

Analyses of Codon Usage Patterns and Codon Usage Bias in Peach (*Prunus persica*)

Li Ruoyu², Zhang Xiaodan⁴, Ma Xinyi³, Guo Rui¹, Yan Shaobin¹, Jin Guang¹, Zhou Ping¹ ✉

¹ Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, 350013, China

² College of Agricultural, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

³ College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

⁴ Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL61801, U.S.A

✉ Corresponding author email: fagaless@163.com

Molecular Plant Breeding, 2022, Vol.13, No.28 doi: [10.5376/mpb.2022.13.0028](https://doi.org/10.5376/mpb.2022.13.0028)

Received: 25 Nov., 2022

Accepted: 02 Dec., 2022

Published: 09 Dec., 2022

Copyright © 2022 Li et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Li R.Y., Zhang X.D., Ma X.Y., Guo R., Yan S.B., Jin G., and Zhou P., 2022, Analyses of codon usage patterns and codon usage bias in peach (*Prunus persica*), Molecular Plant Breeding, 13(28): 1-11 (doi: [10.5376/mpb.2022.13.0028](https://doi.org/10.5376/mpb.2022.13.0028))

Abstract To further study the characteristics of peach codon usage, this study analyzed codon usage biases and codon usage patterns of peach genome, based on the statistical calculations of related GC content, effective number of codons (ENC) and relative synonymous codons (RSCU) from 26 873 coding sequences. The results showed that there were obvious biases in codon usage of peach, and 4 out of 61 codons (UCA, ACA, GCA and GAA) were defined as the optimal codons, all of which end with Adenine at the third codon position. Further analyses of codon usage frequency among peach and other 9 relative species in Rosaceae found that the codon usage patterns in the relative genus were similar. These results also suggested that there was a positive correlation between the copy number of tRNA genes and the occurrence frequency of corresponding amino acids (and specific codons) in the peach genome. These results revealed codon usage patterns in Peach and provide an important reference for the relevant studies on codon usage mechanism and the genetic engineering.

Keywords Peach; Codon usage pattern; Codon usage bias

Synonymous codons correspond to the same amino acid in living organisms. However, the usage frequency of these synonymous codons is different, and several codons are frequently used, which is called "codon usage bias" phenomenon (Plotkin and Kudla, 2011). The preference of codon usage is affected by many factors, such as DNA replication start site (Huang et al., 2009), protein translation efficiency (Zalucki et al., 2007), tRNA abundance (Olejniczak and Uhlenbeck, 2006), gene length (Moriyama and Powell, 1998) et al. Codon preference is determined by complex mechanisms and reflects genetic differences among species. The study of codon bias has important reference value for studying efficient gene expression, predicting new genes and exploring the evolution of related genes (Wu et al., 2015). With the advent of large-scale genome sequencing, more and more animal and plant genomes have been deciphered, making it possible to comprehensively and systematically analyze the codon bias of related species.

Peach (*Prunus Persica* L.) is an important deciduous fruit tree. Because of its diverse agronomic traits, short flowering and fruiting period, and precise genetic linkage map, it has become one of the model tree species in the study of fruit science and genetics. At present, peach genome has been sequenced and published. However, the codon usage bias of peach genome is rarely reported. In this study, sequence characteristics and codon rules of 26 873 annotated coding genes were statistically analyzed to explore the factors affecting codon usage in peach genome. The results provided a reference for studying the molecular evolution of peach genes, and also provided help for further gene structure analysis and genetic engineering improvement.

1 Results and Analysis

1.1 Analysis of codon usage parameters

The content of T3s, C3s, A3s and G3s varied from 5.1% to 70.3% (Mean: 38.9%), 1.7%~83.5% (Mean: 26.8%);

2.7%~57.5% (Mean: 31.8%); 0~72.5% (Mean: 28.1%) respectively. The range of ENC was 22.45 to 61 (Mean: 52.5). 26 873 genes were preliminarily determined to have different degrees of codon usage bias.

The CAI (codon fitness Index) was used to evaluate the matching degree of synonymous codon and codon optimal usage (Table 1). The CAI ranged from 6.2% to 59.2% (Mean: 20.2%). At the same time, the contents of GC3s and codon GC of peach genome were analyzed, and the variation range was 13.8%~90.1% and 23.9%~73.4%, respectively, with an average of 42.8% and 45.4%. Further plots showed the respective GC and GC3s contents of codons of 26 873 coding genes (Figure 1), and the overall trend line deviated from the oblique line $GC = GC3s$, which confirmed that peach codons had a certain bias, which was consistent with previous results of codon preference usage in dicotyledons (Chiapello et al., 1998; Kawabe and Miyashita, 2003).

Table 1 Constitution and usage parameters of codon

Parameter	Variation range	Mean±SD
T3s (%)	5.1~70.3	38.9±7.3
C3s (%)	1.7~83.5	26.8±8.1
A3s (%)	2.7~57.5	31.8±7.1
G3s (%)	0~72.5	28.1±6.7
CAI (%)	6.2~59.2	20.2±3.2
ENC	22.45~61	52.5±4.7
GC3s (%)	13.8~90.1	42.8±9.2
GC content (%)	23.9~73.4	45.4±4.5

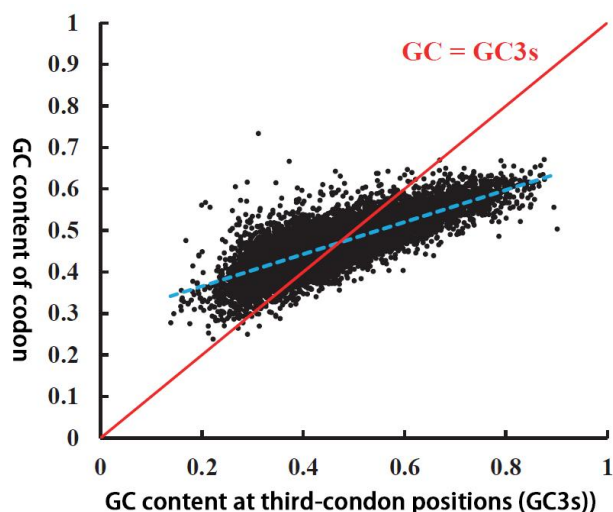


Figure 1 GC and GC3s analyses of codon use features from 26 873 protein-encoding genes

1.2 Using Enc-plot and PR2-plot to analyze codon usage bias

By counting the number of effective codons, ENC-plot can be used to visually show the disequilibrium degree of synonymous codon usage and measure the deviation degree of synonymous codon usage from random selection. The value of ENC ranges from 20 to 61, with the minimum value of 20 indicating that only one codon is used for each amino acid, and the maximum value of 61 indicating that all 61 codons are used. The lower the ENC value is, the more favorable the codon is used. Considering GC3s and ENC as horizontal and vertical coordinates, ENC-plot was used to analyze whether codon usage bias was the result of nucleotide variation (Neutral mutation) at position 3 of the codon (Liu et al., 2012). The results showed that some genes were close to the theoretical trend line, while some genes deviated from the trend line (Figure 2). The result indicate that there are other factors besides the influence of nucleotide neutral mutations that also affect the codon usage bias of these genes.

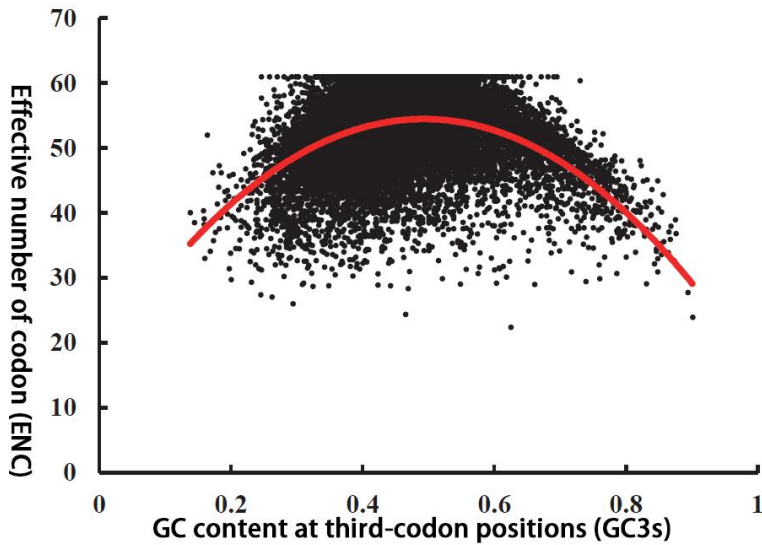


Figure 2 ENC-plot curve of peach genomic coding genes

Note: Black scatter means a single gene, and the curve indicates a theoretic trend

Pr2-plot analysis takes $A3s/(A3s+T3s)$ and $G3s/(G3s+C3s)$ as the horizontal and vertical coordinates to analyze the nucleotide base composition at the third position of the codon. Through the center point ($A=T$; $C=G$) emits a vector that visually shows the type and degree of deviation of the base. In general, if the neutral mutation rate of DNA duplex and the selection pressure are not deviated, $A3s$, $T3s$, $G3s$ and $C3s$ generally follow the law of $A3s = T3s$ and $G3s = C3s$, then the genes are clustered and distributed in the center in PR2-plot (Figure 3). However, the PR2-plot analysis of all peach genes showed that the whole genes showed a tendency to deviate from the center point, that is, the use of the third base of the codon was biased. This indicates that peach codon bias is not only affected by the known nucleotide neutral mutations, but also by other factors, such as selection pressure.

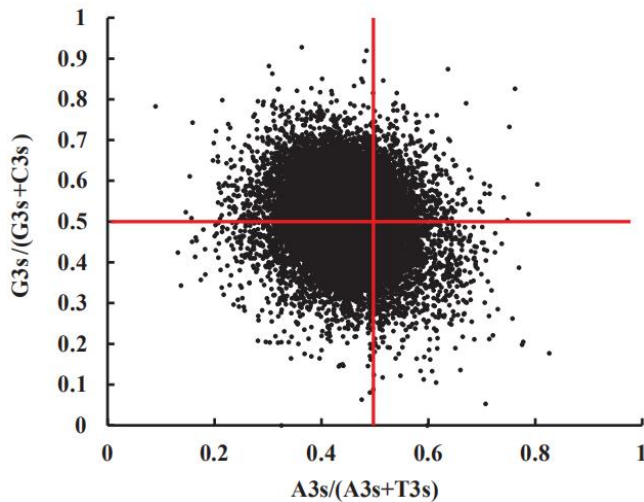


Figure 3 PR2-plot analysis of peach genomic coding genes

1.3 Optimal codon analysis

The RSCU values of all genes were listed in the table (Table 2). There are 27 high-frequency codons in $RSCU > 1$, which are TTT, TTG, CTT, ATT, GTT, GTG, TCT, TCA, CCT, CCA, ACT, ACA, GCT, GCA, TAT, CAT, CAA, AAT, AAG, etc. GAT, GAA, TGT, TGA, AGA, AGG, GGT, GGA. Twenty-three of them ended in A/T and four ended in G/C. The codon with the strongest bias was AGA, whose RSCU value was 1.85. As shown in the Table, the RSCU values of the high and low expression libraries divided by ENC values are listed (Table 3). According to the standard of the difference between the two RSCU values > 0.08 , a total of 16 highly expressed superior

codons were identified, including CTA, CTG, ATA, GTA, TCA, TCG, ACA, ACG, GCA, GAA, TGA, CGT, CGA, CGG, AGT, AGC, etc. Eleven of them ended in A/T and five ended in G/C. Four optimal codons were determined based on the results of high frequency codons and high expression superior codons, which were TCA, ACA, GCA and GAA, and all of them ended in A.

Table 2 Usage of synonymous codon

Amino acid	Codon	RSCU	Percentage (%)	Amino acid	Codon	RSCU	Percentage (%)
Phe	TTT*	1.14	2.46	Tyr	TAT*	1.16	1.62
	TTC	0.86	1.84		TAC	0.84	1.18
Leu	TTA	0.71	1.18	TER	TAA	0.93	0.08
	TTG*	1.53	2.54		TAG	0.76	0.06
	CTT*	1.44	2.39		TGA*	1.31	0.11
	CTC	0.9	1.49	His	CAT*	1.24	1.52
	CTA	0.58	0.96		CAC	0.76	0.92
	Ile	CTG	0.84	1.38	Gln	CAA*	1.09
ATT*		1.41	2.49	CAG		0.91	1.69
ATC		0.81	1.43	Asn	AAT*	1.18	2.66
ATA	0.77	1.36	AAC		0.82	1.84	
Met	ATG	1	2.39	Lys	AAA	0.92	2.81
Val	GTT*	1.54	2.50	Lys	AAG*	1.08	3.28
	GTC	0.71	1.15		Asp	GAT*	1.32
	GTA	0.6	0.98	GAC		0.68	1.76
	GTG*	1.15	1.88	Glu	GAA*	1.03	3.28
Ser	TCT*	1.5	2.23		GAG	0.97	3.06
	TCC	0.87	1.29	Cys	TGT*	1.03	0.99
	TCA*	1.35	2.00		TGC	0.97	0.93
	TCG	0.45	0.66	Trp	TGG	1	1.31
Pro	CCT*	1.4	1.72		Arg	CGT	0.72
	CCC	0.7	0.86	CGC		0.56	0.48
	CCA*	1.46	1.79	CGA		0.67	0.57
	CCG	0.44	0.54	CGG		0.58	0.49
Thr	ACT*	1.35	1.65	Ser	AGT	0.98	1.45
	ACC	0.93	1.15		AGC	0.85	1.27
	ACA*	1.32	1.62	Arg	AGA*	1.86	1.60
	ACG	0.4	0.49		AGG*	1.61	1.38
Ala	GCT*	1.53	2.54	Gly	GGT*	1.14	1.87
	GCC	0.82	1.37		GGC	0.78	1.27
	GCA*	1.32	2.19		GGA*	1.22	1.99
	GCG	0.33	0.55		GGG	0.86	1.40

Note: *: High frequency codon

Table 3 The usage frequency of synonymous codons in high and low genome

Amino acid	Codon	High expressed (RSCU)	Weak expressed (RSCU)	Amino acid	Codon	High expressed (RSCU)	Weak expressed (RSCU)	
Phe	TTT	1.05	1.04	Tyr	TAT	1.07	1	
	TTC	0.95	0.96		TAC	0.93	1	
Leu	TTA	0.69	0.65	TER	TAA	0.94	0.97	
	TTG	1.37	1.44		TAG	0.69	0.78	
	CTT	1.26	1.18		TGA	1.36	1.25	
	CTC	1.1	1.46	His	CAT	1.12	1.07	
	<u>CTA</u>	0.64	0.44		CAC	0.88	0.93	
	<u>CTG</u>	0.93	0.84	Gln	CAA	1.09	1.09	
Ile	ATT	1.25	1.26		CAG	0.91	0.91	
	ATC	0.97	1.12	Asn	AAT	1.05	1.04	
	<u>ATA</u>	0.79	0.61		AAC	0.95	0.96	
Met	ATG	1	1	Lys	AAA	0.89	0.86	
Val	GTT	1.28	1.32		AAG	1.11	1.14	
	GTC	0.9	0.98	Asp	GAT	1.18	1.16	
	<u>GTA</u>	0.63	0.43		GAC	0.82	0.84	
	GTG	1.19	1.26	Glu*	<u>GAA</u>	1	0.9	
Ser	TCT	1.27	1.47		GAG	1	1.1	
	TCC	1.02	1.26	Cys	TGT	0.96	0.99	
	<u>TCA*</u>	1.18	1.08		TGC	1.04	1.01	
	<u>TCG</u>	0.67	0.55	Trp	TGG	1	1	
Pro	CCT	1.26	1.19		Arg	<u>CGT</u>	0.85	0.61
	CCC	0.83	0.92			CGC	0.77	0.86
	CCA	1.29	1.33	<u>CGA</u>		0.84	0.52	
	CCG	0.62	0.56	<u>CGG</u>	0.73	0.49		
Thr	ACT	1.16	1.13	Ser	<u>AGT</u>	0.87	0.76	
	ACC	1.05	1.33		<u>AGC</u>	0.99	0.88	
	<u>ACA*</u>	1.15	1.06	Arg	AGA	1.47	1.82	
	<u>ACG</u>	0.63	0.48		AGG	1.34	1.69	
Ala	GCT	1.31	1.38	Gly	GGT	1.04	1.1	
	GCC	1	1.23		GGC	0.97	1.04	
	<u>GCA*</u>	1.18	0.94		GGA	1.11	1.06	
	GCG	0.5	0.45		GGG	0.88	0.81	

Note: *: Optimal codons; The underlined codons were high expression superior codons

1.4 Comparison of codon usage frequencies between peach and Rosaceae related species

Codon preference varies among species. Pearson correlation was used to analyze the frequency of 61 codon usage in 10 species to measure the similarity of codon usage in each species. The results showed that the correlation coefficient of apple, peach, cherry and plum (codon usage frequency) was higher than 0.99, and the correlation coefficient of five crops in strawberry (codon usage frequency) was higher than 0.95, but the correlation coefficient of apple, peach, cherry, plum and strawberry (codon usage frequency) was generally lower than 0.75. The results indirectly indicated that the codon usage pattern and frequency of apple-peach-cherry-plum were similar, and the codon usage of five crops in the genus strawberry was similar (Figure 4). The results assumed that the *Malus* (e.g. *M. x domestica*), *Prunus* (e.g. *P. Persica*, *P. mume*), and *Cerasus* (e.g. *C. x yedoensis*) had a similar pattern of codon usage, and had some differences in codon usage from *Fragaria* (e.g. *F. iinumae*, *F. Nipponica*, *F. Nubicola*, *F. Orientalis*, *F. xananassa*, *F. F. Vesca*).

<i>M. × domestica</i>	0.99	0.99	0.99	0.72	0.72	0.71	0.71	0.70	0.70
<i>P. persica</i>	1.00	0.99	0.73	0.73	0.73	0.73	0.73	0.73	0.73
<i>P. mume</i>	0.99	0.74	0.74	0.73	0.73	0.73	0.73	0.73	0.73
<i>C. × yedoensis</i>	0.70	0.70	0.69	0.69	0.70	0.69	0.70	0.69	
<i>F. iinumae</i>	1.00	1.00	1.00	0.96	0.98				
<i>F. nipponica</i>		1.00	1.00	0.95	0.98				
<i>F. nubicola</i>			1.00	0.96	0.99				
<i>F. orientalis</i>				0.95	0.99				
<i>E. × ananassa</i>					0.98				
<i>F. vesca</i>									

Figure 4 Pearson correlation analysis of codon usage frequency in 10 relative species

1.5 Analysis of the effects of tRNA gene use

Due to the needs of life activities, organisms need synthesize the required proteins at the fastest speed, and the high-abundance tRNA usually corresponds to the optimal codon (Michaud et al., 2011). Studies have confirmed that the transport of each amino acid requires at most five tRNAs with different anticodons, which are called tRNA isoacceptors. TRNA isoreceptors corresponding to the same amino acid were divided into a family, and the number of tRNA isoreceptors in each species was about 45 to 47. In organisms, the number of tRNA genes can be a good approximation to estimate the abundance of each tRNA isoreceptor, which is correlated with amino acid frequency and the use of related codons (Duret, 2000; Michaud et al., 2011).

The number of tRNA genes in peach genome and the frequency (amino acid frequency) of corresponding amino acids in 26 873 proteins were statistically analyzed, and it was observed that the number of tRNA genes was positively linearly correlated with the frequency of amino acids (Figure 5). Since most tRNAs in organisms can recognize more than one codon, and some codons can be recognized by more than one tRNA isoreceptor, under the assumption that codons are read preferentially by only one tRNA heteroreceptor (i.e., the minimum latent codon recognition pattern), the relationship between the number of tRNA isoreceptor genes and codon usage in peach genome was analyzed statistically. The results showed that, except for some codons, there was a certain linear correlation between the total frequency of most codons in the coding gene and the corresponding tRNA isoreceptor gene (Figure 6; Table 4). Considering the above two results, it can be concluded that the number of tRNA genes in peach genome has a certain influence on the selection of amino acids and codon bias.

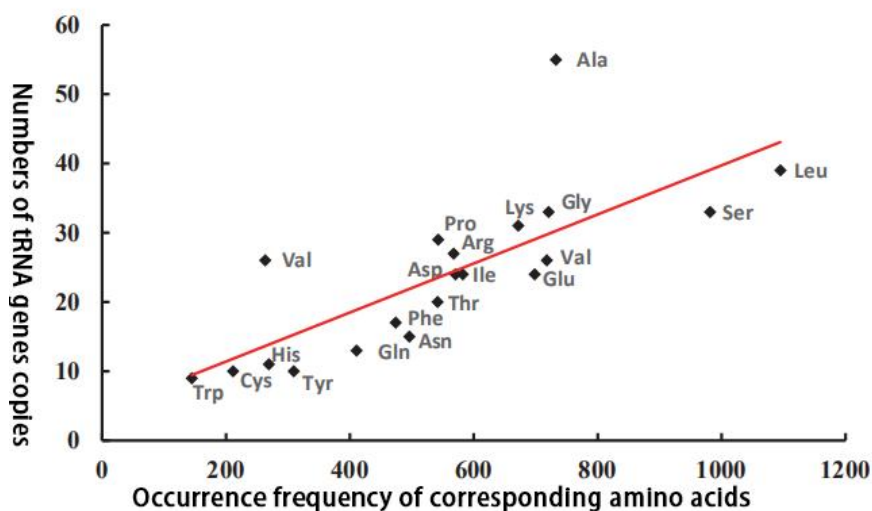


Figure 5 Correlation between the numbers of tRNA gene copies specific for each amino acid and the occurrence frequency of the corresponding amino acids

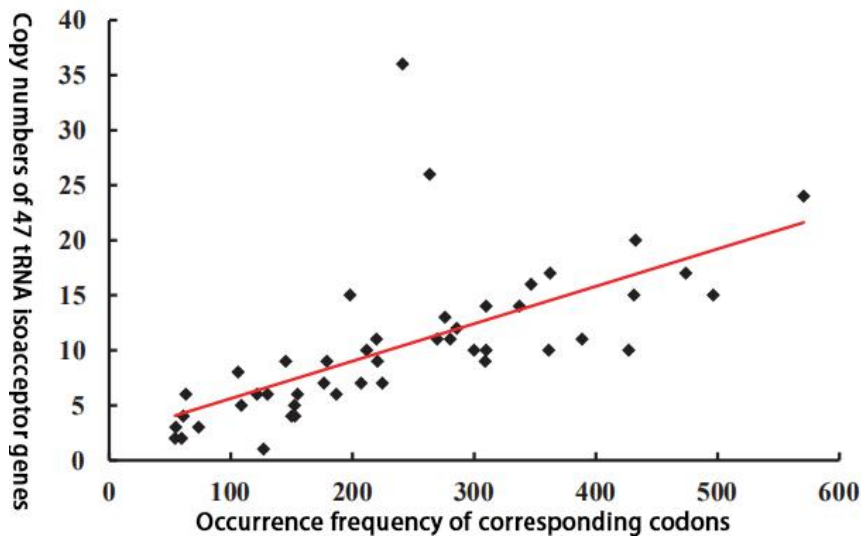


Figure 6 Correlation between the copy numbers of tRNA isoacceptor genes and the occurrence frequency of their corresponding codons

2 Discussion

Codon bias exists widely in many kinds of organisms, which is an inevitable and complicated phenomenon. At present, codon bias has been widely studied in organelles, gene families and the whole genome (Luo et al., 2015; Ye et al., 2018). In this study, codon usage patterns and possible formation factors were analyzed based on peach genome information. The statistical results of codon characteristics of peach genome showed that 26873 genes had different degrees of codon preference. The analysis of ENC-plot and PR2-plot showed that the codon bias of peach was not only affected by nucleotide neutral mutation, but also influenced by selection pressure. At the same time, gene expression level also affects codon preference. The four optimal codons TCA, ACA, GCA and GAA were identified and all ended in A. Therefore, the optimal codons can be selected to obtain better expression results when the relevant amino acids are translated and expressed by genetic engineering. In addition, this study found that the content of tRNA genes in peach genome was correlated with the frequency of use of corresponding amino acids and codons, which confirmed that the content of tRNA genes in genome may also affect the preference of amino acids and codons. This result is consistent with related research reports of *Arabidopsis thaliana* and *Oryza sativa* L. (Michaud et al., 2011).

In conclusion, the analyses presented in this study provide a fundamental and comprehensive understanding of the underlying factors of codon bias in peach genomes. These studies will be of great significance in guiding the heterologous expression of functional genes (Zelasko et al., 2013). The study of codon bias and codon usage frequency is helpful to understand the genetic evolution of related species from another perspective. The results of this study indicate that the codon usage pattern and frequency of apple-peach-cherry - plum (belonging to Drupes, Rinpes, woody perennials) of Rosaceae are similar. The use of codons is similar among the five crops in the Strawberry genus (Berry, perennial herb), but the differences between apple-peach-cherry-plum and strawberry are relatively large, which is in line with the general cognition of plant morphology, phylotaxonomy and molecular evolution. These results indicated that there were specific characteristics of codon usage bias and usage frequency in 10 related species of Rosaceae including peach. The more closely related the species, the more similar the codon usage. The analysis based on codon characteristics also reflects the evolutionary relationship of species at the molecular level. These results provided important references for the prediction of novel coding genes and the improvement of exogenous gene engineering in related plants of Rosaceae and other families (Sharp and Cowe, 1991).

Table 4 The relationship between copy number of tRNA isoacceptor gene and codon frequency under minimal potential codon recognition pattern

Amino acid	Anticodon	Codon	Copy number of tRNA isoacceptor gene	Codon number
Ala	AGC	GCT-GCC	15	430 895
	CGC	GCG	4	60 763
	TGC	GCA	36	240 980
Arg	CCG	CGG	3	54 497
	ACG	CGT-CGC	6	121 369
	CCT	AGG	5	152 281
	TCG	CGA	6	62 938
	TCT	AGA	7	176 349
Asn	GTT	AAT-AAC	15	496 200
Asp	GTC	GAT-GAC	24	570 540
Cys	GCA	TGT-TGC	10	211 534
Gln	TTG	CAA	7	224 251
	CTG	CAG	6	186 553
Glu	TTC	GAA	10	361 155
	CTC	GAG	14	336 945
Gly	GCC	GGT-GGC	16	346 528
	TCC	GGA	11	219 401
	CCC	GGG	6	154 650
His	GTG	CAC-CAT	11	269 356
Ile	AAT	ATT-ATC	20	432 379
	TAT	ATA	4	149 814
Leu	CAG	CTG	4	152 481
	CAA	TTG	11	279 949
	TAA	TTA	6	129 848
	TAG	CTA	8	105 698
	AAG	CTT-CTC	10	426 803
Lys	CTT	AAG	17	362 046
	TTT	AAA	14	309 542
Met	CAT	ATG	26	263 141
Phe	GAA	TTT-TTC	17	473 767
Pro	AGG	CCT-CCC	12	285 305
	CGG	CCG	2	59 352
	TGG	CCA	15	197 855
Ser	AGA	TCT-TCC	11	388 324
	TGA	TCA	9	220 245
	CGA	TCG	3	73 268
	GCT	AGC-AGT	10	299 610
Thr	AGT	ACT-ACC	9	308 733
	CGT	ACG	2	54 088
	TGT	ACA	9	178 685
Trp	CCA	TGG	9	144 820
Tyr	GTA	TAT-TAC	10	309 437
Val	AAC	GTT	13	275 844
	CAC	GTG	7	206 903
	TAC	GTA	5	108 402
	GAC	GTC	1	126 708

3 Materials and Methods

3.1 Information of the gene sequence

The gene coding sequence (CDS) and amino acid sequence selected in this study were obtained from Phytozome JGI database. The sequence files used were *ppersica_298_v2.1.CDs.fa.gz* and *ppersica_298_V2.1.protein.fa.gz*.

3.2 Analysis of codon bias parameters

CondonW 1.4.4 (<http://codonw.sourceforge.net/>) statistics and analyzed the usage of CDS codons in peaches. (1) The codon was evaluated as a whole, and the frequency of A/ T/ C/ G (A3s, T3s, C3s, G3s), the GC content of the codon and the GC content of the codon (GC3s) were analyzed. (2) Relative synonymous codon usage (RSCU) was used to evaluate codon bias, and the effective number of codon (ENC) was calculated accordingly. The codon Adaptation index (CAI) was calculated. The calculation of RSCU and CAI refer to the method of Sharp and Li (1987). ENC was calculated according to Wright's (1990) method.

3.3 Enc-plot and PR2-plot analyzed the influence of factors on codon usage

Enc-plot was used to analyze whether peach codon usage was only affected by nucleotide neutral mutation or other factors. ENC values ranged from 20 to 61 and were negatively correlated with codon usage bias. The relevant GC3s was taken as the abscissa and ENC value was taken as the ordinate to conduct the ENC-plot analysis. It is generally believed that when the codon bias is only affected by neutral mutations, the genes will be evenly distributed along the standard curve or close to the expected curve in the ENC-plot distribution, while if the codon bias is affected by other non-neutral mutations, the distribution of these genes will deviate significantly from the expected curve.

A3s/(A3s+ T3s) and G3s/(G3s + C3s) were used as transverse and vertical coordinates for PR2-plot analysis. Theoretically, when the neutral mutation rate of DNA duplex and the selective pressure are not biased, the frequency of four nucleotides generally follows $A3s = T3s$ and $G3s = C3s$ (where $A3s + T3s + G3s + C3s = 1$) (Sueoka, 1995). At this time, the expression is $A3s+ T3s = G3s + C3s$, which should be at the center point in PR2-plot ($A = T, C = G$), that is, codon usage preference is not affected by selection pressure and is only caused by neutral mutations. On the contrary, the distribution of genes is not uniform, indicating that selection pressure may exist. In this study, the degree of PR2 bias was used to observe whether codon bias was affected by neutral mutations, selective pressure, or both (Sueoka, 2001).

3.4 Determination of the optimal codon

Referring to the references (Liu and Xue, 2005), codons with $RSCU > 1$ were defined as high-frequency codons based on the RSCU values of all genes calculated by CondoW. Then 26873 coding genes were sorted according to the ENC value of the coding genes. The top 5% and bottom 5% genes were selected to form the high expression and low expression libraries respectively, and their RSCU values were calculated respectively. When the difference RSCU between the two genes was greater than 0.08, the codon was defined as the high expression superior codon. The codon that satisfies both the above conditions (high frequency codon and high expression superior codon) is defined as the optimal codon in peach.

3.5 Comparison of genome codon usage between peach and other Rosaceae fruit species

The genome, gene coding sequence and corresponding amino acid sequence of apple, cherry, plum and 5 strawberry species were downloaded, which was released by GDR (GENOME DATADASE FOR ROSACEAE) database (<https://www.rosaceae.org/>). Then the corresponding codon usage frequency was calculated according to the above method. The codon usage frequency matrix of nine important horticultural crops of Rosaceae including peach, apple, cherry, plum and strawberry was constructed. Pearson correlation was used to analyze the correlation of codon usage frequency among different species, and to study whether there was any specific regularity in codon usage of related species.

3.6 Analysis of the effect of tRNA gene copy number on amino acid usage

It has been reported that the abundance of each tRNA isoacceptor can be approximately estimated by calculating

the number of RNA genes, and the latter is related to the frequency of amino acid and codon selection (Duret, 2000; Michaud et al., 2011). In this study, tRNAscan, with default parameter settings, was used to search for tRNA gene sequences in peach genome. The frequency (times) of all codons and corresponding amino acids were counted by Condow, and the correlation between tRNA gene content and corresponding amino acids and codon selection frequency was analyzed by regression analysis.

Authors' contributions

LRY and ZP were the experimental designers and executors of the study. LRY, ZP, ZXD and MXY completed the data analysis and wrote the first draft of the paper. GR, YSB and JG participated in the experimental design and analysis of the experimental results. ZP was the architect and responsible person of the project, and directed the experimental design, data analysis, paper writing and revision. All authors read and approved the final manuscript.

Acknowledgments

This research was co-funded by the Basic Research Project of Fujian Provincial Public Welfare Research Institute (2018R1013-13), the Special Funds for Construction of National Modern Agricultural Industry Technology System (CARS-30-Z-07) and the Innovation Team of Fujian Academy of Agricultural Sciences (Stit2017-1-4).

References

- Chiapello H., Lisacek F., Caboche M., and Hénaut A., 1998, Codon usage and gene function are related in sequences of *Arabidopsis thaliana*, *Gene*, 209(1-2): GC1-GC38.
[https://doi.org/10.1016/S0378-1119\(97\)00671-9](https://doi.org/10.1016/S0378-1119(97)00671-9)
PMid:9583944
- Duret L., 2000, tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes, *Trends in Genetics*, 16(7): 287-289.
[https://doi.org/10.1016/S0168-9525\(00\)02041-2](https://doi.org/10.1016/S0168-9525(00)02041-2)
PMid:10858656
- Huang Y., Koonin E., Lipman D., and Przytycka T., 2009, Selection for minimization of translational frameshifting errors as a factor in the evolution of codon usage, *Nucleic Acids Research*, 37(20): 6799-6810.
<https://doi.org/10.1093/nar/gkp712>
PMid:19745054 PMCID:PMC2777431
- Kawabe A., and Miyashita N., 2003, Pattern of codon usage bias in three dicot and four monocot plant species, *Genes. Genet. Syst.*, 78(5): 343-52.
<https://doi.org/10.1266/ggs.78.343>
PMid:14676425
- Liu H., Huang Y., Du X., Chen Z., Zeng X., Chen Y., and Zhang H., 2012, Patterns of synonymous codon usage bias in the model grass *Brachypodium distachyon*, *Genet. Mol. Res.*, 11(4): 4695-4706.
<https://doi.org/10.4238/2012.October.17.3>
PMid:23096921
- Liu Q., and Xue Q., 2005, Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species, *J. Genet.*, 84(1): 55-62.
<https://doi.org/10.1007/BF02715890>
PMid:15876584
- Luo H., Hu S.S., Wu Q., and Yao H.P., 2015, Analysis of buckwheat chloroplast gene codon bias, *Jiyinzuxue yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 34(11): 2457-2464.
- Michaud M., Cognat V., Duchêne A.M., and Maréchal-Drouard L., 2011, A global picture of tRNA genes in plant genomes, *Plant J.*, 66(1): 80-93.
<https://doi.org/10.1111/j.1365-3113X.2011.04490.x>
PMid:21443625
- Moriyama E.N., and Powell J.R., 1998, Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*, *Nucleic Acids Research*, 26(13): 3188-3193.
<https://doi.org/10.1093/nar/26.13.3188>
PMid:9628917 PMCID:PMC147681
- Olejniczak M., and Uhlenbeck O., 2006, tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition, *Biochimie.*, 88(8): 943-50.
<https://doi.org/10.1016/j.biochi.2006.06.005>
PMid:16828219
- Plotkin J., and Kudla G., 2011, Synonymous but not the same: the causes and consequences of codon bias, *Nat. Rev. Genet.*, 12(1): 32-42.
<https://doi.org/10.1038/nrg2899>
PMid:21102527 PMCID:PMC3074964

- Sharp P., and Cowe E., 1991, Synonymous codon usage in *Saccharomyces cerevisiae*, *Yeast*, 7(7): 657-78.
<https://doi.org/10.1002/yea.320070702>
PMid:1776357
- Sharp P., and Li W., 1987, The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications, *Nucleic Acids Research*, 15(3): 1281-1295.
<https://doi.org/10.1093/nar/15.3.1281>
PMid:3547335 PMCID:PMC340524
- Sueoka N., 1995, Intrastrand parity rules of DNA base composition and usage biases of synonymous codons, *J. Mol. Evol.*, 40(3): 318-25.
<https://doi.org/10.1007/BF00163236>
PMid:7723058
- Sueoka N., 2001, Near homogeneity of PR2-Bias fingerprints in the human genome and their implications in phylogenetic analyses, *J. Mol. Evol.*, 53(4-5): 469-476
<https://doi.org/10.1007/s002390010237>
PMid:11675607
- Wright F., 1990, The effective number of codons used in a gene, *Gene*, 87(1): 23-29.
[https://doi.org/10.1016/0378-1119\(90\)90491-9](https://doi.org/10.1016/0378-1119(90)90491-9)
PMid:2110097
- Wu Y., Zhao D., and Tao J., 2015, Analysis of codon usage patterns in herbaceous peony (*Paeonia lactiflora* Pall.) based on transcriptome data, *Genes*, 6(4): 1125-1139
<https://doi.org/10.3390/genes6041125>
PMid:26506393 PMCID:PMC4690031
- Ye Y.J., Ni Z.X., Bai T.D., and Xu L.A., 2018, The analysis of chloroplast genome codon usage bias in *Pinus massoniana*, *Jiyinzuxue yu Yingyong Shengwuxue (Genomics and Applied Biology)*, 37(10): 4464-4471.
- Zalucki Y., Power P., and Jennings M., 2007, Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins, *Nucleic Acids Res.*, 35(17): 5748-54.
<https://doi.org/10.1093/nar/gkm577>
PMid:17717002 PMCID:PMC2034453
- Zelasko S., Palaria A., and Das A., 2013, Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems, *Protein Expr. Purif.*, 92(1): 77-87.
<https://doi.org/10.1016/j.pep.2013.07.017>
PMid:23973802