

研究报告

Research Report

桃基因密码子使用模式及其偏好性分析

李若愚² 张小丹⁴ 马昕怡³ 郭瑞¹ 颜少宾¹ 金光¹ 周平^{1*}

1 福建省农业科学院果树研究所, 福州, 350013; 2 福建农林大学农学院, 福州, 350002; 3 福建农林大学生命科学学院, 福州, 350002; 4 伊利诺伊大学厄巴纳 - 香槟分校, 厄巴纳, IL61801

* 通信作者, fagaless@163.com

摘要 为深入研究桃密码子使用特征,本研究基于桃(*Prunus persica*)基因组 26 873 个注释基因编码序列的 GC 含量特征、有效密码子数(ENC)和同义密码子相对使用频率(RSCU)等参数的计算,统计分析了桃基因组密码子的偏好性及其密码子使用模式。分析结果表明桃密码子使用具有偏好性,61 个密码子中存在 4 个密码子 TCA, ACA, GCA, GAA 为桃最优密码子,它们均以 A 结尾。进一步统计桃与其它 9 种近缘蔷薇科植物的密码子使用频率,发现近缘种属的密码子使用规律具有一定相似性。研究同时发现桃基因组 tRNA 基因拷贝数与氨基酸及特定密码子的出现频度呈正相关。这些结果揭示了桃基因组密码子的使用模式,为相关密码子使用机制研究及基因工程改良提供了重要的参考。

关键词 桃; 密码子使用模式; 密码子使用偏好性

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 Ping^{1*}

1 Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, 350013; 2 College of Agricultural, Fujian Agriculture and Forestry University, Fuzhou, 350002; 3 College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou, 350002; 4 Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801

* Corresponding author, fagaless@163.com

DOI: 10.5376/mpb.cn.2020.18.0050

Abstract To dissect the codon usage characteristic, our 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 analyses showed that there were obvious biases in codon usage, 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. Our 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

本文首次发表在《分子植物育种》上, 现依据版权所有人授权的许可协议, 采用 Creative Commons Attribution License, 协议对其进行授权, 再次发表与传播

收稿日期: 2020 年 11 月 4 日; 接受日期: 2020 年 11 月 4 日; 发表日期: 2020 年 11 月 4 日

引用格式: 李若愚, 张小丹, 马昕怡, 郭瑞, 颜少宾, 金光, 周平, 2020, 桃基因密码子使用模式及其偏好性分析, 分子植物育种 (网络版) 18(50): 1-9 (doi: 10.5376/mpb.cn.2020.18.0050) (Li R.Y., Zhang X.D., Ma X.Y., Guo R., Yan S.B., Jin G., and Zhou P., 2020, Analyses of codon usage patterns and codon usage bias in peach (*Prunus persica*), Fenzi Zhiwu Yuzhong (Molecular Plant Breeding (online)), 18(50): 1-9 (doi: 10.5376/mpb.cn.2020.18.0050))

在生物体中同义密码子对应同一氨基酸。但有些同义密码子的使用频度不同,个别密码子被频繁使用,称为“密码子使用偏好性”现象(Plotkin and Kudla, 2011)。密码子使用的偏好性受到多种因素的影响,如 DNA 复制起始位点(Huang et al., 2009),蛋白质翻译效率(Zalucki et al., 2007),tRNA 丰度(Olejniczak and Uhlenbeck, 2006),基因长度(Moriyama and Powell, 1998)等。密码子使用偏好性是由复杂的机制决定的,它反映了物种间的遗传差异。研究密码子偏好性对研究基因高效表达,预测新基因以及探究相关基因演化具有重要的参考价值(Wu et al., 2015)。随着基因组大规模测序时代的来临,越来越多的动植物基因组被破译,使得全面、系统地分析相关物种密码子偏好性成为可能。

桃(*Prunus persica* L.)是重要的落叶果树。由于其农艺性状多样,成花结果周期短,遗传连锁图谱精密,已成为果树学和遗传学研究的模式树种之一。目前桃基因组已测序公布,然而其密码子的使用偏好性研究还鲜见报道。本研究统计分析 26 873 个注释编码基因的序列特征和密码子规律,探究了影响桃基因组密码子使用的因素。所得结果为研究桃基因组分子演化提供参考,也为进一步开展基因结构分析及基因工程改良研究提供帮助。

1 结果与分析

1.1 密码子使用参数分析

分析 26 873 个注释基因发现 T3s, C3s, A3s, G3s 含量的变化范围分别为 5.1%~70.3% (平均 38.9%); 1.7%~83.5% (平均 26.8%); 2.7%~57.5% (平均 31.8%); 0~72.5% (平均 28.1%)。ENC 值的变化范围为 22.45~61 (平均 52.5)。初步判定 26 873 个基因具有不同程度的密码子使用偏好性(表 1)。

利用 CAI (密码子适应指数)评估同义密码子和密码子最优使用的匹配程度,结果表明 CAI 变化范围为 6.2%~59.2% (平均 20.2%)。同时统计桃基因组编码基因 GC3s 以及密码子 GC 含量,其变化范围分别为 13.8%~90.1%和 23.9%~73.4%,平均值为 42.8%和 45.4%。进一步绘图展示 26 873 个编码基因密码子各自 GC 和 GC3s 含量(图 1),发现整体趋势线偏离 GC=GC3s 斜线,证实桃密码子具有一定偏好性,这与前人结果中双子叶植物的密码子偏好性使用一致(Chiappello et al., 1998; Kawabe and Miyashita, 2003)。

1.2 ENC-plot、PR2-plot 分析分析密码子使用偏好性

统计有效密码子数,利用 ENC-plot 可以直观地展示了同义密码子使用的非均衡程度,衡量同义密码子的使用与随机选择的偏离程度。ENC 其值介于 20~61,最小值 20 表示每个氨基酸只使用同一种密码子,最大值 61 表示 61 个密码子均被使用,ENC 值越低则密码子使用偏好性越强。以 GC3s 和 ENC 为横纵坐标,ENC-plot 分析密码子使用偏好性是否是密码子第 3 位核苷酸变异(中性突变)的结果(Liu et al., 2012),结果显示有部分基因趋近理论趋势线,部分基因点偏离趋势线(图 2)。该结果表明除核苷酸中性突变影响外还存在其他因素也同时影响着这些基因的密码子使用偏好性。

PR2-plot 分析以 A3s/(A3s+T3s)和 G3s/(G3s+C3s)为横纵坐标,分析密码子第 3 位上核苷酸碱基组成情况,通过图中中心点(A=T; C=G)发出的矢量直观地显示偏离碱基的类型和偏离程度。通常情况下,若 DNA 双链的中性突变率和所受的选择压力没

表 1 密码子组成及使用参数

Table 1 Constitution and usage parameters of codon

参数 Parameter	变化范围 Variation range	平均值±标准差 Mean±SD
T3s (%)	5.10~70.3	38.9±7.3
C3s (%)	1.70~83.5	26.8±8.1
A3s (%)	2.70~57.5	31.8±7.1
G3s (%)	0.00~72.5	28.1±6.7
CAI (%)	6.20~59.2	20.2±3.2
ENC	22.45~61.0	52.5±4.7
GC3s (%)	13.80~90.1	42.8±9.2
GC 含量(%) GC content (%)	23.90~73.4	45.4±4.5

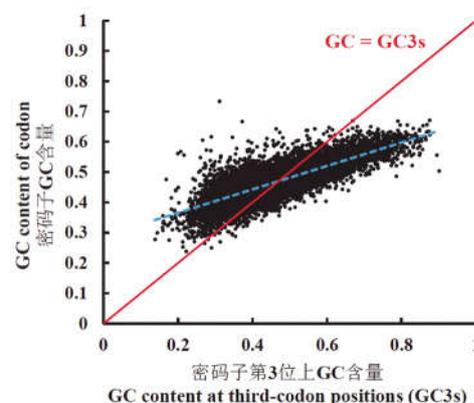


图 1 26 873 个编码基因密码子相应 GC 和 GC3s 分析
Figure 1 GC and GC3s analyses of codon use features from 26 873 protein-encoding genes

有偏差, A3s、T3s、G3s、C3s 总体上遵循 A3s=T3s 和 G3s=C3s 规律, 则在 PR2-plot 图中表现为基因聚集分布于中心处(图 3)。但通过对桃所有基因的 PR2-plot 分析发现, 基因整体表现出偏离中心点的趋势, 即对密码子第 3 位碱基的使用存在偏好性。这表明桃密码子偏好性除了受已知的核苷酸中性突变影响外, 应该还受其它因素影响, 如选择压力等。

1.3 最优密码子分析

表中列出了全部基因的 RSCU 值(表 2), RSCU>1 的高频密码子共有 27 个, 分别为 TTT, TTG, CTT, ATT, GTT, GTG, TCT, TCA, CCT, CCA, ACT, ACA, GCT, GCA, TAT, CAT, CAA, AAT, AAG, GAT, GAA, TGT, TGA, AGA, AGG, GGT, GGA。其中 23 个以 A/T 结尾, 4 个以 G/C 结尾。其中偏好性最强的密码子为 AGA, 其 RSCU 值为 1.85。如表中列出了根据 ENC 值划分的高表达库与低表达库的 RSCU 值(表 3), 根据二者差值 RSCU>0.08 的标准, 共鉴定出 CTA, CTG, ATA, GTA, TCA, TCG, ACA, ACG, GCA, GAA,

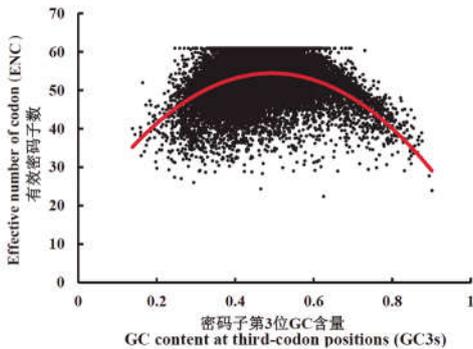


图 2 桃基因组编码基因 ENC-plot 曲线
注: 黑色散点代表单个基因, 曲线为理论趋势线
Figure 2 ENC-plot curve of peach genomic coding genes
Note: Black scatter means a single gene, and the curve indicates a theoretic trend

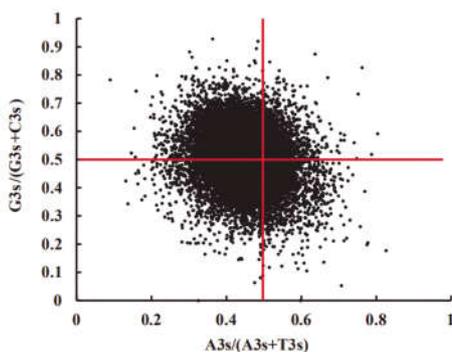


图 3 桃基因组编码基因 PR2-plot 分析
Figure 3 PR2-plot analysis of peach genomic coding genes

TGA, CGT, CGA, CGG, AGT, AGC 等 16 个高表达优越密码子, 其中 11 个以 A/T 结尾, 5 个以 G/C 结尾。依据高频密码子与高表达优越密码子结果确定 4 个最优密码子, 分别为 TCA, ACA, GCA, GAA, 它们均以 A 结尾。

1.4 比较桃与蔷薇科近缘物种的密码子使用频率

不同物种间密码子使用偏好性是不同的。Pearson 相关性分析 10 个物种 61 种密码子使用频度以衡量各物种密码子使用方式的相似性。计算结果表明苹果、桃、樱、梅(密码子使用频率)间的相关系数高于 0.99, 草莓属内 5 种作物(密码子使用频率)间的相关系数高于 0.95, 但苹果、桃、樱、梅与草莓属作物(密码子使用频率)相关系数普遍低于 0.75。该结果间接表明苹果-桃-樱-梅密码子使用方式和频率近似, 草莓属内 5 种作物密码子使用相似(图 4)。根据结果推测苹果属 *Malus* (如苹果 *M. x domestica*)、梅属 *Prunus* (如桃 *P. persica*, 梅 *P. mume*)和樱属 *Cerasus* (如日本樱花 *C. x yedoensis*)密码子使用规律相似, 与草莓属 *Fragaria* (如饭沼草莓 *F. iinumae*, 日本草莓 *F. nipponica*, 西藏草莓 *F. nubicola*, 东方草莓 *F. orientalis*, 凤梨草莓 *F. x ananassa*, 森林草莓 *F. vesca*)密码子使用有一定差别。

1.5 tRNA 基因使用影响分析

生物体由于生命活动的需要能以最快的速度合成需要的蛋白质, 通常高丰度 tRNA 与最优密码子具有对应性(Michaud et al., 2011)。研究证实每种氨基酸的转运最多需要五个具有不同反密码子的 tRNA, 它们被称为 tRNA 同工受体(tRNA isoacceptor), 对应相同氨基酸的 tRNA 同工受体被划分为一个家族, 每个物种 tRNA 同工受体的数量约为 45~47 个。在生物中, tRNA 基因的数量可以很好地近似估算每个

苹果	0.99	0.99	0.99	0.72	0.72	0.71	0.71	0.70	0.70
<i>M. x domestica</i>									
桃	1.00	0.99	0.73	0.73	0.73	0.73	0.73	0.73	0.73
<i>P. persica</i>									
梅	0.99	0.74	0.74	0.73	0.73	0.73	0.73	0.73	0.73
<i>P. mume</i>									
日本樱花	0.70	0.70	0.69	0.69	0.70	0.69	0.70	0.69	0.69
<i>C. x yedoensis</i>									
饭沼草莓	1.00	1.00	1.00	0.96	0.98				
<i>F. iinumae</i>									
日本草莓	1.00	1.00	0.95	0.98					
<i>F. nipponica</i>									
西藏草莓	1.00	0.96	0.99						
<i>F. nubicola</i>									
东方草莓	0.95	0.99							
<i>F. orientalis</i>									
凤梨草莓	0.98								
<i>F. x ananassa</i>									
森林草莓									
<i>F. vesca</i>									

图 4 基于 Pearson 相关性分析 10 个近缘种的密码子使用频度
Figure 4 Pearson correlation analysis of codon usage frequency in 10 relative species

表 2 同义密码子使用情况

Table 2 Usage of synonymous codon

氨基酸	密码子	同义密码子的相对使用度	百分比(%)	氨基酸	密码子	同义密码子的相对使用度	百分比(%)
Amino acid	Codon	RSCU	Percentage (%)	Amino acid	Codon	RSCU	Percentage (%)
Phe	TTT*	1.14	2.46	Tyr	TAT*	1.16	1.62
Leu	TTC	0.86	1.84		TAC	0.84	1.18
	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
	CTG	0.84	1.38	Gln	CAA*	1.09	2.03
Ile	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		AAG*	1.08	3.28
	GTC	0.71	1.15	Asp	GAT*	1.32	3.41
	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	0.62
	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

注: *: 高频密码子

Note: *: High frequency codon

tRNA 同工受体的丰度,进而与氨基酸频率及相关密码子的使用相关(Duret, 2000; Michaud et al., 2011)。

统计桃基因组中 tRNA 基因数量与 26 873 个蛋白中相应氨基酸的出现次数(氨基酸频度),观察到 tRNA 基因数量与氨基酸频度呈线性正相关(图 5)。由于在生物体内大多数 tRNA 可以识别不止一个密码子,而某些密码子可以被不止一个同位受体 tRNA 识别,在假设密码子仅优先被一个 tRNA 同源受体读取的情况(即最小潜在密码子识别模式),统计分析桃

基因组中 tRNA 同工受体基因数量与密码子使用之间的关系。结果表明,除去个别密码子,大部分密码子在编码基因中出现的总次数与相应的 tRNA 同工受体基因之间也存在一定线性相关性(图 6; 表 4)。综合考虑上述两个结果认为桃基因组 tRNA 基因的数量对氨基酸选用及密码子的偏好性具有一定的影响。

2 讨论

密码子偏好性广泛存在于多种生物中,是一种

表 3 高, 低表达密码子的同义密码子的使用频率

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

氨基酸	密码子	高表达(RSCU)	低表达(RSCU)	氨基酸	密码子	高表达(RSCU)	低表达(RSCU)
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
	CTA	0.64	0.44		CAC	0.88	0.93
Ile	CTG	0.93	0.84	Gln	CAA	1.09	1.09
	ATT	1.25	1.26		CAG	0.91	0.91
	ATC	0.97	1.12	Asn	AAT	1.05	1.04
Met	ATA	0.79	0.61		AAC	0.95	0.96
	Val	ATG	1	1	Lys	AAA	0.89
GTT		1.28	1.32	AAG		1.11	1.14
GTC		0.9	0.98	Asp	GAT	1.18	1.16
GTA		0.63	0.43		GAC	0.82	0.84
GTG		1.19	1.26		Glu*	GAA	1
Ser	TCT	1.27	1.47	GAG		1	1.1
	TCC	1.02	1.26	Cys	TGT	0.96	0.99
	TCA*	1.18	1.08		Trp	TGC	1.04
	TCG	0.67	0.55	Arg	TGG	1	1
Pro	CCT	1.26	1.19		CGT	0.85	0.61
	CCC	0.83	0.92	CGC	0.77	0.86	
	CCA	1.29	1.33	CGA	0.84	0.52	
	CCG	0.62	0.56	CGG	0.73	0.49	
	Thr	ACT	1.16	1.13	Ser	AGT	0.87
ACC		1.05	1.33	AGC		0.99	0.88
ACA*		1.15	1.06	Arg	AGA	1.47	1.82
ACG		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
	GCA*	1.18	0.94		GGA	1.11	1.06
	GCG	0.5	0.45		GGG	0.88	0.81

注: *: 最优密码子; 加下划线的密码子为高表达优越密码子

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

不可避免的复杂现象。目前密码子偏好性已经在细胞器, 基因家族和全基因组范围内广泛被研究(罗洪等, 2015; 叶友菊等, 2018)。本研究基于桃基因组信息分析了密码子使用模式和可能的形成因素。桃基因组密码子特征的统计结果表明 26 873 个基因具有不同程度的密码子使用偏好性。ENC-plot 与 PR2-plot 分析表明桃密码子偏好性在受到核苷酸中性突变影响的同时, 选择压力也可能是影响密码子偏好性的重

要因素。同时, 基因的表达水平同样影响密码子的偏好性, 分析鉴定得出的 4 个最优密码子 TCA、ACA、GCA、GAA, 均以 A 结尾, 因此在基因工程翻译表达相关氨基酸时, 可以针对性地选用最优密码子以获得更好的表达效果。此外, 本研究发现桃基因组 tRNA 基因含量与相应氨基酸和密码子的使用频度具有一定相关性, 从而证实基因组 tRNA 基因的含量也可能影响氨基酸及密码子的偏好性, 该结果与拟南芥、

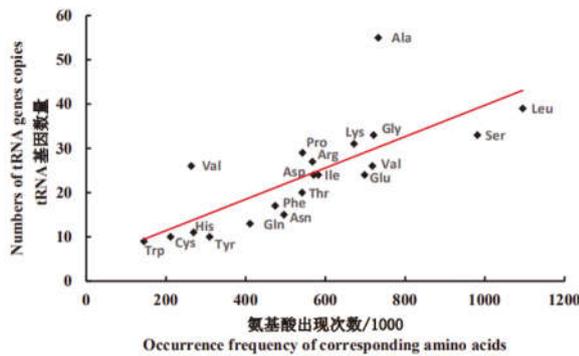


图5 每种氨基酸特异的 tRNA 基因拷贝数与相应氨基酸出现次数之间的相关性

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

水稻的相关研究报道一致(Michaud et al., 2011)。

总而言之,本研究的一系列分析为研究桃基因组密码子偏好性的潜在因素提供了基础且全面的理解。这些研究将对指导功能基因的外源表达具有重要意义(Zelasko et al., 2013)。研究密码子的偏好性和密码子使用频率有助于从另一视角了解相关物种的遗传演化规律。本研究结果表明蔷薇科苹果-桃-樱-梅(属核果类、仁果类,多年生木本)密码子使用方式和频率近似,草莓属内5种作物(属浆果类,多年生草本)密码子使用相似,但苹果-桃-樱-梅与草莓类差异相对较大,这符合植物形态学、系统分类学和分子进化学的一般认知,说明所检测的桃等10种蔷薇科植物近缘物种密码子使用偏好性和使用频率存在特定特征,亲缘性越近,物种密码子使用规律越类似。基于密码子特征的分析同样在分子水平也反映了物种演化关系。这些结果对蔷薇科及其它科属内近缘植物新编码基因预测以及外源基因工程改良提供重要的参考(Sharp and Cowe, 1991)。

3 材料与方法

3.1 基因序列信息

本研究选取的基因编码序列(CDS)和氨基酸序列均自 Phytozome JGI 数据库获得。所用序列文件为 Ppersica_298_v2.1.cds.fa.gz 与 Ppersica_298_v2.1.protein.fa.gz。

3.2 密码子偏好性参数的分析

CondonW 1.4.4 (<http://codonw.sourceforge.net/>) 统计分析桃子 CDS 密码子的使用情况。(1)对密码子进行整体评价,分析密码子第3位上4种碱基 A/T/

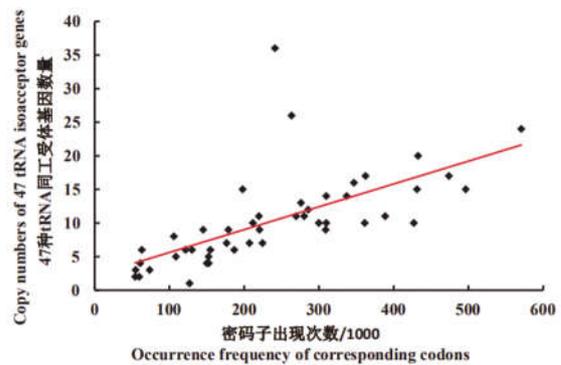


图6 桃基因组中 tRNA 同工受体的基因拷贝数与相应密码子出现次数之间的相关性分析

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

C/G 的出现频率(A3s, T3s, C3s, G3s)、密码子的 GC 含量和第3位上 GC 含量(GC3s)等。(2)采用同义密码子的相对使用度(relative synonymous codon usage, RSCU)对密码子偏好性进行评价,并针对性统计有效密码子数(effective number of codon, ENC),计算密码子适应指数(codon adaptation index, CAI)。其中 RSCU 和 CAI 的计算参照 Sharp 和 Li (1987)的方法。ENC 的计算参照 Wright (1990)的方法。

3.3 ENC-plot 和 PR2-plot 分析密码子使用所受因素影响

ENC-plot 分析桃密码子使用是否只受核苷酸中性突变影响,或者还存在其它因素影响。ENC 值范围为 20 至 61,与密码子使用偏差负相关。以相关 GC3s 为横坐标,ENC 值为纵坐标进行 ENC-plot 作图分析。一般认为,当密码子偏好性仅受中性突变影响时,在 ENC-plot 分布图上基因将沿标准曲线或靠近预期曲线均衡分布,而如果密码子偏好性受其它非中性突变影响时,这些基因分布将显著偏离预期曲线。

以 $A3s/(A3s+T3s)$ 和 $G3s/(G3s+C3s)$ 为横纵坐标进行 PR2-plot 作图分析。理论上认为:当 DNA 双链的中性突变率和所受选择压力无偏差时,四个核苷酸的出现频率总体上遵循 $A3s=T3s$ 和 $G3s=C3s$ (其中 $A3s+T3s+G3s+C3s=1$) (Sueoka, 1995)。此时表现为 $A3s+T3s=G3s+C3s$, 在 PR2-plot 图中应该处于中心点(A=T, C=G),即密码子使用偏好不受选择压力影响仅由中性突变引起。反之,基因分布不均匀,则说明选择压力可能存在。本研究利用 PR2 的偏差程度观察密码子偏好性是否受中性突变,选择压力或两者共同影响(Sueoka, 2001)。

表 4 最小潜在密码子识别模式下 tRNA 同工受体基因数与密码子使用频率的关系

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

氨基酸 Amino acid	反密码子 Anticodon	密码子 Codon	tRNA 同工受体基因拷贝数 Copy number of tRNA isoacceptor gene	密码子数 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
Asn	TCT	AGA	7	176 349
	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

3.4 最优密码子的确定

参照文献(Liu and Xue, 2005)。根据 CondoW 计算获得的全部基因的 RSCU 值, 将 RSCU>1 的密码子定义为高频密码子。此后根据编码基因的 ENC 值对 26 873 个编码基因进行排序, 取前 5% 和后 5% 的基因分别组成高表达和低表达库, 分别计算其 RSCU 值, 当二者差值 RSCU>0.08 时, 则定义该密码子为高表达优越密码子。同时满足上述条件(高频密码子和高表达优越密码子)的密码子定义为桃的最优密码子。

3.5 桃与其他蔷薇科果树物种基因组密码子使用情况比较

下载 GDR (GENOME DATADASE FOR ROSACEAE) 数据库 (<https://www.rosaceae.org/>) 公布的苹果、樱、梅及 5 种草莓属物种基因组、基因编码序列和相应氨基酸序列, 依照上文方法计算相应的密码子使用频率。构建桃、苹果、樱、梅及草莓等 9 个蔷薇科重要园艺作物的密码子使用频度矩阵, 皮尔森 (Pearson) 相关性分析不同物种间密码子使用频度的相关性, 研究相关物种密码子使用是否存在特定的规律性。

3.6 tRNA 基因拷贝数对氨基酸使用影响分析

报道表明可以通过计算 RNA 基因的数量近似估算每个 tRNA 同工受体(tRNA isoacceptor)丰度, 而后者与氨基酸以及密码子的选用频率相关(Duret, 2000; Michaud et al., 2011)。本研究利用 tRNAscan, 默认参数设定, 搜索桃基因组的 tRNA 基因序列。CondoW 统计全部密码子及对应氨基酸的使用频次(次数), 回归分析 tRNA 基因含量与对应氨基酸以及密码子选用频率的相关性。

作者贡献

李若愚、周平是本研究的实验设计和实验研究的执行人; 李若愚、周平、张小丹及马昕怡完成数据分析, 论文初稿的写作; 郭瑞, 颜少宾和金光参与实验设计, 试验结果分析; 周平是项目的构思者及负责人, 指导实验设计, 数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由福建省省属公益类科研院所基本科研专项(2018R1013-13)、国家现代农业产业技术体系建设专项资金(CARS-30-Z-07)和福建省农业科学院

创新团队(STIT2017-1-4)共同资助。

参考文献

- 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
- 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
- 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
- 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
- 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
- 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
- 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 (罗洪, 胡莎莎, 吴琦, 姚慧鹏, 2015, 甜荞叶绿体基因密码子偏爱性分析, *基因组学与应用生物学*, 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
- 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
- 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
- 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
- Sharp P., and Cowe E., 1991, Synonymous codon usage in *Saccharomyces cerevisiae*, *Yeast*, 7(7): 657-78
- 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
- Sueoka N., 1995, Intrastrand parity rules of DNA base composi-

- tion and usage biases of synonymous codons, *J. Mol. Evol.*, 40(3): 318-25
- 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
- Wright F., 1990, The effective number of codons used in a gene, *Gene*, 87(1): 23-29
- 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
- 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 (叶友菊, 倪州献, 白天道, 徐立安, 2018, 马尾松叶绿体基因组密码子偏好性分析, *基因组学与应用生物学*, 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
- 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