academicJournals

Vol. 15(8), pp. 236-251, 24 February, 2016
DOI: 10.5897/AJB2015.15111
Article Number: 535FEC657329
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Codon usage bias analysis for the coding sequences of *Camellia sinensis* and *Brassica campestris*

Prosenjit Paul and Supriyo Chakraborty*

Department of Biotechnology, Assam University, Silchar-788011, Assam, India.

Received 20 November, 2015; Accepted 1 February, 2016

Codon usage bias plays an important role in the regulation of gene expression. A couple of measures are widely used to quantify the codon usage in genes. On the other hand, no quantitative endeavour has been made to compare the pattern of codon usage diversity within and between different genes of *Camellia sinensis* and *Brassica campestris.* Nucleotide composition and its relationship with codon usage bias were analyzed. Additionally, the rare codons were identified by computing the recurrence of event of all codons in coding sequences of *C. sinensis* and *B. campestris.* The host cell, *Escherichia coli* used universally, failed to express smoothly many eukaryotic genes. For this, the authors prognosticated the codons showing the highest and the lowest expressivity of the coding sequences of *C. sinensis* and *B. campestris,* in *E. coli K12* strain to improve the expression level of the genes.

Key words: Codon usage bias, gene expressivity, codon usage pattern, residual value, synonymous codon.

INTRODUCTION

Gene expression is a fundamental cellular process by which proteins are synthesized in a cell based on the information encoded in the genes. Most amino acids can be encoded by more than one codon; such codons are depicted as being synonymous, and mostly vary by one nucleotide in the third position. Synonymous codons are not used uniformly, varies across species and within genome in the same species, the phenomenon is called codon usage bias (CUB) (Akashi, 1994; Behura and Severson, 2013). Molecular evolutionary investigations on codon bias suggest that recurrence of codon use changes between genes from the same genome and also between genomes (Hooper and Berg, 2000). Highly expressed genes are more biased in terms of their codon usage as compared to low expressed genes, and provide differential efficiency as well as accuracy in the translation of genes (Rocha, 2004; Hershberg and Petrov, 2008). The selection associated with translational efficiency/ accuracy is often termed as 'translation selection'. During the last two decades, numerous lines of evidence suggested that codon usage bias is driven by selection, particularly for species of fungi (Bennetzen and Hall, 1982; Ikemura, 1985), bacteria (Ikemura, 1981; Sharp and Li, 1987a) and insects (Akashi, 1997; Moriyama and Powell, 1997).

Soon, after the discovery of whole genome sequencing

*Corresponding author. E-mail: supriyoch2008@gmail.com. Tel: +919435700831. Fax: 03842-270802.

technology, codon usage bias was analysed for numerous organisms (Plotkin and Kudla, 2011). Numerous factors have been shown to influence codon usage bias: (i) genomic composition (Supek and Vlahovicek, 2005); (ii) selective forces (Ikemura, 1985); and (iii) horizontal gene transfer, with transferred genes retaining the codon frequencies of their former host (Lawrence and Ochman, 1998). Connections have also been demonstrated between codon usage and several factors namely: (a) gene length (Lawrence and Ochman, 1998); (b) gene translation initiation signal (Ma et al., 2002); (c) expression level (Gouy and Gautier, 1982; Sharp and Li, 1986; Sharp et al., 1986; Sharp and Li, 1987b); (d) protein amino acid composition (Lobry and Gautier, 1994); (e) protein structure (D'Onofrio et al., 2002); (f) tRNA abundance (Ikemura, 1981, 1982); (g) mutation frequency and patterns (Sueoka, 1999); and (h) GC composition (Sueoka and Kawanishi, 2000). Besides, the relative impact of each of these factors varies from genome to genome, and from gene to gene. Despite the fact that there is still no final result on the formation mechanism, codon bias has been widely used to predict the exogenous and endogenous gene expression level (Lee et al., 2007; Yu et al., 2007; Zheng et al., 2007), identify horizontally transferred genes (Goldman et al., 2007), evolutionary relationship (Ram et al., 2007), and confirm the coding sequences. Researchers proposed that in some prokaryotes, many indices exhibit a positive correlation with the gene expression level, such as codon adaptation index (CAI) (Sharp and Li, 1987b), codon bias index (CBI) (Bennetzen and Hall, 1982), and frequency of optimal codons (Kanaya et al., 2001). Then again, in a few eukaryotes there is no confirmation to bolster this, particularly for higher eukaryotes where, the correlation between codon bias and expression level is extremely weak (Murray et al., 1989; Kanaya et al., 2001).

In this study, we investigated the codon usage bias (CUB) for *Camellia sinensis* and *Brassica campestris* by analyzing the codon adaptation index (CAI), relative codon usage bias (RCBS), effective number of codons (ENc), synonymous codon usage order (SCUO), and GC/AT content at each codon position. The purpose of this study was to perform a comparative analysis of codon usage bias and codon contexts pattern among the coding sequences (cds) of *C. sinensis* and *B. campestris.*

Escherichia coli cells were, as often as possible, utilized as host cells as a part of the investigation of exogenous protein expression. Many eukaryotic genes cannot be efficiently expressed in a prokaryote like *E. coli*. One of the effective methods for improving the expression level of a eukaryotic gene in a prokaryote is to replace the usage of 'rare codons' with synonymous codons showing highest expressivity in prokaryotes. While replacing the rare codons, the stability of genes at genomic or transcriptional level should be taken into consideration. Here, a novel computational method to identify the codons of *C. sinensis* and *B. campestris* was introduced exhibiting the highest and lowest expressivity in *E. coli* k12 strain.

MATERIALS AND METHODS

The complete coding sequences of the thirty genes from *C. sinensis* and forty seven genes from *B. campestris* were retrieved from the National Centre of Biotechnology (NCBI) nucleotide database accessible from the website www.ncbi.nlm.nih.gov. Each of those cds were devoid of any unknown base (N), intercalary stop codon and possessed the start and stop codons.

Relative codon usage bias and codon adaptation index were used to study the overall codon usage variation among the genes. RCBS is the overall score of a gene indicating the influence of RCB of each codon in a gene. RCB reflects the level of gene expression. RCBS was calculated as by Roymondal et al. (2009). Gene expressivity was again measured by calculating the codon adaptation index as per Sharp and Li (1986). It essentially measures the distance from a given gene to a reference gene with respect to their amino-acid codon usages. CAI defines translational optimal codons as those that appear frequently in highly expressed genes that is:

$$CAI(L(g)) = \exp\left(\frac{1\sum_{l=1}^{L} \log w_{c(l)}}{L}\right) = \left(\prod_{l=1}^{L} w_{c(l)}\right)^{1/L}$$

Where, L is the length of gene g and w_c (l) is the relative adaptiveness of the codon c in the reference genes (not g).

Certain codons will appear multiple times in the gene. Hence we can rewrite the equation to sum up codons rather than length, and use counts rather than frequencies. This makes the dependence on the actual gene more clear. The more usual form is:

$$CAI(o(g)) = \exp\left(\frac{1}{o_{tot}} \sum_{c \in C} o_c \log w_c\right) = \left(\prod_{c \in C} o_c \log w_c\right)^{\frac{1}{o_{tot}}}$$

The effective number of codons (ENc) is the total number of different codons used in a sequence. The values of ENc for standard genetic code range from 20 (where only one codon is used per amino acid) to 61 (where all possible synonymous codons are used with equal frequency). ENc measures bias toward the use of a smaller subset of codons, away from equal use of synonymous codons. For example, as mentioned above, highly expressed genes tend to use fewer codons due to selection. The underlying idea of ENc is similar to the concept of zygosity from population genetics, which refers to the similarity for a gene from two organisms. ENc value was calculated as per Wright (1990). The measure of codon usage, synonymous codon usage orders (SCUO) of genes was computed as per Wan et al. (2004). GC3s is the frequency of (G+C) and A3s, T3s, G3s, and C3s are the distributions of A, T, G and C bases at the third codon position (Gupta and Ghosh, 2001). A series of scripts (programs) were writen in Perl language and run in Windows for analysis. These programs were used to estimate the above mentioned genetic parameters.

The correlations between all the above mentioned parameters were measured with the gene expressivity to find out the genetic factors playing major role in the genes of *C. sinensis* and *B. campestris.* All codon quantifications were performed using the Anaconda software (Moura et al. 2007). The residual values of each codon pair were also quantified from the coding sequences of each plant species by the Anaconda program. The occurrence frequency of each codon for a particular amino acid was also calculated and compared with their expressivity values to identify

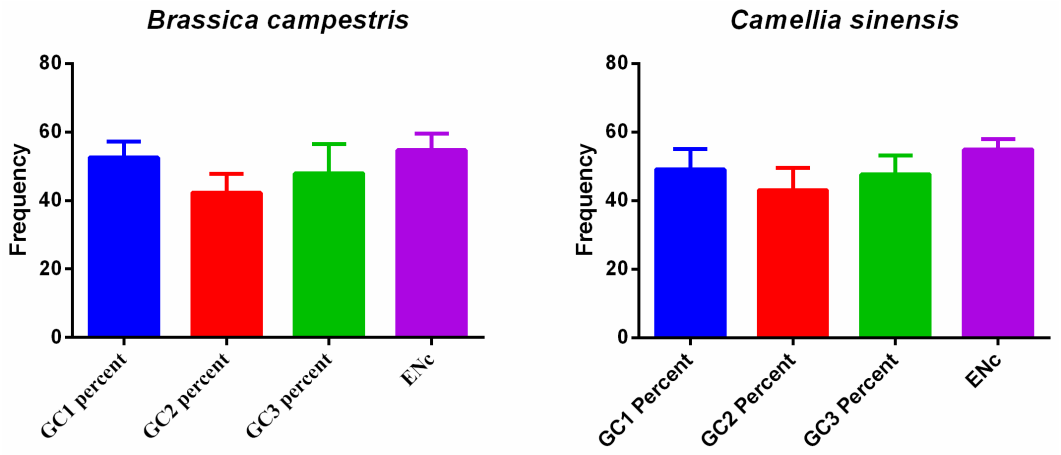

Figure 1. Effective number of codon (ENc) distribution for the genes of *B. campestris* and *C. sinensis.* GC% at third codon position for *C. sinensis* and for first codon position for *B. campestris* showed strong correlation (0.3, 0.4) respectively with the ENc among all the codon positions.

the codons playing a prominent role in determining the level of gene expression.

RESULTS AND DISCUSSION

The present study was carried out to assess the codon usage pattern and gene expressivity for the genes of *C. sinensis* and *B. campestris*. In numerous microscopic organisms, intragenomic diversity in codon usage among genes has been reported. The genes selected for the present study from the two plants with their accession numbers together with the overall AT and GC%, RCBS, CAI, ENc, SCUO, GC1, GC2 and GC3 are given in the supplementary file. It was found that the codons of *C. sinensis* and *B. campestris* are rich in A and/or T. Yet, on account of *Homo sapiens*, it has been shown that the codons ending in G and/or C are dominating in the whole coding region.

Due to the difference in mutational bias, the GC percentage among different species varies to a great extent, even for the species within the same order. To determine if GC bias among *C. sinensis* and *B. campestris* has an association with codon bias, the non-directional codon bias measure effective number of codons (ENc) was resorted to. The effective number of codons used by a gene and GC% at the three different synonymous codon positions (GC1s, GC2s and GC3s) are used to study the codon usage variation among the genes of *B. campestris* and *C. sinensis* (Figure 1). To quantify the level of diversity in the synonymous codon usage among all the selected cds within the genome of *B. campestris* and *C. sinensis*, the mean distance between the pairs of cds was estimated. The mean distance was found to be 0.07 with a median of 0.06 for *C. sinensis* and a mean distance of 0.09 with the median of 0.07 for the cds sequence of *B. campestris*. When focusing on the previously studied genomes (Lafay et al., 2000; Grocock and Sharp, 2002; Wu et al., 2005), the mean values for *Bacillus subtilis* 168 (0.60), *E. coli* K12 MG1655 (0.47), *Helicobacter pylori* 26695 (0.38), and *Haemophilus influenzae* Rd KW20 (0.37) indicated that the mean values varied widely among species. ENc is a widely accepted measure of codon usage bias that quantifies the degree of deviation from equal use of synonymous codons. It has been suggested that ENc may be dependent on the strength of the codon bias discrepancy (Fuglsang, 2004). The coefficient of determination (denoted as R^2) indicates how well the data points fit a straight line or curve. From the analysis, it is apparent that the coefficient of determination is 0.37 and 0.15 for the genes of *C. sinensis* and *B. campestris*, respectively (Figure 2). This reveals that 37% of the variation in expressivity for the cds of *C. sinensis* and 15% for the cds of *B. campestris* could be explained by the ENc. The remaining percentage of the variation in expressivity could be attributed to unknown factors, that is, genetic variation and/or other external factors.

Synonymous codon usage orders (SCUO) of genes of each species were further analyzed. SCUO is a relatively easier approach as compared to RSCU and is considered as more robust for comparative analysis of codon usage. The SCUO analysis shows that a majority of the genes selected for the present study are associated with high codon usage bias (43% cds in *C. sinensis* and 68% in *B. campestris* have SCUO≥0.5). This outcome proposes that these genes are associated with specific functions such as translational processes, ribosomes (mostly ribosomal protein genes), intracellular activities, transport, oxidation-reduction process and others (Supplementary Tables 1 and 2).

The Anaconda software was used to determine the adjusted residual values for association of each codon pair in genome-wide manner for the two plant species.

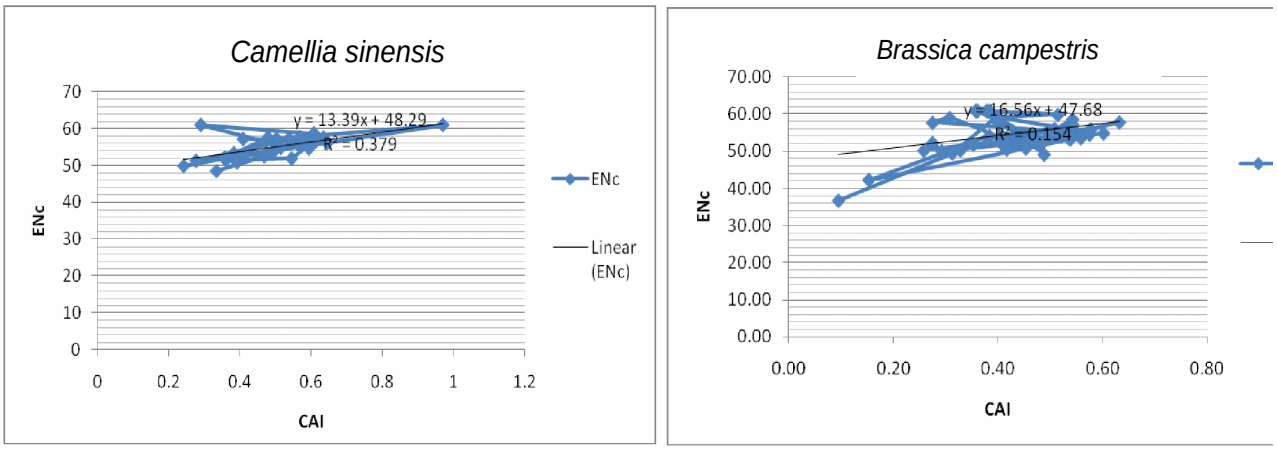

Figure 2. ENc values plotted against the CAI for the cds of *Camellia sinensis* and *Brassica campestris*. The coefficient of determination (denoted as R^2) is 0.37 and 0.15 for the genes of *Camellia sinensis* and *Brassica campestris* respectively suggesting that 37% of the variation in expressivity for the cds of *Camellia sinensis* and 15% for the cds of *Brassica campestris* could be explained by the ENc.

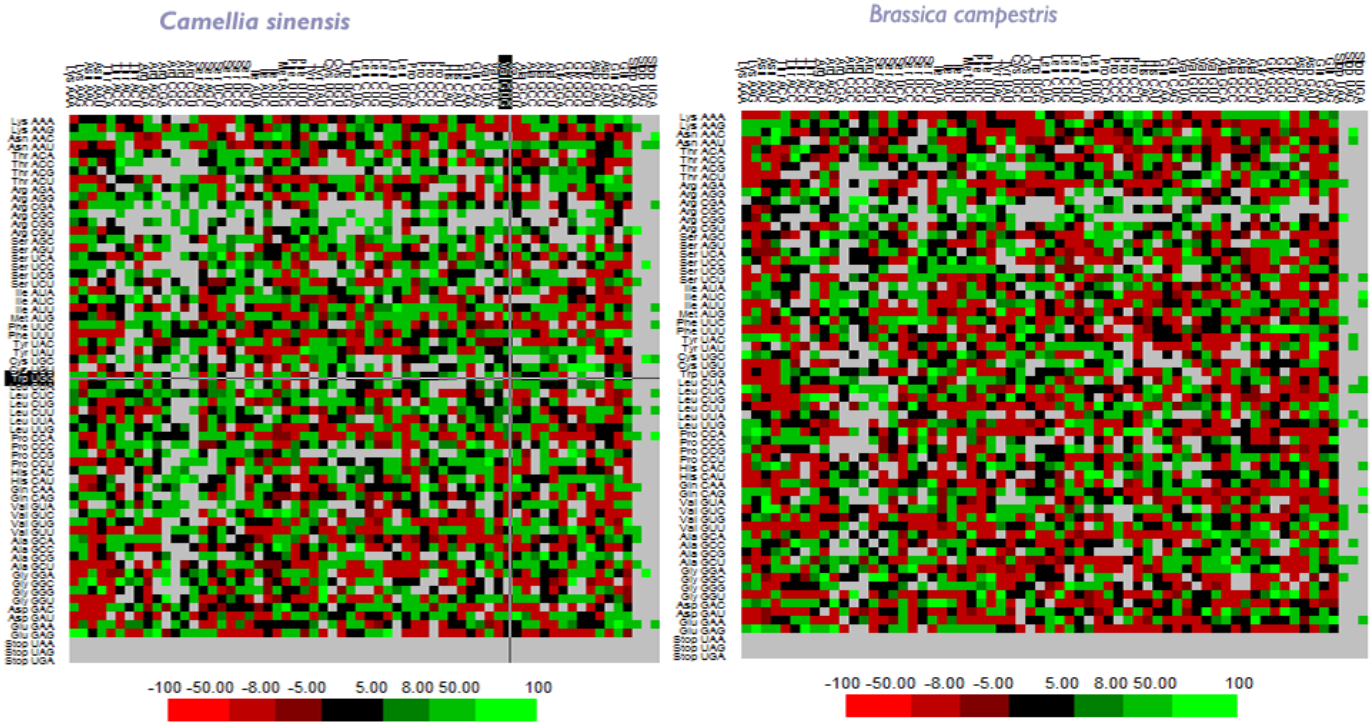

Figure 3. Patterns of codon context variation in *C. sinensis* and *B. campestris.* The green colour represents the highest number of the contexts and red colour represents the lowest number of contexts. The 59 codons are in rows and the 39 codons in columns. The colour intensity corresponds to the residual value present in each cell of the contingency table.

The residual values signify the Chi-square test association between the two codons of each context (Moura et al., 2007). Furthermore, based on the average cluster patterns of adjusted residual values of codon pair frequencies among the *C. sinensis* and *B. campestris*, it was found that specific contexts were represented more often than other contexts. The cluster patterns revealed distinctions as well as commonalities of codon context variations between *C. sinensis* and *B. campestris*. The codon contexts are localized diagonally from left top to right bottom. Being in the diagonal positions, they represent contexts of the same triplet sequences suggesting that these contexts (homogenous codon contexts) are generally frequent in these plants (Figure 3). The cluster pattern is based on the average matrix of residuals of each codon context among the species of

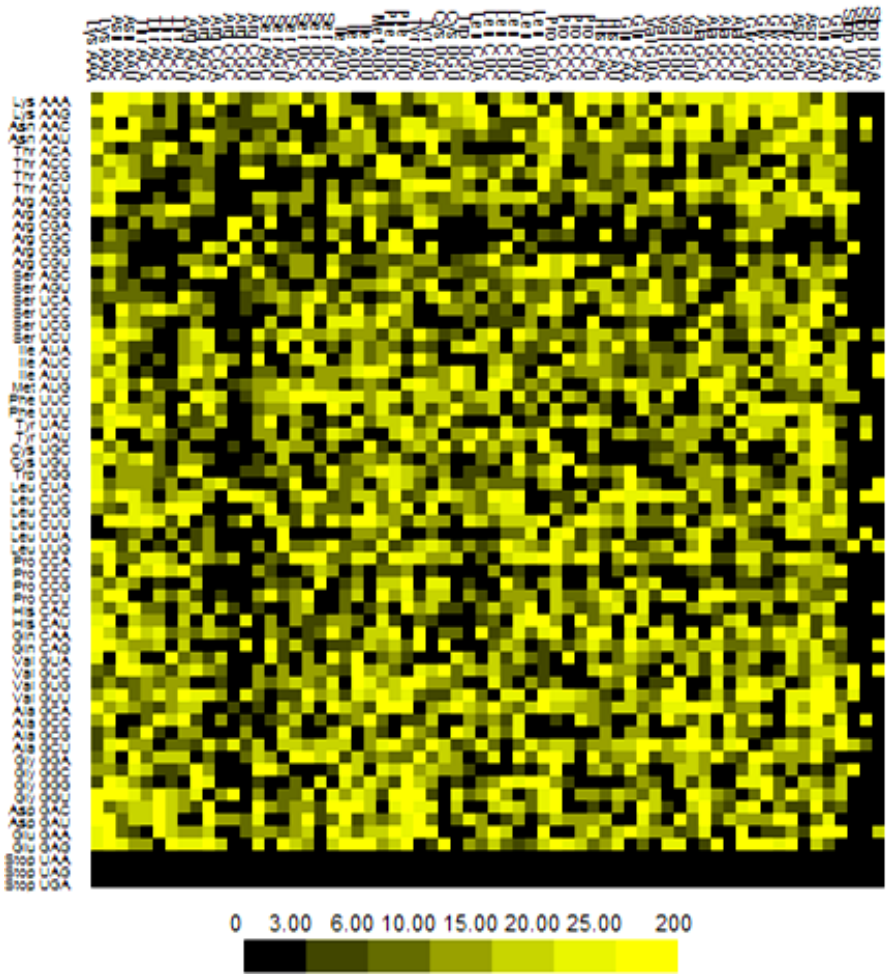

Figure 4. Figure 4: Comparison of codon context pattern between Camellia sinensis and Brassica campestris. Differential display map was obtained by calculating the module of the difference between the residuals of each map. The yellow cells indicate the highest context difference and the black cells represent pairs of codons that have similar residual values between two species.

each order. The map constructed for the two plants was, again, compared in one single display to allow detection of overall patterns of codon context. Differential Display Map (DDM) was constructed from the absolute value by subtracting both maps cell-by-cell (Figure 4).

Researchers proposed that codons which are utilized less as often as possible all through the genome are rate limiting factors of exogenous gene expression supported by experimental verification (Garcia et al., 1986; Zhang et al., 2004). In *C. sinensis* and *B. campestris*, the 'rare codon' was defined by calculating the recurrence of event of all codons (Threshold selected: 10/1000) in coding sequences (Figure 5). In the meantime, our examination demonstrated that many of these rare codon pairs contain termination codons (Table 1). Based on the hypothesis that gene expressivity and codon composition are strongly correlated, the codon adaptation index has been defined to provide an intuitively meaningful measure of the extent of the codon preference in a gene. We have estimated the CAI and RCBS for each cds as a measure of gene expressivity (supplementary material). The CAI with RCBS were compared and it was observed that both showed a similar pattern. In concurrence with different past studies (Ikemura, 1981, 1982; Moriyama and Powell, 1997), it was observed that RCBS decreased with the length of the encoded cds. Since the RCBS value depends on cds length, CAI was used as a central measure for expressivity analysis.

Gene expression studies are essential for predicting the expression potentiality of a particular gene of interest. This will help in the discovery of new coding sequences of genes for most elevated protein expression in a cell so that these man-made proteins can be synthesized and used for therapeutic drives world-wide. Along these lines, it is important to find the codons that dictate the highest and the lowest expressivity of a cds within a particular

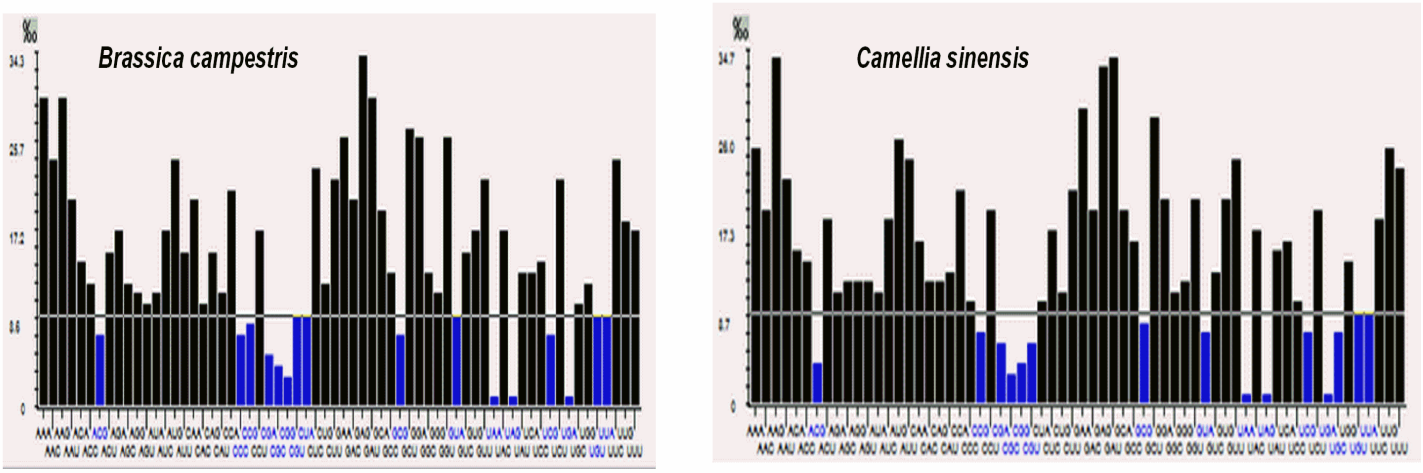

Figure 5. Rare codons for the cds of *B. Campestris* and *C. sinensis*. The 'rare codon' was defined by calculating the frequency of occurrence of all codons in coding sequences (threshold selected 10/1000).

Table 1. Rare codons for the cds of *Brassica Campestris* and *Camellia sinensis.*

	Brassica campestris	***Camellia sinensis***
Rare codons	ACG, CCG, CGA, CGG, CUA, CCC, CGC, CGU, GCG, GUA, UCG, UGU	ACG, CCG, CGA, CGG, CGC, CGU, GCG, GUA, UCG, UUA, UGU, UGC

expression system. The DDM analysis results suggested that both plants showed similar codon context pattern to some extent. For confirmation, the pattern of synonymous codons usage for both plants were compared. In support of our previous study on cereals (Chakraborty and Paul, 2015), both plants selected for the present study also maintained more or less similar pattern of synonymous codon usage (Figure 6). These result indicated that througout the evolution, both plants maintained a precise pattern of codon usage, may be due to the natural selection, mutation or any other external factors. Again, the role of each codon in terms of expressivity within the two plants were analyzed. The occurrence frequency of 59 codons (except stop codons and codons for Met & Trp) were calculated for each cds of *C. sinensis and B. campestris* and predicted their expression level in *E. coli* K12 strain. The occurrence frequency for each codon in cds was again allied with their expressivity values. Using the criterion derived from statistical analysis (positive and negative codon bias relating to the gene expression level), the codons showing the highest and lowest expressivity in *E. coli* k12 we obtained (Table 2). *E. coli* genome tRNA copy number data sets available in the genomic tRNA database (http://gtrnadb.ucsc.edu/) also support the results of highest and lowest productive codons.

To confirm the results of this analysis, we changed the original cds downloaded from the database to the highest expressive and the lowest expressive cds sequence by replacing the codons with highest and lowest expressive codons, respectively. The expressivity values for all the three sets of a cds sequence (original, highest was lowest cds) was calculated by using codonW. These results revealed that the highest as well as the lowest coding sequences significantly differed in expression level from the original cds downloaded from the NCBI database.

Conclusion

A novel method for identification of codons showing the highest and the lowest expressivity was introduced, in view of their recurrence of event. The event recurrence for every codon/cds was again allied with their expressivity values. Using the criterion derived from statistical analysis, the codons showing the highest and the lowest expressivity in *E. coli* k12 were obtained. The natural codons present in cds were replaced by the predicted codons of this study showing the lowest and the highest expressivity using a Perl program developed by the authors of this study. By comparing the expressivity values of our cds with that of original cds downloaded from NCBI, we have established that our method is a general one, not connected with the adjustments in gene length and overall nucleotide

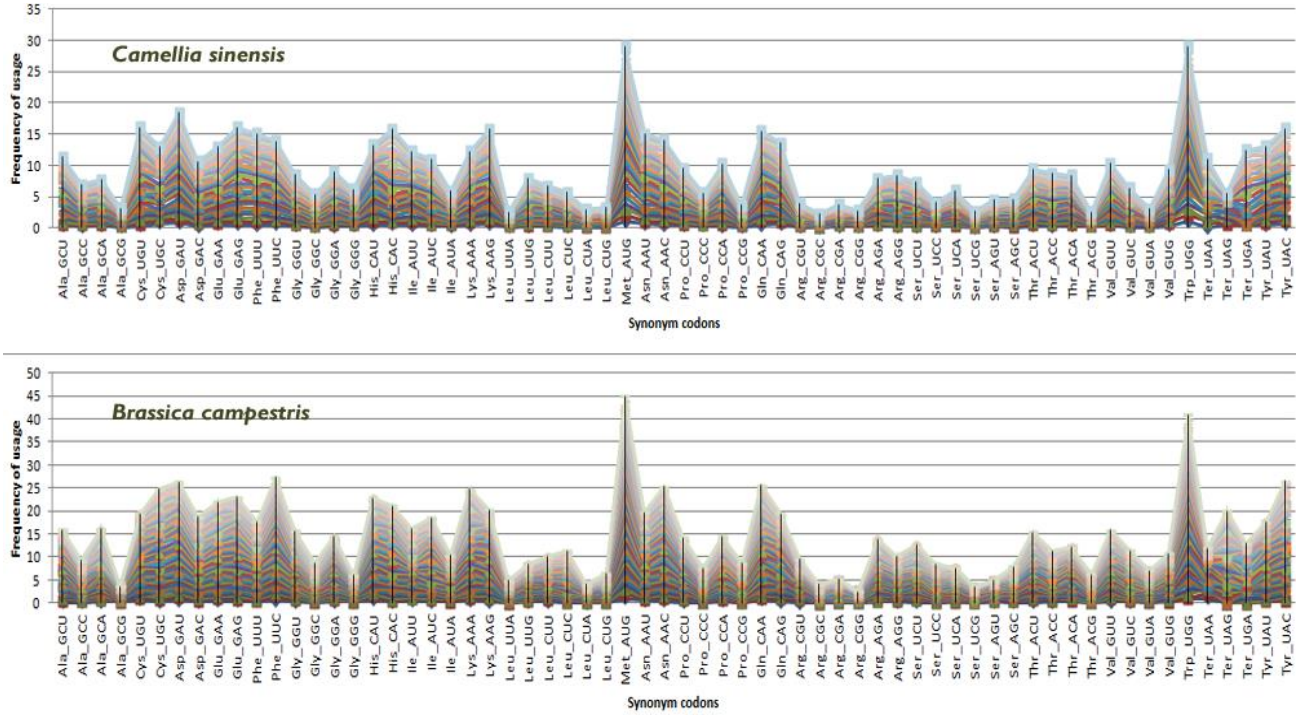

Figure 6. Comparison of the pattern of synonymous codons usage for *C. sinensis* and *B. campestris*. Synonymous codons were placed in the x-axis and their usage frequency in the y-axis. Both plants showed the almost similar pattern of synonymous codon usage with little variation in the usage frequency.

Table 2. Codons for highest and lowest expressivity for the genes of *C. sinensis* and *B. campestris*.

Amino acids	Codons showing lowest expressivity		Codons showing highest expressivity	
	Camellia sinensis	***Brassica campestris***	***Camellia sinensis***	***Brassica campestris***
Serine	TCG, TCC	TCA, TCG	TCT, AGC	AGC
Phenylalanine	TTT	TTT	TTC	TTC
Leucine	CTA, CTG	CTA	CTT, TTG	CTC, TTG
Tyrosine	TAT	TAT	TAC	TAC
Cysteine	TGC	TGT	TGT	TGC
Proline	CCA	CCA	CCC, CCT	CCC, CCT
Histidine	CAC	CAT	CAT	CAC
Glutamine	CAG	CAG	CAA	CAG
Arginine	AGG,CGA, CGT	AGA	AGA	CGT
Isoleucine	ATT, ATA	ATT, ATA	ATC	ATC
Threonine	ACG	ACG	ACA, ACC	ACC
Asparagine	AAT	AAT	AAC	AAC
Lysine	AAA	AAA	AAG	AAG
Valine	GTA, GTC	GTA	GTG	GTC
Alanine	GCG	GCG	GCT, GCC	GCC
Aspartic acid	GAC	GAC	GAT	GAT
Glutamic acid	GAA	GAA	GAG	GAG
Glycine	GGC	GGG	GGA	GGC

composition, with a little noise in measurements. To design the highest and lowest expressive cds of the genes of *C. sinensis* and *B. campestris* in *E. coli* K12 strain, the restriction sites in the bacterium were not

considered.

Availability

The coding sequences for both plants are available in the nucleotide database of NCBI. The softwares used, that is, codon W and Anaconda for the present study are freely available, downloaded from (http://codonw.sourceforge.net/) and (http://bioinformatics.ua.pt/software/anaconda/), respectively.

Conflicts of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGEMENTS

The authors are thankful to Assam University, Silchar, Assam, India for providing the necessary facilities in carrying out this research work.

REFERENCES

Akashi H (1994). Synonymous codon usage in Drosophila melanogaster: natural selection and translational accuracy. Genetics 136(3):927-935.

Akashi H (1997). Codon bias evolution in Drosophila. Population genetics of mutation-selection drift. Gene 205(1-2):269-278.

Behura SK, Severson DW (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. Biol. Rev. Camb. Philos. Soc. 88(1):49-61.

Bennetzen JL, Hall BD (1982). Codon selection in yeast. J. Biol. Chem. 257(6):3026-3031.

Chakraborty S, Paul P (2015). Guanine and Cytosine at the Second Codon Position Influence Gene Expression in Cereals. Proc. Natl. Acad. Sci. India B Biol. Sci. 1-11.

D'Onofrio G, Ghosh TC, Bernardi G (2002). The base composition of the genes is correlated with the secondary structures of the encoded proteins. Gene 300(1-2):179-187.

Fuglsang A (2004). Bioinformatic analysis of the link between gene composition and expressivity in Saccharomyces cerevisiae and Schizosaccharomyces pombe. Antonie Van Leeuwenhoek 86(2):135-147.

Garcia GM, Mar PK, Mullin DA, Walker JR, Prather NE (1986). The E. coli dnaY gene encodes an arginine transfer RNA. Cell 45(3):453-459.

Goldman B, Bhat S, Shimkets LJ (2007). Genome evolution and the emergence of fruiting body development in Myxococcus xanthus. PLoS One 2(12):e1329.

Gouy M, Gautier C (1982). Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10(22):7055-7074.

Grocock RJ, Sharp PM (2002). Synonymous codon usage in Pseudomonas aeruginosa PA01. Gene 289(1-2):131-139.

Gupta SK, Ghosh TC (2001). Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa. Gene 273(1):63-70.

Hershberg R, Petrov DA (2008). Selection on codon bias. Annu. Rev. Genet. 42:287-299.

Hooper SD, Berg OG (2000). Gradients in nucleotide and codon usage along *Escherichia coli* genes. Nucleic Acids Res. 28(18):3517-3523.

Ikemura T (1981). "Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system." J. Mol. Biol. 151(3):389-409.

Ikemura T (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158(4):573-597.

Ikemura T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2(1):13-34.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T (2001). Codon usage and tRNA genes in eukaryotes: correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. J. Mol. Evol. 53(4-5):290-298.

Lafay B, Atherton JC, Sharp PM (2000). Absence of translationally selected synonymous codon usage bias in Helicobacter pylori. Microbiology 146(4):851-860.

Lawrence JG, Ochman H (1998). Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA 95(16):9413-9417.

Lee MH, Yang SJ, Kim JW, Lee HS, Kim JW, Park KH (2007). Characterization of a thermostable cyclodextrin glucanotransferase from Pyrococcus furiosus DSM3638. Extremophiles 11(3):537-541.

Lobry JR, Gautier C (1994).Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes. Nucleic Acids Res. 22(15):3174-3180.

Ma J, Campbell A, Karlin S (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. 184(20):5733-5745.

Moriyama EN, Powell JR (1997). Codon usage bias and tRNA abundance in Drosophila. J. Mol. Evol. 45(5):514-523.

Moura G, Pinheiro M, Arrais J, Gomes AC, Carreto L, Freitas A, Oliveira JL, Santos MA (2007). Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. PLoS One 2(9):e847.

Murray EE, Lotzer J, Eberle M (1989). Codon usage in plant genes. Nucleic Acids Res. 17(2):477-498.

Plotkin JB, Kudla G (2011). Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12(1):32-42.

Ram MR, Beena G, Ragunathan P, Malathi R (2007). Analysis of structure, function, and evolutionary origin of the ob gene product-leptin. J. Biomol. Struct. Dyn. 25(2):183-188.

Rocha EP (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14(11):2279-2286.

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 PM, Li WH (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24(1-2):28-38.

Sharp PM, Li WH (1987b). The codon Adaptation Index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15(3):1281-1295.

Sharp PM, Li WH (1987a). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4(3):222-230.

Sharp PM, Tuohy TM, Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14(13):5125-5143.

Sueoka N (1999). Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C. J. Mol. Evol. 49(1):49-62.

Sueoka N, Kawanishi Y (2000). DNA G+C content of the third codon position and codon usage biases of human genes. Gene 261(1):53-62.

Supek F, Vlahovicek K (2005). Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. BMC Bioinformatics 6:182.

Wan XF, Xu D, Kleinhofs A, Zhou J (2004). Quantitative relationship between synonymous codon usage bias and GC composition across

unicellular genomes. BMC Evol. Biol. 4:19.
Wright F (1990). The 'effective number of codons' used in a gene. Gene 87(1):23-29.
Wu G, Culley DE, Zhang W (2005). Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology 151(7):2175-2187.
Yu X, Li Z, Xia X, Fang H, Zhou C, Chen H (2007). Expression and purification of ancrod, an anticoagulant drug, in Pichia pastoris. Protein Exp. Purif. 55(2):257-261.
Zhang R, Ou HY, Zhang CT (2004). DEG: a database of essential genes. Nucleic Acids Res. 32(Database issue):D271-272.
Zheng Y, Zhao WM, Wang H, Zhou YB, Luan Y, Qi M, Cheng YZ, Tang W, Liu J, Yu H, Yu XP (2007). Codon usage bias in Chlamydia trachomatis and the effect of codon modification in the MOMP gene on immune responses to vaccination. Biochem. Cell Biol. 85(2):218-226.

Supplementary Table 1. *Camellia sinensis.*

Gene Names with their Accession numbers	A	T	G	C	CDS length	AT percent	GC percent
>gi|90968983|gb|DQ366599.2| *Camellia sinensis* acetyl CoA carboxylase mRNA, complete cds	484	460	489	367	1800	52.44	47.56
>gi|224037813|gb|FJ656220.1| *Camellia sinensis* polyphenol oxidase (PPO) mRNA, complete cds	502	415	439	444	1800	50.94	49.06
>gi|89111302|dbj|AB247284.1| *Camellia sinensis* pRB mRNA for retinoblastoma related protein, complete cds	877	879	695	627	3078	57.05	42.95
>gi|89111300|dbj|AB247283.1| *Camellia sinensis* cycD3-2 mRNA for cyclin D3-2, complete cds	298	337	268	216	1119	56.75	43.25
>gi|89111298|dbj|AB247282.1| *Camellia sinensis* cycD3 mRNA for cyclin D3-1, complete cds	303	292	241	280	1116	53.32	46.69
>gi|89111294|dbj|AB247280.1| *Camellia sinensis* cycb mRNA for cyclin B, complete cds	388	334	329	272	1323	54.57	45.43
>gi|472247167|gb|KC242133.1| *Camellia sinensis* stearoyl-acyl carrier protein desaturase mRNA, complete cds	332	319	295	245	1191	54.66	45.34
>gi|529205468|gb|KC920896.1| *Camellia sinensis* cultivar Longjing43 glycerol-3-phosphate acyltransferase mRNA, complete cds	349	376	317	311	1353	53.59	46.42
>gi|523713207|gb|KC847167.1| *Camellia sinensis* omega-3 fatty acid desaturase (FAD8) mRNA, complete cds	353	372	323	311	1359	53.35	46.65
>gi|480359963|gb|KC700025.1| *Camellia sinensis* AMP deaminese mRNA, complete cds	717	718	605	531	2571	55.82	44.19
>gi|76177060|gb|DQ194356.1| *Camellia sinensis* cultivar UPASI-10 4-coumaroyl CoA ligase mRNA, complete cds	433	440	420	473	1766	49.43	50.57
>gi|449082926|dbj|AB741571.1| *Camellia sinensis* CsFT1 mRNA for flowering locus T, complete cds	116	144	148	117	525	49.52	50.48
>gi|357966792|gb|JN392472.1| *Camellia sinensis* chitinase (CHIT1) mRNA, complete cds	250	237	244	238	969	50.26	49.74
>gi|398025483|gb|JX042312.1| *Camellia sinensis* clone 111/1 ricin B lectin domain protein II mRNA, complete cds	103	83	82	68	336	55.36	44.64
>gi|339232482|gb|JN024667.1| *Camellia sinensis* anthocyanidin reductase 2 (ANR2) mRNA, complete cds	297	276	249	222	1044	54.89	45.12
>gi|161789847|gb|EU284131.1| *Camellia sinensis* glutamine synthetase mRNA, complete cds	297	265	288	221	1071	52.47	47.53
>gi|307090029|gb|HM204933.1| *Camellia sinensis* isolate yuanxiaolv caffeic acid O-methyltransferase (comt) mRNA, complete cds	272	278	280	262	1092	50.37	49.63
>gi|50841418|gb|AY574920.1| *Camellia sinensis* dihydroflavonol 4-reductase mRNA, complete cds	291	270	260	223	1044	53.74	46.26
>gi|294847479|gb|GU944768.1| *Camellia sinensis* anthocyanidin reductase (ANR) mRNA, complete cds	276	256	242	240	1014	52.47	47.54

Supplementary Table 1. Contd.

>gi|117622287|gb|EF055882.1| *Camellia sinensis* cytosolic glutamine synthetase mRNA, complete cds	295	267	288	221	1071	52.47	47.53
>gi|91992505|gb|DQ461974.1| *Camellia sinensis* ATPase mRNA, complete cds	61	87	62	48	258	57.36	42.64
>gi|76177136|gb|DQ194358.1| *Camellia sinensis* cultivar UPASI-10 flavonoid 3',5'-hydroxylase mRNA, complete cds	397	397	376	363	1533	51.79	48.21
>gi|76152008|gb|DQ120521.2| *Camellia sinensis* cultivar UPASI-10 chalcone isomerase mRNA, complete cds	181	166	176	170	693	50.07	49.93
>gi|59611828|gb|AY907710.1| *Camellia sinensis* caffeine synthase mRNA, complete cds	334	303	253	208	1098	58.02	41.99
>gi|532212606|gb|KF006992.1| *Camellia sinensis* cultivar Longjing43 S-adenosylmethionine decarboxylase mRNA, complete cds	265	333	259	223	1080	55.37	44.63
>gi|428135437|gb|JQ790527.1| *Camellia sinensis* caffeoyl-CoA-O-methyltransferase (CCoAOMT) mRNA, complete cds	198	165	189	186	738	49.19	50.81
>gi|326380569|gb|GU992402.1| *Camellia sinensis* anthocyanidin reductase 1 mRNA, complete cds	299	275	248	222	1044	54.98	45.02
>gi|308943876|gb|HM440161.1| *Camellia sinensis* lipoxygenase mRNA, complete cds	769	703	654	580	2706	54.40	45.60
>gi|62955863|gb|AY945842.1| *Camellia sinensis* flavonoid 3',5'-hydroxylase mRNA, complete cds	401	398	374	360	1533	52.12	47.88
>gi|76786310|gb|DQ198089.1| *Camellia sinensis* flavonol synthase (FLS) mRNA, complete cds	288	259	236	213	996	54.92	45.08

All the cds selected for the present study from the two plants, this file provides the gene name, accession numbers along with the overall AT and GC percentage, RCBS, CAI, ENc, SCUO, GC1, GC2 & GC3 for all the genes. These data allow for the reconstruction of all the analyses.

Supplementary Table 1. Contd.

Gene Names with their Accession numbers	CAI	ENc	RCBS	SCUO	GC1 Percent	GC2 percent	GC3 percent
>gi|90968983|gb|DQ366599.2| *Camellia sinensis* acetyl CoA carboxylase mRNA, complete cds	0.64	57.55	0.041	0.055	55	42.3	45.3
>gi|224037813|gb|FJ656220.1| *Camellia sinensis* polyphenol oxidase (PPO) mRNA, complete cds	0.61	56.63	0.041	0.050	52.3	40.5	54.3
>gi|89111302|dbj|AB247284.1| *Camellia sinensis* pRB mRNA for retinoblastoma related protein, complete cds	0.55	51.89	0.025	0.090	47.9	43.9	37.1
>gi|89111300|dbj|AB247283.1| *Camellia sinensis* cycD3-2 mRNA for cyclin D3-2, complete cds	0.36	51.48	0.064	0.111	50.4	34.3	45
>gi|89111298|dbj|AB247282.1| *Camellia sinensis* cycD3 mRNA for cyclin D3-1, complete cds	0.52	54.79	0.063	0.080	53	38.2	48.9

Supplementary Table 1. Contd.

>gi|89111294|dbj|AB247280.1| *Camellia sinensis* cycb mRNA for cyclin B, complete cds	0.47	52.52	0.054	0.115	53.7	39.5	43.1
>gi|472247167|gb|KC242133.1| *Camellia sinensis* stearoyl-acyl carrier protein desaturase mRNA, complete cds	0.33	48.51	0.060	0.164	54.2	37.8	44.1
>gi|529205468|gb|KC920896.1| *Camellia sinensis* cultivar Longjing43 glycerol-3-phosphate acyltransferase mRNA, complete cds	0.49	53.64	0.053	0.102	52.5	42.1	44.6
>gi|523713207|gb|KC847167.1| *Camellia sinensis* omega-3 fatty acid desaturase (FAD8) mRNA, complete cds	0.57	57.70	0.054	0.059	53	42.2	44.8
>gi|480359963|gb|KC700025.1| *Camellia sinensis* AMP deaminese mRNA, complete cds	0.59	54.63	0.030	0.064	52.5	37.5	42.6
>gi|76177060|gb|DQ194356.1| *Camellia sinensis* cultivar UPASI-10 4-coumaroyl CoA ligase mRNA, complete cds	0.54	57.23	0.045	0.055	39.4	54.8	57.5
>gi|449082926|dbj|AB741571.1| *Camellia sinensis* CsFT1 mRNA for flowering locus T, complete cds	0.28	51.27	0.123	0.175	55.4	48	48
>gi|357966792|gb|JN392472.1| *Camellia sinensis* chitinase (CHIT1) mRNA, complete cds	0.61	58.72	0.081	0.062	43.2	51.2	55.1
>gi|398025483|gb|JX042312.1| *Camellia sinensis* clone 111/1 ricin B lectin domain protein II mRNA, complete cds	0.29	61.00	0.194	0.060	43.8	35.7	54.5
>gi|339232482|gb|JN024667.1| *Camellia sinensis* anthocyanidin reductase 2 (ANR2) mRNA, complete cds	0.47	56.44	0.067	0.077	50.1	41.7	43.4
>gi|161789847|gb|EU284131.1| *Camellia sinensis* glutamine synthetase mRNA, complete cds	0.40	52.40	0.066	0.106	53.5	44.3	44.8
>gi|307090029|gb|HM204933.1| *Camellia sinensis* isolate yuanxiaolv caffeic acid O-methyltransferase (comt) mRNA, complete cds	0.49	53.45	0.070	0.105	38.4	58.6	51.9
>gi|50841418|gb|AY574920.1| *Camellia sinensis* dihydroflavonol 4-reductase mRNA, complete cds	0.50	57.27	0.067	0.070	46.8	38.8	53.2
>gi|294847479|gb|GU944768.1| *Camellia sinensis* anthocyanidin reductase (ANR) mRNA, complete cds	0.97	61.00	0.068	0.159	51.6	41.1	49.7
>gi|117622287|gb|EF055882.1| *Camellia sinensis* cytosolic glutamine synthetase mRNA, complete cds	0.41	52.71	0.066	0.106	53.2	44.3	45.1
>gi|91992505|gb|DQ461974.1| *Camellia sinensis* ATPase mRNA, complete cds	0.57	55.31	0.241	0.050	48.8	40.7	38.4
>gi|76177136|gb|DQ194358.1| *Camellia sinensis* cultivar UPASI-10 flavonoid 3',5'-hydroxylase mRNA, complete cds	0.61	56.57	0.049	0.072	54	40.9	49.7
>gi|76152008|gb|DQ120521.2| *Camellia sinensis* cultivar UPASI-10 chalcone isomerase mRNA, complete cds	0.24	49.87	0.095	0.147	51.5	43.3	55
>gi|59611828|gb|AY907710.1| *Camellia sinensis* caffeine synthase mRNA, complete cds	0.36	52.11	0.064	0.118	48.9	35.8	41.3
>gi|532212606|gb|KF006992.1| *Camellia sinensis* cultivar Longjing43 S-adenosylmethionine decarboxylase mRNA, complete cds	0.39	50.87	0.065	0.123	46.9	42.8	44.2

Supplementary Table 1. Contd.

>gi|428135437|gb|JQ790527.1| *Camellia sinensis* caffeoyl-CoA-O-methyltransferase (CCoAOMT) mRNA, complete cds	0.38	53.34	0.097	0.160	34.9	62.7	54.8
>gi|326380569|gb|GU992402.1| *Camellia sinensis* anthocyanidin reductase 1 mRNA, complete cds	0.48	57.58	0.067	0.076	50.3	41.4	43.4
>gi|308943876|gb|HM440161.1| *Camellia sinensis* lipoxygenase mRNA, complete cds	0.65	56.13	0.029	0.056	52.1	40.4	44.3
>gi|62955863|gb|AY945842.1| *Camellia sinensis* flavonoid 3',5'-hydroxylase mRNA, complete cds	0.59	56.85	0.049	0.067	53.4	40.3	49.9
>gi|76786310|gb|DQ198089.1| *Camellia sinensis* flavonol synthase (FLS) mRNA, complete cds	0.41	57.19	0.076	0.059	33.9	45.6	55.7

Supplementary Table 2. *Brassica campestris.*

Gene names with their accession numbers	A	T	G	C	CDS length	AT percent	GC percent
receptor protein kinase SRK12 (dbj-D38564)	758	697	617	499	2571	56.59	43.41
oleifera copper and zinc superoxide dismutase mRNA(gb-KF356248)	111	122	121	105	459	50.76	49.24
Brassica rapa cultivar Samjin Col-2-like protein mRNA (gb-AY356370)	288	235	237	206	966	54.14	45.86
Brassica rapa cultivar Samjin reduced vernalization response 1 mRNA (gb-AY356368)	288	248	238	216	990	54.14	45.86
Brassica rapa plastid-lipid associated protein PAP3 mRNA (gb-AF290565)	245	269	269	300	1083	47.46	52.54
Brassica rapa subsp. campestris glutathione reductase mRNA (gb-JN795550)	403	388	437	275	1503	52.63	47.37
Brassica rapa subsp. pekinensis cultivar Huangya 14 MF21 mRNA (gb-JF437596)	129	123	105	99	456	55.26	44.74
Brassica rapa subsp. chinensis exocyst subunit EXO70A1 mRNA (gb-JX997396)	572	488	478	379	1917	55.30	44.71
Brassica rapa subsp. chinensis male sterility 2 mRNA (gb-EF093533)	567	491	467	326	1851	57.16	42.84
Brassica rapa subsp. chinensis cultivar Aikangqing MF21 mRNA (gb-JF437595)	128	123	106	99	456	55.04	44.96
Brassica rapa var. purpuraria cultivar Zitai 1 MF21 mRNA(gb-JF437594)	129	122	105	100	456	55.04	44.96
Brassica rapa subsp. chinensis receptor-like kinase SSP mRNA(gb-KC576523)	427	399	312	260	1398	59.08	40.92
Brassica rapa subsp. chinensis MLPK mRNA(gb-KC576522)	370	324	318	245	1257	55.21	44.79
Brassica rapa subsp. chinensis kinase-associated protein phosphatase mRNA(gb-KC576521)	447	452	439	312	1650	54.49	45.52
Brassica rapa subsp. chinensis aspartic proteinase mRNA(gb-KC576520)	359	434	412	316	1521	52.14	47.86
Brassica rapa subsp. chinensis senescence-associated cysteine protease mRNA(gb-KC576519)	282	266	292	240	1080	50.74	49.26
Brassica rapa subsp. chinensis ARC1 mRNA(gb-KC576518)	522	484	484	496	1986	50.66	49.35
Brassica rapa subsp. pekinensis beta-carotene hydroxylase mRNA(gb-GQ178285)	211	242	239	229	921	49.19	50.81
Brassica rapa subsp. pekinensis cultivar Huangya14 profilin mRNA(gb-EU163278)	107	88	119	91	405	48.15	51.85
Brassica rapa subsp. rapa cultivar Wenzhoupancai profilin mRNA(gb-EU163276)	109	89	117	90	405	48.89	51.11
Brassica rapa subsp. pekinensis cultivar Huangya14 anther-specific proline rich protein mRNA(gb-EF101148)	518	402	314	512	1746	52.69	47.31
Brassica rapa subsp. pekinensis cultivar Xiaoqingko anther-specific proline rich protein mRNA(gb-EF101141)	513	398	312	508	1731	52.63	47.37

Supplementary Table 2. Contd.

Brassica rapa subsp. pekinensis flowering locus C3 mRNA(gb-DQ866876)	172	143	144	135	594	53.03	46.97
Brassica rapa subsp. pekinensis flowering locus C2 mRNA(gb-DQ866875)	177	140	138	136	591	53.64	46.36
Brassica rapa subsp. pekinensis flowering locus C1 mRNA(gb-DQ866874)	177	151	161	132	621	52.82	47.18
Brassica rapa subsp. pekinensis cold-regulated protein mRNA(gb-DQ491005)	130	80	114	78	402	52.24	47.76
Brassica rapa subsp. pekinensis MORN mRNA(gb-FJ460465)	371	382	391	365	1509	49.90	50.10
Brassica rapa subsp. pekinensis cultivar Matsushima SUR1 mRNA(gb-EF611271)	353	306	349	354	1362	48.39	51.62
Brassica rapa subsp. pekinensis cultivar Kwan-Hoo Choi ST5a mRNA(gb-EF611266)	267	236	264	253	1020	49.31	50.69
Brassica rapa subsp. pekinensis cultivar Matsushima ST5a mRNA(gb-EF611263)	268	236	263	253	1020	49.41	50.59
Brassica rapa subsp. pekinensis CYP83B1 mRNA(gb-EF611260)	404	379	343	374	1500	52.20	47.80
Brassica rapa subsp. pekinensis MYB mRNA(gb-DQ903665)	284	268	248	247	1047	52.72	47.28
Brassica rapa subsp. pekinensis acyl desaturase mRNA(gb-DQ886528)	319	300	323	288	1230	50.33	49.68
Brassica rapa subsp. pekinensis geranylgeranyl reductase mRNA(gb-DQ886527)	385	338	434	334	1491	48.49	51.51
Brassica rapa subsp. pekinensis biotin synthase mRNA(gb-DQ886525)	317	273	279	259	1128	52.31	47.70
Raphanus sativus cultivar Sakurashimadakon MF21 mRNA(gb-JF437605)	129	120	107	100	456	54.61	45.40
Brassica nigra cultivar 071-01 MF21 mRNA(gb-JF437604)	131	117	107	101	456	54.39	45.61
Brassica juncea var. rugosa cultivar Yamakada MF21 mRNA(gb-JF437603)	128	123	106	99	456	55.04	44.96
Brassica carinata cultivar 079-01 MF21 mRNA(gb-JF437602)	129	120	107	100	456	54.61	45.40
Brassica napus cultivar 071-02 MF21 mRNA(gb-JF437601)	131	120	109	102	462	54.33	45.67
Brassica napus var. napobrassica cultivar Datou1 MF21 mRNA(gb-JF437600)	120	121	105	95	441	54.65	45.35
Brassica oleracea var. capitata cultivar Jixin MF21 mRNA(gb-JF437599)	119	119	108	95	441	53.97	46.03
Brassica juncea var. multiceps cultivar Maertou MF21 mRNA(gb-JF437598)	129	123	105	99	456	55.26	44.74
Brassica rapa subsp. nipposinica cultivar Wanshengwanye MF21 mRNA(gb-JF437597)	129	123	105	99	456	55.26	44.74
Brassica rapa subsp. chinensis napin mRNA(gb-HM027884)	158	111	130	156	555	48.47	51.53
Mesostigma viride photosystem II PsbR protein mRNA(gb-DQ370085)	76	84	115	139	414	38.65	61.35
Brassica rapa BcNDK 2 mRNA for nucloside diphosphate kinase 2(dbj-AB078008)	164	190	179	160	693	51.08	48.92

All the cds selected for the present study from the two plants, this file provides the gene name, accession numbers along with the overall AT and GC percentage, RCBS, CAI, ENc, SCUO, GC1, GC2 & GC3 for all the genes. These data allow for the reconstruction of all the analyses.

Supplementary Table 2. Contd.

Gene names with their accession numbers	CAI	ENc	RCBS	SCUO	GC1 Percent	GC2 percent	GC3 percent
receptor protein kinase SRK12 (dbj-D38564)	0.60	54.90	0.03	0.051	44.8	40.1	45.3
oleifera copper and zinc superoxide dismutase mRNA(gb-KF356248)	0.15	42.28	0.14	0.093	62.7	51	34
Brassica rapa cultivar Samjin Col-2-like protein mRNA (gb-AY356370)	0.38	54.15	0.07	0.077	51.2	44.7	41.6
Brassica rapa cultivar Samjin reduced vernalization response 1 mRNA (gb-AY356368)	0.48	51.75	0.07	0.097	50	40.3	47.3
Brassica rapa plastid-lipid associated protein PAP3 mRNA (gb-AF290565)	0.51	56.13	0.06	0.127	55.4	48.2	54

Supplementary Table 2. Contd.

Brassica rapa subsp. campestris glutathione reductase mRNA (gb-JN795550)	0.53	56.00	0.05	0.126	54.9	42.7	44.5
Brassica rapa subsp. pekinensis cultivar Huangya 14 MF21 mRNA (gb-JF437596)	0.40	59.19	0.14	0.053	48	42.8	43.4
Brassica rapa subsp. chinensis exocyst subunit EXO70A1 mRNA (gb-JX997396)	0.48	52.26	0.04	0.104	53.1	36.6	44.4
Brassica rapa subsp. chinensis male sterility 2 mRNA (gb-EF093533)	0.49	49.16	0.04	0.085	48.3	37.8	42.5
Brassica rapa subsp. chinensis cultivar Aikangqing MF21 mRNA (gb-JF437595)	0.40	58.37	0.14	0.080	48	43.4	43.4
Brassica rapa var. purpuraria cultivar Zitai 1 MF21 mRNA(gb-JF437594)	0.40	59.19	0.14	0.072	48	42.8	44.1
Brassica rapa subsp. chinensis receptor-like kinase SSP mRNA(gb-KC576523)	0.41	52.12	0.05	0.104	40.9	35.6	37.8
Brassica rapa subsp. chinensis MLPK mRNA(gb-KC576522)	0.42	50.53	0.06	0.086	51.6	43.7	39.1
Brassica rapa subsp. chinensis kinase-associated protein phosphatase mRNA(gb-KC576521)	0.63	57.87	0.04	0.124	54.4	39.6	42.5
Brassica rapa subsp. chinensis aspartic proteinase mRNA(gb-KC576520)	0.45	50.83	0.05	0.087	55.2	41.4	46.9
Brassica rapa subsp. chinensis senescence-associated cysteine protease mRNA(gb-KC576519)	0.54	58.69	0.07	0.086	53.6	42.2	51.9
Brassica rapa subsp. chinensis ARC1 mRNA(gb-KC576518)	0.56	53.59	0.04	0.077	52.9	40	55.1
Brassica rapa subsp. pekinensis beta-carotene hydroxylase mRNA(gb-GQ178285)	0.40	53.54	0.08	0.073	54.7	43.6	54.1
Brassica rapa subsp. pekinensis cultivar Huangya14 profilin mRNA(gb-EU163278)	0.31	59.09	0.16	0.088	60.7	40	54.8
Brassica rapa subsp. rapa cultivar Wenzhoupancai profilin mRNA(gb-EU163276)	0.27	57.77	0.16	0.103	60	40	53.3
Brassica rapa subsp. pekinensis cultivar Huangya14 anther-specific proline rich protein mRNA(gb-EF101148)	0.57	54.61	0.04	0.109	54.5	54.1	33.3
Brassica rapa subsp. pekinensis cultivar Xiaoqingko anther-specific proline rich protein mRNA(gb-EF101141)	0.58	54.70	0.04	0.117	54.6	54.1	33.4
Brassica rapa subsp. pekinensis flowering locus C3 mRNA(gb-DQ866876)	0.35	51.77	0.11	0.091	59.1	33.3	48.5
Brassica rapa subsp. pekinensis flowering locus C2 mRNA(gb-DQ866875)	0.26	50.23	0.11	0.119	59.4	32	47.7
Brassica rapa subsp. pekinensis flowering locus C1 mRNA(gb-DQ866874)	0.27	52.26	0.10	0.157	60.4	30.4	50.7
Brassica rapa subsp. pekinensis cold-regulated protein mRNA(gb-DQ491005)	0.29	50.10	0.15	0.133	55.2	36.6	51.5
Brassica rapa subsp. pekinensis MORN mRNA(gb-FJ460465)	0.53	53.53	0.05	0.167	53.9	47.5	48.9
Brassica rapa subsp. pekinensis cultivar Matsushima SUR1 mRNA(gb-EF611271)	0.54	57.12	0.05	0.163	56.4	38.5	59.9
Brassica rapa subsp. pekinensis cultivar Kwan-Hoo Choi ST5a mRNA(gb-EF611266)	0.38	53.50	0.07	0.153	52.6	36.5	62.9
Brassica rapa subsp. pekinensis cultivar Matsushima ST5a mRNA(gb-EF611263)	0.41	53.89	0.07	0.280	52.6	36.5	62.6
Brassica rapa subsp. pekinensis CYP83B1 mRNA(gb-EF611260)	0.58	55.99	0.05	0.127	54.4	35.8	53.2
Brassica rapa subsp. pekinensis MYB mRNA(gb-DQ903665)	0.53	54.51	0.07	0.118	52.4	50.1	39.3
Brassica rapa subsp. pekinensis acyl desaturase mRNA(gb-DQ886528)	0.50	55.46	0.06	0.127	53.7	38.8	56.6
Brassica rapa subsp. pekinensis geranylgeranyl reductase mRNA(gb-DQ886527)	0.54	53.35	0.05	0.127	51.5	51.5	62
Brassica rapa subsp. pekinensis biotin synthase mRNA(gb-DQ886525)	0.46	53.28	0.06	0.144	52.9	44.7	45.5
Raphanus sativus cultivar Sakurashimadakon MF21 mRNA(gb-JF437605)	0.36	61.00	0.14	0.118	48	44.7	43.4
Brassica nigra cultivar 071-01 MF21 mRNA(gb-JF437604)	0.51	59.90	0.14	0.117	48	46.1	42.8
Brassica juncea var. rugosa cultivar Yamakada MF21 mRNA(gb-JF437603)	0.40	59.22	0.14	0.127	48.7	42.8	43.4
Brassica carinata cultivar 079-01 MF21 mRNA(gb-JF437602)	0.36	61.00	0.14	0.118	48	44.7	43.4
Brassica napus cultivar 071-02 MF21 mRNA(gb-JF437601)	0.38	61.00	0.14	0.118	48.7	44.2	44.2
Brassica napus var. napobrassica cultivar Datou1 MF21 mRNA(gb-JF437600)	0.38	61.00	0.15	0.133	48.3	42.9	44.9
Brassica oleracea var. capitata cultivar Jixin MF21 mRNA(gb-JF437599)	0.36	61.00	0.15	0.111	48.3	44.9	44.9

Supplementary Table 2. Contd.

Brassica juncea var. multiceps cultivar Maertou MF21 mRNA(gb-JF437598)	0.40	59.19	0.14	0.474	48	42.8	43.4
Brassica rapa subsp. nipposinica cultivar Wanshengwanye MF21 mRNA(gb-JF437597)	0.40	59.19	0.14	0.163	48	42.8	43.4
Brassica rapa subsp. chinensis napin mRNA(gb-HM027884)	0.31	49.50	0.12	0.192	59.5	37.8	57.3
Mesostigma viride photosystem II PsbR protein mRNA(gb-DQ370085)	0.09	36.65	0.16	0.116	54.3	50	79.7
Brassica rapa BcNDK 2 mRNA for nucloside diphosphate kinase 2(dbj-AB078008)	0.33	50.33	0.10	0.116	54.5	48.5	43.7