

几种植物花分生组织决定基因 *APETALA1* 密码子使用模式比较

吴彦庆^{1,2}, 李志远^{1,2}, 赵大球², 陶俊^{2,*}

(¹扬州大学动物科学与技术学院, 江苏扬州 225009; ²扬州大学园艺与植物保护学院, 江苏省作物遗传生理重点实验室, 江苏扬州 225009)

摘要: 为了揭示花分生组织决定基因 *API* (*APETALA1*) 的分子机制以及物种间的进化关系, 从 GenBank 数据库挑选 8 个不同植物 *API* 基因编码序列, 利用 R 软件分别计算 GC 含量、有效密码子数 (Effective number of codons, ENC)、同义密码子相对使用度 (Relative synonymous codon usage, RSCU)、相对密码子使用偏好性 (Relative codon usage bias, RCBS) 等参数, 分析比较不同植物 *API* 基因密码子使用模式中密码子偏好性和碱基组成动力学。结果表明, 不同植物 *API* 基因密码子使用受到 GC 偏好性 (GC bias) 影响, 特别是 GC3s; 大多数最优密码子偏好使用以 A/U 结尾, 其中不同植物 *API* 基因中 ACA、AUA 和 CAG 均表现为较高的 RSCU 值, 为最优势密码子; 芍药 (*Paeonia lactiflora*) 和牡丹 (*Paeonia suffruticosa*) *API* 基因的密码子使用特点相似, 碧桃 (*Prunus persica*) 和苹果 (*Malus × domestica*) 相似; 较低的最优密码子频率 Fop 值和较高的 ENC 值暗示了不同植物 *API* 基因在密码子使用中具有一个较弱的偏好性。

关键词: *APETALA1*; 植物; 密码子使用偏好性; 碱基组成动力学

中图分类号: S 68

文献标志码: A

文章编号: 0513-353X (2017) 03-0504-11

Comparative Analysis of Codon Usage Patterns in *APETALA1* Gene Across Several Plant Species

WU Yanqing^{1,2}, LI Zhiyuan^{1,2}, ZHAO Daqiu², and TAO Jun^{2,*}

(¹College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu 225009, China; ²Key Laboratory of Crop Genetics and Physiology, College of Horticulture and Plant Protection, Yangzhou University, Yangzhou, Jiangsu 225009, China)

Abstract: To reveal the molecular mechanism of *APETALA1* (*API*) gene and the evolution among different plants, we selected the complete nucleotide coding sequences of *API* gene from 8 different plant species for CUB analysis. Using R software, we performed a comparative analysis of codon bias and compositional dynamics of codon usage patterns in *API* gene across eight different plant species by determining the GC contents, several genetic indices namely effective number of codons (ENC), relative synonymous codon usage (RSCU), and relative codon usage bias (RCBS) etc. Results showed codon usage in *API* gene in plants were influenced by GC bias, mainly due to GC3s. The majority of optimal codons were A/U ending for the corresponding amino acid, of which ACA, AUA and CAG were over-represented

收稿日期: 2016-11-05; **修回日期:** 2017-03-14

基金项目: 国家自然科学基金项目 (31372097, 31400592); 江苏省高校自然科学研究重大项目 (13KJA210005)

* 通信作者 Author for correspondence (E-mail: taojun@yzu.edu.cn)

codons in the *API* gene of all the selected plant species. The codon usage pattern for *API* in *Paeonia lactiflora* showed resemblance to that of *Paeonia suffruticosa*; similarly, and *Prunus persica* to *Malus × domestica*. Low Fop (Frequency of optimal codon) and high ENC values suggested a weaker codon bias observed for the *API* gene in different plant species.

Keywords: *APETALA1*; plant; codon usage bias; compositional dynamics

某一特定基因在翻译中编码相同氨基酸的同义密码子 (synonymous codon) 不均一使用情况被称为密码子使用偏好性 (codon usage bias, CUB)。CUB 是基因组的一个重要特征, 它可以随着某一基因组中大部分基因或者特定基因变化而改变 (Duret & Mouchiroud, 1999; Supek & Vlahovicek, 2005)。随着新一代高通测技术的出现, 不同物种的全基因组信息在美国国家生物技术信息中心 (NCBI) GenBank 数据库中已经公布, 吸引了许多科学家开始研究某一物种或者特定基因的 CUB。以往, 少数研究主要分析原核生物及低等真核生物中密码子偏性对基因表达的影响 (Gustafsson et al., 2004), 随后开始关注高等动植物以及利用密码子使用模式解释不同物种多样性 (蒋玮 等, 2014; Ma et al., 2015; 杨国锋 等, 2015; 张涛 等, 2015)。密码子偏好性研究具有重要的生物学意义, 其中基于核苷酸组成动力学和基因水平转移分析 CUB 不仅可以了解分子水平下的进化过程, 也可以设计转基因用于提高其外源表达和发现新基因 (Carbone et al., 2003), 分析基因表达的功能保守性 (Lithwick & Margalit, 2005)。由于同义密码子的转录特性不均一导致了 mRNA 转录本产生的蛋白数量可能发生显著变化 (齐婷 等, 2014), 由此进一步表明 CUB 与高表达基因的密码子使用情况密切相关 (Sueoka, 1988)。此外, 研究表明一个基因密码子使用模式受氨基酸序列、突变压力以及自然选择共同影响 (Mandlik et al., 2014)。受选择压力作用, 基因的密码子偏好性变异可能出现在相同属的不同物种之间。

花的发生式样、发育过程以及结构形成都具有较强的遗传稳定性, 是被子植物系统进化与发育研究的主要依据, 也是经典分类学家所关注的重要特征。因此, 花器官是研究植物发育、基因和进化三者关联的理想模型系统 (曾英 等, 2001)。目前已有研究主要集中在芍药 (*Paeonia lactiflora*)、拟南芥 (*Arabidopsis thaliana*)、牡丹 (*Paeonia suffruticosa*)、草莓 (*Fragaria × ananassa*) 等模式植物上, 已分离到一系列调节花发育的重要基因 (Mandel et al., 1992; 任磊, 2011; 邹冬梅 等, 2012; Ge et al., 2014), 其中花分生组织决定基因 *API* (*APETALA1*) 属于花器官发育 ABCDE 模型成员之一, 可以调控花序分生组织转变为花分生组织, 激活 B 类基因的表达, 决定花瓣和萼片的形成, 具有影响花器官形态特征的功能 (Mandel et al., 1992)。因此, 本研究中通过测定核苷酸组成 (GC 含量) 和遗传指数, 如有效密码子数 (Effective number of codons, ENC)、同义密码子相对使用度 (Relative synonymous codon usage, RSCU)、相对密码子使用偏好性 (Relative codon usage bias, RCBS) 等, 分析和比较 8 个不同植物 *API* 基因密码子使用模式和碱基组成动力学, 以期理解基因结构和功能以及不同物种基因的进化关系提供参考。

1 材料与方法

1.1 序列数据来源

API 基因完整的编码序列 (以 AUG 为起始密码子, 以 UAA、UAG 或 UGA 为终止密码子) 取自 GenBank 数据库 (<http://www.ncbi.nlm.nih.gov>), 用于密码子偏好性 CUB 分析 (表 1)。

表 1 8 个不同植物 *API* 基因登陆号以及编码序列长度
Table 1 Eight plant species with accession number and length (bp) of coding sequences for *API* gene

编号 Coding sequence number	植物 Plant	登录号 Accession No.	长度/bp Length
1	芍药 <i>Paeonia lactiflora</i>	KC354376	729
2	牡丹 <i>Paeonia suffruticosa</i>	HM143943	729
3	杨树 <i>Populus trichocarpa</i>	XM_002316040	747
4	碧桃 <i>Prunus persica</i>	EU079377	717
5	苹果 <i>Malus × domestica</i>	EU672877	720
6	葡萄 <i>Vitis vinifera</i>	GU133634	726
7	月季 <i>Rosa hybrida</i>	FJ970027	744
8	芒果 <i>Mangifera indica</i>	GQ152892	744

1.2 碱基组成偏好性分析

利用 Mobyle 软件 (Peden, 1999) (<http://www.molbiol.ox.ac.uk/cu>, version 1.4.2) 计算每个基因序列的碱基组成, 并统计: (I) 第 3 位密码子各碱基含量 (A3s、U3s、C3s、G3s); (II) 第 1 密码子 GC 含量 (GC1) 与第 2 位密码子 GC 含量 (GC2); (III) 密码子整体 GC 含量 (GC)、第 1 与 2 位密码子的平均 GC 含量 (GC12); (IV) 第 3 位密码子的 GC 含量 (GC3s)。此外, 分析每个编码序列的 AU、GC 和 GC3s 偏度 (skewness), 来评估碱基组成偏好性与基因表达水平的关系。

1.3 有效密码子分析

有效密码子 (Effective Number of Codon, ENC) 是一个基因的密码子使用频率与同义密码子平均使用频率偏差的量化值, 范围在 20 到 61 之间, 越靠近 20 偏好性越强。此值可以反映同义密码子非均衡使用的偏好程度 (Wright, 1990; Fuglsang, 2004)。高表达基因其密码子偏爱程度也大, 从而 ENC 值较小; 低表达基因则含有较多种类的稀有密码子, ENC 值也较大, 所以, 当前普遍通过比较 ENC 来确定内源基因表达量的相对高低。ENC 值越小, 对应的内源基因往往表达量也越高。

1.4 相对同义密码子使用度分析

相对同义密码子使用度 (Relative Synonymous Codon Usage, RSCU) 是该密码子的实际观察值与理论观察值之间的比值。理论观察值为各同义密码子使用频率相同时的观察值, 此时不存在密码子偏性。RSCU = 1, 密码子使用无偏性; RSCU > 1, 密码子出现的频率比其它同义密码子高, 反之, 出现频率低 (Sharp et al., 1986)。

1.5 最优密码子频率分析

最优密码子是指在某物种高表达基因中使用频率最高的密码子, 也有人将一个氨基酸的最优密码子定义为具有最大数量的反密码子 tRNA 基因的密码子。最优密码子频率 (Frequency of Optimal Codon, Fop) 是衡量基因密码子使用偏好性的一个常用指标 (Ikemura, 1985), 代表最优密码子数占有同义密码子的比例 (Ikemura, 1981), 其数值在 0.36 (基因具有相同的密码子偏好性) 到 1 (基因具有很强的密码子偏好性) 范围内变化 (Stenico et al., 1994)。每个编码序列的 Fop 值按照 Lavner 和 Kotler (2005) 提供的公式计算。

1.6 基因表达评估分析

基因表达通过相对密码子使用偏好性 (Relative Codon Usage Bias, RCBS) 进行评估 (Karlin & Mrazek, 1996)。每个 *API* 基因序列的 RCBS 值按照 Fox 和 Erill (2010) 提供的公式进行计算。

1.7 统计分析

利用 Microsoft Excel 和 SPSS 18.0 (<http://www.spss.com/>) 软件基于 Spearman's rank 方法进行相关性分析和方法显著性比较; 基于不同物种 *API* 基因编码序列的 Kimura 2-parameter (K2P) 距离构建邻近进化树 (Neighbor Joining, NJ), Bootstrap 检验采用 1 000 重复, 总共 8 个 CDS 序列, 包含空缺口和丢失数据的序列被淘汰, 进化分析采用 MEGA5.0 软件; 利用 R 软件基于层次聚类 (Hierarchical Clustering) 方法进行聚类热图 (Heat Map) 绘制, 分析密码子与 GC3 相关性系数和不同植物 RSCU 值。

2 结果与分析

2.1 不同植物 *API* 的密码子与 GC 偏好性 (GC3s) 的关系

利用聚类热图分析了密码子碱基组成与 GC 偏好性 (GC3s) 的相关性 (图 1)。

结果中, 大多数以 G/C 结尾的密码子的碱基组成与 GC 偏好性表现为正相关, 而大多数以 A/U 结尾的密码子与 GC 偏好性表现为负相关。

有趣的是, 3 个密码子 ACA (苏氨酸)、AUA (异亮氨酸) 和 CAG (谷氨酰胺) 均与 GC3s 表现为极显著的相关性 ($P < 0.01$), 表明该密码子使用受 GC3s 偏好性影响。

2.2 不同植物 *API* 基因的密码子碱基组成和最优密码子测定

分析了不同植物 *API* 编码序列的碱基组成情况 (表 2), 其中 A 平均数量最高 (233.3), 其次是 G (184.1), U (165.6) 和 C (148.9)。平均 AU 和 GC 含量分别为 54.5% 和 45.5%, 因此, 整体碱基组成分析暗示不同植物 *API* 基因编码序列密码子中 AU 使用频率高于 GC。密码子第 3 位碱基平均含量 (A3、U3、C3、G3) 结果显示 G3 比例最高, 其次是 U3 和 A3。GC3 值变化范围 0.453 ~ 0.552, 平均值为 0.504, 标准差 *SD* 为 0.034。第 1 与 2 位密码子的平均 GC 含量 (GC12) 变化范围为 0.421 ~ 0.448, 平均值为 0.431%, 标准差 *SD* 为 0.008。

此外, 按照 Lavner 和 Kotler (2005) 方法计算了每个氨基酸最优密码子发生频率 *Fop*, 进一步通过统计分析找出高低频率使用的密码子, 结果 (图 2) 显示, 不同物种 *API* 基因中存在 18 个最优密码子, 并且大部分以 A/U 结尾。

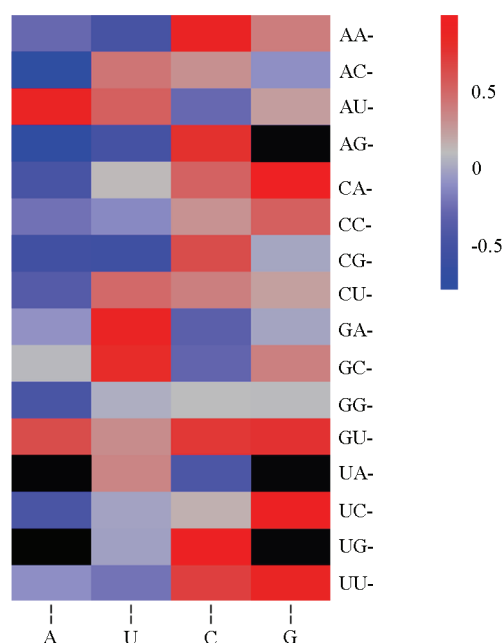


图 1 密码子与 GC3 相关系数热图

红色代表正相关, 蓝色代表负相关, 黑色表示终止密码子 (UAA、UAG、UGA) 和起始密码子 (AUG、UGG)。

Fig. 1 Heat maps of correlation coefficient of codons with GC3

The color coding red represents the positive correlation, green as negative correlation. The black fields are stop codons (UAA, UAG, UGA) and start codons (AUG, UGG) codons.

表 2 *API* 基因编码序列的核苷酸组成分析
Table 2 Nucleotide composition analysis in the coding sequences of *API* gene

No.	A	U	G	C	A3	U3	G3	C3	AU	GC	GC1	GC2	GC3	AU3	GC12
1	247	164	176	142	74	59	71	39	0.564	0.436	0.514	0.342	0.453	0.547	0.428
2	246	161	175	147	74	58	69	42	0.558	0.442	0.531	0.337	0.457	0.543	0.434
3	236	169	192	150	59	63	80	47	0.542	0.458	0.526	0.337	0.510	0.490	0.432
4	223	160	182	151	52	54	78	54	0.534	0.464	0.519	0.322	0.552	0.444	0.421
5	227	165	177	151	56	62	75	47	0.544	0.456	0.550	0.308	0.508	0.492	0.429
6	225	170	195	154	54	67	76	51	0.531	0.469	0.532	0.363	0.512	0.488	0.448
7	231	166	181	148	58	63	81	40	0.547	0.453	0.512	0.347	0.500	0.500	0.430
8	231	170	195	148	53	62	90	43	0.539	0.461	0.540	0.306	0.536	0.464	0.423
平均 Mean	233.3	165.6	184.1	148.9	60	61	77.5	45.4	0.545	0.455	0.528	0.333	0.504	0.496	0.431
SD	9.11	3.89	8.56	3.56	8.96	3.93	6.52	5.32	0.011	0.011	0.013	0.020	0.034	0.035	0.008

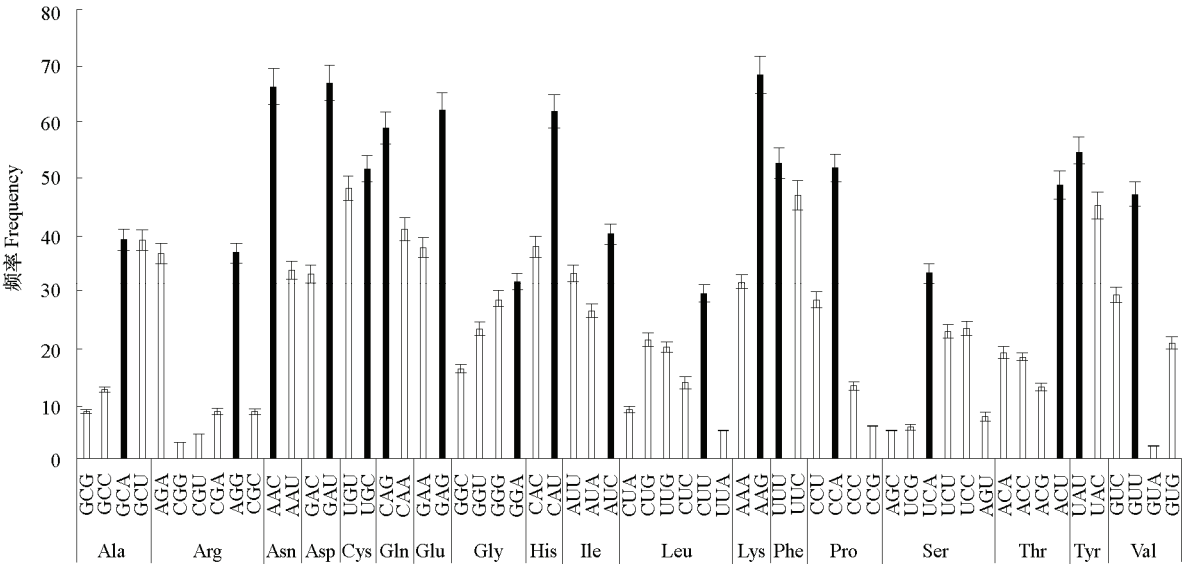


图 2 不同植物 *API* 基因最优密码子和其他密码子的整体频率
黑色代表相应氨基酸的最优密码子。

Fig. 2 Overall frequency of optimal and non-optimal codon used in *API* genes among plants
Black color coding represents optimal used codons with corresponding amino acid.

2.3 不同植物 *API* 基因的相对同义密码子使用度

8 个物种 *API* 基因 59 个密码子(除 AUG 蛋氨酸和 UGG 色氨酸)的相对同义密码子使用度 RSCU 值计算,RSCU 大于 1 代表了使用频率较高的密码子,RSCU 值大于 1.6 表示了偏好性较强的密码子。挑选的不同植物 *API* 基因 RSCU 分析表明, 28 个高频密码子几乎同等偏好以 G/C 和 A/U 结尾(表 3)。

RSCU 值聚类分析显示(图 3), 在不同植物 *API* 基因中 AGG(除苹果), AGA(除了芍药和牡丹), GUU(除了葡萄和月季)等密码子 RSCU 值大于 1.6, 表现为偏好性较强的密码子。

表 3 不同植物 *API* 基因相对同义密码子使用度 (RSCU) 整体值
Table 3 Overall relative synonymous codon usage patterns (RSCU) for *API* gene among 8 plant species

氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU	氨基酸 Amino acid	密码子 Codon	数量 Number	RSCU
Ala	GCA*	43	1.550	Gln	UGU	15	0.938		CUC	34	0.829		UCA*	45	2.061
	GCC	14	0.505		CAA	71	0.821		CUG*	53	1.293		UCC*	30	1.374
	GCG	10	0.360		CAG*	102	1.179		CUU*	73	1.780		UCG	8	0.366
	GCU*	44	1.586		GAA	72	0.754		UUA	13	0.317		UCU*	30	1.374
Arg	AGA*	51	2.217	Glu	GAG*	119	1.246		UUG*	50	1.220	Thr	ACA	14	0.800
	AGG*	51	2.217		GGA*	31	1.265		AAA	48	0.632		ACC	12	0.686
	CGA	12	0.522		GGC	16	0.653		AAG*	104	1.368		ACG	10	0.571
	CGC	12	0.522		GGG*	27	1.102		UUC	22	0.936		ACU*	34	1.943
	CGG	5	0.217	Gly	GGU	24	0.980	Phe	UUU*	25	1.064	Tyr	UAC	27	0.900
	CGU	15	0.523		CAC	19	0.760		CCA*	23	2.000		UAU*	33	1.100
	AAC*	70	1.333		CAU*	31	1.240		CCC	6	0.522		GUA	2	0.114
	AAU	35	0.667		AUA	24	0.818		CCG	3	0.261		GUC*	20	1.143
Asn	GAC	24	0.658	Ile	AUC*	34	1.159	Pro	CCU*	14	1.217		GUG	15	0.857
	GAU*	49	1.342		AUU*	30	1.023		AGC	7	0.321		GUU*	33	1.886
Asp				Leu	CUA	23	0.561	Ser	AGU	11	0.504				
	UGC*	17	1.062												

*RSCU > 1.

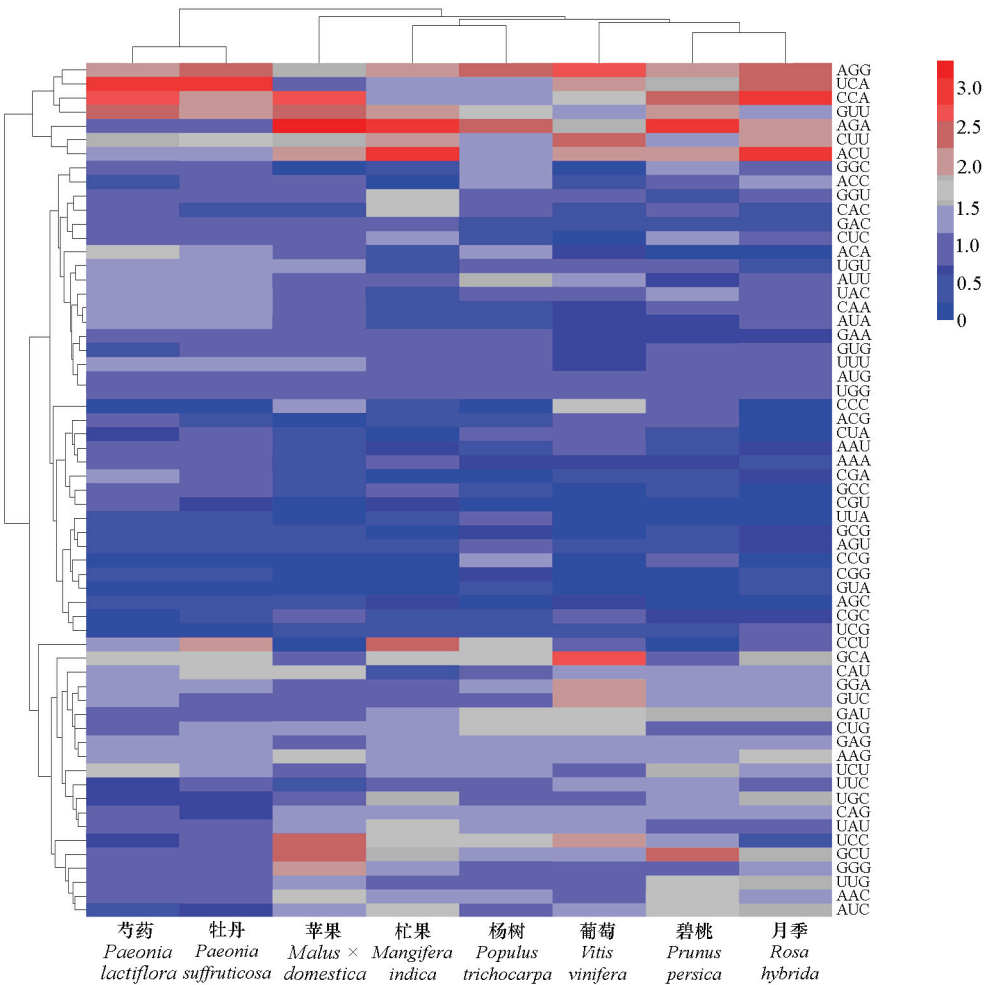


图 3 不同植物 *API* 基因密码子相对使用度 (RSCU) 聚类

每个长方形代表了一个密码子的 RSCU 值, 颜色强度代表了不同 RSCU 值。

Fig. 3 Clustering of RSCU values of each codon among *API* gene across plants

Each rectangular box on the map represents the RSCU value of a codon corresponding to the *API* gene across plant species.

2.4 不同植物 *API* 基因密码子使用模式的系统关系

构建了不同植物 *API* 基因进化树 (图 4)。一般来说, 具有相似功能的基因表现出相似的密码子使用模式 (Tatarinova et al., 2010)。本研究中结果发现, 8 个不同植物之间 *API* 基因密码子使用模式存在一定的差异, 比较密切接近的物种 *API* 基因具有显著相似的密码子使用模式, 芍药 (*Paeonia lactiflora*) 和牡丹 (*Paeonia suffruticosa*) 的 *API* 基因使用相似, 碧桃 (*Prunus persica*) 和苹果 (*Malus × domestica*) 相似。

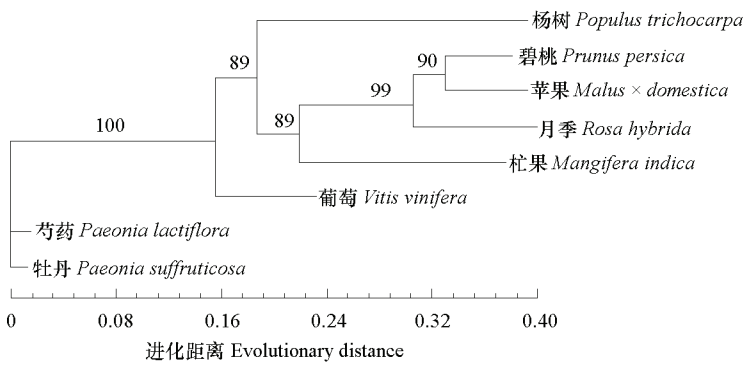


图 4 不同植物 *API* 基因编码序列基于 Kimura 2 参数距离 (K2P) 的进化分析
分支上的数字代表支持率。

Fig. 4 Phylogenetic analysis of the Kimura 2-parameter (K2P) distances of the selected coding sequences among *API* genes of different plant species
The number on the branch represents favorability rating.

2.5 不同植物 *API* 基因选择压力的影响

如表 4 所示, ENC 值变化范围从 53 到 57 (平均值 55 ± 1.5), 表明 8 个不同植物 *API* 基因密码子使用具有较小的变异。然而, GC3 值变化范围从 0.453 到 0.552 (平均值 0.504 ± 0.034)。此外, 计算了 GC3 偏斜值 (GC3 skew), 其变化范围为 0.182 到 0.353。此外, 最优密码子频率 Fop 值较低和 ENC 值较高暗示不同植物 *API* 基因在密码子使用中具有一个较弱的偏好性 (Grantham et al., 1981)。

表 4 不同植物 *API* 基因密码子使用偏好性指标
Table 4 Codon usage bias indices for *API* gene across plant species

植物 Plants	RCBS	ENC	CG3	Fop	Highest_RSCU	GC 偏度 GC skew	GC3 偏度 GC3 skew
杨树 <i>Populus trichocarpa</i>	0.222	57	0.510	0.415	ACU (Thr)	0.123	0.260
牡丹 <i>Paeonia suffruticosa</i>	0.242	57	0.457	0.352	AGA (Arg)	0.087	0.243
芍药 <i>Paeonia lactiflora</i>	0.250	56	0.453	0.346	UCA (Ser)	0.107	0.291
苹果 <i>Malus × domestica</i>	0.348	54	0.508	0.427	AGA (Arg)	0.079	0.230
月季 <i>Rosa hybrida</i>	0.349	53	0.536	0.387	GCA (Ala)	0.137	0.353
碧桃 <i>Prunus persica</i>	0.353	55	0.552	0.430	UCA (Ser)	0.093	0.182
葡萄 <i>Vitis vinifera</i>	0.355	54	0.500	0.386	CCA (Pro)	0.100	0.339
芒果 <i>Mangifera indica</i>	0.355	54	0.512	0.445	AGA (Arg)	0.117	0.197

注: RCBS 表示相对密码子使用偏好性, ENC 表示有效密码子数, GC3 表示第 3 位密码子 GC 含量, Fop 表示最优密码子频率, RSCU 表示相对同义密码子使用度。
Note: RCBS-Relative codon usage bias, ENC-Effective number of codons, GC3-GC contents at third positions of codon, Fop-Frequency of optimal codons, RSCU-Relative synonymous codon usage.

基因 RCBS 值可以被用于衡量基因表达的有效指标, 其数值主要受基因碱基组成偏好性的影响 (Wright, 1990)。不同物种 *API* 基因 RCBS 值的变化范围为 0.222 到 0.355 (表 4)。平均 RCBS 值较低, 表明具有较低表达水平的 *API* 基因密码子偏好性表现较弱。

3 讨论

研究发现, 不同生物体 CUB 现象与许多基因组特征有关, 如基因长度、GC 含量、重组率、基因表达水平、遗传密码修复等 (Karlin & Mrazek, 1996; Duret, 2000; Urrutia & Hurst, 2003; Roymondal et al., 2009; Palidwor et al., 2010)。此外, 相关研究揭示突变压力、翻译选择、蛋白二级结构、复制和选择性转录、蛋白亲疏水性以及外界环境也会影响生物体密码子使用模式 (Butt et al., 2014)。在原核和单细胞真核生物体中密码子使用主要受到突变偏差 (mutational bias) 和选择 (selection) 的共同影响 (Gouy & Gautier, 1982; Sharp et al., 2005; Bragg et al., 2012); 在多细胞真核生物中如果蝇 (*Drosophila melanogaster*) 和秀丽隐杆线虫 (*Caenorhabditis elegans*), 主要受到翻译效率的选择作用影响 (Shields et al., 1988; Stenico et al., 1994; Vicario et al., 2007); 在病毒中如细小病毒科 (Parvoviridae) 主要受到突变压力和自然选择的相互作用影响 (Shi et al., 2013)。此外, 在单细胞和多细胞生物体中研究还发现, 含有大量 tRNA 基因拷贝数的最优密码子 (optimal codons) 数量随着基因表达水平提高而上升, 表明最优密码子和 tRNA 丰度之间存在明显的正相关 (Ikemura, 1981; Akashi, 1995; Duret, 2000)。然而, 在一些植物中如拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*)、玉米 (*Zea*) 以及芍药 (*Paeonia lactiflora*), 密码子偏好性主要受到基因碱基组成和基因表达水平的影响 (Chiapello et al., 1998; Liu et al., 2004, 2010; Morton & Wright, 2007; Wu et al., 2015)。综合以上研究发现, 每个生物在长期进化过程中都会形成一种特定的密码子使用模式, 并且受到不同因素的影响, 其中碱基组成是基因组 DNA 的一个重要特征, 不同核苷酸的使用频率是由突变和回复突变之间的平衡决定的。

一般而言, 自然选择和突变压力情况下碱基组成被认为是不同生物体密码子使用变异的主要因素 (Ikemura, 1981; Li, 1987; Xu et al., 2011), GC 含量往往反映了方向性突变的强弱, 尤其是同义密码子的主要差别体现在第 3 位碱基上 (GC3s), 由于密码子第 3 位上碱基突变受到的选择压较小, GC3 常被作为分析密码子使用模式的一个重要依据 (Hassan et al., 2009)。因此, 本研究中通过分析了不同植物 *API* 基因密码子碱基组成动力学 (compositional dynamics), 进一步发现 *API* 基因大多数以密码子与 GC3s 存在较高的相关性, 表明植物 *API* 基因密码子偏好性受到碱基组成影响较大, 特别是 GC3 含量。此外, 大部分 ENC 值与 GC3s 含量呈现负相关, 并且 GC3 偏斜值 (GC3 skew) 较小, 进一步暗示了第 3 位碱基 GC3s 组成在 *API* 基因密码子使用偏好性中起重要作用 (Yang & Nielsen, 2008)。

此外, ENC 值是指一个基因的密码子使用频率与同义密码子平均使用频率偏差的量化值, 是目前公认的评价基因密码子偏好性的最具参考价值的参数, 普遍通过比较 ENC 来确定内源基因表达量的相对高低, ENC 值越小则对应的内源基因往往表达量越高。本研究中发现 8 个植物 *API* 基因 ENC 值均大于 50, 其较大的 ENC 值表明 *API* 基因在不同植物中表达水平平均很低, 可能与 *API* 基因功能机制有关。

不同物种或同一物种的不同基因对密码子的偏好性有所不同 (赖瑞联 等, 2016; 罗火林 等, 2016)。本研究发现不同植物 *API* 基因整体来看, 对密码子 G/C 和 A/U 结尾使用偏好性基本相同,

然而大多数最优密码子偏好以 A/U 结尾, 符合双子植物密码子的使用特性。本研究中仅仅选取了 8 个不同物种 *API* 基因进行比较, 并没有对单子叶植物和双子叶植物、草本植物和木本植物进行系统分析, 今后需要进一步深入分析与验证。密码子偏好性是生物进化过程中逐步形成的, 是对宿主基因组环境的一种适应, 本研究中揭示了芍药 (*Paeonia lactiflora*) 和牡丹 (*Paeonia suffruticosa*)、碧桃 (*Prunus persica*) 和苹果 (*Malus × domestica*) *API* 基因的密码子使用特点比较接近, 同时本研究确定了 *API* 基因最优势密码子 ACA、AUA 和 CAG, 这为今后设计密码子优化 (Codon optimization) 和转基因工程进行 *API* 基因功能验证提供指导。

References

- Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics*, 139 (2): 1067 – 1076.
- Behura S K, Severson D W. 2012. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One*, 7 (8): e43111.
- Bragg J G, Quigg A, Raven J A, Wagner A. 2012. Protein elemental sparing and codon usage bias are correlated among bacteria. *Molecular Ecology*, 21 (10): 2480 – 2487.
- Butt A M, Nasrullah I, Tong Y. 2014. Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One*, 9 (3): e90905.
- Carbone A, Zinovyev A, Kepes F. 2003. Codon adaptation index as a measure of dominating codon bias. *Bioinformatics*, 19 (16): 2005 – 2015.
- Chiapello H, Lisacek F, Caboche M, 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.
- Duret L, Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 96 (8): 4482 – 4487.
- Fox J M, Erill I. 2010. Relative codon adaptation: a generic codon bias index for prediction of gene expression. *DNA Research*, 17 (3): 185 – 196.
- Fuglsang A. 2004. The ‘effective number of codons’ revisited. *Biochemical and Biophysical Research Communications*, 317 (3): 957 – 964.
- Ge J T, Zhao D Q, Han C X, Wang J, Hao Z J, Tao J. 2014. Cloning and expression of floral organ development-related genes in herbaceous peony (*Paeonia lactiflora* Pall.). *Molecular Biology Reports*, 41 (10): 6493 – 6503.
- Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research*, 10 (22): 7055 – 7074.
- Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Research*, 9 (1): R43 – R74.
- Gustafsson C, Govindarajan S, Minshull J. 2004. Codon bias and heterologous protein expression. *Trends in Biotechnology*, 22 (7): 346 – 353.
- Hassan S, Mahalingam V, Kumar V. 2009. Synonymous codon usage analysis of thirty two mycobacteriophage genomes. *Advances in Bioinformatics*, 316936.
- Hooper S D, Berg O G. 2000. Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Research*, 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. *Journal of Molecular Biology*, 151 (3): 389 – 409.
- Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2 (1): 13 – 34.

- Jiang Wei, Lü Bei-bei, He Jian-hua, Wang Jing-bin, Wu Xiao, Wu Guo-gan, Bao Da-peng, Chen Ming-jie, Zhang Jin-song, Tan Qi, Tang Xue-ming. 2014. Codon usage bias in the straw mushroom *Volvariella volvacea*. Chinese Journal of Biotechnology, 30 (9): 1424 - 1435. (in Chinese)
- 蒋 玮, 吕贝贝, 何建华, 王金斌, 吴 潇, 武国干, 鲍大鹏, 陈明杰, 张劲松, 谭 琦, 唐雪明. 2014. 草菇密码子偏好性分析. 生物工程学报, 30 (9): 1424 - 1435.
- Karlin S, Mrazek J. 1996. What drives codon choices in human genes. Journal of Molecular Biology, 262 (4): 459 - 472.
- Lai Rui-lian, Lin Yu-ling, Zhong Chun-shui, Lai Zhong-xiong. 2016. Analysis of codon bias of auxin receptor gene *TIR1* in *Dimocarpus* Longan. Acta Horticulturae Sinica, 43 (4): 771 - 780. (in Chinese)
- 赖瑞联, 林玉玲, 钟春水, 赖钟雄. 2016. 龙眼生长素受体基因 *TIR1* 密码子偏好性分析. 园艺学报, 43 (4): 771 - 780.
- Li W H. 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. Journal of Molecular Evolution, 24 (4): 337 - 345.
- Lithwick G, Margalit H. 2005. Relative predicted protein levels of functionally associated proteins are conserved across organisms. Nucleic Acids Research, 33 (3): 1051 - 1057.
- Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J. 2010. Analysis of synonymous codon usage in *Zea mays*. Molecular Biology Reports, 37 (2): 677 - 684.
- Liu Q P, Feng Y, Zhao X A, Dong H, Xue Q Z. 2004. Synonymous codon usage bias in *Oryza sativa*. Plant Science, 167 (1): 101 - 105.
- Luo Huo-lin, Zhang Ya-nan, Luo Li-ping, Xiong Dong-jin, Yang Bo-yun. 2016. The codon usage bias of *FLOWERING LOCUS T* gene (*FT*) in orchidaceae. Molecular Plant Breeding, 14 (1): 51 - 58. (in Chinese)
- 罗火林, 张亚楠, 罗丽萍, 熊冬金, 杨柏云. 2016. 兰科植物 *FT* 基因的密码子偏性分析. 分子植物育种, 14 (1): 51 - 58.
- Mandel M A, Gustafson-Brown C, Savidge B, Yanofsky M F. 1992. Molecular characterization of the *Arabidopsis* floral homeotic-gene *APETALAI*. Nature, 360 (6401): 273 - 277.
- Mandlik V, Shinde S, Singh S. 2014. Molecular evolution of the enzymes involved in the sphingolipid metabolism of *Leishmania*: selection pressure in relation to functional divergence and conservation. BMC Evolutionary Biology, 14: 142.
- Ma Q P, Li C, Wang J, Wang Y, Ding Z T. 2015. Analysis of synonymous codon usage in *FAD7* genes from different plant species. Genetics and Molecular Research, 14 (1): 1414 - 1422.
- Morton B R, Wright S I. 2007. Selective constraints on codon usage of nuclear genes from *Arabidopsis thaliana*. Molecular Biology and Evolution, 24 (1): 122 - 129.
- Palidwor G A, Perkins T J, Xia X. 2010. A general model of codon bias due to GC mutational bias. PLoS One, 5 (10): e13431.
- Peden J F. 1999. Analysis of codon usage[Ph. D. Dissertation]. UK: University of Nottingham.
- Qi Ting, Liu Jing, Li Qian, Yu Jin-hui, Yuan Yong-ze, Liu De-li. 2014. The analysis of rare codons and secondary structure of mRNA for the optimization expression about *Magnaporthe oryzae* CYP51B. Journal of Huazhong Normal University (Natural Sciences), 48 (3): 392 - 398. (in Chinese)
- 齐 婷, 刘 婧, 李 倩, 余金辉, 袁永泽, 刘德立. 2014. 稻瘟菌 *MgCYP51B* 基因稀有密码子和 mRNA 二级结构分析与表达优化. 华中师范大学学报 (自然科学版), 48 (3): 392 - 398.
- Ren Lei. 2011. Cloning and expression of floral organ development related genes in tree peony[Ph. D. Dissertation]. Beijing: Chinese Academy of Forestry. (in Chinese)
- 任 磊. 2011. 牡丹花器官发育相关基因的克隆与表达[博士论文]. 北京: 中国林业科学研究院.
- Roymondal U, Das S, Sahoo S. 2009. Predicting gene expression level from relative codon usage bias: an application to *Escherichia coli* genome. DNA Research, 16 (1): 13 - 30.
- Sharp P M, Bailes E, Grocock R J, Peden J F, Sockett R E. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Research, 33 (4): 1141 - 1153.
- Sharp P M, Tuohy T M, Mosurski K R. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Research, 14 (13): 5125 - 5143.
- Shields D C, Sharp P M, Higgins D G, Wright F. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous

- codons. *Molecular Biology and Evolution*, 5 (6): 704 – 716.
- Shi S L, Jiang Y R, Liu Y Q, Xia R X, Qin L. 2013. Selective pressure dominates the synonymous codon usage in Parvoviridae. *Virus Genes*, 46 (1): 10 – 19.
- Stenico M, Lloyd A T, Sharp P M. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Research*, 22 (13): 2437 – 2446.
- Sueoka N. 1988. Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 85 (8): 2653 – 2657.
- Supek F, Vlahovick K. 2005. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC Bioinformatics*, 6: 182.
- Tatarinova T V, Alexandrov N N, Bouck J B, Feldmann K A. 2010. GC3 biology in corn, rice, sorghum and other grasses. *BMC Genomics*, 11: 308.
- Urrutia A O, Hurst L D. 2003. The signature of selection mediated by expression on human genes. *Genome Research*, 13 (10): 2260 – 2264.
- Vicario S, Moriyama E N, Powell J R. 2007. Codon usage in twelve species of *Drosophila*. *BMC Evolutionary Biology*, 7: 226 – 226.
- Wright F. 1990. The ‘effective number of codons’ used in a gene. *Gene*, 87 (1): 23 – 29.
- Wu Y Q, Zhao D Q, Tao J. 2015. Analysis of codon usage patterns in herbaceous peony (*Paeonia lactiflora* Pall.) based on transcriptome data. *Genes*, 6 (4): 1125 – 1139.
- Xu C, Cai X, Chen Q, Zhou H, Cai Y, Ben A. 2011. Factors affecting synonymous codon usage bias in chloroplast genome of oncidium gower ramsey. *Evolutionary Bioinformatics Online*, 7: 271 – 278.
- Yang Guo-feng, Su Kun-long, Zhao Yi-ran, Song Zhi-bin, Sun Juan. 2015. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*. *Acta Prataculturae Sinica*, 24 (12): 171 – 179. (in Chinese)
- 杨国锋, 苏昆龙, 赵怡然, 宋智斌, 孙娟. 2015. 蒺藜苜蓿叶绿体密码子偏好性分析. *草业学报*, 24 (12): 171 – 179.
- Yang Z, Nielsen R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Molecular Biology and Evolution*, 25 (3): 568 – 579.
- Zeng Ying, Hu Jin-yong, Li Zhi-jian. 2001. MADS-box genes in the evolution of flower development. *Plant Physiology Communications*, 37 (4): 281 – 287. (in Chinese)
- 曾英, 胡金勇, 李志坚. 2001. 植物 MADS 盒基因与花器官的进化发育. *植物生理学通讯*, 37 (4): 281 – 287.
- Zhang Tao, Zhang Gen-xi, Han Kun-peng, Wang Jin-yu, Fan Qing-can, Li Ting-ting, Duan Lian, Wang Yong-juan. 2014. Characterization of codon usage of *MyoG* gene in Jinghai Yellow Chicken. *Acta Agriculturae Boreali-Sinica*, 29 (4): 71 – 79. (in Chinese)
- 张涛, 张跟喜, 韩昆鹏, 王金玉, 樊庆灿, 李婷婷, 段炼, 王永娟. 2014. 京海黄鸡 *MyoG* 基因密码子特性分析. *华北农学报*, 29 (4): 71 – 79.
- Zhou Dong-mei, Liu Yue-xue, Zhang Zhi-hong, Li He, Ma Yue, Dai Hong-yan. 2012. Cloning, expression and promoter analysis of *API* homologous gene from strawberry (*Fragaria × ananassa*). *Scientia Agricultura Sinica*, 45 (10): 1972 – 1981. (in Chinese)
- 邹冬梅, 刘月学, 张志宏, 李贺, 马跃, 代红艳. 2012. 草莓 *API* 同源基因的克隆、表达及启动子分析. *中国农业科学*, 45 (10): 1972 – 1981.