPKK and MAPK. The string of reactions in the stress signalling pathways involve these three MAPK families; MAPKKKs phosphorylate the serine/threonine of MAPKKs at their activation loop which in turn double phosphorylate the MAPKs at their T-D-Y motif in the activation loop. The MAPK genes get activated by the stress stimuli and form various cascade which form the metabolic pathways to regulate the stress response. The cascade actions acts downstream of the receptors on extracellular surface or acts as a sensor that's transduces the extracellular signals into intracellular responses [8].

Arabidopsis genome has been reported to harbour 80 MAPKKKs, 10 MAPKKs and 20 MAPKs genes. Whole genome

analysis of rice (*Oryza sativa*) has revealed 75 MAPKKs, 8 MAPKKs and 15 MAPKs [9, 10]. *In-silico* studies of the Soybean (*Glycine max*) has identified 38 MAPKs, 11 MAPKKs, and 150 MAPKKs [11]. Chilling stress is one of the serious problems associated with the production of major crops such as rice and maize [12] in temperate zone. The *Arabidopsis* MAPKs, based on their structural motif and sequence similarities, is classified into four groups (A-D). The group A, B and C possesses a T-E-Y motif whereas the fourth group A possesses a T-D-Y motif [13, 14]. A signalling pathway comprising of MEKK1-MKK2-MPK4/MPK6 is reported to regulate the response to abiotic stresses in *Arabidopsis* [15]. The abiotic stresses induce the production of ROS (reactive oxygen species). Several MAPK signalling pathways are induced only by ROS and in turn regulate the ROS production. Many MAPK pathways in response to abiotic stresses have been studied in rice. Experiment on long term exposure of rice plants to cooling stress revealed the involvement of several MAPKs namely MEK1 (MAPKK) and MAP1 (MAPK) [16]. MAPK5 gene in rice has shown multiple activities, responding to both biotic and abiotic stress. Resistivity to drought in rice plants is provided by an MAPKK of the B3 subgroup named DSM1 [17]. MAPK33 has also shown activity withstanding drought [18].

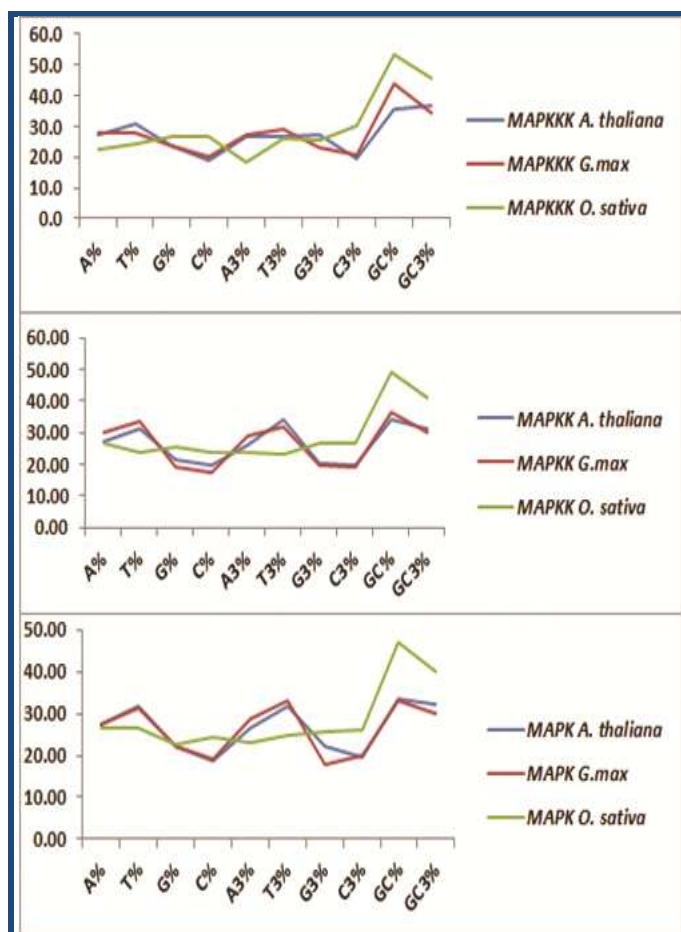


Figure 1: Nucleotide composition (%) of the three MAPK gene families, in Arabidopsis, soybean and rice

Microarray analysis of MPK4 activity in soybean revealed that it negatively regulates SA and H₂O₂ accumulation. Therefore silencing of MPK4 in soybean significantly increases SA and H₂O₂ accumulation, up-regulating genes involved in defense

response and providing the plants a better resistance to downy mildew and soybean mosaic virus as compared to vector controlled plants. MPK4 has been reported to down-regulate genes involved in growth and development, such as those in auxin signalling pathways and in cell cycle and proliferation [19]. Very little information is available on the codon usage pattern of MAPK genes across plant species despite substantial literature on physiological mechanisms. The present study has therefore been undertaken to elucidate the codon usage of MAPK genes. Overall comparative analysis of the codon usage patterns of the MAPK genes reported in the three organisms - *A. thaliana*, *G. max* and *O. sativa* could be beneficial to analyse the conservative nature of the genes, codon usage patterns and the compositional role in determining the optimal codons. Information regarding the codon usage patterns could help reveal the evolutionary history of individual genes within or between organisms, and the expression of genes as well.

Methodology:

Coding DNA sequence data

Coding sequence data (a total of 127 cds) of the different genes of the MAPK families of *Arabidopsis*, soybean and rice were retrieved from NCBI (www.ncbi.nlm.nih.gov). The genes with different accession numbers are listed in **Table 1** (see **supplementary material**). Different analytical parameters for codon usage bias of the MAPK genes with different accession number were estimated and analysed.

Analysis of the codon usage profile

Several parameters have been used to characterize the sequence data of the genes. Gene members of each of the three MAPK families are analysed and compared family-wise across the three organisms.

RSCU

The relative synonymous codon usage (RSCU) measures the frequencies of optimal codons of each of the synonymous codons encoding an amino acid. It assists in characterising the codons in a genetic sequence, whether it follows unbiased pattern of the codons being used or certain codons are more preferred. Codons with RSCU values greater than 1 are generalised to possess a positive codon usage bias and those less than 1 are considered to possess negative codon usage bias [20]. Bioinformatics tools available online, codonW (mobyle.pasteur.fr/cgi-bin/portal.py#forms::codonW) have been used to estimate the RSCU which is mathematically expressed as: **(For equation-> (1) please see supplementary material)**, where, x_i is the number of frequency of j^{th} codon for i^{th} amino acid and n_i being the number of alternative synonymous codons available for the i^{th} amino acid.

GC content

It is a measure of the occurrence of the nucleotide bases guanine (G) and cytosine (C) in the entire genetic sequence [21]. The GC content can be measured partially at the 1st, 2nd and 3rd synonymous codon position throughout the sequence, designated as GC_{1S}, GC_{2S}, and GC_{3S} respectively [22]. Other indices for nucleotide bias such as A%, T%, G%, and C% are measured. The nucleotide measure at the third codon positions designated as A_{3S}, T_{3S}, G_{3S} and C_{3S} are also calculated. Mathematically GC% is expressed as: **(For equation-> (2) please see supplementary material)**.

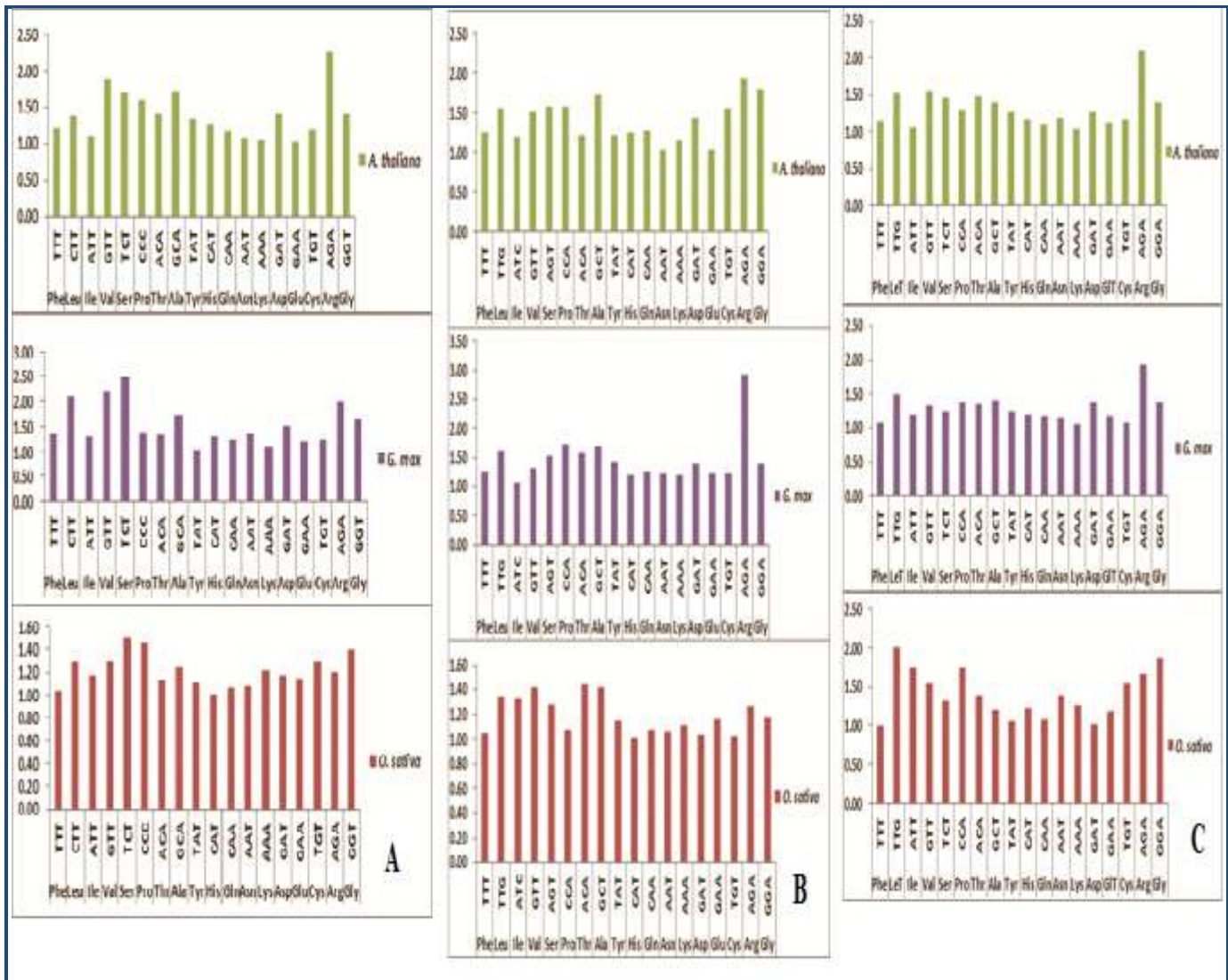


Figure 2: RSCU values of the Arabidopsis, soybean and rice showing three families: (A) MAPK; (B) MAPKK and (C) MAPKKK

Effective number of codons (N_c)

As proposed by Wright in the year 1990, N_c measures the absolute synonymous codon usage bias [23]. It measures the total number of synonymous codons used for each amino acid. The values of N_c ranges from 20 (when only one codon is used for an amino acid), to 61 (when all the codons are used with equal frequency) [1]. It may be expressed as: (For equation-> (3) please see supplementary material). Where, F_2 is the probability that two randomly chosen codon for an amino acid, possibly encoded by two distinct codons, are identical, F_3 is the probability that three randomly chosen codons for an amino acid with three synonymous codons are identical and so on [1, 23].

Gene expression analysis

The expressivity of a gene is characterized on the basis of CAI value for a gene. It is measured as calculating the geometric means of the relative adaptiveness (w) of all the codons in a gene. The CAI value ranges from 0 (when less frequent codons are used) to 1 (when most frequent codons are used) [24]. Highly expressed genes tend to have high CAI values whereas less expressive genes possess low CAI values. CAI is

mathematically expressed as: (For equation-> (4) please see supplementary material), where, L is the number of codons in the gene and $w_{c(k)}$ is the w value for the k_{th} codon in the gene. Relative adaptiveness (w) can be calculated as: (For equation-> (5) please see supplementary material). Where X_{ij} is the number of occurrence of j^{th} codons in the reference set of highly expressed genes and X_{max} is the maximum X_{ij} for i^{th} amino acid.

Results:

Characteristic patterns of nucleotide composition

The nucleotide composition of the MAPK genes in three organisms shows a clear characteristic pattern of resemblance with a few exceptions in rice (Figure 1). In case of MAPKKK family, the overall GC% of Arabidopsis is 35.6%, which is low, revealing that the organism has high overall AT content. The overall GC content in rice and soybean gene sequences is 43.99% and 53.10% respectively. This result suggests that there exists a balance between the AT and GC content in the soybean genes, while in rice the overall AT content slightly exceeds the overall GC content for MAPK genes. From the GC₃% in all the three gene families across three species, it is evident that the

GC% is markedly suppressed in *Arabidopsis* and soybean at the third synonymous codon position. The rice genes show an overall consistent pattern from the other two species in respect of the GC₃%. The overall GC₃ content in all the gene families of rice is slightly lower (40.02%-45.39%) as against the AT₃ content (54.61%-59.08%). Thus the comparative study of the MAPKKK, MAPKK and the MAPK family genes show that *Arabidopsis* and soybean have resemblance in their pattern of nucleotide usage whereas rice genes deviate from the other two species (Table 1).

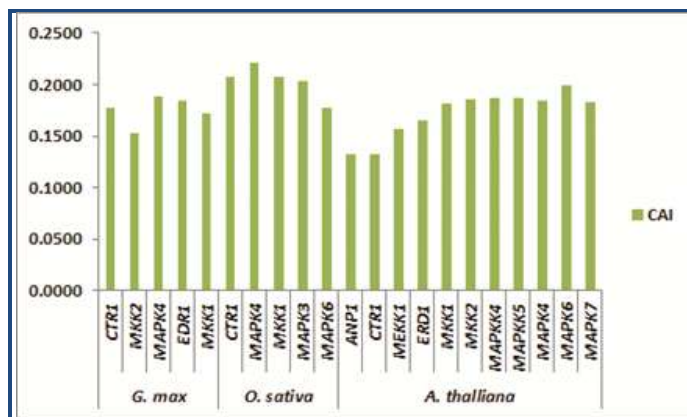


Figure 3: Frequency of estimated CAI values of different genes in *Arabidopsis*, rice and soybean

Synonymous codon usage pattern

Codons with RSCU values greater than 1 are considered to have positive codon usage bias [22]. The most preferred codon for each of 18 amino acids bearing the highest RSCU values are shown in Table 2 (see supplementary material). Inspection of the overall RSCU values reveal that the codons TTT, GTT, AAT, and GAT coding for phenylalanine, valine, asparagine and aspartic acid respectively, have got the highest preference in all the three organisms. Besides this, the comparative study between the *Arabidopsis* and soybean (*G. max*) reveals significant resemblance in the preference of codons; 11 out of 18 amino acids show the same preferred codon. But rice does not show any significant trend: only two amino acids resemble with soybean, and three with *Arabidopsis* for preferred codon. The resemblance between *Arabidopsis* and soybean is further evident from the codon preference in the MAPKK family; 12 out of 18 amino acids show the same preference of codons. Only one amino acid i.e. valine is encoded by the most preferred codon GTT in three MAPK gene families across all the species. The MAPKKK family has shown the extreme resemblance for the most preferred codons between *Arabidopsis* and soybean; 15 out of 18 amino acids have the same preferred codon. But in rice the most preferred codon for most amino acids differs from *Arabidopsis* and soybean. The comparison of RSCU values for most preferred codon for each amino acid in three organisms is shown in Figure 2 for each gene family. The third codon position of all the preferred codons predominantly possesses the nucleotide T followed by A in all the three gene families indicating that this preference could be due to translational selection or mutation bias.

Expressivity of genes

Codon adaptation index (CAI), as proposed by Sharp and Li in the year 1987, is a measure to predict the expressivity of genes

[24]. The CAI value of each gene belonging to different families was found to be very low, ranging from 0.147 to 0.208 (Table 1), which indicates that the genes are not possibly optimized for high expression (Figure 3). This could be due to the fact that stress induces the MAPK gene expression and that MAPK genes are not house-keeping genes by nature. Hence their CAI values are usually low indicating low expression under non-stress environment.

Biasness in codon usage

The N_c values of all genes are towards higher side, ranging from 50.83 to 56.55. The higher values of N_c signify that the gene sequences show low biasness in the codon usage profile.

Discussion:

The comparative study of the codon usage for MAPK genes in three species i.e. *Arabidopsis*, soybean and rice have shown that the first two organisms share a high degree of similarity between them. The MAPK genes of these two organisms have shown very high resemblance in codon preference in the coding sequences. The gene family wise comparison across the three organisms revealed that *Arabidopsis* and soybean are compositionally alike. In contrast, rice showed a somewhat different nucleotide composition from the other two species. This could be due to the fact that *Arabidopsis* and soybean are dicots unlike rice which is a monocot. *Arabidopsis* and soybean genes are overall AT-rich as well as at the third synonymous codon position. The rice MAPK sequences, on the other hand, are overall GC-rich and at the third synonymous position. Guo and his co-workers (2007) also reported a high GC content in rice genes [25]. The results suggest that the codon usage pattern in all the MAPK genes in the three species used in the study might be influenced by their base compositional properties. The general trend of the N_c values in the three species is consistently higher side for all the genes indicating a low bias in the codon usage pattern in these genes.

The overall perusal of the RSCU values for the three gene families revealed a high similarity between the codon usage in *Arabidopsis* and *G. max*. Majority of the preferred codons in the genes of these two species are T-redundant at the silent codon position. This may be due to the high prevalence of AT content in these genes. However, in case of *O. sativa*, the preference for the nucleotides at the third synonymous codon position is balanced in MAPK and MAPKK genes, which is evident from the nearly equal distribution of overall GC and AT content in these two gene families. In MAPKKK gene family of *O. sativa*, however, the preferential codons showed the increased tendency of using T at the third position. Valine is encoded by the most preferred codon GTT in three species. Highly expressed genes generally show a tendency of using a limited number of codons which they use preferentially [24]. The level of gene expression as estimated by the CAI values has shown that the MAPK genes are not possibly highly expressed. Since stress induces the expression of MAPK genes, these genes might be evolutionarily so organized as to give low expression under non-stress or normal conditions. CAI values of all the genes have shown a close proximity to 0, suggesting that they have less expressivity [24]. Translational selection might have played a role in this context, rendering the MAPK genes to be optimized for low expression under non-stress environment.

Conclusion:

This work is the first attempt to gain insight into the codon usage profiles of stress induced MAPK genes across three plant species. Mutational pressure and natural selection have been projected as the main impetus behind the codon usage bias in various organism ranging from small prokaryotes to large plants and animals [21, 26, 27]. In the present study, the results indicate that apart from the mutational pressure, translational selection might be playing a pivotal role in order to make these genes optimized for translating efficiently. MAPK families comprise of a huge number of genes interacting with each other in order to combat different environmental stresses. Based on the type of stress and the species involved, the codon usage as well as the expression of genes may vary. Thus, it is necessary to carry out further detailed analysis of the codon usage pattern in MAPK and the associated genes involved in the cascade of actions under biotic and abiotic stress environments in different species.

Acknowledgement:

The authors are grateful to Assam University, Silchar-788011, Assam, India for providing the research facility.

Disclosure:

The authors do not have any competing interest. No financial aid from any funding agency was obtained for undertaking the present study.

References:

- [1] Zhou T *et al.* *Biosystems*. 2005 **81**: 77 [PMID: 15917130]
- [2] Grantham R *et al.* *Nucleic Acids Res.* 1980 **8**: r49 [PMID: 6986610]
- [3] Marin A *et al.* *Nucleic Acids Res.* 1989 **17**: 6181 [PMID: 2570402]
- [4] Warnecke T & Hurst LD, *Mol Biol Evol.* 2007 **24**: 2755 [PMID: 17905999]
- [5] Stoletzki N & Eyre-Walker A, *Mol Biol Evol.* 2007 **24**: 374 [PMID: 17101719]
- [6] Qin F *et al.* *Plant Cell Physiol.* 2011 **52**: 1569 [PMID: 21828105]
- [7] Tuteja N, *Plant Signal Behav.* 2007 **2**: 135 [PMID: 19516981]
- [8] Cai G *et al.* *Plant Sci.* 2014 **214**: 57 [PMID: 24268164]
- [9] Hamel LP *et al.* *Trends Plant Sci.* 2006 **11**: 192 [PMID: 16537113]
- [10] Kumar K *et al.* *Plant Physiol Biochem.* 2008 **46**: 891 [PMID: 18619847]
- [11] Neupane *et al.* *Evol Bioinform Online.* 2013 **9**: 363 [PMID: 24137047]
- [12] Einset J *et al.* *Plant Signal Behav.* 2007 **2**: 365 [PMID: 19704600]
- [13] Cheong YH *et al.* *Plant Physiol.* 2003 **132**: 1961 [PMID: 12913152]
- [14] He C *et al.* *Mol Plant Microbe Interact.* 1999 **12**: 1064 [PMID: 10624015]
- [15] Teige M *et al.* *Mol Cell.* 2004 **15**: 141 [PMID: 15225555]
- [16] Wen JQ, *Plant Physiol.* 2002 **129**: 1880 [PMID: 12177502]
- [17] Ning J *et al.* *Plant Physiol.* 2010 **152**: 876 [PMID: 20007444]
- [18] Lee SK *et al.* *J Biosci.* 2011 **36**: 139 [PMID: 21451255]
- [19] Liu JZ *et al.* *Plant Physiol.* 2011 **157**: 1363 [PMID: 21878550]
- [20] Nakamura Y *et al.* *Nucleic Acids Res.* 1998 **26**: 334 [PMID: 9399868]
- [21] Qiu S *et al.* *Mol Biol Evol.* 2011 **28**: 771 [PMID: 20855431]
- [22] Bulmer M, *Genetics* 1991 **129**: 897 [PMID: 1752426]
- [23] Wright F, *Gene* 1990 **87**: 23 [PMID: 2110097]
- [24] Sharp PM & Li WH, *Nucleic Acids Res.* 1987 **15**: 1281 [PMID: 3547335]
- [25] Guo X *et al.* *FEBS Lett.* 2007 **581**: 1015 [PMID: 17306258]
- [26] Sharp PM *et al.* *Philos Trans R Soc Lond B Biol Sci.* 2010 **365**: 1203 [PMID: 20308095]
- [27] Liu, Q *PLoS One.* 2012 **7**: e48295 [PMID: 23144748]

Edited by P Kanguane

Citation: Singha *et al.* *Bioinformation* 10(7): 436-442 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Methodology:

Eq.No	Equation	Explanation
(1)	$RSCU_{ij} = \frac{x_{ij}}{\frac{1}{n_i} \sum_{i=1}^{n_i} x_{ij}}$	where, x_i is the number of frequency of j^{th} codon for i^{th} amino acid and n_i being the number of alternative synonymous codons available for the i^{th} amino acid.
(2)	$GC = \frac{G + C}{A + T + G + C} \times 100$	GC content can be measured partially at the 1 st , 2 nd and 3 rd synonymous codon
(3)	$N_C = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$	F_2 is the probability that two randomly chosen codon for an amino acid, possibly encoded by two distinct codons, are identical, F_3 is the probability that three randomly chosen codons for an amino acid with three synonymous codons are identical and so on
(4)	$CAI = \exp \frac{1}{L} \sum_{k=1}^L \ln w_{c(k)}$	L is the number of codons in the gene and $w_{c(k)}$ is the w value for the k_{th} codon in the gene.
(5)	$w_j = X_{ij} / X_{imax}$	Where X_{ij} is the number of occurrence of j^{th} codons in the reference set of highly expressed genes and X_{max} is the maximum X_{ij} for i^{th} amino acid.

Table 1: Nucleotide composition of the MAPK gene families in *A. thaliana*, *G. max* and *O. sativa*

Gene family	Organism	A%	T%	G%	C%	A3%	T3%	G3%	C3%
MAPKKK	<i>A. thaliana</i>	27.19	30.56	23.47	18.95	26.66	26.66	27.05	19.64
	<i>G.max</i>	28.09	27.92	23.71	20.29	27.03	28.79	23.21	20.97
	<i>O. sativa</i>	22.68	24.18	26.48	26.66	18.24	26.14	25.23	30.39
Mean		25.98	27.55	24.55	21.96	23.98	27.20	25.16	23.66
SD		2.366	2.618	1.368	3.364	4.060	1.148	1.570	4.788
MAPKK	<i>A. thaliana</i>	27.31	31.12	21.60	19.95	25.95	34.05	20.23	19.77
	<i>G.max</i>	29.89	33.55	19.07	17.49	29.21	31.62	19.81	19.37
	<i>O. sativa</i>	26.79	23.87	25.69	23.65	23.65	23.01	26.86	26.48
Mean		28.00	29.51	22.12	20.36	26.27	29.56	22.30	21.87
SD		1.354	4.115	2.729	2.532	2.282	4.735	3.229	3.264
MAPK	<i>A. thaliana</i>	27.30	31.79	22.14	18.79	26.43	31.56	22.25	19.76
	<i>G.max</i>	27.47	31.26	22.01	19.22	28.80	33.15	18.04	20.01
	<i>O. sativa</i>	26.48	26.37	22.80	24.35	23.08	24.90	25.79	26.22
Mean		27.08	29.81	22.32	20.79	26.10	29.87	22.03	22.00
SD		0.433	2.437	0.345	2.525	2.344	3.575	3.171	2.987
Gene family	Organism	GC%	AT%	GC3%	AT3%	Nc	CAI		
MAPKKK	<i>A. thaliana</i>	35.60	64.40	36.87	36.87	56.27	0.1466		
	<i>G.max</i>	43.99	56.01	34.12	34.12	54.97	0.1807		
	<i>O. sativa</i>	53.10	46.90	45.39	45.39	56.55	0.2080		
Mean		44.23	55.77	38.79	38.79	55.93	0.1785		
SD		7.145	7.145	4.800	4.80	0.688			
MAPKK	<i>A. thaliana</i>	34.31	65.69	31.30	31.30	52.61	0.1853		
	<i>G.max</i>	36.57	63.43	29.96	29.96	51.56	0.1627		
	<i>O. sativa</i>	49.38	50.63	40.89	40.89	54.09	0.2075		
Mean		40.08	59.92	34.05	34.05	52.75	0.1852		
SD		6.635	6.635	4.868	4.868	1.036			
MAPK	<i>A. thaliana</i>	33.46	66.54	32.19	32.19	55.29	0.1912		
	<i>G.max</i>	32.92	67.08	29.98	29.98	53.42	0.1825		
	<i>O. sativa</i>	47.14	52.86	40.02	40.02	50.83	0.2005		
Mean		37.84	62.16	34.07	34.07	53.18	0.1914		
SD		6.579	6.579	4.309	4.309	1.826			

*SD stands for standard deviation

Table 2: RSCU values of the most preferred codons for MAPK genes in *Arabidopsis*, soybean and rice

MAPKKK						
<i>Amino acids</i>	<i>Codons</i>	<i>G. max</i>	<i>Codons</i>	<i>A. thaliana</i>	<i>Codons</i>	<i>O. sativa</i>
<i>Phe</i>	TTT	1.08	TTT	1.13	TTT	1.00
<i>Leu</i>	TTG	1.50	TTG	1.53	CTC	2.00
<i>Ile</i>	ATT	1.18	ATT	1.04	ATT	1.74
<i>Val</i>	GTT	1.34	GTT	1.55	GTT	1.55
<i>Ser</i>	TCT	1.23	TCT	1.45	TCC	1.32
<i>Pro</i>	CCA	1.38	CCA	1.30	CCA	1.74
<i>Thr</i>	ACA	1.37	ACA	1.47	ACC	1.38
<i>Ala</i>	GCT	1.40	GCT	1.39	GCG	1.21
<i>Tyr</i>	TAT	1.25	TAT	1.26	TAC	1.05
<i>His</i>	CAT	1.20	CAT	1.16	CAC	1.22
<i>Gln</i>	CAA	1.18	CAA	1.09	CAG	1.08
<i>Asn</i>	AAT	1.15	AAT	1.19	AAT	1.38
<i>Lys</i>	AAA	1.06	AAG	1.03	AAG	1.27
<i>Asp</i>	GAT	1.38	GAT	1.27	GAC	1.02
<i>Glu</i>	GAA	1.18	GAG	1.11	GAG	1.17
<i>Cys</i>	TGT	1.07	TGT	1.16	TGC	1.54
<i>Arg</i>	AGA	1.93	AGA	2.09	AGG	1.67
<i>Gly</i>	GGA	1.37	GGT	1.40	GGC	1.87
MAPKK						
<i>Amino acid</i>	<i>Codons</i>	<i>G. max</i>	<i>Codons</i>	<i>A. thaliana</i>	<i>Codons</i>	<i>O. sativa</i>
<i>Phe</i>	TTT	1.27	TTT	1.25	TTC	1.04
<i>Leu</i>	TTG	1.61	CTT	1.54	CTG	1.35
<i>Ile</i>	ATC	1.08	ATT	1.18	ATC	1.33
<i>Val</i>	GTT	1.31	GTT	1.51	GTT	1.42
<i>Ser</i>	AGT	1.52	TCT	1.58	TCC	1.28
<i>Pro</i>	CCA	1.71	CCT	1.56	CCA	1.08
<i>Thr</i>	ACA	1.58	ACT	1.20	ACA	1.45
<i>Ala</i>	GCT	1.68	GCT	1.72	GCA	1.42
<i>Tyr</i>	TAT	1.41	TAT	1.20	TAC	1.15
<i>His</i>	CAT	1.22	CAT	1.26	CAT	1.01
<i>Gln</i>	CAA	1.27	CAA	1.27	CAG	1.08
<i>Asn</i>	AAT	1.22	AAT	1.04	AAC	1.06
<i>Lys</i>	AAA	1.22	AAA	1.14	AAG	1.12
<i>Asp</i>	GAT	1.38	GAT	1.43	GAT	1.04
<i>Glu</i>	GAA	1.23	GAG	1.04	GAG	1.17
<i>Cys</i>	TGT	1.23	TGT	1.55	TGT	1.02
<i>Arg</i>	AGA	2.92	AGA	1.93	CGC	1.26
<i>Gly</i>	GGA	1.38	GGA	1.79	GGT	1.18
MAPK						
<i>Amino acid</i>	<i>Codons</i>	<i>G. max</i>	<i>Codons</i>	<i>A. thaliana</i>	<i>Codons</i>	<i>O. sativa</i>
<i>Phe</i>	TTT	1.36	TTT	1.23	TTT	1.04
<i>Leu</i>	CTT	2.08	TTG	1.38	CTG	1.28
<i>Ile</i>	ATT	1.30	ATC	1.09	ATT	1.18
<i>Val</i>	GTT	2.21	GTT	1.89	GTT	1.30
<i>Ser</i>	TCT	2.48	TCT	1.70	TCC	1.51
<i>Pro</i>	CCC	1.39	CCA	1.60	CCA	1.45
<i>Thr</i>	ACA	1.33	ACT	1.40	ACT	1.13
<i>Ala</i>	GCA	1.69	GCT	1.74	GCA	1.25
<i>Tyr</i>	TAT	1.04	TAT	1.34	TAC	1.11
<i>His</i>	CAT	1.29	CAT	1.27	CAC	1.01
<i>Gln</i>	CAA	1.22	CAA	1.16	CAG	1.06
<i>Asn</i>	AAT	1.36	AAT	1.07	AAT	1.07
<i>Lys</i>	AAA	1.09	AAA	1.05	AAG	1.21
<i>Asp</i>	GAT	1.50	GAT	1.41	GAT	1.17
<i>Glu</i>	GAA	1.19	GAG	1.02	GAG	1.14
<i>Cys</i>	TGT	1.23	TGT	1.21	TGC	1.28
<i>Arg</i>	AGA	2.00	AGA	2.28	CGC	1.21
<i>Gly</i>	GGT	1.64	GGA	1.41	GGC	1.40